PROSIDING SEMINAR NASIONAL STATISTIKA
UNIVERSITAS DIPONEGORO 2011
ISBN: 978-979-097-142-4

INTERVAL KONFIDENSI SPLINE KUADRAT DENGAN PENDEKATAN *PIVOTAL QUANTITY*

Rowan Daflix Syaranamual[1], I Nyoman Budiantara[2]

[1)] Mahasiswa Magister Jurusan Statistika ITS

[2)] Dosen Jurusan Statistika ITS

Abstrak

Spline merupakan salah satu model dalam regresi nonparametrik yang mempunyai interpretasi statistik secara visual sangat khusus dan sangat baik. Disamping itu spline mampu menangani karakter data atau fungsi yang bersifat mulus (*smooth*). Spline juga memiliki kemampuan yang sangat baik untuk menangani data yang perilakunya berubah-ubah pada sub-sub interval tertentu. Umumnya dalam regresi nonparametrik, estimator spline sangat tergantung pada parameter penghalus. Untuk memilih parameter penghalus optimal dalam estimator spline, biasanya menggunakan metode GCV (*Generalized Cross Validation*). Sedangkan untuk persoalan inferensi khususnya estimasi interval (interval konfidensi) biasanya menggunakan pendekatan Bayesian. Tetapi pendekatan ini memerlukan pengetahuan Matematika yang relatif tinggi dan sulit dipahami oleh banyak pengguna Statistika. Pada penelitian ini akan dikaji interval konfidensi dengan spline kuadratik berdasarkan pada *Pivotal Quantity* yang merupakan generalisasi regresi parametrik. Selanjutnya, dirancang interval konfidensi spline kuadratik menggunakan data berat badan balita di kabupaten Blitar yang diperoleh dari Dinas Kesehatan propinsi Jawa Timur tahun 2009.

Kata kunci: *Regresi Nonparametrik, Spline, Pivotal Quantity, Interval Konfidensi*

1. Pendahuluan

Analisis regresi merupakan suatu metode yang biasanya digunakan untuk memodelkan sebuah persoalan riil ke dalam bentuk persamaan matematis yang dapat menjelaskan hubungan antara variabel respon dan variabel prediktor. Pendekatan yang digunakan dalam analisis regresi ada tiga jenis, yaitu pendekatan parametrik, nonparametrik dan semiparametrik. Bila kurva regresi membentuk pola tertentu, maka pendekatannya berupa pendekatan parametrik. Sedangkan bila pola kurva regresi tidak diketahui atau tidak membentuk sebuah pola yang jelas, maka digunakan pendekatan nonparametrik. Jika digabungkan keduanya, digunakan pendekatan regresi semiparametrik. Persoalan inferensi yang sangat penting dalam regresi *spline* adalah interval konfidensi. Konsep interval konfidensi terpendek menggunakan metode optimasi yang sudah dikenal, adalah *Lagrange Multiple*. Wahba (1983;1990) dan Wang (1998) menggunakan pendekatan Bayesian dalam merancang interval konfidensi untuk kurva regresi. Pendekatan Bayesian ini memerlukan pemahaman tentang distribusi *prior improper*, fungsi risiko Bayes, distribusi *posterior*, proses stokastik *Winner*, mean *posterior* dan lainnya, yang umumnya kurang dimengerti oleh pengguna Statistika.

Dalam kehidupan sehari-hari, terdapat beberapa kasus yang harus diselesaikan menggunakan *spline* polinomial *truncated* derajat dua dimana $p = 2$. Akibatnya, perlu diturunkan suatu model *spline* kuadrat untuk menyelesaikan persoalan tersebut. Salah satu contoh aplikasi *spline* kuadrat sangat sesuai digunakan pada Pertumbuhan Balita di Amerika Serikat (Eubank, 1988). Dalam ilmu kesehatan anak, istilah pertumbuhan dan perkembangan menyangkut semua aspek kemajuan yang dicapai oleh manusia sejak lahir sampai dewasa (Soetjiningsih, 1995). Pertumbuhan berarti bertambah besar dalam aspek fisis akibat multiplikasi sel dan bertambahnya jumlah zat interseluler. Oleh karena itu, pertumbuhan dapat diukur dalam sentimeter atau inci dan dalam kilogram atau *pound*. Bawah lima tahun atau biasanya disebut balita adalah periode penting dalam tumbuh kembang anak. Balita memiliki rentang usia dimulai dari nol sampai dengan lima tahun atau 0-60 bulan. Pada masa balita merupakan usia penting dalam tumbuh kembang anak secara fisik karena masa ini pertumbuhan dasar akan mempengaruhi dan menentukan perkembangan anak selanjutnya. Salah satu ukuran perbandingannya adalah ukuran berat badan terhadap usia.

Pada penelitian ini variabel yang digunakan untuk menilai pertumbuhan balita adalah berat badan menurut usia yang merupakan indikator global pertumbuhan karena berat badan menurut usia mengindikasikan keadaan gizi saat ini (Soetjiningsih, 1995). Aritonang (2000) menyatakan berat badan menurut usia lebih sensitif terhadap perubahan keadaan gizi yang kecil. Kemudian menurut Soetjiningsih (1995) berat badan

ISBN: 978-979-097-142-4

merupakan ukuran antopometrik terpenting yang dipakai pada setiap pemeriksaan kesehatan anak pada semua kelompok usia dengan usia sebagai indikator pertumbuhan balita adalah Eubank (1988). Penelitian ini akan mengkonstruksi interval konfidensi untuk regresi nonparametrik dengan menggunakan *spline* kuadratik. Selanjutnya, diaplikasikan untuk memodelkan berat badan terhadap usia balita di kabupaten Blitar propinsi Jawa Timur dengan menggunakan pendekatan *Pivotal Quantity*.

2. Tinjauan Pustaka

2.1. *Spline* Dalam Regresi Nonparametrik

Spline dalam regresi nonparametrik mempunyai sifat fleksibilitas yang tinggi dan mempunyai kemampuan mengestimasi perilaku data yang cenderung berbeda pada interval yang berlainan (Eubank, 1988 dan Budiantara, 2006). Kemampuan ini ditunjukkan oleh fungsi *truncated* (potongan-potongan) yang melekat pada estimator dan potongan-potongan tersebut yang disebut titik *knots*. Titik *knots* merupakan titik perpaduan bersama yang menunjukkan perubahan pola perilaku fungsi pada selang yang berbeda (Hardle, 1990).

Secara umum fungsi G dalam *spline* berorde *m-1* dengan titik *knots* $k_1, k_2, \ldots, kj$ adalah sembarang fungsi yang dapat dinyatakan menjadi

$$G(t_i) = \sum_{j=0}^{m-1} \gamma_j t_i^j + \sum_{j=1}^{J} \gamma_{j+m-1} (t_i - k_j)_+^{m-1} \tag{1}$$

dimana γ_j merupakan konstanta yang bernilai real dan

$$(t - k_j)_+^{m-1} = \begin{cases} (t - k_j)^{m-1} & , t - k_j \geq 0 \\ 0 & , t - k_j < 0 \end{cases}$$

Misalkan terdapat n pasangan data observasi $\{\mathbf{t_i}, \mathbf{y_i}\}$ dengan $i = 1, 2, \ldots n$. dan diasumsikan hubungan antara t_i dan y_i mengikuti model:

ISBN: 978-979-097-142-4

$$y_i = f(t_i) + \varepsilon_i \text{ , } a < t_1 < \ldots < t_n < b \text{ , } \varepsilon_i \sim \text{IIDN}(0, \sigma^2). \quad (2)$$

dimana $f(t_i)$ adalah fungsi yang tidak diketahui yang akan diduga, dan diasumsikan merupakan fungsi yang kontinu dan diferensiabel. Sebuah penduga *smoothing spline* $\hat{f}$ untuk f didefinisikan sebagai yang meminimumkan Kriteria *penalized least square* (PLS):

$$n^{-1}\sum_{i=1}^{n}\left(y_i - f(t_i)\right)^2 + \lambda\int_a^b\left(f^{(m)}(x)\right)^2 dx \quad (3)$$

2.2. Interval Konfidensi

Estimasi interval adalah suatu interval tertentu yang memuat parameter dengan probabilitas tertentu. Misalkan $X_1, X_2, \ldots, X_n$ adalah sampel random yang diambil dari populasi dengan parameter θ, maka interval konfidensi $1-\alpha$ untuk θ adalah $P(a \le \theta \le b) = 1-\alpha$.

Diberikan sampel random $X_1, X_2, \ldots, X_n$ yang diambil dari populasi yang berdistribusi $N(\mu, \sigma^2)$ dengan σ^2 diketahui. Untuk mendapatkan interval konfidensi $1-\alpha$ untuk parameter μ dapat menggunakan *Pivotal Quantity* :

$$Z = \frac{\bar{X} - \mu}{\frac{\sigma}{\sqrt{n}}} \sim N(0,1).$$

Apabila ditentukan nilai $1-\alpha = 95\%$ maka interval konfidensi 95% untuk parameter μ diberikan oleh :

$$P(-1{,}96 \le Z \le 1{,}96) = 95\%$$

ISBN: 978-979-097-142-4

$$\Leftrightarrow \quad P\left(-1{,}96 \le \frac{\overline{X} - \mu}{\frac{\sigma}{\sqrt{n}}} \le 1{,}96 \right) = 95\%$$

$$\Leftrightarrow \quad P\left(\overline{X} - 1{,}96.\frac{\sigma}{\sqrt{n}} \le \mu \le \overline{X} + 1{,}96.\frac{\sigma}{\sqrt{n}} \right) = 95\%$$

Jika σ tidak diketahui, diganti dengan *s* yaitu standar deviasi sampel. Dengan demikian, untuk dapat mendapatkan interval konfidensi $1-\alpha$ untuk μ digunakan *pivotal Quantity* :

$$T = \frac{\overline{X} - \mu}{\frac{s}{\sqrt{n}}} \sim t_{(n-1)}$$

Dengan demikian, interval konfidensi $1-\alpha$ dapat diperoleh dari :

$$P\left(-t_{\left(n-1;\alpha/2\right)} \le T \le t_{\left(n-1;\alpha/2\right)} \right) = 1-\alpha$$

$$\Leftrightarrow \qquad P\left(\overline{X} - t_{\left(n-1;\alpha/2\right)} \frac{s}{\sqrt{n}} \le \mu \le \overline{X} + t_{\left(n-1;\alpha/2\right)} \frac{s}{\sqrt{n}} \right) = 1-\alpha.$$

2.3. Berat Badan Menurut Usia

Berat badan menurut Usia (BB/U) mencerminkan status gizi saat ini, karena berat badan menggambarkan masa tubuh (otot dan lemak) yang sensitif terhadap perubahan yang mendadak, seperti oleh sakit infeksi dan tidak cukup makan. Oleh karena itu, indikator BB/U dapat memberikan gambaran masalah gizi masa lalu atau kronis. Di samping itu, karena berat badan juga labil terhadap perubahan yang terjadi, maka BB/U juga memberikan gambaran masalah gizi akut. Indeks ini cukup sensitif untuk menilai status gizi kurang yang akut sebagai akibat memburuknya situasi, baik pada masyarakat miskin maupun pada masyarakat yang keadaan sosial ekonominya lebih baik.

ISBN: 978-979-097-142-4

3. Metodologi

Data yang digunakan pada penelitian ini adalah data sekunder yang terdiri atas berat badan balita usia 0-60 bulan di kabupaten Blitar yang berasal dari Dinas Kesehatan propinsi Jawa Timur tahun 2009.

Variabel yang digunakan dalam penelitian ini adalah berat badan menurut usia. Variabel prediktor (*t*) adalah usia balita yaitu 0-60 bulan dan variabel respon (*y*) adalah berat badan balita.

Langkah-langkah analisis adalah sebagai berikut :

1. Mengkaji interval konfidensi *spline* kuadratik dengan pendekatan *Pivotal Quantity*.

   Langkah-langkahnya adalah sebagai berikut :

   a. Menentukan basis fungsi *spline* kuadratik.

   b. Melakukan estimasi parameter dari model regresi nonparametrik *spline* $Y_i = G(t_i) + \varepsilon_i$, dimana

   $$G(t_i) = \gamma_0 + \gamma_1 t_i + \gamma_2 t_i^2 + \sum_{j=1}^{K} \gamma_{j+2} (t_i - k_j)_+^{m-1}$$

   dengan menggunakan metode MLE.

   c. Menghitung panjang interval konfidensi $1-\alpha$ dengan menggunakan konsep interval konfidensi terpendek.

   d. Menghitung optimasi interval dengan menggunakan metode *Lagrange Multiple*.

   e. Mendefinisikan fungsi Lagrange .

   f. Menderivatifkan fungsi Lagrange.

g. Menyelesaikan secara simultan langkah f.

2. Mengaplikasikan interval konfidensi *spline* kuadrat pada data berat badan menurut usia balita di kabupaten Blitar yang diperoleh dari Dinas Kesehatan propinsi Jawa Timur tahun 2009. Langkah-langkahnya adalah sebagai berikut :

4. Membuat *scatter plot* antara usia balita (t) dan berat badan (y) untuk mengetahui pola hubungan antara kedua variabel tersebut.
5. Memodelkan berat badan dan usia balita dengan menggunakan regresi *spline*.
6. Menentukan penaksir parameter model regresi *spline*.
7. Memilih model *spline* terbaik dengan memilih titik *knots* optimum dilihat dari nilai GCV yang paling minimum.
8. Berdasarkan model *spline* yang diperoleh pada langkah d, dilakukan pengujian asumsi residual.
9. Menghitung nilai koefisien determinasi (R^2).
10. Menghitung interval konfidensi kurva regresi dengan menggunakan pendekatan *spline* kuadrat.

4. Hasil Dan Pembahasan

4.1. Estimasi Titik dan Interval Konfidensi untuk Kurva Regresi *G*

Diberikan suatu basis untuk ruang *spline* kuadrat berbentuk

$$\left\{1, t, \cdots, t^2, (t-\lambda_1)_+^2, \cdots, (t-\lambda_K)_+^2\right\}$$

dengan

$$(t-\lambda)_+^2 = \begin{cases} (t-\lambda)^2, & t \geq \lambda \\ 0 & t < \lambda \end{cases} \tag{4}$$

dan $\lambda_1, \ldots, \lambda_K$ merupakan titik-titik *knots*. Untuk setiap fungsi G dalam ruang *spline* dapat dinyatakan sebagai :

ISBN: 978-979-097-142-4

$$G(t) = \gamma_0 + \gamma_1 t + \gamma_2 t^2 + \sum_{k=1}^{K} \gamma_{k+2} (t - \lambda_k)_+^2$$

dengan γ_j merupakan konstanta yang bernilai riil. Model regresi *spline* dapat ditulis menjadi :

$$z_i = \gamma_0 + \gamma_1 t_i + \gamma_2 t_i^2 + \sum_{k=1}^{K} \gamma_{k+2} (t_i - \lambda_k)_+^2 + \varepsilon_{i.}$$

Jika kita asumsikan sesatan random ε_i berdistribusi normal independen dengan mean nol dan variansi σ^2, maka z_i juga berdistribusi normal dengan mean $G(t_i)$ dan variansi σ^2. Akibatnya, diperoleh fungsi *Likelihood* :

$$L(z,G) = (2\pi\sigma^2)^{-n/2} Exp\left(-\frac{1}{2\sigma^2}(z_i - G(t_i))\right)^2$$

Estimasi titik untuk *G* diperoleh dengan menyelesaikan Optimasi *Likelihood* :

$$\underset{G}{Max}\{L(z,G)\} = \underset{\gamma \in R^{2+1+K}}{Max}\left\{(2\pi\sigma^2)^{-1/2} Exp\left(-\frac{1}{2\sigma^2}\sum_{i=1}^{n} z_i - \gamma_0 - \gamma_1 t_i - \gamma_2 t_i^2 - \sum_{k=1}^{K}\gamma_{k+2}(t_i - \lambda_k)_+^2\right)^2\right\}$$

Jika diambil transformasi logaritma dan mengingat persamaan (4) maka diperoleh fungsi:

$$Log\ L(z,\lambda,\gamma) = -\frac{n}{2}\log(2\pi\sigma^2) - \frac{1}{2\sigma^2}\sum_{i=1}^{n}\left(z_i - \gamma_0 - \gamma_1 t_i - \gamma_2 t_i^2 - \sum_{k=1}^{K}\gamma_{k+2}(t_i - \lambda_k)_+^2\right)^2$$

Dengan penyajian secara matriks, diperoleh :

$$Log\ L(z,\lambda,\gamma) = -\frac{n}{2}\log(2\pi\sigma^2) - \frac{1}{2\sigma^2}(z - T(t,\lambda)\gamma)'(z - T(t,\lambda)\gamma) \qquad (5)$$

dengan $\gamma = (\gamma_0, \gamma_1, \gamma_2, \ldots, \gamma_{K+2})'$, $z = (z_1, \ldots, z_n)'$, dan $T(t,\lambda)$ matriks berukuran $n \times (K+2)$, diberikan oleh :

ISBN: 978-979-097-142-4

$$T(t,\lambda)=\begin{pmatrix} 1 & t_1 & t_1^2 & (t_1-\lambda_1)_+^2 & \cdots & ((t_1-\lambda_K)_+^2) \\ 1 & t_2 & t_2^2 & (t_2-\lambda_1)_+^2 & \cdots & (t_2-\lambda_K)_+^2 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & t_n & t_n^2 & (t_n-\lambda_1)_+^2 & \cdots & (t_n-\lambda_K)_+^2 \end{pmatrix}$$

Jika persamaan (5) diderivatifkan parsial terhadap γ kemudian hasilnya disamakan dengan nol, diperoleh :

$$\frac{\partial\, Log\; L(z,\lambda,\gamma)}{\partial\gamma}=\frac{\partial}{\partial\gamma}\left[-\frac{1}{2\sigma^2}(z-T(t,\lambda)\gamma)'(z-T(t,\lambda)\gamma)\right]=0$$

Dengan sedikit penjabaran dan mengingat $T(t,\lambda)$ merupakan matriks dengan *rank* penuh, maka diperoleh estimasi *Likelihood* untuk γ adalah :

$$\hat{\gamma}(t,\lambda)=[T'(t,\lambda)T(t,\lambda)]^{-1}T'(t,\lambda)z$$

Estimator kurva regresi $G(t)$ diberikan oleh :

$$\begin{aligned}\hat{G}(t,\lambda)&=T(t,\lambda)[T'(t,\lambda)T(t,\lambda)]^{-1}T'(t,\lambda)z\\ &=W(t,\lambda)z\end{aligned}$$

Dengan $\quad W(t,\lambda)=T(t,\lambda)[T'(t,\lambda)T(t,\lambda)]^{-1}T'(t,\lambda)$

Terlihat bahwa $\hat{G}(t,\lambda)$ merupakan estimator linier dalam observasi *z* dan sangat tergantung pada titik *knots* $\lambda=\{\lambda_1,\cdots,\lambda_K\}$. Dalam model *spline*, titik *knots* dipilih dengan berbagai metode diantaranya dengan metode *Generalized Cross Validation* (GCV).

Diketahui $\varepsilon=(\varepsilon_1,...,\varepsilon_n)'$ berdistribusi $N(0,\sigma^2 I)$, maka :

Z berdistribusi $N(G(t),\sigma^2 I)$

ISBN: 978-979-097-142-4

Selanjutnya variabel random $\hat{\gamma}(t,\lambda)$ berdistribusi $N\left(G(t),[T'(t,\lambda)T(t,\lambda)]^{-1}\sigma^2\right)$. Ekspektasi dan variansi dari $\hat{G}(t,\lambda)$ berturut-turut diberikan oleh :

$$E\left(\hat{G}(t,\lambda)\right)=G(t)$$

dan

$$Var\left(\hat{G}(t,\lambda)\right)=\sigma^2 W(t,\lambda)$$

karena sifat linearitas dari distribusi normal maka variabel random $\hat{G}(t,\lambda)$ berdistribusi $N\left(G(t),\sigma^2 W(t,\lambda)\right)$. Jika diambil transformasi $Z(t,\lambda)=(\sigma W(t,\lambda))^{-1}\left(\hat{G}(t,\lambda)-G(t)\right)$, atau

$$Z(t_i,\lambda_1,\ldots,\lambda_K)=\frac{\gamma_0+\gamma_1 t_i+\gamma_2 t_i^2+\sum_{k=1}^{K}\hat{\gamma}_{k+2}(t_i-\lambda_k)_+^2-G(t_i)}{\sqrt{\sigma^2\omega_i(t_i,\lambda_1,\ldots,\lambda_K)}},\quad i=1,2,\ldots,n$$

$$=\frac{\hat{\gamma}_0+\hat{\gamma}_1 t_i+\hat{\gamma}_2 t_i^2+\sum_{k=1}^{K}\hat{\gamma}_{k+2}(t_i-\lambda_k)_+^2-G(t_i)}{\sqrt{\sigma^2\omega_i(t_i,\lambda_1,\ldots,\lambda_K)}}$$

dengan $\omega_i(t_i,\lambda_1,\ldots,\lambda_K)$ elemen diagonal ke-*i* dari matriks $W(t,\lambda)$. Variabel random $Z(t_i,\lambda_1,\ldots,\lambda_K)$ berdistribusi $N(0,1)$. Dengan demikian $Z(t_i,\lambda_1,\ldots,\lambda_K)$ merupakan *Pivotal Quantity* untuk kurva regresi $G(t_i)$. Interval konfidensi $1-\alpha$ diperoleh dari menyelesaikan persamaan probabilitas :

$$P\left(x\le U(t_i,\lambda_1,\ldots,\lambda_K)\le y\right)=1-\alpha$$

dengan $x\in R,\ y\in R,$ dan $x<y$, *i*=1,2,…,*n*. Persamaan diatas dapat dinyatakan menjadi :

ISBN: 978-979-097-142-4

$$P\left(x \le \frac{\hat{\gamma}_0 + \hat{\gamma}_1 t_i + \hat{\gamma}_2 t_i^2 + \sum_{k=1}^{K} \hat{\gamma}_{k+2} (t_i - \lambda_k)_+^2 - G(t_i)}{\sqrt{\sigma^2 \omega_i (t_i, \lambda_1, \ldots, \lambda_K)}} \le y \right) = 1 - \alpha$$

Dengan sedikit penjabaran, diperoleh interval konfidensi $1-\alpha$ untuk $G(t_i)$, $i = 1,2,\ldots,n$:

$$P\left(\begin{array}{l} \left[\hat{\gamma}_0 + \hat{\gamma}_1 t_i + \hat{\gamma}_2 t_i^2 + \sum_{k=1}^{K} \hat{\gamma}_{k+p} (t_i - \lambda_k)_+^p \right] - y\sqrt{\sigma^2 \omega_i (t_i, \lambda_1, \ldots, \lambda_K)} \le G(t_i) \le \\ \left[\hat{\gamma}_0 + \hat{\gamma}_1 t_i + \hat{\gamma}_2 t_i^2 \sum_{k=1}^{K} \hat{\gamma}_{k+p} (t_i - \lambda_k)_+^p \right] - x\sqrt{\sigma^2 \omega_i (t_i, \lambda_1, \ldots, \lambda_K)} \end{array} \right) = 1 - \alpha \quad (6)$$

Dengan menggunakan konsep Interval konfidensi terpendek, harus ditentukan nilai $x \in R$ dan $y \in R$, sehingga panjang interval $\ell(x,y)$ pada persamaan (6) terpendek. Untuk tujuan ini, dicari penyelesaian optimasi bersyarat berikut :

$$\underset{x \in R, y \in R}{Min} \{\ell(x,y)\} = \underset{x \in R, y \in R}{Min} \left\{ (y - x)\sqrt{\sigma^2 \omega_i (t_i, \lambda_1, \ldots, \lambda_K)} \right\}, \quad (7)$$

Dengan syarat :

$$\int_x^y \psi(u)\, du = 1 - \alpha \quad \text{atau} \quad \Phi(y) - \Phi(x) - (1-\alpha) = 0 \quad (8)$$

Fungsi ψ merupakan distribusi probabilitas N(0,1) dan Φ merupakan distribusi probabilitas Kumulatif N(0,1). Optimasi (7) dan (8) dapat diselesaikan dengan menggunakan metode *Lagrange Multiple*. Dibentuk fungsi *Lagrange* :

$$U(x,y,c) = (y - x)\sqrt{\sigma^2 \omega_i (t_i, \lambda_1, \ldots, \lambda_K)} + c\left[\int_x^y \psi(u)\, du - 1 - \alpha \right]$$

Selanjutnya dengan menderivatifkan fungsi $U(x,y,c)$ terhadap *x*, *y* dan *c* diperoleh :

$$\frac{\partial U(x,y,c)}{\partial x} = 0 \;\Rightarrow\; -\sqrt{\sigma^2 \omega_i (t_i, \lambda_1, \ldots, \lambda_K)} - c\psi(x) = 0 \quad (9)$$

ISBN: 978-979-097-142-4

$$\frac{\partial U(x,y,c)}{\partial y} = 0 \;\Rightarrow\; \sqrt{\sigma^2 \omega_i(t_i, \lambda_1, \ldots, \lambda_K)} + c\psi(y) = 0 \tag{10}$$

$$\frac{\partial U(x,y,c)}{\partial c} = 0 \;\Rightarrow\; \Phi(y) - \Phi(x) - (1-\alpha) = 0 \tag{11}$$

Persamaan (9) dan (10) menghasilkan penyelesaian :

$$\psi(x) = \psi(y) \tag{12}$$

Mengingat persamaan (11) dan $Z \sim N(0,1)$ maka penyelesaian (12) adalah $x = y$ (tidak memenuhi) atau $x = -y$. Jadi agar diperoleh interval konfidensi terpendek harus diambil nilai *x* dan *y* yang memenuhi persamaan :

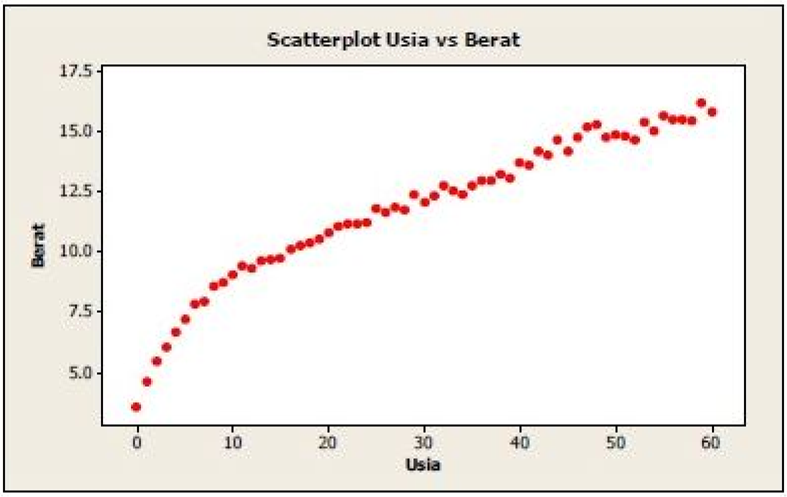
$$\int_{-\infty}^{x} \psi(u)\, du = \frac{\alpha}{2} = \int_{y}^{\infty} \psi(u)\, du \tag{13}$$

Jika tingkat konfidensi $1-\alpha$ diberikan, maka nilai *x* dan *y* dapat dilihat dalam tabel distribusi N(0,1). Interval konfidensi $1-\alpha$ untuk kurva regresi $G(t_i), i = 1,2,\ldots,n$ diberikan oleh persamaan (6), dengan *x* dan *y* memenuhi persamaan (13).

4.2. Aplikasi Model dan Interval Konfidensi

Pola pertumbuhan secara umum pada balita juga terjadi pada balita yang berada di kabupaten Blitar Propinsi Jawa Timur. Plot antara berat badan balita (*y*) dalam kilogram dan usia (*t*) dalam bulan balita di kabupaten Blitar Propinsi Jawa Timur diberikan dalam Gambar 1.

ISBN: 978-979-097-142-4



Gambar 1. Plot Berat Badan Balita di kabupaten Blitar Propinsi Jawa Timur.

Berdasarkan Gambar 1 diatas nampak bahwa terjadi perubahan pola pertumbuhan balita di kabupaten Blitar Propinsi Jawa Timur pada interval usia tertentu. Oleh sebab itu digunakan model *spline* polinomial *truncated* untuk memodelkan pola hubungan antara usia dan berat badan balita ini yaitu :

$$G(t) = \gamma_0 + \gamma_1 t + \gamma_2 t^2 + \ldots + \gamma_p t^p + \sum_{k=1}^{K} \gamma_{k+p} (t - \lambda_k)_+^2 ,$$

untuk bermacam nilai p yang menunjukkan orde *spline* dan bermacam k yang menunjukan banyaknya titik *knots*. Pemilihan titik *knots* optimal dalam model *spline* dapat menggunakan metode GCV. *Knots* yang optimal berkaitan dengan nilai GCV yang terkecil. Tabel 1 menunjukkan berbagai model *spline*, seperti *spline* linear (p = 1), *spline* kuadratik (p = 2), dan *spline* kubik (p = 3) beserta nilai GCV-nya masing-masing.

Tabel 1. Model-model *spline* dengan 1,2 dan 3 Titik *Knots* serta Nilai GCV-nya

Model	Titik	Nilai	Titik	Nilai	Titik	Nilai

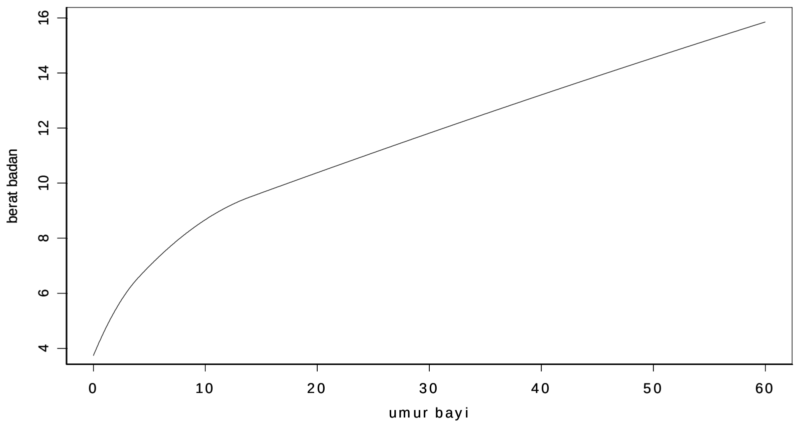
ISBN: 978-979-097-142-4

spline	***Knots***	**GCV**	***Knots***		**GCV**	***Knots***			**GCV**
	λ_1		λ_1	λ_2		λ_1	λ_2	λ_3	
Spline linier	7	0,09773	**8**	**11**	**0,06025**	8	10	20	0,07914
	8	0,08891	8	12	0,06029	8	10	21	0,07835
	9	0,08917	8	13	0,06050	8	10	22	0,07793
	10	0,09477	8	14	0,06078	8	10	23	0,07772
	11	0,10479	8	15	0,06111	8	10	24	0,07743
Spline kuadratik	12	0,11927	9	11	0,06036	9	11	16	0,06254
	13	0,13384	9	12	0,06038	9	11	17	0,06254
	14	0,14973	9	13	0,06046	9	11	18	0,06255
	15	0,16537	9	14	0,06056	9	11	19	0,06255
	16	0,17968	9	15	0,06068	9	11	20	0,06255
Spline Kubik	17	0,19408	17	26	0,08013	11	17	25	0,06273
	18	0,20832	17	27	0,08053	11	17	26	0,06273
	19	0,22224	17	28	0,08080	11	17	27	0,06272
	20	0,23573	17	29	0,08099	11	17	28	0,06268
	21	0,24931	17	30	0,08108	11	17	29	0,06261

Berdasarkan Tabel 1 terlihat bahwa nilai GCV terkecil adalah 0,06025 terdapat pada model *spline* kuadratik dengan dua titik *knots* yaitu pada balita yang berusia 8 bulan dan

ISBN: 978-979-097-142-4

balita yang berusia 11 bulan. Model *spline* kuadratik dengan dua *knots* tersebut ditunjukkan pada Gambar 2.



Gambar 2. *Spline* Kuadratik Dengan *Knots* Pada Balita Usia 8 Bulan dan 11 Bulan.

Pengujian parameter secara simultan pada model spline dengan *knots* optimum menggunakan hipotesis sebagai berikut.

$H_0 : \sigma_i^2 = \sigma^2$

$H_1 :$ minimal terdapat satu σ_i^2 yang tidak sama, i=1,2,3,...,*m*

Selanjutnya, pada tingkat signifikansi 5%, diperoleh F_{hitung} sebesar 89,48463. Karena $F_{\text{hitung}} > F_{0.05;(5,61)} = 2{,}536579$ maka H_0 ditolak. Hal ini memberikan kesimpulan bahwa dalam kasus ini terjadi heteroskedastisitas. Sehingga, model *spline* yang sebaiknya digunakan dalam kasus ini adalah *spline* terbobot. Langkah selanjutnya yang dilakukan adalah menentukan bobot yang digunakan dalam model *spline* terbobot. Bobot diperoleh dengan menggunakan metode *Generalized Moving Average* (GMA) (Silverman, 1985). Nilai bobot yang diperoleh dengan menggunakan metode GMA secara lengkap dapat dilihat pada lampiran 1. Selanjutnya ditentukan model *spline* terbobot terbaik dengan memilih titik *knots* optimal, menggunakan metode GCV dan diperoleh bahwa titik *knots* optimal terjadi pada model spline terbobot kuadratik dua titik knot (t = 8 dan t = 13), adalah model spline terbaik, karena memiliki nilai GCV terkecil yaitu 0,1039. Model *spline* terbobotnya diberikan oleh :

ISBN: 978-979-097-142-4

$$\hat{G}(t) = 3{,}612 + 0{,}944t - 0{,}041t^2 + 0{,}031(t-8)_+^2 + 0{,}01(t-13)_+^2$$

Dengan sedikit penjabaran model *spline* terbobot ini dapat disajikan dalam bentuk :

$$\hat{G}(t) = \begin{cases} 3{,}612 + 0{,}944t - 0{,}041t^2 & ;\ t \le 8 \\ 5{,}596 + 0{,}448t - 0{,}01t^2 & ;\ 8 \le t \le 13 \\ 7{,}286 + 0{,}188t & ;\ t \ge 13 \end{cases}$$

Dari model yang terbentuk diatas, maka dapat dikatakan bahwa pertumbuhan balita pada periode usia $\le$ 8 bulan, pertumbuhannya sangat cepat dan memiliki pola kuadratik. Pertumbuhan balita pada periode usia antara 8 dan 13 bulan juga mengalami pertumbuhan yang cepat namun tak secepat periode $\le$ 8 bulan, namun masih berpola kuadratik. Tetapi pertumbuhan balita pada periode $\ge$ 13 bulan mengalami pertumbuhan yang cenderung lambat dan memiliki pola linier. Kemudian, dengan menggunakan tingkat signifikansi 5%, diperoleh nilai distribusi F dengan derajat bebas pembilang 4 dan derajat bebas penyebut 56, adalah sebesar 32170,22. Untuk mengetahui parameter mana yang signifikan dilakukan pengujian parameter secara parsial pada tingkat signifikansi 5% menggunakan hipotesis sebagai berikut :

$$H_0 : \gamma_j = 0$$

$$H_1 : \gamma_j \neq 0, \text{ untuk } j = 1,2,3,4.$$

Hasil hipotesis ini diperoleh dalam perhitungan statistik, yang dapat dilihat pada Tabel 2.

Tabel 2. Analisis Variansi Model Spline Terbobot Kuadrat Dua Knot.

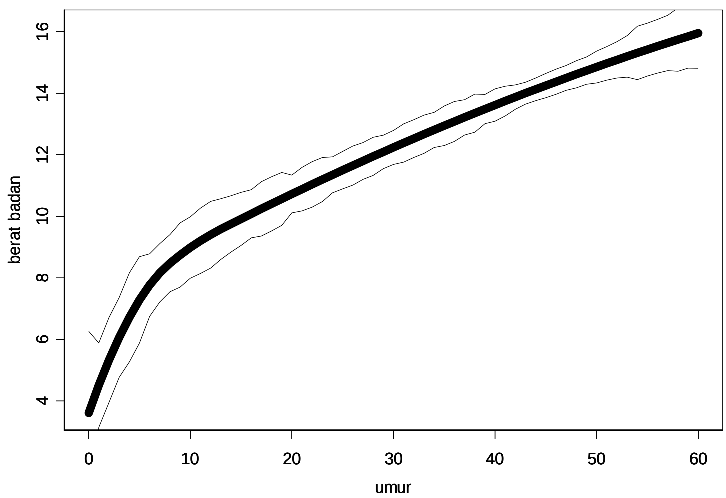
Parameter	Estimasi Parameter	S. Dev	T hitung
γ_1	0,944	0,018	51,8563
γ_2	-0,041	0,001	-22,7909
γ_3	0,031	0,003	9,1528

γ_4	0,01	0,001	5,2834

Setelah mendapatkan model *spline* terbaik dengan model *spline* kuadratik dua titik *knots* yaitu 8 dan 13 bulan, dibangun interval konfidensi 95% yaitu :

$$P\left(\begin{array}{l}\left(\left[3{,}612+0{,}944t_i-0{,}041t_i^2+0{,}031(t-8)_+^2+0{,}01(t-13)_+^2\right]-1{,}96\sqrt{\sigma^2\omega_i(t_i,\lambda_1,\ldots,\lambda_K)}\right)\le G(t_i)\\ \le\left(\left[3{,}612+0{,}944t_i-0{,}041t_i^2+0{,}031(t-8)_+^2+0{,}01(t-13)_+^2\right]+1{,}96\sqrt{\sigma^2\omega_i(t_i,\lambda_1,\ldots,\lambda_K)}\right)\end{array}\right)=0{,}95$$

Interval konfidensi 95% untuk kurva regresi $G(t_i)$ ditunjukkan oleh Gambar 3.



Gambar 3. Interval Konfidensi *Spline* Kuadratik dengan Titik *Knots* 8 dan 13

5. Kesimpulan dan Saran

Berdasarkan hasil dan pembahasan yang telah dilakukan, maka dapat diambil kesimpulan sebagai berikut :

4 Estimasi titik kurva regresi *spline* kuadratik dalam regresi nonparametrik dapat diperoleh dengan menggunakan optimasi *Likelihood*. Untuk membangun interval

ISBN: 978-979-097-142-4

konfidensi dalam regresi nonparametrik dapat digunakan pendekatan *Pivotal Quantity* dengan konsep interval konfidensi terpendek.

5 Model *spline* kuadratik merupakan model yang sangat baik untuk memodelkan pola hubungan berat badan dan usia balita di kabupaten Blitar dengan model *spline* terbobotnya diberikan oleh.

$$\hat{G}(t) = 3{,}612 + 0{,}944t - 0{,}041t^2 + 0{,}031(t-8)_+^2 + 0{,}01(t-13)_+^2$$

Sedangkan saran yang dapat diberikan adalah sebagai berikut :

2. Pada penelitian berikutnya dapat dibuat interval konfidensi untuk kurva regresi nonparametrik menggunakan *spline* umum berorde $m > 3$.
3. Interval konfidensi dengan *pivotal Quantity* dapat digeneralisasikan pada model-model regresi nonparametrik yang lebih rumit, seperti multirespon, data longitudinal dan lain-lain.

Daftar Pustaka

Akaike, H., 1974, A New Look at the Statistical Model Identification, *IEEE Transactions on Automatic Control*, Vol.19, hal. 716-723.

Aritonang, I., 2000, *Pemantauan Pertumbuhan Balita (Petunjuk Praktis Menilai Status Gizi & Kesehatan)*, Kanisius, Yogyakarta.

Budiantara, I.N., 2000, Metode U, GML, CV dan GCV Dalam Regresi Nonparametrik Spline, *Majalah Ilmiah Himpunan Matematika Indonesia (MIHMI), 6,285-290.*

_________, 2006, Model Spline Dengan Knots Optimal, *Jurnal Ilmu Dasar, FMIPA Universitas Jember,7,77-85*.

Cantoni, E., Roncheti, E., (2001). Resistant Selection of The Smoothing Parameter for Smoothing Spline, *Statistics. Comput.* 11,141-146.

Drapper, N.R and Smith, H., 1996, *Applied Regression Analysis, 2nd edition*, John Wiley & Sons, Chapman and Hall, New York.

Eubank, R.L., 1988, *Spline Smoothing and Nonparametric Regression*, Mercel Dekker, New York.

Gujarati, D., 1992, *Essentials of Econometrics*, McGRAW-Hill. Inc, New York.

Hardle, W., 1990, *Applied Nonparametric Regression*, Cambridge University Press, New York.

Khair, A., Budiantara, I.N., dan Fitriasari, K., 2006, Spline Polinomial Truncated untuk Interval Konfidensi Kurva Regresi Nonparametrik, *Prosiding Seminar Nasional Statistika VII*, ITS, Surabaya.

Lee, C.M., Thomas, 2003, On Spline Regression with Spatially-Adaptive Penalties. Unpublished Manuscript.

Soetjiningsih, 1995, *Tumbuh Kembang Anak*, Laboratorium Ilmu Kesehatan Anak Universitas Airlangga, Surabaya.

Supariasa, I.N., Bakri, B., dan Fajar, I., (2002), *Penilaian Status Gizi*, Penerbit Buku Kedokteran EGC, Jakarta.

Wahba, G., 1983, Bayesian Confidence Interval for the Cross Validated Smoothing Spline, *Journal of the Royal Statistical Society, series B*, 45, 133-150.

Wang, N., 2003, Marginal Nonparametric Kernel Regression Accounting for Within-Subject Correlation, *Biometrika*, 90, 43-52.

Wang, Y., 1998, Spline Smoothing Models with Correlated Errors, *Journal of the American Statistical Association*, 93, 341-348.

Wu, H. and Zhang, J.T., 2006, *Nonparametric Regression Method for Longitudinal Data Analysis : Mixed Effects Modeling Approaches*, John Wiley & Sons, New York.

ISBN: 978-979-097-142-4

Zhang, H.P., 1997, Multivariate Addaptive Spline for the Analysis Longitudinal Data, *Journal of Computational and Graphical Statistics*, 6, 74-91.