Statistika, Vol. 6 No. 1, 47 – 54
Mei 2006

Pendugaan Regresi Spline Terpenalti dengan Pendekatan Model Linear Campuran

ANIK DJURAIDAH[1] DAN AUNUDDIN[2]

[1)] Mahasiswa S3 program studi Statistika IPB
aniksb@plasa.com
[2)] Dosen Departemen Statistika FMIPA IPB

ABSTRAK

Regresi spline terpenalti, atau P-spline, adalah regresi yang ditentukan dengan kuadrat terkecil dan penalti kekasaran. P-spline dapat direpresentasikan dalam bentuk model linear campuran dengan komponen ragam mengontrol tingkat ketidaklinearan dari penduga fungsi mulusnya. Pendugaan P-spline dengan pendekatan model linear campuran mempunyai tiga keuntungan. Keuntungan pertama adalah P-spline dapat diduga dengan metode kemungkinan maksimum (ML) atau dengan metode kemungkinan maksimum berkendala (REML). Keuntungan kedua adalah komputasi lebih cepat karena menggunakan basis pemulus berdimensi rendah. Keuntungan ketiga adalah P-spline dapat dikembangkan untuk model dengan peubah penjelas lebih dari satu.
Kata kunci: spline terpenalti, P-spline, pemulus spline, parameter pemulus, penalti kekasaran, model linear campuran, REML, komponen ragam

1. Pendahuluan

Analisis regresi digunakan untuk memodelkan hubungan antara peubah respon dengan satu atau lebih peubah penjelas. Metode ini berkembang dari model parametrik sampai dengan model nonparametrik. Regresi parametrik memerlukan asumsi yang ketat, sebaliknya regresi nonparametrik tidak memerlukan asumsi. Sehingga regresi nonparametrik bersifat fleksibel dan menghasilkan model yang berbasis data. Beberapa metode regresi nonparametrik antara lain metode kernel, regresi spline, pemulus spline, dan ekspansi deret wavelet dan Fourier.

Pendekatan model nonparametrik dengan spline ada dua macam yaitu regresi spline dan pemulus spline. Regresi spline memerlukan jumlah simpul (*knot*) yang relatif sedikit dan dapat diduga dengan metode kuadrat terkecil. Pemulus spline memerlukan jumlah simpul yang banyak dan kemulusan kurva ditentukan oleh parameter pemulus dan fungsi penalti. Eilers dan Mark (1996) menggabungkan kedua pendekatan spline di atas menjadi P-spline. Pendekatan yang sama juga dikemukakan Ruppert dan Carroll (1997) dengan nama regresi spline terpenalti (*penalized spline regression*) dan disebut juga dengan P-spline. P-spline menggabungkan dua keuntungan, yaitu dari pendugaan parametrik pada regresi spline dan penyesuaian yang fleksibel terhadap tingkat kehalusan kurva yang dihasilkan dari penalti kekasaran (*roughness penalty*) pada pemulus spline.

Regresi spline terpenalti mempunyai hubungan matematis yang sederhana dengan model linear campuran seperti yang dibahas oleh Wang (1998) Fan dan Zhang (1998), Brumback *et al* (1999), Vebyla (1999), French *et al* (2001), Kamman dan Wand (2003), dan Wand (2003). Metode ini dapat diformulasikan sebagai penduga kemungkinan maksimum dalam kerangka model linear campuran. Efek samping yang menarik dari hubungan ini adalah parameter pemulus berhubungan secara langsung dengan komponen ragam dari model linear campuran. Hubungan antara pemulusan dengan model linear campuran baru terbuka setelah munculnya paket program model campuran di SAS dan S-PLUS

2. Regresi Spline Terpenalti

Misalkan (x_i, y_i) adalah pengukuran pada peubah penjelas x dan peubah respon y untuk $1 \le i \le n$. Misalkan hubungan fungsional antara x dengan y dimodelkan sebagai

$$y_i = s(x_i) + \varepsilon_i \quad (1)$$

dengan s adalah fungsi mulus, ε_i bebas stokastik dengan ragam σ^2. Model (1) adalah bentuk regresi nonparametrik yang paling sederhana dan banyak metode pendekatan yang dapat digunakan seperti yang dibahas oleh Eubank (1988), Green dan Silverman (1994), dan Simonoff (1996). Misalkan fungsi mulus s diduga dengan model regresi spline yaitu:

$$s(\mathbf{x};\boldsymbol{\beta}) = \beta_0 + \beta_1 x + \ldots + \beta_p x^p + \sum_{k=1}^{K} u_{pk}(x-\kappa_k)_+^p \tag{2}$$

dengan $\boldsymbol{\beta} = (\beta_0, \ldots, \beta_p, u_{p1}, \ldots, u_{pK})'$ adalah vektor koefisien regresi spline, $p \geq 1$ adalah bilangan bulat positif, $(w)_+^p = w^p \mathbf{I}(w \geq 0)$ adalah basis fungsi pangkat terpotong berderajat-p (*truncated power function* selanjutnya disingkat dengan FPT), dan $\kappa_1 < \ldots < \kappa_K$ adalah simpul tetap. Untuk peubah tunggal, selain basis FPT dapat juga digunakan basis natural kubik spline, atau basis B-spline, sedangkan untuk peubah ganda umumnya digunakan fungsi basis radial.

Penduga parameter $\hat{\boldsymbol{\beta}}$ ditentukan dengan minimisasi jumlah kuadrat terpenalti, yaitu $\mathrm{J}(s)$ yang didefinisikan sebagai:

$$\mathrm{J}(s) = \sum_{i=1}^{n} (y_i - s(x_i;\boldsymbol{\beta}))^2 + \lambda\, \boldsymbol{\beta}'\,\mathbf{D}\,\boldsymbol{\beta} \tag{3}$$

dengan λ adalah parameter pemulus, dan $\mathbf{D} = \mathrm{diag}(\mathbf{0}_{p+1}, \mathbf{1}_K)$. Suku pertama pada $\mathrm{J}(s)$ adalah jumlah kuadrat galat dan suku keduanya adalah penalti kekasaran. Kriteria penentuan model pada persamaan (3) merupakan gabungan antara kriteria pada model regresi spline dengan kriteria dari pemulus spline. Sehingga minimisasi $\mathrm{J}(s)$ pada nilai λ tertentu akan memberikan kompromi antara kebaikan pengepasan dengan kehalusan kurva. Model aditif degan kriteria pendugaan pada persamaan (3) disebut juga dengan regresi spline terpenalti (Ruppert dan Carroll, 1997).

Parameter pemulus $\lambda \geq 0$ menggambarkan tingkat pertukaran antara jumlah kuadrat galat dengan keragaman lokal. Bila λ bernilai besar maka komponen utama dalam $\mathrm{J}(s)$ adalah penalti kekasaran sehingga kurva s akan tampak mulus Sebaliknya bila λ bernilai kecil maka komponen utama dalam $\mathrm{J}(s)$ adalah komponen jumlah kuadrat galat sehingga kurva s akan tampak kasar. Misalkan $\mathbf{T}$ adalah matriks desain untuk regresi spline dengan baris ke-i dari matriks $\mathbf{T}$ yaitu

$$\mathbf{T}_i = \left(1,\ x_i,\ \ldots x_i^p,\ (x_i-\kappa_1)_+^p,\ \ldots (x_i-\kappa_K)_+^p\right)'$$

maka dalam notasi matriks $\mathrm{J}(s)$ dinyatakan sebagai

$$\left\| \mathbf{y} - \mathbf{T}\,\boldsymbol{\beta} \right\|^2 + \lambda\, \boldsymbol{\beta}'\,\mathbf{D}\,\boldsymbol{\beta} \tag{4}$$

Minimisasi persamaan (4) menghasilkan penduga bagi parameter $\boldsymbol{\beta}$ yaitu

$$\hat{\boldsymbol{\beta}} = (\mathbf{T}'\,\mathbf{T} + \lambda \mathbf{D})^{-1}\mathbf{T}'\,\mathbf{y}$$

sehingga penduga regresi spline terpenalti adalah

$$\hat{\mathbf{y}} = \mathbf{T}\hat{\boldsymbol{\beta}} = \mathbf{T}(\mathbf{T}'\,\mathbf{T} + \lambda\ \mathbf{D})^{-1}\mathbf{T}'\,\mathbf{y} \tag{5}$$

3. Model Linear Campuran

Model linear campuran banyak digunakan untuk analisa data dengan galat yang berkorelasi. Model linear campuran juga dikenal dengan model komponen ragam secara umum didefinisikan sebagai:

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{u} + \boldsymbol{\varepsilon}\text{, dengan } \begin{bmatrix}\mathbf{u}\\ \boldsymbol{\varepsilon}\end{bmatrix} \sim \left(\begin{bmatrix}0\\0\end{bmatrix}, \begin{bmatrix}\mathbf{G} & 0\\ 0 & \mathbf{R}\end{bmatrix}\right) \tag{6}$$

dimana **X** adalah matriks desain dari efek tetap yang teramati, **β** adalah vektor parameter pengaruh efek tetap yang tidak diketahui, **Z** adalah matriks desain efek acak yang teramati, **u** adalah vektor efek acak yang tidak diketahui, dan **ε** juga vektor galat acak yang tidak diketahui. Sehingga nilaitengah dan matriks ragam-peragam untuk **y** adalah $E(\mathbf{y}) = \mathbf{X\beta}$ dan $\text{var}(\mathbf{y}) = \mathbf{V} = \mathbf{ZGZ'} + \mathbf{R}$.

Persamaan model linear campuran didefinisikan sebagai (Christensen, 1987) :

$$\begin{bmatrix} \mathbf{X'R^{-1}X} & \mathbf{X'R^{-1}Z} \\ \mathbf{Z'R^{-1}X} & \mathbf{Z'R^{-1}Z + G^{-1}} \end{bmatrix} \begin{bmatrix} \boldsymbol{\beta} \\ \mathbf{u} \end{bmatrix} = \begin{bmatrix} \mathbf{X'R^{-1}y} \\ \mathbf{Z'R^{-1}y} \end{bmatrix}$$

Bila matriks **G** dan **R** diketahui, maka penduga bagi parameter **β** dan **u** dalah :

$$\begin{bmatrix} \hat{\boldsymbol{\beta}} \\ \hat{\mathbf{u}} \end{bmatrix} = \begin{bmatrix} \mathbf{X'R^{-1}X} & \mathbf{X'R^{-1}Z} \\ \mathbf{Z'R^{-1}X} & \mathbf{Z'R^{-1}Z + G^{-1}} \end{bmatrix}^{-1} \begin{bmatrix} \mathbf{X'} \\ \mathbf{Z'} \end{bmatrix} \mathbf{R^{-1}y}$$

Dari persamaan matriks di atas, penduga efek tetap **β** dapat dinyatakan sebagai

$$\hat{\boldsymbol{\beta}} = \left(\mathbf{X'V^{-1}X}\right)^{-1}\mathbf{X'V^{-1}y} \qquad (7)$$

Penduga ini disebut juga dengan penduga kuadrat terkecil terampat (GLS). Sedangkan vektor efek acak **u** diprediksi melalui prediksi linear terbaik (BLP). Bila **β** diketahui, maka BLP(**u**) adalah $\hat{\mathbf{u}} = \mathbf{GZ'V^{-1}}(\mathbf{y} - \mathbf{X\beta})$.

Penduga takbias linear terbaik (BLUP) untuk **β** identik dengan solusi GLS, dan BLUP untuk **u** adalah BLP(**u**) dengan subsitusi $\hat{\boldsymbol{\beta}}$ pada **β**. Salah satu cara yang paling sederhana untuk mendapatkan BLUP adalah menggunakan justifikasi Henderson (*Henderson's justification)* dengan menggunakan asumsi sebaran (Speed, 1991), yaitu $\mathbf{y}|\mathbf{u} \sim N(\mathbf{X\beta} + \mathbf{Zu}, \mathbf{R})$ dan $\mathbf{u} \sim N(\mathbf{0}, \mathbf{G})$

Maksimisasi fungsi kemungkinan bersama (**y,u**) pada **β** dan **u** yang tidak diketahui akan menghasilkan kriteria :

$$(\mathbf{y} - \mathbf{X\beta} - \mathbf{Zu})'\mathbf{R}^{-1}(\mathbf{y} - \mathbf{X\beta} - \mathbf{Zu}) + \mathbf{u'G^{-1}u} \qquad (8)$$

Suku pertama pada persamaan (8) adalah jumlah kuadrat terampat dan suku keduanya adalah penalti. Hal ini menunjukkan bahwa pendugaan BLUP dari (**β**,**u**) terdiri dari GLS ditambah dengan satu suku penalti. BLUP dari (**β**,**u**) dapat ditulis sebagai

$$\begin{bmatrix} \hat{\boldsymbol{\beta}} \\ \hat{\mathbf{u}} \end{bmatrix} = \left(\mathbf{C'R^{-1}C} + \mathbf{B}\right)^{-1}\mathbf{C'R^{-1}y}$$

dimana $\mathbf{C} \equiv [\mathbf{X}\ \mathbf{Z}]$ dan $\mathbf{B} \equiv \begin{bmatrix} \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{G}^{-1} \end{bmatrix}$

sehingga $\text{BLUP}(\mathbf{y}) = \mathbf{X}\hat{\boldsymbol{\beta}} + \mathbf{Z}\hat{\mathbf{u}} = \mathbf{C}\left(\mathbf{C'R^{-1}C} + \mathbf{B}\right)^{-1}\mathbf{C'R^{-1}y}$

Metode yang paling banyak digunakan dalam pendugaan matriks ragam-peragam pada model linear campuran adalah metode ML dan REML. Metode REML menghasilkan penduga takbias bagi parameter matriks ragam-peragam, sedangkan ML menghasilkan penduga yang bias. Penduga ML dari **V** ditentukan berdasarkan model sebaran

$$\mathbf{y} \sim N(\mathbf{X\beta} + \mathbf{Zu}, \mathbf{V})$$

Fungsi log-kemungkinan (*log-likelihood*) dari **y** berdasarkan model sebaran di atas adalah:

$$\ell_{ML}(\boldsymbol{\beta}, \mathbf{V}) = -\tfrac{1}{2}\left\{\log|\mathbf{V}| + (\mathbf{y} - \mathbf{X\beta})'\mathbf{V}^{-1}(\mathbf{y} - \mathbf{X\beta}) + n\log(2\pi)\right\} \qquad (9)$$

Penduga ML bagi $(\boldsymbol{\beta}, \mathbf{V})'$ diperoleh dengan memaksimumkan $\ell_{ML}(\boldsymbol{\beta}, \mathbf{V})$. Jika dilakukan optimasi terhadap **β** dengan menganggap **V** tetap, maka akan diperoleh bahwa

$\ell_{ML}(\boldsymbol{\beta}, \mathbf{V})$ maksimum pada $\hat{\boldsymbol{\beta}} = \left(\mathbf{X}'\mathbf{V}^{-1}\mathbf{X}\right)^{-1}\mathbf{X}'\mathbf{V}^{-1}\mathbf{y}$, yang sama dengan BLUP pada persamaan (7).

Substitusi $\hat{\boldsymbol{\beta}}$ pada persamaan (9) menghasilkan profil log-kemungkinan (*profile log-likelihood*) untuk **V** yaitu:

$$\begin{aligned}\ell_P(\mathbf{V}) &= -\tfrac{1}{2}\left\{\log|\mathbf{V}| + \left(\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}}\right)'\mathbf{V}^{-1}\left(\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}}\right) + n\ \log(2\pi)\right\} \\ &= -\tfrac{1}{2}\left\{\log|\mathbf{V}| + \mathbf{y}'\mathbf{V}^{-1}\left(\mathbf{I} - \mathbf{X}\left[\mathbf{X}'\mathbf{V}^{-1}\mathbf{X}\right]^{-1}\mathbf{X}'\mathbf{V}^{-1}\right)\mathbf{y}\right\} - \tfrac{n}{2}\log(2\pi)\end{aligned} \tag{10}$$

Penduga ML bagi parameter-parameter dalam **V** dapat diperoleh dengan maksimisasi persamaan (10) terhadap parameter-parameternya.

Penurunan kriteria REML lebih rumit, yaitu meliputi maksimisasi fungsi kemungkinan dari kombinasi linear elemen **y** yang tidak tergantung pada **β**. Uraian lengkap tentang metode REML dapat dijumpai pada Searle *et al* (1992). Fungsi kriteria dari REML adalah

$$\begin{aligned}\ell_{REML}(\mathbf{V}) &= -\tfrac{1}{2}\left\{\log|\mathbf{V}| + \log\left|\mathbf{X}'\mathbf{V}^{-1}\mathbf{X}\right| + \left(\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}}\right)'\mathbf{V}^{-1}\left(\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}}\right) + (n-p)\log(2\pi)\right\} \\ &= \ell_P(\mathbf{V}) - \tfrac{1}{2}\log\left|\mathbf{X}'\mathbf{V}^{-1}\mathbf{X}\right| + \tfrac{p}{2}\log(2\pi)\end{aligned}$$

Keuntungan utama dari REML dibandingkan ML adalah REML memperhitungkan derejat bebas dari efek tetap dalam model linear campuran.

4. Hubungan antara Regresi Spline Terpenalti dengan Model linear Campuran

Hubungan antara spline dengan model linear campuran telah dibahas oleh banyak peneliti, seperti Vebyla (1999) yang menghubungkan antara pemulus spline dari Green dan Silverman (1994) dengan peubah penjelas tak bias terbaik (BLUP) dari model linear campuran. Wang (1998) juga menjelaskan hubungan pemulus spline dengan model linear campuran melalui perluasan metode kemungkinan maksimum terampat (*generalized maximum likelihood*) untuk data yang berkorelasi. Sedangkan hubungan antara regresi spline terpenalti dengan model linear campuran antara lain diuraikan oleh Fan dan Zhang (1998), Brumback *et al* (1999), French *et al* (2001), Kamman dan Wdan (2003), dan Wdan (2003).

Kunci hubungan antara regresi spline terpenalti dengan model linear campuran adalah koefisien penalti u_{pk} pada model (2) ekivalen dengan memperlakukan koefisien ini sebagai efek acak pada model linear campuran (6). Misalkan didefinisikan parameter $\boldsymbol{\beta}^* = \left(\beta_0, \beta_1, \cdots, \beta_p\right)'$, $\mathbf{u} = \left(u_{p1}, \ldots, u_{pK}\right)'$, dan matriks desain

$$\mathbf{X} = \begin{bmatrix} 1 & t_1 & \cdots & t_1^p \\ \vdots & \vdots & \ddots & \vdots \\ 1 & t_1 & \cdots & t_n^p \end{bmatrix} \text{ dan } \mathbf{Z} = \begin{bmatrix} (t_1 - \kappa_1)_+^p & \cdots & (t_1 - \kappa_K)_+^p \\ \vdots & \ddots & \vdots \\ (t_n - \kappa_1)_+^p & \cdots & (t_n - \kappa_K)_+^p \end{bmatrix}$$

Kriteria spline terpenalti pada persamaan (4) jika dibagi dengan σ_ε^2 dapat ditulis sebagai

$$\frac{1}{\sigma_\varepsilon^2}\ \left\| \mathbf{y} - \mathbf{X}\boldsymbol{\beta}^* - \mathbf{Z}\mathbf{u} \right\|^2 + \frac{\lambda_s}{\sigma_\varepsilon^2}\ \left\| \mathbf{u} \right\|^2 \tag{11}$$

Persamaan (11) sama dengan kriteria BLUP dari model linear campuran pada persamaan (8) dengan memperlakukan **u** sebagai koefisien dari efek acak dengan

$$\mathrm{cov}(\ \mathbf{u}\ ) = \sigma_u^2\mathbf{I} \text{ dimana } \sigma_u^2 = \frac{\sigma_\varepsilon^2}{\lambda_s}$$

Sehingga representasi regresi spline terpenalti dalam bentuk model linear campuran adalah $\mathbf{y} = \mathbf{X}\boldsymbol{\beta}^* + \mathbf{Z}\mathbf{u} + \boldsymbol{\varepsilon}$, dengan $\mathrm{cov}\begin{bmatrix}\mathbf{u} \\ \boldsymbol{\varepsilon}\end{bmatrix} = \begin{bmatrix}\sigma_u^2\mathbf{I} & \mathbf{0} \\ \mathbf{0} & \sigma_\varepsilon^2\mathbf{I}\end{bmatrix}$

BLUP untuk fungsi $s(\mathbf{x}) = (s(x_1),..., s(x_n))'$ diberikan oleh

$$\hat{\mathbf{y}} = \mathbf{X}\hat{\boldsymbol{\beta}}^* + \mathbf{Z}\hat{\mathbf{u}} \tag{12}$$

dimana $\hat{\boldsymbol{\beta}}^* = \left(\mathbf{X}'\left(\sigma_u^2\mathbf{Z}\mathbf{Z}' + \sigma_\varepsilon^2\mathbf{I}\right)^{-1}\mathbf{X}\right)^{-1}\mathbf{X}'\left(\sigma_u^2\mathbf{Z}\mathbf{Z}' + \sigma_\varepsilon^2\mathbf{I}\right)^{-1}\mathbf{y}$

dan $\hat{\mathbf{u}} = \sigma_u^2\mathbf{Z}'\left(\sigma_u^2\mathbf{Z}\mathbf{Z}' + \sigma_\varepsilon^2\mathbf{I}\right)^{-1}\left(\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}}^*\right)$.

Solusi $\hat{\mathbf{y}}$ di atas dapat dinyatakan dalam bentuk

$$\hat{\mathbf{y}} = \mathbf{C}'\left(\mathbf{C}'\mathbf{C} + \lambda\mathbf{D}\right)^{-1}\mathbf{C}'\mathbf{y} \tag{13}$$

dimana $\mathbf{C} = [\mathbf{X} \quad \mathbf{Z}]$, $\mathbf{D} = \mathrm{diag}(\mathbf{0}_{p+1}, \mathbf{1}_K)$ dan $\lambda_s = \frac{\sigma_\varepsilon^2}{\sigma_u^2}$.

Persamaan (13) ekivalen dengan solusi regresi spline terpenalti pada persamaan (5). Bukti ini menunjukkan bahwa BLUP bagi $s(\mathbf{x})$ pada model linear campuran ekivalen dengan penduga regresi spline terpenalti.

5. Penerapan pada Data Pencemar PM_{10}

PM_{10} adalah partikulat debu dengan diameter aerodinamik <10 mikron. Partikulat debu dalam bentuk tersuspensi merupakan campuran yang sangat rumit dari berbagai senyawa organik dan anorganik yang tersebar di udara. Partikulat ini berada di udara dalam waktu yang relatif lama dalam keadaan melayang-layang di udara. Data konsentrasi PM_{10} (µg/m^3) yang digunakan pada penelitian ini diukur oleh stasiun pemantau kualitas udara kota Surabaya. Lokasi pengamatan di Taman Prestasi Surabaya pada waktu bulan Mei 2002 sampai Agustus 2002. Data lebih dulu ditransformasi logaritma agar bentuk sebarannya simetrik.

Pemodelan pada data dilakukan dengan pemulus kubik spline dan regresi spline terpenalti (untuk selanjutnya disebut sebagai P-spline). Pada pemulus kubik spline, penduga parameter diseleksi berdasarkan nilai GCV (*General Cross Validation*) yang paling kecil. Derajat dari basis FPT yang digunakan pada P-spline adalah 1, 2, dan 3. Jumlah simpul yang digunakan pada P-spline sebesar 36 mengikuti aturan Ruppert (2002). Pendugaan model P-spline menggunakan pendekatan model linear campuran.

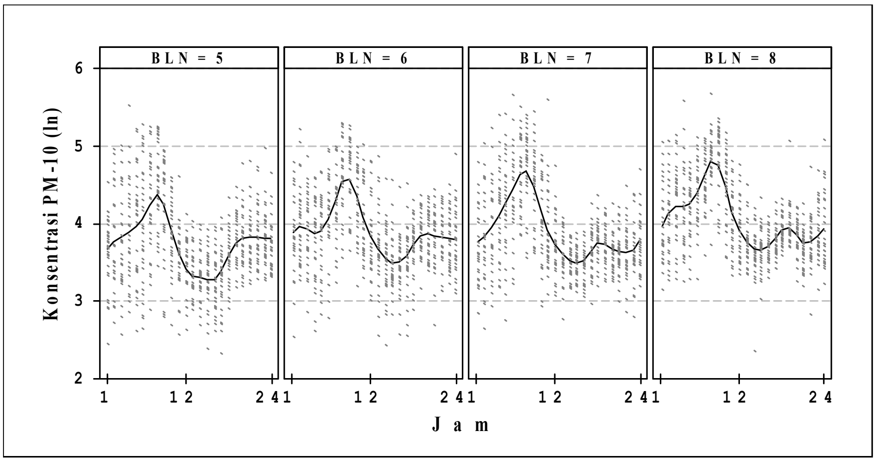
Hasil pendugaan model disarikan pada Tabel 1. Pada Tabel 1 tampak perbedaan nilai penduga parameter pemulus, penalti kekasaran, dan ragam galat antara pemulus kubik spline dengan ketiga model P-spline tidak terlalu besar. Demikian juga perbedaan nilai AIC antara ketiga model P-spline tidak besar. Pemulus kubik spline mempunyai penduga parameter pemulus dan ragam galat terkecil. Sedangkan dari ketiga model P-spline tampak kecenderungan nilai penalti kekasaran semakin kecil dengan bertambahnya derajat basis P-spline.

Tabel 1. Nilai Penduga Model

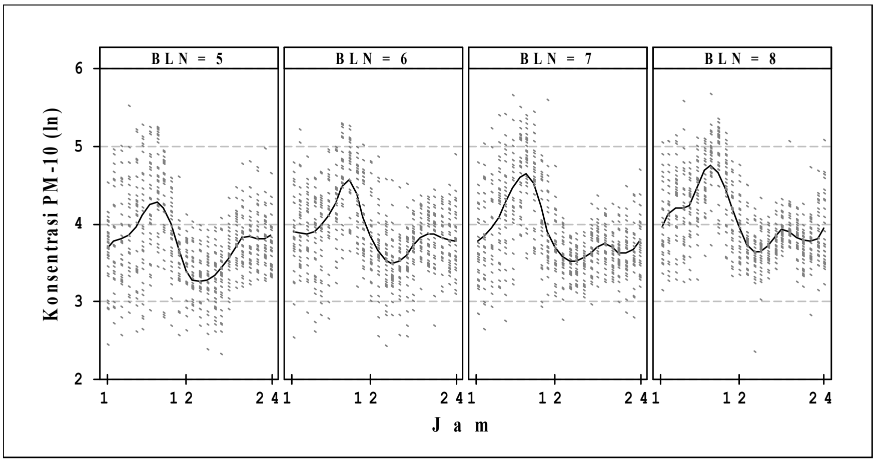
Penduga	Pemulus Kubik Spline	Regresi Spline Terpenalti Basis Berderajat		
		1	2	3
Parameter Pemulus	1.7504	2.1227	2.2651	2.8293
Penalti kekasaran	8.4919	6.3664	5.9948	5.3733
Ragam galat	0.1927	0.1943	0.1952	0.1949
AIC	-	3678.0	3734.7	3770.9

*) *Akaike Information Criteria*

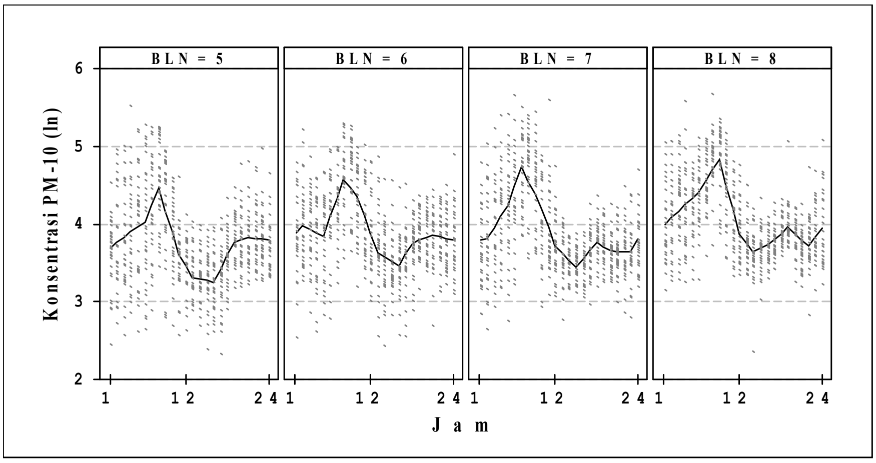
Kurva penduga dari pemulus kubik spline untuk PM10 disajikan pada Gambar 1, sedangkan kurva penduga P-spline dengan basis FPT berderajat 1, 2, dan 3, masing-masing disajikan pada Gambar 2, 3, dan 4. Secara umum kurva pada Gambar 1 sampai Gambar 4 mempunyai pola yang sangat mirip. Perbedaan antara ketiga model P-spline terdapat pada lengkungan ujung-ujung kurva. Kurva P-spline berderajat-1 mempunyai ujung lengkungan yang lancip, sedangkan kurva pada P-spline berderajat-2 dan P-spline berderajat-3 mempunyai ujung lengkungan kurva tumpul. Di antara ketiga model P-spline, tampak kurva P-spline berderajat-3 paling mulus, karena mempunyai nilai penalti kekasaran yang paling kecil. Pola lengkungan pada kurva P-spline berderajat-2 mirip dengan pola lengkungan pada kurva pemulus kubik spline, hal ini karena derajat polinomial kedua model sama.



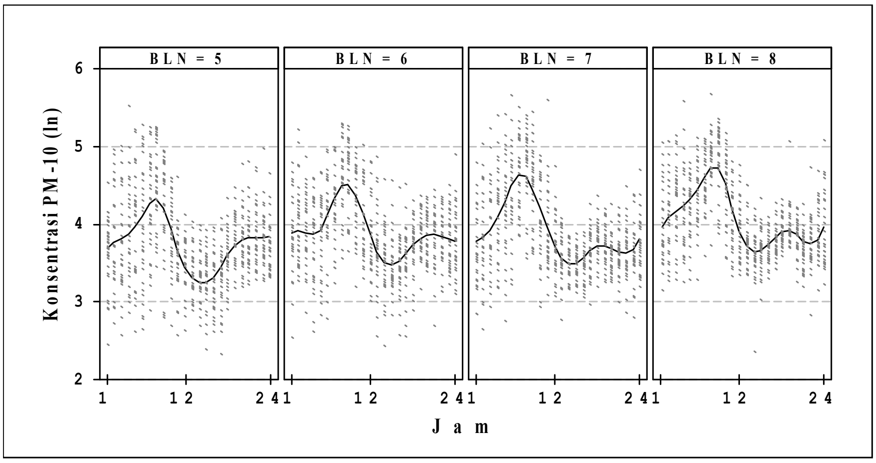
Gambar 1. Kurva Pemulus Kubik Spline



Gambar 3. Kurva P-spline Berderajat-2



Gambar 2. Kurva P-spline Berderajat-1



Gambar 4. Kurva P-spline Berderajat-3

6. Simpulan

Pendugaan model regresi spline terpenalti dengan pendekatan model linear campuran menghasilkan nilai dugaan yang hampir sama dengan pemulus kubik spline. Pendekatan ini memberikan kemudahan dalam pendugaan model regresi spilne terpenalti, sehingga dapat dikembangkan untuk memodelkan respon dengan peubah penjelas lebih dari satu. Komputasi model P-spline menjadi lebih cepat untuk data yang besar karena menggunakan basis berdimensi rendah. Kurva P-spline berderajat-2 sangat mirip dengan kurva pemulus spline.

7. Daftar Pustaka

Brumback BA, Ruppert D, Wand MP. 1999. Comment on Variable selection and function estimation in additive nonparametric regression using a data-based prior by Shively, Kohn and Wood. *Journal of the American Statistical Association* 94: 794-797.

Christensen, R. 1984. *Plane Answers to Complex Questions. The Theory of Linear Models.* New York : Springer-Verlag.

Eilers PHC, Marx BD. 1996. Flexible smoothing with B-splines and penalties (with discussion). *Statistical Science* 11: 89-121.

Eubank RL. 1988. *Spline Smoothing and Nonparametric Regression.* New York : Marcel Deker.

Fan J, Zhang JT. 1998. Comment on Smoothing spline models for the analysis of nested and crossed samples of curves by Brumback and Rice. *Journal of the American Statistical Association* 93: 961-994.

French JL, Kammann EE, Wand MP. 2001. Comment on Semiparametric nonlinear mixed-effects models and their applications by Ke and Wang. *Journal of the American Statistical Association* 96:1285-1288.

Green PJ, Siverman BW. 1994. *Nonparametric Regression and Generalized Linear Models : a Roughness Penalty Approach.* London : Chapman & Hall.

Kammann EE, Wand MP. 2003. Geoadditive models. *Applied Statistics* 52:1-18.

Ruppert D, Carroll RJ. 1997. Penalized regression splines. Unpublished manuscript. [terhubung berkala]. http://www.orie.cornell.edu/~davidr/ papers/ Index /index.html

Ruppert D. 2002. Selecting the number of knots for penalized splines. *Journal of Computational and Graphical Statistics* 11: 735–757.

Searle SR, Casella G, McCulloch CE. 1992. *Variance Component*. New York : John Wiley and Sons.

Simonoff JS.1996. *Smoothing Methods in Statistics*. New York : Springer-Verlag.

Speed T. 1991. Comment on That BLUP is a good thing : the estimation of random effects by Robinson. *Statistical Science* 6 :42-44.

Wand M. 2003. Smoothing and mixed models. *Computational Statistics* 18:223–249.

Wang Y, 1998. Mixed effects smoothing spline analysis of variance. *Journal of the Royal Statistical Society*, Series B 60:159-174.

Verbyla AP, Cullis BR, Kenward, MG, Welham SJ. 1999. The analysis of designed experiments and longitudinal data by using smoothing splines (with discussion). *Journal of the Royal Statistics Society*, Series C 48: 269-312.