 ISBN: 978-979-16353-3-2

S-12

METODE KUADRAT TERKECIL UNTUK ESTIMASI KURVA REGRESI SEMIPARAMETRIK SPLINE

Wahyu Wibowo
Mahasiswa S3 Jurusan Matematika
Universitas Gadjah Mada Yogyakarta
Sri Haryatmi
Jurusan Matematika
Universitas Gadjah Mada Yogyakarta
I Nyoman Budiantara
Jurusan Statistika
Institut Teknologi Sepuluh Nopember Surabaya

Abstrak

Pada regresi semiparametrik, untuk memperoleh estimator spline pada dasarnya terdapat dua pendekatan optimasi, yaitu estimator spline yang diperoleh berdasarkan optimasi *penalized least square* (PLS) dan estimator spline yang diperoleh berdasarkan optimasi *least square* (LS) dengan menggunakan fungsi keluarga yang memuat titik-titik knots. Apabila estimator spline yang diperoleh berdasarkan optimasi PLS, maka persoalan utama dalam estimator ini adalah pemilihan parameter penghalus yang optimal. Sedangkan apabila estimator spline yang diperoleh dengan optimasi LS, maka persoalan utama dalam estimator ini adalah pemilihan titik-titik knot yang optimal Selain itu pendekatan basis spline *truncated* memberikan perhitungan matematik yang relatif lebih mudah dan sederhana. Selain itu optimasinya dapat dikerjakan tanpa melibatkan *penalty,* yaitu dengan menggunakan optimasi kuadrat terkecil (*least square*). Makalah ini akan mengilustrasikan optimasi kuadrat terkecil dengan fungsi keluarga polinomial spline *truncated* yang memuat titik knots pada estimasi parameter regresi semiparametrik.

Kata kunci : Regresi semiparametrik, kuadrat terkecil, polinomial spline truncated, GCV

1. Pendahuluan

Model semiparametrik, dalam beberapa literatur disebut juga dengan model linear parsial (*partial linear model, partially linear model, partly linear model, partial spline model*), memuat dua komponen sekaligus, yaitu komponen parametrik dan nonparametrik. Model ini lebih fleksibel daripara model linear karena keberadaan dua komponen ini akan mengakomodasi hubungan antara respon dengan prediktor yang bersifat linear, dan hubungan antar respon dengan prediktor yang bersifat nonlinear. Model regresi semiparametrik merupakan pendekatan pemodelan yang diawali penggunaannya oleh Engle (1986), dan selama beberapa dekade terakhir banyak dikembangkan baik secara teori maupun aplikasi. Beberapa contoh aplikasi diantaranya dalam bidang ekonomi (Schmalensee and Stoker, 1999), biometrik (Zeger and Diggle, 1994), dan ilmu lingkungan (Prada-Sánchez et al., 2000). Lebih lengkap tentang ilustrasi aplikasi-aplikasi regresi semiparametrik dapat diilhat pada Hardle et al (2000). Lin dan Carroll (2001) juga menggunakan model semiparametrik untuk pengelompokan data dengan mengembangkan *semiparametric partially generalized linear models* untuk pengelompokan data. Chen dan Jin (2006) menggunakan model semiparametrik untuk pengelompokan data. Selanjutnya, Qin et al (2008, 2009) menggunakan regresi semiparametrik pada data longitudinal. You dan Zhou (2009) menggunakan polinomial spline pada regresi semiparametrik untuk data panel.

Salah satu pendekatan untuk komponen nonparametrik dalam regresi semiparametrik adalah spline. Penggunaan spline karena mempunyai interpretasi statistik dan interpretasi visual sangat khusus dan sangat baik (Eubank,1988). Selain itu spline juga mampu menangani karakter data/fungsi yang bersifat mulus (*smooth*) dan memiliki kemampuan yang sangat baik untuk menangani data yang perilakunya berubah-ubah pada sub-sub interval tertentu (Cox dan O'Sullivan,1996 dan Budiantara,2006). Penggunaan spline dalam regresi semiparametrik diantaranya dapat dilihat pada Engle (1986), Heckman (1986), Rice (1986), Wahba (1990), Green dan Silverman (1994), Eubank et al (1999), Aneiros et al (2004), Yu dan Ruppert (2002,2004), Delecroix et al. (2006), Liang (2006) dan Pitrun et al (2006). Aneiros et al (2004) menggunakan spline untuk mengestimasi dan menguji parameter pada regresi semiparametrik dengan error yang berkorelasi. He et al (2005) dan Qin et al (2008) mengembangkan metode estimasi yang robust pada regresi semiparametrik. Liang et al (2006) membandingkan secara numerik estimator spline dan kernel dalam regresi semiparametrik. Pitrun et al (2006) menggunakan spline berdasarkan uji nonlinearitas pada regresi semiparametrik. He dan Shi (1996), Shi dan Li (1994) He et al (2005) dan

 ISBN: 978-979-16353-3-2

Qin et al (2008) telah mengembangkan metode estimator spline parsial untuk mengestimasi kurva regresi yang bersifat robust. Budiantara (2006c) juga telah memberikan pendekatan keluarga spline untuk mengestimasi kurva regresi semiparametrik dan menghasilkan estimator spline parsial *truncated* dan membandingkan hasilnya dengan pendekatan B-spline dan MARS. Kelebihan estimator spline parsial *truncated* adalah memiliki interpretasi statistik yang sangat baik dan sangat sesuai untuk mengestimasi data yang mempunyai perubahan perilaku yang berubah-ubah pada sub-sub interval yang berlainan.

Banyak penelitian yang membahas metode estimasi parameter model regresi linear parsial. Diantaranya penelitian tersebut adalah Robinson (1988) yang mengkonstruksi estimator kuadrat terkecil untuk β berdasar estimasi komponen nonparametrik dengan estimator kernel Nadaraya-Watson. Haryatmi dan Subanar (2003) melakukan kajian tentang kernel pada model linier parsial dengan kasus heteroskedatisitas. Hardle, Liang dan Gao (2000) menyatakan bahwa jika komponen nonparametrik diketahui, estimator kuadrat terkecil dapat diterapkan dengan mengasumsikan komponen nonparametrik *g* sebagai parameter pengganggu (*nuisance parameter*). Selanjutnya *g* diestimasi dengan metode pemulusan (smoothing). Sedangkan Engle, et al (1986), Heckman (1986), Rice (1986), Wahba (1990), Green dan Silverman (1994), Eubank, et al (1998) menggunakan smoothing pline yang dipenalti untuk estimator pada regresi semiparametrik.

Pada regresi semiparametrik, untuk memperoleh estimator spline pada dasarnya terdapat dua pendekatan optimasi, yaitu estimator spline yang diperoleh berdasarkan optimasi *penalized least square* (PLS) dan estimator spline yang diperoleh berdasarkan optimasi *least square* (LS) dengan menggunakan fungsi keluarga yang memuat titik-titik knots. Apabila estimator spline yang diperoleh berdasarkan optimasi PLS, maka persoalan utama dalam estimator ini adalah pemilihan parameter penghalus yang optimal. Sedangkan apabila estimator spline yang diperoleh dengan optimasi LS, maka persoalan utama dalam estimator ini adalah pemilihan titik-titik knot yang optimal. Budiantara (2006b) memperlihatkan dengan menggunakan basis

 ISBN: 978-979-16353-3-2

fungsi keluarga spline *truncated* bahwa pemilihan parameter penghalus dan pemilihan titik knot optimal estimator spline dalam regresi nonparametrik adalah ekuivalen. Selain itu pendekatan basis spline *truncated* memberikan perhitungan matematik yang relatif lebih mudah dan sederhana. Disamping itu optimasinya dapat dikerjakan tanpa melibatkan *penalty* yaitu dengan menggunakan optimasi kuadrat terkecil (*least square*). Budiantara, dkk (2009) juga menggunakan spline truncated polinomial pada regresi nonparametrik heteroskedastik data longitudinal. Makalah ini akan mengilustrasikan optimasi kuadrat terkecil dengan fungsi keluarga spline *truncated* yang memuat titik knots pada estimasi parameter regresi semiparametrik.

Bentuk kurva regresi spline dangat dipengaruhi oleh parameter penghalus. Craven dan Wahba (1979) memberikan metode *cross validation* (CV) untuk memilih parameter penghalus optimal dalam estimator spline. Wang (1998) memberikan metode *unbiased risk* (UBR) untuk memilih parameter penghalus estimator spline. Wahba (1990) memberikan metode untuk memilih parameter penghalus optimal dalam estimator spline, yaitu metode *generalized cross validation* (GCV). Wahba (1990) memperlihatkan secara teoritis bahwa GCV mempunyai sifat optimal asimtotik, yang tidak dimiliki oleh metode lain. Karena kelebihannya ini metode GCV menjadi sangat terkenal dalam regresi nonparametrik dan semiparametrik dan sering digeneralisasikan serta disesuaikan bentuk formulanya oleh para peneliti dalam estimator spline lain untuk memilih parameter penghalus optimal. Budiantara (2000a) menggeneralisasikan metode GCV dari Wahba (1990) untuk estimator spline terbobot dan memperlihatkan sifat optimal asimtotik metode ini masih berlaku untuk estimator spline terbobot.

2. Hasil Penelitian

Pandang *n* sampel random dengan variabel hasil pengukuran pada masing-masing sampel $(y_i, x_i, t_i), i = 1, 2, \ldots, n$. Selanjutnya dibentuk model semiparametrik

$$y_i = \beta_0 + \beta_1 x_i + f(t_i) + \varepsilon_i, \quad i = 1, 2, \ldots, n \tag{1}$$

(β_0, β_1) parameter untuk komponen parametrik, dan *f* komponen nonparametrik yang merupakan fungsi yang tidak diketahui dan termuat di dalam ruang *Sobolev* $W_2^m[a,b]$, dengan :

$$W_2^m[a,b] = \{ f \,;\, \int_a^b (f^{(m)}(t))^2 dt < \infty \}$$

. $\varepsilon_1, \varepsilon_2, ..., \varepsilon_n$ merupakan error random yang saling independen dengan mean nol dan varian σ^2. Selanjutnya akan didapatkan estimator untuk (β_0, β_1) dan *f*. Kurva *f* dalam hal ini akan dihampiri dengan fungsi spline polinomial *truncated* derajat *p* dengan *k* titik knots $\kappa_1, \kappa_2, ..., \kappa_k$ dan disajikan dalam bentuk,

$$f(t_i) = \alpha_0 + \alpha_1 t_i + \alpha_2 t_i^2 + ... + \alpha_p t_i^p + \alpha_{p1}(t_i - \kappa_1)_+^p + ... + \alpha_{pk}(t_i - \kappa_k)_+^p \qquad (2)$$

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

Sehingga model (1) dapat dinyatakan menjadi

$$y_i = \beta_0 + \beta_1 x_i + \alpha_0 + \alpha_1 t + \alpha_2 t^2 + ... + \alpha_p t^p + \alpha_{p1}(t - \kappa_1)_+^p + ... + \alpha_{pk}(t - \kappa_k)_+^p + \varepsilon_i, \quad i = 1, 2, ..., n \quad (3)$$

Apabila dinyatakan dalam notasi matriks, diperoleh

$$\mathbf{y} = \mathbf{x}\boldsymbol{\beta} + \mathbf{z}\boldsymbol{\alpha} + \boldsymbol{\varepsilon} \qquad (4)$$

dengan

$$\mathbf{y} = \begin{bmatrix} y_1 \\ \vdots \\ y_n \end{bmatrix}, \mathbf{x} = \begin{bmatrix} 1 & x_1 \\ \vdots & \vdots \\ 1 & x_n \end{bmatrix}_{n\text{x}2}, \boldsymbol{\beta} = \begin{bmatrix} \beta_0 \\ \beta_1 \end{bmatrix}_{2\text{x}1}, \mathbf{z} = \begin{bmatrix} 1 & x & \cdots & x^p & (t-\kappa_1)_+^p & \cdots & (t-\kappa_k)_+^p \\ \vdots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ 1 & x & \cdots & x^p & (t-\kappa_1)_+^p & \cdots & (t-\kappa_1)_+^p \end{bmatrix}_{n\text{x}(1+p+k)}$$

$$\boldsymbol{\alpha} = \begin{bmatrix} \alpha_0 & \alpha_1 & \cdots & \alpha_p & \alpha_{p1} & \cdots & \alpha_{pk} \end{bmatrix}_{1\text{x}(1+p+k)}^{\text{T}}$$

Selanjutnya, ditentukan $\mathbf{C} = [\mathbf{x} \;\; \mathbf{z}]$, $\boldsymbol{\omega} = \begin{bmatrix} \boldsymbol{\beta} \\ \boldsymbol{\alpha} \end{bmatrix}$, sehingga (4) dapat dinyatakan dengan

 ISBN: 978-979-16353-3-2

$$\mathbf{y} = \mathbf{C}\boldsymbol{\omega} + \boldsymbol{\varepsilon} \tag{5}$$

Estimator $\boldsymbol{\omega}$ diperoleh dengan meminimumkan

$$= (\mathbf{y} - \mathbf{C}\boldsymbol{\omega})^{\mathbf{T}}(\mathbf{y} - \mathbf{C}\boldsymbol{\omega}) = \mathbf{y}^{\mathbf{T}}\mathbf{y} - 2\boldsymbol{\omega}^{\mathrm{T}}\mathbf{C}^{\mathrm{T}}\mathbf{y} + \boldsymbol{\omega}^{\mathrm{T}}\mathbf{C}^{\mathrm{T}}\mathbf{C}\boldsymbol{\omega} \tag{6}$$

Untuk meminimumkan, diturunkan terhadap $\boldsymbol{\omega}$ dan disamadengankan nol, sehingga diperoleh :

$$\begin{aligned} \frac{\partial \boldsymbol{\varepsilon}^{\mathrm{T}}\boldsymbol{\varepsilon}}{\partial \boldsymbol{\omega}} &= -2\mathbf{C}^{\mathrm{T}}\mathbf{y} + 2\mathbf{C}^{\mathrm{T}}\mathbf{C}\boldsymbol{\omega} = \mathbf{0} \\ \mathbf{C}^{\mathrm{T}}\mathbf{C}\boldsymbol{\omega} &= \mathbf{C}^{\mathrm{T}}\mathbf{y} \end{aligned} \tag{6}$$

yang merupakan persamaan normal. Penyelesaian persamaan ini akan merupakan estimator $\boldsymbol{\omega}$. Sesuai aljabar matrik, karena matrik **C** mempunyai rank (3 + *p* + *k*) dan $\mathbf{C}^{\mathrm{T}}\mathbf{C}$ matrik *positive-definite*, maka $\mathbf{C}^{\mathrm{T}}\mathbf{C}$ akan merupakan matrik nonsingular. Sehingga persamaan (6) akan mempunyai penyelesaian tunggal,

$$\hat{\boldsymbol{\omega}} = (\mathbf{C}^{\mathrm{T}}\mathbf{C})^{-1}\mathbf{C}^{\mathrm{T}}\mathbf{y} \tag{7}$$

Dalal hal ini, $\hat{\boldsymbol{\omega}}$ merupakan estimator kuadrat terkecil $\boldsymbol{\omega}$. Mengingat persamaan (2), estimator (7) berlaku hanya untuk derajat polinomial *p* dan banyak knots *k* yang tertentu. Sehingga lebih tepat kalau estimator ini dinyatakan dengan

$$\hat{\boldsymbol{\omega}}_{p,\kappa} = (\mathbf{C}^{\mathrm{T}}\mathbf{C})^{-1}\mathbf{C}^{\mathrm{T}}\mathbf{y} \tag{8}$$

$$\hat{\mathbf{y}} = C\hat{\omega}_{p,\kappa} = \mathbf{C}(\mathbf{C}^{\mathrm{T}}\mathbf{C})^{-1}\mathbf{C}^{\mathrm{T}}\mathbf{y} \tag{9}$$

$$\hat{\mathbf{y}} = \mathbf{H}(p;\kappa_1,\ldots,\kappa_k)\mathbf{y} \tag{10}$$

dengan $\mathbf{H}(p;\kappa_1,\ldots,\kappa_k) = \mathbf{C}(\mathbf{C}^{\mathrm{T}}\mathbf{C})^{-1}\mathbf{C}^{\mathrm{T}}$

Permasalahan selanjutnya adalah bagaimana menentukan derajat polinomial *p* dan banyak knots *k* yang akan digunakan dalam estimator. Untuk keperluan ini akan dipergunakan kriteria GCV, yang didefinisikan :

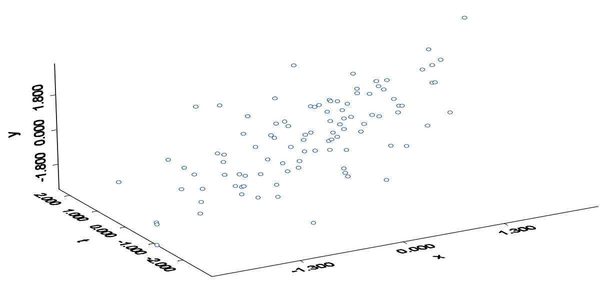
$$GCV(p;\kappa_1,\kappa_2,...,\kappa_k) = \frac{n^{-1}\left\|(\mathbf{I}-\mathbf{H}(p;\kappa_1,\kappa_2,...,\kappa_k)\mathbf{y}\right\|^2}{\left[n^{-1}trace(\mathbf{I}-\mathbf{H}(p;\kappa_1,\kappa_2,...,\kappa_k)\right]^2} \tag{11}$$

Nilai *p* dan *k* dipilih dengan menyelesaikan optimasi

$$GCV(p_{opt};\kappa_{1_{opt}},\kappa_{2_{opt}},...,\kappa_{k_{opt}}) = \min_{\substack{p\in\Re^+ \\ \kappa_1\in\Re^+ \\ \vdots \\ \kappa_k\in\Re^+}} \left(GCV(p;\kappa_1,\kappa_2,...,\kappa_k)\right) \tag{12}$$

Selajutnya metode kuadrat terkecil di atas akan diimplementasikan dengan data simulasi dan data real. Untuk simulasi, ditentukan n = 100, kemudian dibangkitkan x~Normal(0,1), t~Normal(0,1), e~Normal(0,0.01), b0=1 , b1=1.2, dan

$$y = b_0 + b_1 x + \sin^3(2\pi t) + e$$



Gambar 1. Plot data hasil simulasi

Berdasarkan kriteria GCV, diperoleh nilai p dan K yang meminimumkan GCV yaitu *p*=2 dan *k*=4 (Tabel 1). Hal ini berarti fungsi nonparametrik yang diperoleh adalah fungsi polinomial spline *truncated* berderajat 2 dengan banyak titik knots 4.

 ISBN: 978-979-16353-3-2

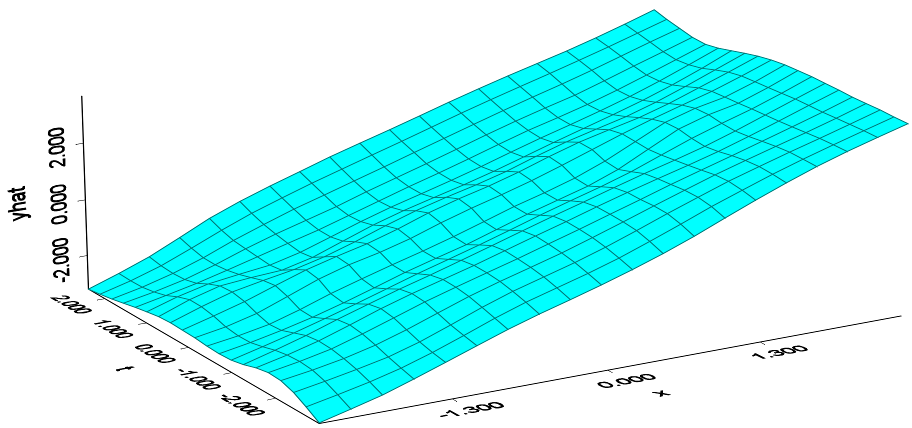
Tabel 1 : Nilai Fungsi $G(p,K)$ Untuk Berbagai Nilai *k* dan *p*

k	***p***	$G(p,\kappa)$	***k***	***p***	$G(p,\kappa)$
1	1	0.2936040	3	1	0.2955018
1	2	0.2893703	3	2	0.2944539
1	3	0.2856304	3	3	0.2896809
1	4	0.2881142	3	4	0.2858854
2	1	0.2832623	4	1	0.2919151
2	2	0.2954097	**4**	**2**	**0.2769566**
2	3	0.2854055	4	3	0.2898823
2	4	0.2876135	4	4	0.2870036

Selanjutnya parameter model diestimasi dengan metode kuadrat terkecil yang diimplementasikan dengan sofware *Splus*, dan diperoleh model sebagai berikut,

$$\hat{y}=-0.336+1.258x -0.336 -1.222t-0.35t^2 +3.343(t+0.546)_+^2-7.369(t+0.069)_+^2 + 6.275(t-0.375)_+^2 -1.639(t-0.891)_+^2 \qquad (13)$$

Pada Gambar 2 disajikan kurva regresi semiparametrik dengan polinomial spline *truncated* berderajat 2 dengan banyak titik knots 4. Pada gambar tersebut sumbu-sumbunya adalah *x*, *t*, dan *y* taksiran yang diperoleh dari persamaan (13). Selanjutnya dengan menjumlahkan taksiran parametrik dan nonparametrik akan diperoleh nilai taksiran respon, yang disajikan pada Gambar 3. Pada gambar tersebut nampak jika nilai aktual bertambah, maka nilai taksiran juga akan naik. Hal ini mengindikasikan model yang diperoleh dapat menjelaskan perubahan pada repon. Jika dievaluasi, model ini mempunyai R^2=0.81 dan MSE = 0.281.



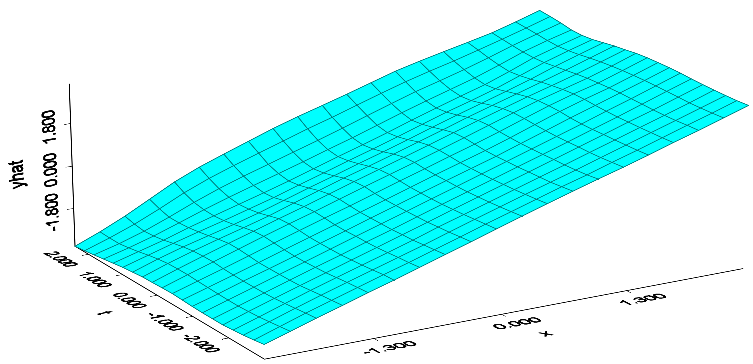
Gambar 2.Kurva regresi semiparametrik dengan polinomial spline *truncated* derajat 2 dan 4 titik knots

Sebagai perbandingan, pada Gambar 3 disajikan kurva regresi semiparametrik dengan kondisi tidak optimum, diambil spline polinomial *truncated* derajat 2 dengan banyak titik knots 2. Gambar yang dihasilkan lebih mulus, cekungan-cekungannya lebih sedikit yang disebabkan banyak titik knots hanya dua, sehingga hanya ada dua perubahan perilaku kurva regresi semiparametrik.

3. Kesimpulan

Untuk model regresi semiparametrik dengan komponen nonparametrik yang didekati dengan fungsi spline polinomial *truncated*, metode kuadrat terkecil dapat diterapkan untuk mengestimasi parameter model tersebut. Namun, estimator yang diperoleh masih tergantung pada berapa derajat polinomial dan banyak titik knots. Penentuan derajat polinomial dan banyak knots yang tepat dapat diselesaikan dengan cara meminimumkan GCV.

ISBN: 978-979-16353-3-2



Gambar 3.Kurva regresi semiparametrik dengan polinomial spline

truncated derajat 2 dan 2 titik knots

Daftar Pustaka

Agustini, T., Budiantara, I.N., Wibowo, W., Ihsan., 2009, "On Choosing of Optimal Bandwidth For Fourier Series Estimator in Multiresponse Nonparametric Regression", IndoMS International Conference On Mathematics and Its Application, Gadjah Mada University, Yogyakarta, 12-13 Oktober 200

Aneiros Pe´ rez, G., Gonza´ lez Manteiga, W., Vieu, P., 2004. Estimation and testing in a partial linear regression model under long-memory dependence. Bernoulli 10 (1), 49–78.

Budiantara, I.N., 2000a, Metode U, GML, CV, dan GCV dalam Regresi Nonparametrik Spline, *Majalah Ilmiah Himpunan Matematika Indonesia (MIHMI),* 6, 41-45

Budiantara, I.N., 2006a, Regresi Nonparametrik Dalam Statistika, *Makalah Pembicara Utama pada Seminar Nasional Matematika, Jurusan Matematika,* FMIPA. Universitas Negeri Makassar, Makassar

Budiantara, I.N., 2006b, Model Spline dengan Knots Optimal, *Jurnal Ilmu Dasar*, FMIPA Universitas Jember, 7, 77-85

Budiantara, I.N., Suryadi, F., Otok, B.W., dan Guritno, S., 2006c, Pemodelan B-pline dan MARS pada Nilai Ujian Masuk Terhadap IPK Mahasiswa Jurusan Disain Komunikasi UK Petra Surabaya, *Jurnal Teknik Industri*, 8, 1-13

Budiantara, I.N., Lestari, B., Islamiyati, A., Wibowo, W., 2009,“Pemilihan Knot Optimal Dalam Estimator Spline Terbobot Pada Regresi Nonparametrik Heteroskedastik Data Longitudinal”, Prosiding Seminar Nasional Statistika XI, Institut Teknologi Sepuluh Nopember, Surabaya

Chen, K., Jin, Z., 2006. Partial linear regression models for clustered data, *J. Amer. Statist. Assoc*. 101, 195–204.

Cox, D.D., dan O’ullivan, F., 1996, Penalized Type Estimator for Partially Linear Model, *The Annal of Statistics, 22, 211-237*

Craven, P., and Wahba, G., 1979, "Smoothing Noisy Data With Spline Functions," *Numerische Mathematik*, 31, 377-403

Delecroix, M., Hristache, M., Patilea, V., 2006, On semiparametric M-estimation in single-index regression, *Journal of Statistical Planning and Inference*, 136, 730 – 769

Engle, R.F., Granger, C.W.J., Rice, J., Weiss, A., 1986. Semiparametric estimates of the relation between weather and electricity sales. *J. Amer. Statist. Assoc*. 81 (394), 310–320.

Eubank, R.L., 1999, *Nonparametric Regression and Spline Smoothing,* New York, Marcel Dekker

Green, P.J. and Silverman, B.W., 1994, *Nonparametric Regression and Generalized Linear Model,* Chapman & Hall, London

Ha¨ rdle, W., 1990. *Applied Nonparametric Regression*, Cambridge University Press, Cambridge.

Ha¨ rdle, W., Liang, H., Gao, J., 2000, *Partially Linear Models,* Physica-Verlag, Wurzburg.

Haryatmi, S. dan Subanar, 2003, Partially Linear Model With Heteroscedastic Errors. Presented at SEAMS – Gadjah Mada University, International Conference on Mathematics and Its Applications, Yogyakarta

Hardle, W., and Muller, M., 1997, Multivariate and Semiparametric Kernel Regression, *Technical Report*, Humboldt Universit, Berlin, Germany

He, X., and Shi, 1996, Bivariate Tensor Product B-spline in a Partly Lienar Model Linear, *Journal of Multivariate Analysis,* 58, 162-181

Heckman, N.E., 1986, Spline smoothing in partly linear model, *Journal of the Royal Statistical Society, Series B*, 48, 244-248

Liang, H., 2006, Estimation in partially linear models and numerical comparisons, *Computational Statistics & Data Analysis*, 50, 675 – 687

Lin, X.H., Carroll, R.J., 2001. Semiparametric regression for clustered data. *Biometrika* 88, 1179–1185

Pitrun, I., King, M.L., Zhang X., 2006, Smoothing spline based tests for non-linearity in a partially linear model, *Journal of Statistical Planning and Inference*, 136, 2446 – 2469

Prada-Sánchez, J.M., Febrero-Bande, M., Cotos-Yáñez, T., González-Manteiga,W., Bermúdez-Cela, J.L., Lucas-Dominguez, T., 2000. Prediction of SO2 pollution incidents near a power station using partially linear models and an historical matrix of predictor-response vectors, *Environmetrics,* 11, 209–225.

Qin, G.Y., Zhu, Z.Y., 2008. Robust estimation in partial linear mixed model for longitudinal data. *Acta Mathematica Scientia* 28B (2), 334–348.

Qin, G., Zhu, Z., Fung, W.K., 2009, Robust estimation of covariance parameters in partial linear model for longitudinal data, *Journal of Statistical Planning and Inference*, 139, 558 – 570

Rice, J., 1986, Convergence rates for partially splines models, *Statistics &Probability Letters,* 4, 203-208

Robinson, P.M., 1988. Root-N-consistent semiparametric regression. *Econometrica* 56 (4), 931–954.

Schmalensee, R., Stoker, T.M., 1999. Household gasoline demand in the United States. Econometrica 67, 645–662.

Shi, P., and Li, G., 1994, On ther Rate Convergence of Minimum L_1-Norm Estimates in Partly Linear Models, *Communication in Statistics, Theory and Methods, 23, 175-196*

Silverman, B. W., 1986, *Density Estimation for Statistics and Data Analysis*, Vol 26 of Monographs on Statistics and Applied Probability, Chapman and Hall, London

Speckman, P., 1988. Kernel smoothing in partial linear models. J. Roy. Statist. Soc. Ser. B 50 (3), 413–436.

Wahba, G., 1990, *Spline Model for Observational Data*, SIAM, XII, Philadelphia

Wang, Y., 1998, Spline Smoothing Models with Correlated Error, *Journal of the Royal Statistical Sociaty, Series B, 5r0, 341-348*

Xia, Y, Härdle. W, 2006, Semi-parametric estimation of partially linear single-index models, Journal of Multivariate Analysis, 97, 1162 – 1184

You, J., Zhou, X., 2009, Partially linear models and polynomial spline approximations for the analysis of unbalanced panel data*, Journal of Statistical Planning and Inference*, 139, 679 -- 695

Yu, Y. and Ruppert, D., 2002, Penalized spline estimation for partially linear single-index models. J. Amer. Statist. Assoc. 97, 1042-1054.

Yu, Y. and Ruppert, D., 2004, Root-N Consistency Of Penalized Spline Estimator For Partially Linear Single-Index Models Under General Euclidean Space, *Statistica Sinica*, 14, 449-455

Zeger, S.L., Diggle, P.J., 1994. Semiparametric models for longitudinal data with application to CD4 cell numbers in HIV seroconverters. *Biometrics,* 50, 689–699.