

# Weight Estimation Using Generalized Moving Average

Jerry Dwi Trijoyo Purnomo<sup>1</sup>, I Nyoman Budiantara<sup>1</sup>, dan Kartika Fitriasari<sup>1</sup>

**Abstract** — Estimation of regression curve usually conducted using three methods; parametric method, non-parametric method, and semi-parametric method. Non-parametric method has several techniques, which are histogram, kernel, and spline. From various types of spline techniques, weighted partial spline is developed to solve heterokedasticity problem, this is due to the inability of original partial spline model in handling the heterokedasticity problem. Different techniques are used in choosing the weighted criteria, one of the technique is Generalized Moving Average (GMA).

Study about the amount of electricity power loss in PT. PLN East Java Province, North Surabaya Region, resulted that there was a tendency of heterogeneous variance. Using weighted partial spline model with GMA method give better result than original partial spline model. This finding indicates the model of weighted partial spline using GMA method is better than original partial spline model in explaining the heterogeneity of variance.

**Keywords** — Weighted partial spline, Generalized Moving Average, Original spline

## I. PENDAHULUAN

Estimasi kurva regresi umumnya dilakukan dengan pendekatan parametrik yang mulai diperkenalkan oleh Laplace sejak abad XVIII dan juga Boscovich pada tahun 1757. Dalam regresi parametrik diasumsikan bahwa bentuk kurva regresi  $f$  diketahui. Pembuatan asumsi tersebut berdasarkan pada teori, pengalaman masa lalu atau tersedianya sumber-sumber lain yang dapat memberi pengetahuan atau informasi yang terperinci.

Estimasi dapat juga dilakukan berdasarkan pendekatan yang tidak terikat dengan asumsi bentuk kurva regresi tertentu, yang memberikan fleksibilitas yang lebih besar dari kurva regresi. Metode pendekatan seperti ini dinamakan pendekatan nonparametrik yang mulai dikenal sejak abad XIX. Ada beberapa teknik untuk mengestimasi dalam regresi nonparametrik, antara lain histogram, estimator Kernel, Spline, dan lain-lain.

Masalah yang sering muncul dalam regresi adalah tidak semua variabel penjelas dapat didekati dengan pendekatan parametrik, karena tidak adanya informasi tentang bentuk hubungan variabel penjelas tersebut dengan variabel responnya, sehingga harus digunakan pendekatan nonparametrik. Dengan menggabungkan dua pendekatan tersebut dalam suatu pendekatan regresi akan didapatkan suatu model semiparametrik. Estimasi mo-

del semiparametrik ekuivalen dengan mengestimasi parameter-parameter pada komponen parametrik dan estimasi kurva pada komponen nonparametrik.

Dalam penelitian Dewayani [1], terdapat enam variabel prediktor, yaitu jumlah gangguan jenis benda ( $X_1$ ), jumlah gangguan jenis alam ( $X_2$ ), jumlah gangguan jenis unit pembangkit ( $X_3$ ), jumlah gangguan jenis peralatan ( $X_4$ ), jumlah pekerjaan oleh pihak PT PLN ( $X_5$ ), jumlah gangguan jenis sebab lain ( $X_6$ ), serta variabel respon, jumlah energi listrik yang hilang di Surabaya Utara. Dalam penelitian ini ada kecenderungan varian yang tidak homogen. Hal ini terlihat dari visualisasi data antara variabel-variabel prediktor tersebut dengan variabel respon. Adanya kecenderungan adanya varian yang tidak homogen ini menyebabkan hasil yang diperoleh kurang dapat menjelaskan hubungan antara variabel respon dan prediktor di atas. Karena itu digunakan pembobot untuk mengatasi ketidakseimbangan varian, karena spline original yang digunakan kurang dapat menjelaskan adanya ketidakseimbangan varian.

Model semiparametrik diperkenalkan oleh He dan Shi [2], Engle, Granger, Rice, dan Weiss [3], Green, Jennison, dan Scheult [4], serta Heckman [5]. Model ini dirumuskan sebagai:

$$y_i = x_i' \gamma + f(t_i) + \varepsilon_i, \quad i=1,2,\dots,n \tag{1}$$

dengan  $x_i' = (x_{i1}, \dots, x_{ip})$  dan  $t_i, i=1,2,\dots,n$  merupakan variabel-variabel prediktor. Vektor  $\gamma = (\gamma_1, \dots, \gamma_p)' \in R^p$  tidak diketahui dan  $f$  diasumsikan merupakan anggota anggota ruang Sobolev  $W_2^m[0,1] = \{f \mid f^{(k)}, k = 0, 1, \dots, m-1 \text{ kontinu pada } [0,1], \text{ dan } \int_0^1 [f^{(m)}(t)]^2 dt < \infty\}$ . Residual  $\varepsilon_i$  berdistribusi

independen dengan mean nol dan variansi  $\sigma^2$ . Estimator  $f$  diperoleh dari meminimumkan *Penalized Least Square* (PLS):

$$l_\lambda(f) = R(f) + \lambda J(f), \quad f \in W_2^m[0,1] \tag{2}$$

Fungsional  $l_\lambda(f)$  memuat tiga komponen, yaitu komponen likelihood  $R(f)$ , fungsional penalti  $J(f)$  dan parameter penghalus  $\lambda$ .

Estimator tipe PLS ini dikembangkan oleh Heckman [5], Eubank [6], Wahba [7], serta Chen dan Shiau [8] untuk estimator spline parsial original, dengan mengambil kesamaan-kesamaan:

$$R(f) = n^{-1} \sum_{i=1}^n (y_i - x_i' \gamma - f(t_i))^2 \tag{3}$$

$$J(f) = \int_0^1 [f^{(m)}(t)]^2 dt \tag{4}$$

Bentuk estimator  $\gamma_\lambda$  dan  $f_\lambda$  diperoleh dengan meminimumkan PLS:

Naskah diterima pada tanggal 3 September 2008, selesai revisi pada tanggal 22 Desember 2008

<sup>1</sup> Jerry Dwi T.P., I Nyoman B., dan Kartika F. adalah dosen jurusan Statistika, FMIPA, Institut Teknologi Sepuluh Nopember, Surabaya, INDONESIA. E-mail : [jerry.purnomo@yahoo.com](mailto:jerry.purnomo@yahoo.com)

$$I_\lambda(f) = n^{-1} \sum_{i=1}^n (y_i - x_i' - f(t_i))^2 + \lambda \int_0^1 [f^{(m)}(t)]^2 dt \quad (5)$$

Pada sisi lain, Shi dan Li [9] mengembangkan model semiparametrik untuk estimator spline parsial tipe-norm  $L_1$ . Estimator  $\gamma_\lambda$  dan  $f_\lambda$  diperoleh dengan meminimumkan *Penalized Least Square* (PLS) dan mengganti persamaan (3):

$$R(f) = n^{-1} \sum_{i=1}^n |y_i - x_i' \gamma - f_\lambda(t_i)| \quad (6)$$

Lebih lanjut He dan Shi [9] mengusulkan penggunaan estimator spline parsial tipe-M (median) untuk pendekatan mean respon regresi semiparametrik. Estimator diperoleh dengan meminimumkan PLS dan mengganti persamaan (3) :

$$R(f) = n^{-1} \sum_{i=1}^n \rho(y_i - x_i' \gamma - f_\lambda(t_i)) \quad (7)$$

$\rho$  fungsi konveks.

Beberapa macam tipe spline yang dikembangkan para peneliti di atas, dikembangkan model spline parsial terbobot untuk mengatasi ketidaksamaan variansi, dikarenakan model spline parsial kurang sesuai untuk menangani ketidaksamaan variansi regresi semiparametrik. Model ini dikembangkan oleh Budiantara [10] serta Subanar dan Budiantara [11]. Bentuk estimator diperoleh dengan meminimumkan *Penalized Least Square* Terbobot (PLST):

$$I_\lambda(f) = n^{-1} \sum_{i=1}^n w_i (y_i - x_i' - f(t_i))^2 + \lambda \int_0^1 [f^{(m)}(t)]^2 dt \quad (8)$$

untuk setiap  $f$  anggota ruang sobolev  $w_2^m[0,1]$ . Bentuk estimator spline parsial terbobot dipengaruhi oleh parameter penghalus  $\lambda$ . Untuk  $\lambda \rightarrow \infty$ , estimator spline parsial terbobot merupakan estimator polinomial [10]. Estimator komponen parametrik merupakan estimator yang konsisten dan berdistribusi normal asimtotik [11].

## II. SPLINE DALAM REGRESI NONPARAMETRIK DAN ESTIMASI BOBOT

Estimasi kurva regresi umumnya dilakukan dengan tiga pendekatan, yaitu pendekatan parametrik dengan asumsi bentuk kurva regresi diketahui, pendekatan nonparametrik, dengan asumsi pendekatan kurva regresi tidak diketahui, serta pendekatan semiparametrik yang merupakan gabungan antara pendekatan parametrik dan nonparametrik.

### A. Spline Dalam Regresi Nonparametrik

Fungsi spline berorde  $(m-1)$  dengan titik-titik knots  $S_1, S_2, \dots, S_k$  adalah sebarang fungsi yang dapat disajikan dalam bentuk [5]:

$$S(t) = \sum_{i=1}^{k-1} \alpha_i t^i + \sum_{j=1}^h \delta_j (t - S_j)_+^{k-1} \quad (8)$$

$$(t - S_j)_+^{k-1}, t \geq S_j$$

dengan  $(t - S_j)_+^{k-1} =$

$$0, t < S_j$$

$$j = 1, 2, \dots, k$$

$\alpha$  dan  $\delta$  adalah konstanta real dan  $S_1, S_2, \dots, S_k$  adalah titik-titik knots.

### B. Estimasi Bobot

Pada dasarnya ada dua macam cara mendapatkan estimasi bobot. Pertama dengan coba-coba (*trial error*). Sistem *trial error* ini adalah mendapatkan bobot dengan menggunakan fungsi prediktornya (Montgomery dan Peck [12]). Namun demikian sulit untuk menemukan bobot yang optimal karena banyaknya kombinasi dari fungsi prediktor tersebut. Cara kedua adalah dengan estimasi menggunakan *Moving Average* Silverman [13]. Metode estimasi dengan *Moving Average* ini salah satunya adalah *Generalized Moving Average* (GMA), dengan mengambil persamaan:

$$\hat{W}_i^{-1} = (n_i - m_i + 1)^{-1} \sum_{i=m_i}^{n_i} r_i^{*2}$$

dengan:  $m_i = \max(1, i - k)$

$$n_i = \min(n, i + k)$$

$$r_i^* = \frac{w_i^{1/2} \{y_i - f_\lambda(t_i)\}}{\hat{\sigma} \{1 - n^{-1} \text{tr}A(\lambda)\}^{1/2}}$$

$$\hat{\sigma}^2 = \frac{\sum_{i=1}^n w_i \{y_i - f_\lambda(t_i)\}^2}{n - \text{tr}A(\lambda)}$$

## III. DATA PENELITIAN

Regresi semiparametrik digunakan untuk menganalisis data, dimana terjadi perubahan perilaku dari data tersebut pada interval-interval tertentu. Salah satu metode regresi semiparametrik yang digunakan adalah spline.

### A. Spline Parsial Original

Sebagai visualisasi spline parsial original digunakan data penelitian Dewayani [1] tentang jumlah energi listrik yang hilang di PT PLN distribusi Jawa Timur Wilayah Surabaya Utara. Dalam penelitian Dewayani [1] ini terdapat dua variabel prediktor, yaitu gangguan jenis benda ( $x_1$ ), dan jenis gangguan jenis unit pembangkit ( $x_2$ ), serta variabel respon jumlah energi listrik yang hilang. Visualisasi spline parsial jumlah energi listrik yang hilang dapat dilihat pada Gambar 1 dan Gambar 2.

Gambar 1 memperlihatkan hubungan antara jumlah energi listrik yang hilang di PT PLN dengan gangguan jenis benda ( $x_1$ ) memperlihatkan adanya perubahan pola perilaku data disekitar titik 9 dan 14,1. Berdasarkan Tabel 1, didapatkan bahwa hubungan antara keduanya dapat dijelaskan dengan model spline linear.

Gambar 2 memperlihatkan estimasi spline original hubungan antara jumlah energi listrik yang hilang dengan gangguan jenis benda ( $x_1$ ) adalah spline linear dengan dua titik knots. Gambar 3 menunjukkan hubungan antara jumlah energi listrik yang hilang dengan gangguan jenis unit pembangkit ( $x_2$ ). Dari sini terlihat hubungan keduanya adalah linear. Sedangkan Gambar 4 menunjukkan regresi linear jumlah energi listrik yang hilang dengan gangguan jenis unit pembangkit ( $x_2$ ). Dari keempat gambar tersebut terlihat adanya kasus heterokedastisitas, karena makin besar nilai  $x$ , variansi juga makin besar. Model spline parsial original untuk data ini adalah:

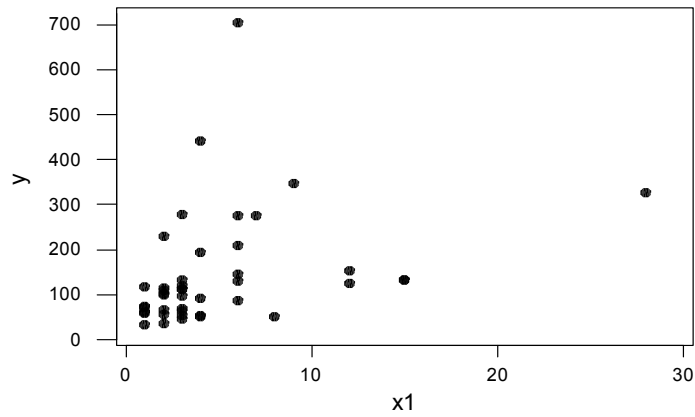
$$y = 1,09 + 23,25x_1 - 54,73(x_1 - 9)_+^1 + 50,02(x_1 - 14,1)_+^1 + 4,73x_2 + \epsilon$$

Nilai koefisien determinasi untuk model ini sebesar 66,39%.

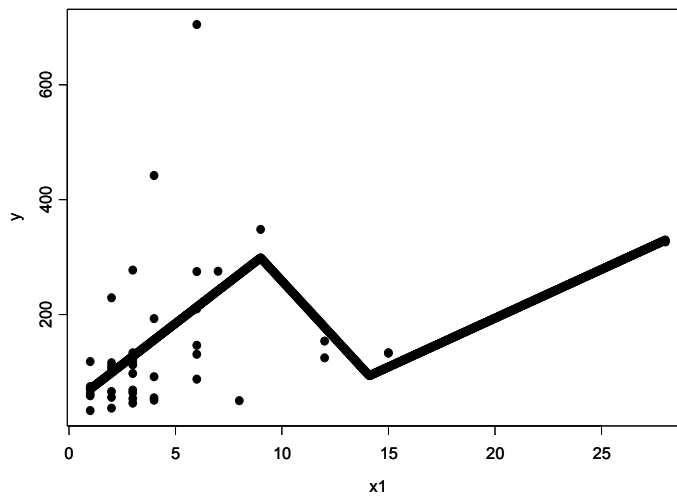
Berdasarkan model spline parsial original, memperlihatkan kasus heterokedastisitas (varian residual tidak konstan). Ini terlihat dari plot antara residual dengan  $\hat{y}$  (Gambar 5) yang bentuknya tidak random (*horizontal band*) melainkan berbentuk corong. Hal ini mengindikasikan bahwa model spline parsial original kurang dapat menjelaskan adanya kasus heterokedastisitas.

TABEL 1  
SKOR GCV SPLINE PARSIAL ORIGINAL DENGAN SATU TITIK KNOTS

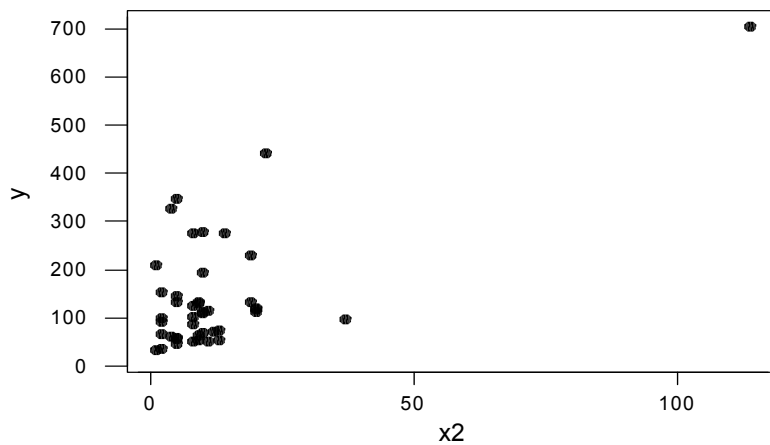
No	Orde	Titik Knot	GCV
1	2	9,3 14,1	6893,625
2	2	9,2 14,1	6892,278
3	2	9,1 14,1	6891,460
<b>4</b>	<b>2</b>	<b>9,0 14,1</b>	<b>6891,108</b>
5	2	9,0 14,2	6891,621
6	2	9,0 14,3	6892,763



Gambar 1 . Plot antara y dan x<sub>1</sub>



Gambar 2 Plot antara y dan x<sub>1</sub> dan estimasi spline original

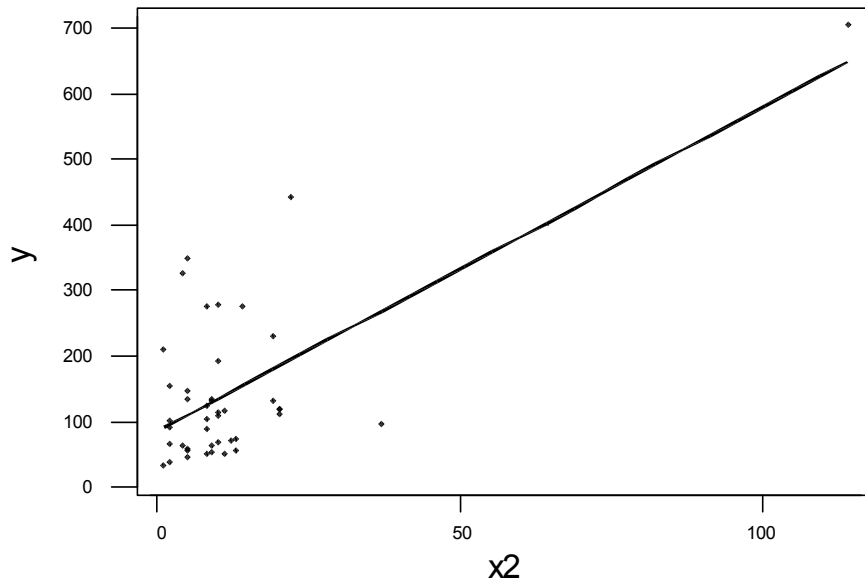


Gambar 3. Plot antara y dan x<sub>2</sub>

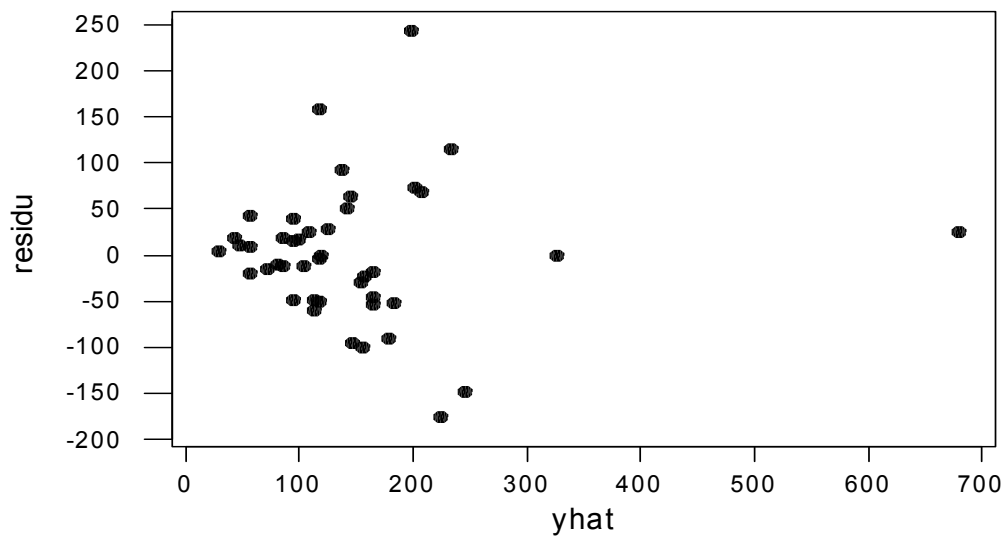
### Regression Plot

$$y = 85,4292 + 4,93934 x_2$$

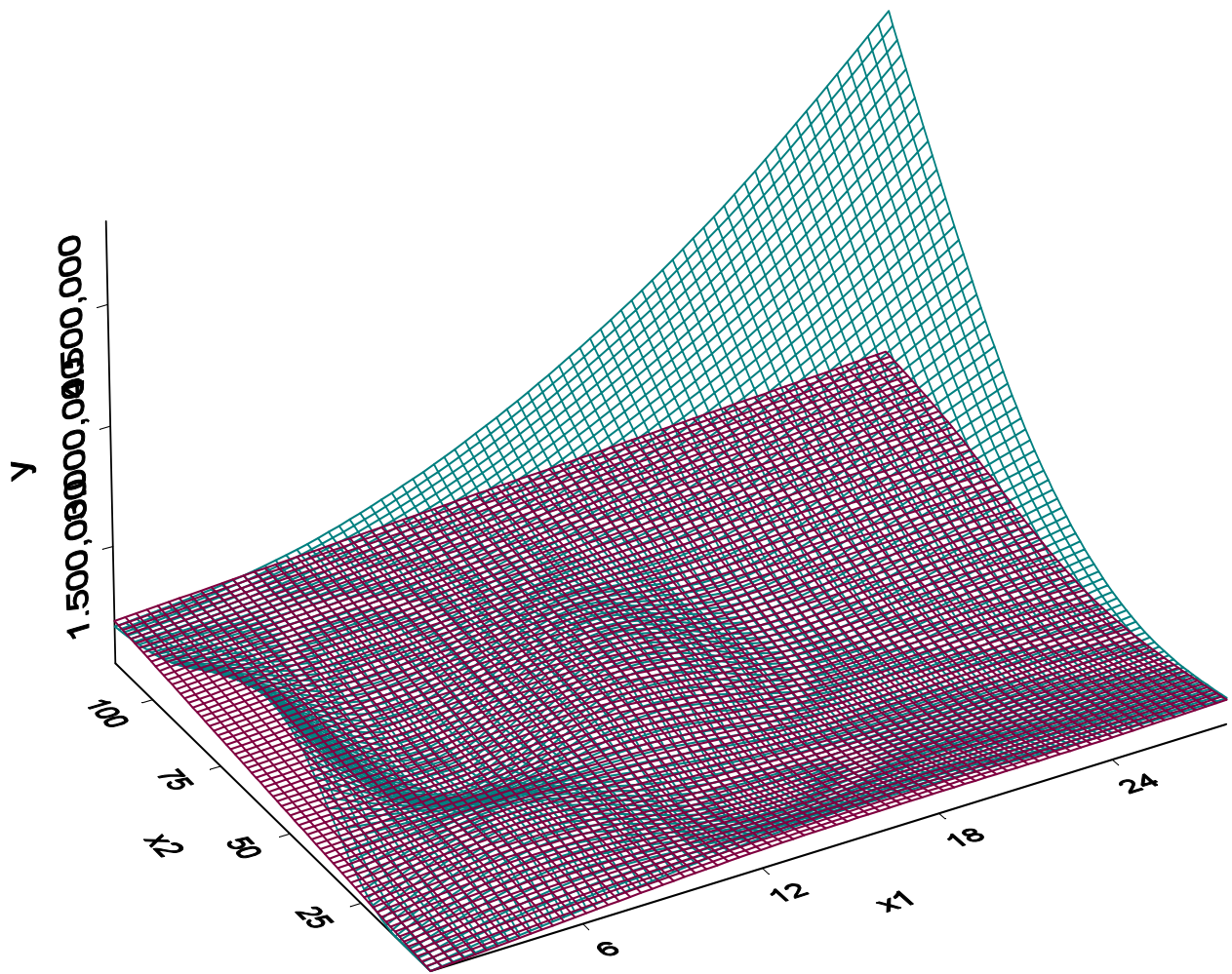
S = 94,8883   R-Sq = 45,7 %   R-Sq(adj) = 44,4 %



Gambar 4. Plot y dan  $x_2$  dan regresi linear



Gambar 5 Plot antara residual dan  $\hat{y}$



Gambar 6 Plot tiga dimensi spline parsial original antara respon (y = hijau) dan nilai taksiran respon ( $\hat{y}$  = merah)

**B. Spline Parsial Terbobot**

Pada analisis di atas terlihat bahwa spline parsial original kurang mampu menjelaskan adanya heterokedastisitas. Untuk mengatasi adanya heterokedastisitas diberikan suatu bobot yang sesuai. Salah satu metode untuk mendapatkan bobot adalah dengan *Generalized Moving Average (GMA)* [13].

TABEL 2  
SKOR GCV SPLINE PARSIAL TERBOBOT DENGAN SATU KNOTS

No	Orde	Titik Knots	GCV
1	3	6	10109,50
2	3	8	9505,72
3	3	9	9161,36
4	3	10	8842,28
5	3	12	8468,83
6	3	14	8338,75
7	3	15	7322,85

Tabel 2 memperlihatkan model spline kuadratik terbobot dengan titik knots 15 mempunyai nilai GCV terkecil, yaitu 7322,85. Estimasi model spline parsial terbobot diberikan oleh:

$$\hat{y} = \gamma_0 + \gamma_1 x_1 + \gamma_2 x_1^2 + \gamma_3 (x_1 - S)_+^2 + \gamma_4 x_2$$

Berdasarkan analisis diagnostik residual didapatkan model spline kuadratik terbobot mempunyai variansi

data yang konstan dan berdistribusi normal. Selanjutnya dilakukan uji hipotesis koefisien-koefisien regresi. Pertama dilakukan uji hipotesis serentak.

$$H_0 : \gamma_0 = \gamma_1 = \gamma_2 = \gamma_3 = \gamma_4 = 0$$

$$H_1 : \text{paling sedikit ada satu } \gamma_i \neq 0$$

Ringkasan analisis variansi model spline terbobot diberikan dalam Tabel 3.

Dengan tingkat signifikansi  $\alpha = 0,05$  diperoleh kesimpulan bahwa paling sedikit ada satu  $\gamma_i \neq 0, i = 0, 1, 2, 3, 4$ . Selanjutnya dilakukan uji koefisien-koefisien regresi, terutama fungsi *truncated* ( $\gamma_3$ ) yang memberikan pengaruh signifikan terhadap model.

$$H_0 : \gamma_0 = 0, H_1 : \gamma_0 \neq 0$$

$$H_0 : \gamma_1 = 0, H_1 : \gamma_1 \neq 0$$

$$H_0 : \gamma_2 = 0, H_1 : \gamma_2 \neq 0$$

$$H_0 : \gamma_3 = 0, H_1 : \gamma_3 \neq 0$$

$$H_0 : \gamma_4 = 0, H_1 : \gamma_4 \neq 0$$

TABEL 3  
ANALISIS VARIANSI MODEL SPLINE PARSIAL TERBOBOT

Sumber Variasi	Derajat Bebas	Jumlah Kuadrat (JK)	Rata-rata JK	F
Regresi	4	862401	215600	29,31315
Residual	38	279492	7355	
Total	42	1141894		

Nilai F-tabel: 2,618988

TABEL 4  
ESTIMASI MODEL SPLINE PARSIAL TERBOBOT DENGAN METODE GMA

Koefisien	Estimasi	St Dev	t-hitung
$\gamma_0$	-35	25	-1
$\gamma_1$	45	10	4
$\gamma_2$	-2	0	-3
$\gamma_3$	6	1	3
$\gamma_4$	4	0	8

Nilai t-tabel : 2,024394

Tabel 4 memperlihatkan  $\gamma_0$  saja yang tidak signifikan terhadap model, sehingga  $\gamma_0$  tidak dimasukkan ke dalam model.

TABEL 5  
ESTIMASI MODEL PARSIAL TERBOBOT DENGAN METODE GMA

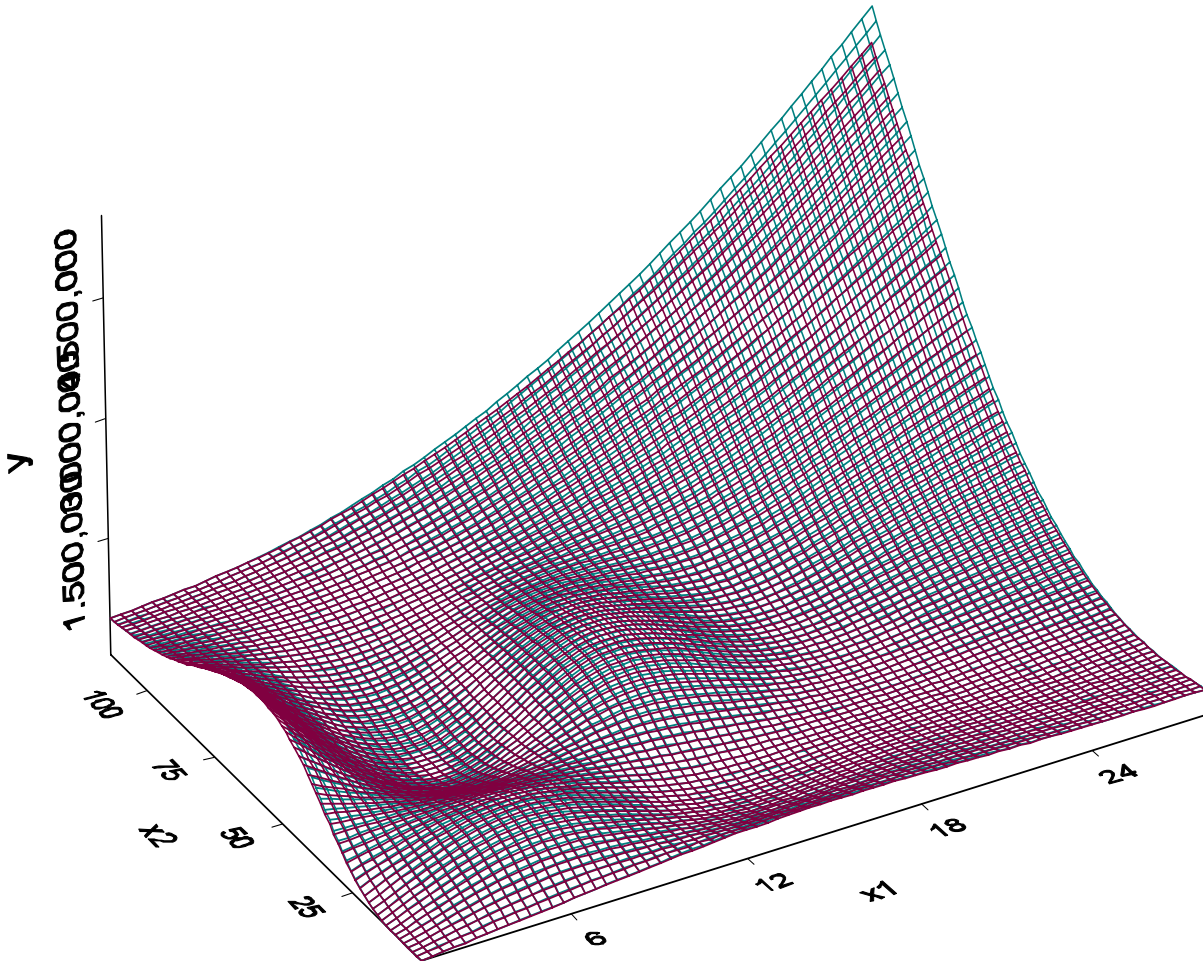
Koefisien	Estimasi	St Dev	t-hitung
$\gamma_1$	68,776	22	3
$\gamma_2$	-17	6	-2
$\gamma_3$	17,038	6	2
$\gamma_4$	10,135	2	4

Nilai t-tabel: 2,022691

Tabel 5 di atas memperlihatkan model spline parsial terbobot dengan menggunakan metode GMA adalah:

$$\hat{y} = 68,78x_1 - 17x_1^2 + 17,04(x_1 - 15)_+^2 + 10,14x_2$$

Nilai koefisien determinasi untuk model ini sebesar 71,92%.



Gambar 7. Plot tiga dimensi spline parsial terbobot dengan metode GMA. (y =hijau ;  $\hat{y}$  =merah)

Berdasarkan analisis di atas dapat disimpulkan bahwa spline parsial dengan bobot lebih baik dibandingkan spline parsial original. Hal ini terlihat dari nilai  $R^2$  spline dengan bobot yang lebih besar dibandingkan spline tanpa bobot, disamping gambar tiga dimensi antara  $\hat{y}$  (merah) dan y (hijau) yang menunjukkan bahwa untuk spline dengan bobot nilai  $\hat{y}$  merupakan estimasi yang baik untuk y, karena plot antara keduanya berimpit.

IV. KESIMPULAN

Spline parsial original kurang sesuai untuk permasalahan ketidaksamaan varian (heterokedastisitas). Model spline dengan bobot layak dipertimbangkan se-

bagai suatu model pendekatan, karena memberikan hasil yang lebih dibandingkan spline parsial original. Ada dua cara dalam penentuan bobot, pertama dengan *trial error*, dan kedua menggunakan estimasi bobot dengan *Moving Average*. Metode estimasi dengan *Moving Average* ini salah satunya adalah *Generalized Moving Average* (GMA). Model spline terbobot dengan menggunakan GMA adalah:

$$\hat{y} = 68,78x_1 - 17x_1^2 + 17,04(x_1 - 2)_+^2 + 10,14x_2$$

Nilai koefisien determinasi untuk model ini sebesar 71,92%.

## V. DAFTAR PUSTAKA

- [1] Dewayani, I. "Penerapan Model Nonparametrik Dengan Metode Spline Pada Jumlah Energi Listrik Yang Hilang di PT PLN Distribusi Jawa Timur Wilayah Surabaya Utara", *Tugas Akhir*, ITS, Surabaya. 2004.
- [2] He, X. dan Shi, P. "Bivariate Tensor Product B-Spline in a Partly Linear Models", *Journal of Multivariate Analysis*, 58, 162-181. 1996.
- [3] Engle, R.L., Granger, C., Rice, J. and Weiss, A. "Semiparametric Estimates of Relation Between Weather and Electricity Sales", *Journal of The American Statistical Association*, 81, 310-320. 1986.
- [4] Green, P., Jennison, C., Seheult, A. Analysis of Field Experiments by Least Square Smoothing, *Journal of The Royal Statistical Society*, Ser. B, 47, 299-314. 1985.
- [5] Heckman, N. "Spline Smoothing in a Partly Linear Models", *Journal of The Royal Statistical Society*, ser B, 48, 244-248. 1986.
- [6] Eubank, R.L. A Note on Smoothness Priors and Nonlinear Regression., *Journal of the American Statistical Association*, 81, 514-517. 1986.
- [7] Wahba, G. "Spline Models for Observasional Data", *SIAM*, Pennsylvania. 1990.
- [8] Chen, H. dan Shiau, J.J.H. "Data Driven Efficient Estimators for a Partially Linear Model". *The Annals of Statistics*, 22, 211-237. 1994.
- [9] Shi, P., dan Li, G. On the Rate Convergence of "Minimum  $L_1$ -Norm" Estimates in a partly Linear Model, *Communication in Statistics, Theory and Methods*, 23, 175-196. 1994.
- [10] Budiantara, I.N. "Estimator Spline Terbobot Dalam Regresi Semiparametrik", *Majalah Ilmu Pengetahuan dan Teknologi*, 10, 103-109. 1999.
- [11] Subanar dan Budiantara, I.N. "Weighted Spline Estimator in a Partially Linear Models", *Proceeding of the SEAMS-GMU International Conference 1999 on Mathematics and Its Applications*, 61-70. 1999.
- [12] Montgomery, D.C and Peck, E.A. *Introduction to Linear Regression Analysis*, New York. John Wiley and Sons. 1982.
- [13] Silverman, B.W. "Some Aspect of The Spline Smoothing Approach to Nonparametric Regression Curve Fitting (With Discussion)", *Journal of The Royal Statistical Society*, ser B, 47, 1-52. 1985.