

ANALISIS REGRESI NONPARAMETRIK DENGAN TEKNIK *SMOOTHING*

Khoirin Nisa, Netti Herawati, Eri Setiawan
Department of Mathematics, University of Lampung
Jl. Soemantri Brojonegoro No. 1 Bandar Lampung, Indonesia
Penulis Korespondensi : khoirin.nisa@fmipa.unila.ac.id

Abstract

Smoothing has become a common technique in non-parametric methods which is used to estimate a function. This paper discuss three smoothing techniques for nonparametric regression that are often studied in literatures, i.e. kernel smoothing, smoothing splines, and the locally estimated scatterplot smoothing methods. The three methods will be examined empirically by using four generated data. Using the generated data, the result shows that the smoothing spline gives better performance than the other two methods.

Keywords: Kernel, smoothing spline, loess.

PENDAHULUAN

Analisis regresi adalah salah satu metode statistika yang dapat digunakan untuk menyelidiki hubungan antara dua peubah atau lebih. Metode ini berkaitan dengan masalah pendugaan atau peramalan nilai peubah tak bebas Y berdasarkan nilai-nilai peubah bebas X . Dalam analisis regresi sederhana, setiap anggota populasi dapat dinyatakan dalam koordinat (x,y) sehingga suatu sampel acak berukuran n dari populasi tersebut dapat dilambangkan sebagai $\{(x_i, y_i); i = 1, 2, \dots, n\}$. Jika sampel acak tersebut ditebarkan atau diplotkan dalam koordinat kartesius maka akan diperoleh apa yang disebut diagram pencar (*scatterplot*). Dengan mengamati diagram pencar dari data dapat dilihat apakah data mengikuti suatu pola tertentu, misalnya jika titik-titik dalam diagram pencar mengikuti pola garis lurus, hal ini menunjukkan bahwa kedua peubah tersebut berhubungan secara linier, dan jika titik-titik dalam diagram pencar menunjukkan pola kuadratik maka ini menunjukkan bahwa kedua peubah tersebut berhubungan secara kuadratik, dst.

Dalam regresi parametrik dengan bentuk $y = f(x) + \varepsilon$, f adalah suatu fungsi yang diketahui dan kita harus menentukan bentuk yang tepat bagi f . Sedangkan dalam regresi nonparametrik, f adalah fungsi pemulus yang tidak diketahui dan tidak ditentukan di awal. Mirip dengan regresi parametrik, pembobotan terhadap pengamatan y digunakan untuk memperoleh nilai dugaan. Namun regresi nonparametrik tidak menggunakan bobot yang sama seperti dalam kuadrat terkecil biasa (*ordinary least square*) atau bobot yang proporsional dengan kebalikan dari ragam seperti yang sering terjadi dalam kuadrat terkecil terboboti (*weighted least square*), pemikiran rasional yang berbeda menentukan pemilihan bobot dalam regresi nonparametrik. Dalam kasus regressor tunggal, pengamatan dengan informasi terbesar mengenai $f(x_0)$ adalah pengamatan x_i yang berada pada lokasi terdekat dengan x_0 . Oleh karena itu, suatu fungsi jarak antara lokasi x_i dengan x_0 menentukan bobot bagi y_i . Titik yang lebih dekat dengan x_0 menerima bobot yang lebih besar dari titik-titik yang lebih jauh dari x_0 . Seringkali, titik terjauh dari x_0 menerima bobot sedikit atau tidak sama sekali.

Pemulusan (*smoothing*) merupakan salah satu pendekatan yang digunakan dalam analisis regresi nonparametrik. Tujuan dari pemulusan adalah untuk memperkecil keragaman dari data yang tidak memiliki pengaruh sehingga ciri-ciri dari data akan tampak lebih jelas. Pemulusan telah menjadi teknik umum di dalam metode-metode non parametrik yang digunakan untuk menduga fungsi. Tiga teknik pemulusan yang cukup populer adalah *kernel smoothing*, *smoothing splines*, dan metode *locally estimated scatterplot smoothing* (Loess). Masing-masing teknik memiliki cara yang berbeda dalam menduga kurva regresi.

Makalah ini membahas hasil studi pendahuluan yang kami lakukan untuk menyelidiki dan membandingkan ketiga metode serta aplikasinya dalam berbagai bidang. Dalam studi pendahuluan ini kami menyajikan ketiga teknik di atas untuk menduga kurva regresi nonparametrik dengan menggunakan data bangkitan. Untuk melihat hasil dugaan kurva regresi dengan berbagai pola, kami membangkitkan data dengan menggunakan empat fungsi regresi berbeda yang kami adopsi dari studi yang

dilakukan oleh Berry *et.al* (2002), Lee (2003), Aydin & Omay (2006), dan Faraway (2006).

REGRESI NONPARAMETRIK

Misalkan terdapat n pasang pengukuran $\{(x_i, y_i); i = 1, 2, \dots, n\}$ yang memenuhi model berikut;

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

dimana galat ε_i berdistribusi normal dan saling bebas dengan mean 0 dan varians σ^2 yaitu, $\varepsilon_i \sim \text{iid } N(0, \sigma^2)$, dan $f(x)$ adalah fungsi yang tidak diketahui. Tujuan dasar dari regresi nonparametrik adalah menduga fungsi f yang tidak diketahui (dari semua kemungkinan fungsi f dengan turunan pertama dan kedua kontinu) dalam model (1). Dalam regresi nonparametrik, f adalah suatu fungsi pemulus yang tidak diketahui dan tidak ditentukan oleh pembuat model. Suatu teknik berbasis data akan menentukan bentuk kurva fungsional antara X dan Y .

Pada bagian ini akan disajikan tinjauan teoritis tiga metode pemulusan regresi nonparametrik seperti yang telah kami sebutkan di atas, yaitu : *kernel smoothing*, *smoothing spline* dan *loess*.

1. Kernel Smoothing

Tujuan dari regresi *kernel smoothing* adalah untuk mendapatkan dan menggunakan bobot yang sesuai $w_{ij(\text{ker})}$ untuk menghasilkan nilai dugaan melalui:

$$\hat{y}_{(\text{ker})} = \sum_{j=1}^n w_{ij(\text{ker})} y_j \quad (2)$$

Perhatikan bahwa persamaan (2) juga dapat dianggap atau sama saja dengan menghitung $\hat{f}(x_i)$ untuk $i = 1, 2, \dots, n$. Setiap titik dari n data diberi bobot $w_{ij(\text{ker})}$, $j = 1, 2, \dots, n$ untuk setiap titik dugaan x_i (atau bobot $w_{ij(\text{ker})}$ berbeda untuk suatu titik prediksi x_0).

Misal K adalah kernel simetrik yang berpusat di 0, sebuah kernel esensinya adalah suatu pembobot, dalam kasus ini ia merupakan pembobot dalam keanggotaan lingkungan (*neighborhood membership*). Kernel yang biasa dipilih adalah kernel sebaran normal standar $K(z) = \frac{1}{\sqrt{2\pi}} \exp\{-z^2/2\}$. Maka pembobot bagi y_i diberikan oleh:

$$W_{ij(\text{ker})} = \frac{K\left(\frac{x-x_i}{h}\right)}{\sum_{j=1}^n K\left(\frac{x-x_j}{h}\right)}$$

Sehingga persamaan (2) dapat ditulis menjadi :

$$\hat{y}_{(\text{ker})} = \sum_{j=1}^n \frac{K\left(\frac{x-x_i}{h}\right)}{\sum_{j=1}^n K\left(\frac{x-x_j}{h}\right)} y_j$$

Parameter pemulus h disebut sebagai lebar jendela. Catat bahwa untuk kernel-kernel yang biasa seperti kernel normal, semua titik pengamatan digunakan untuk menghitung pemulusan pada setiap titik, tetapi titik-titik data yang terletak sangat jauh mendapatkan pembobot yang sangat kecil. Nilai lebar jendela yang besar akan memberikan hasil yang sangat mulus karena pembobotan pada titik-titik data hanya berubah sedikit sepanjang interval pemulusan. Sedangkan nilai lebar jendela yang kecil akan memberikan hasil yang lebih kasar (*wiggles*) karena titik-titik data yang berdekatan mendapatkan bobot yang jauh lebih besar dibandingkan dengan titik yang jauh (Rizzo, 2008). Kajian lengkap tentang regresi *kernel smoothing* dapat dilihat dalam berbagai literatur misalnya Gijbels *et al.* (1999), Aydin (2007), Sheng (2009), dan sebagainya.

2. *Smoothing Spline*

Regresi *smoothing spline* menduga fungsi regresi dengan meminimumkan kriteria kuadrat terkecil tersanksi (*penalized least squares*). Suatu besaran kasar (*roughness penalty*) dikalikan dengan parameter pemulus λ positif ditambahkan ke dalam jumlah kuadrat sisaan dalam regresi *smoothing spline*. Sehingga pendugaan terhadap fungsi f

bergantung kepada parameter pemulus λ yang nilainya secara umum diperoleh dari data.

Penduga *smoothing spline* kubik \hat{f}_λ untuk f didefinisikan sebagai sesuatu yang meminimumkan fungsi *Penalized Least Squares* (PLS) berikut :

$$Q_\lambda(s) = \sum_{i=1}^n (y_i - f(x_i))^2 + \lambda \int (f'(x))^2 dx \quad (3)$$

Bagian pertama dari persamaan (3) merupakan jumlah kuadrat sisaan atau fungsi jarak antara data dan dugaan, dan bagian kedua merupakan *roughness penalty*, yaitu ukuran kemulusan kurva dalam memetakan data, dan $\lambda > 0$ adalah parameter pemulus, yaitu pengontrol keseimbangan antara kecocokan terhadap data (*Goodness-of-Fit*) dan kemulusan kurva (*penalty*). Dengan kata lain, ia mengontrol keseimbangan antara bias dan ragam dari \hat{f}_λ . Apabila nilai λ besar maka akan memberikan bobot *penalty* (kemulusan) yang besar dan mempunyai ragam yang kecil.

Dengan data amatan sebanyak n maka untuk menyelesaikan persamaan (3) digunakan algoritma numerik dari persamaan matriks berikut:

$$PLS(\mathbf{f}) = (\mathbf{Y} - \mathbf{f})^T (\mathbf{Y} - \mathbf{f}) + \lambda \mathbf{f}^T \mathbf{K} \mathbf{f}$$

dimana $\mathbf{Y}^T = (y_1, \dots, y_n)$, $\mathbf{f}^T = (f(x_1), \dots, f(x_n))$, sedangkan \mathbf{K} adalah matriks *penalty* yang mempunyai struktur spesifik, yaitu:

$$\mathbf{K} = \mathbf{D}^T \mathbf{C}^{-1} \mathbf{D}$$

dimana $\mathbf{D} = (d_{ij})$ adalah matriks *upper triangular* $[n-2, n]$ yang memiliki struktur sebagai berikut:

$$\begin{pmatrix} \lambda_1^{-1} & -(\lambda_1^{-1} + \lambda_2^{-1}) & \lambda_2^{-1} & 0 & \dots & 0 \\ 0 & \lambda_2^{-1} & -(\lambda_2^{-1} + \lambda_3^{-1}) & \lambda_3^{-1} & \dots & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & \dots & \dots & \lambda_{n-2}^{-1} & -(\lambda_{n-2}^{-1} + \lambda_{n-1}^{-1}) & \lambda_{n-1}^{-1} \end{pmatrix}$$

dan $C = (c_{ij})$ adalah matriks simetrik tridiagonal $[n-2, n-2]$ yang memiliki struktur sebagai berikut:

$$\begin{pmatrix} 2(\lambda_1 + \lambda_2) & \lambda_2 & 0 & 0 & \dots & 0 \\ \lambda_2 & \dots & \dots & \dots & \dots & \dots \\ \vdots & \vdots & \vdots & \vdots & \ddots & \lambda_{n-2} \\ 0 & \dots & \dots & \dots & \lambda_{n-2} & 2(\lambda_{n-2} + \lambda_{n-1}) \end{pmatrix}$$

Pemilihan parameter pemulus yang optimal dapat dilakukan dengan menggunakan berbagai kriteria, diantaranya dengan menggunakan *Cross Validation* (CV), *Generalized Cross Validation* (GCV), kriteria Cp Mallow (*Mallows' Cp criterion*) dan *Improved akaike information criterion* (AIC). (Lee, 2003).

Misal h_{ii} adalah elemen diagonal dari matriks hat $H = X(X^T X)^{-1} X^T$. Untuk *smoothing spline*, fungsi skor *leave-one-out* CV diberikan oleh :

$$CV(\lambda) = \frac{1}{n} \sum_{i=1}^n \left[\frac{y_i - f(x_i)}{1 - h_{ii}} \right]^2$$

Sedangkan dalam GCV, penyebut $(1-h_{ii})$ dalam $CV(\lambda)$ digantikan oleh $(1-\text{tr}(H)/n)$, sehingga menghasilkan fungsi skor GCV :

$$GCV(\lambda) = \frac{1}{n} \sum_{i=1}^n \left[\frac{y_i - f(x_i)}{1 - n^{-1} \text{tr}(H)} \right]^2$$

Fungsi skor untuk kriteria Cp Mallow diberikan oleh :

$$C_p(\lambda) = \frac{1}{n} \left\{ \|(H - I)y\|^2 + 2\sigma^2 \text{tr}(H) + \sigma^2 \right\}$$

Sedangkan fungsi skor AIC :

$$AIC_c(\lambda) = \log \frac{\|(H - I)y\|^2}{n} + \frac{2\{\text{tr}(H) + 1\}}{n - \text{tr}(H) - 2} + 1$$

Kriteria-kriteria ini merupakan metode klasik dalam penentuan parameter pemulus, metode lainnya adalah dengan berdasarkan fungsi resiko dan fungsi *loss*. Pembahasan detail tentang hal ini dapat dilihat dalam berbagai literatur misalnya dalam Hardle & Marron (1995), Berry et.al (2002), Lee (2003), Keele (2008), dsb.

3. Loess

Metode *loess* merupakan metode pemulus yang secara luas sering digunakan dengan sifat-sifat ketegaran (*robustness*) yang baik. Ia merupakan *running-line smoother* terboboti, dan setiap garis lokal diduga menggunakan suatu metode *robust* menggantikan kuadrat terkecil. Sehingga *loess* menjadi pemulus yang tegar.

Pendugaan *loess* dilakukan secara iteratif. Misalkan t merupakan bilangan index iterasi, mulai dengan $t=0$, kita misalkan $d_k(x_i)$ meotasikan jarak anatar x_i dengan tetangga terdekatnya yang ke- k , di mana k (atau k/n) adalah parameter pemulus. Kernel yang digunakan untuk pembobotan lokal di sekitar x_i adalah :

$$K_i(x) = K\left(\frac{x - x_i}{d_k(x_i)}\right) \quad \text{dengan} \quad K_i(x) = \begin{cases} (1 - |z|^3)^3 & , \text{ untuk } |z| \leq 1 \\ 0 & , \text{ selainnya} \end{cases}$$

merupakan kernel *tricube*.

Parameter dugaan dari regresi terboboti lokal untuk titik ke- i pada iterasi ke- t diperoleh dengan meminimumkan jumlah kuadrat terboboti :

$$\sum_{j=1}^n (y_j - (\beta_{0,i}^{(t)} + \beta_{1,i}^{(t)} x_j))^2 K_i(x_j) \quad (4)$$

Kita notaskan dugaan ini seagai $\hat{\beta}_{m,i}^{(t)}$ untuk $m = 0, 1$ dan $i = 1, \dots, n$. Catat bahwa nilai dugaan untuk peubah respon yang diberikan oleh regresi lokal adalah $\hat{y}_i^{(t)} = \hat{\beta}_{0,i}^{(t)} + \hat{\beta}_{1,i}^{(t)} x_i$.

Untuk menyiapkan iterasi berikutnya, setiap pengamatan diberi bobot baru berdasarkan ukuran sisaannya, hal ini untuk menurunkan bobot (*downweight*) data pencilan. Jika $e_i^{(t)} = y_i - \hat{y}_i^{(t)}$, maka definisikan bobot ketegaran sebagai berikut:

$$r_i^{(t+1)} = B\left(\frac{e_i^{(t)}}{6x\text{median} |e_i^{(t)}|}\right), \text{ dimana } B \text{ merupakan kernel } \textit{biweight} \text{ yang diberikan}$$

oleh :

$$B(x) = \begin{cases} (1 - z^2)^2 & \text{untuk } |z| \leq 1 \\ 0 & \text{selainnya} \end{cases}$$

Kemudian bobot K_i pada persamaan (4) digantikan dengan $r_i^{(t+1)}K_i(x_j)$ dan dugaan bobot lokal baru diperoleh. Dugaan yang dihasilkan untuk setiap i memberikan $y_i^{(t+1)}$.

DATA DAN METODE

1. Data

Data yang kami gunakan adalah data simulasi yang dibangkitkan dan diolah menggunakan perangkat lunak R. Data dibangkitkan dengan menggunakan beberapa fungsi non linier. Fungsi yang digunakan untuk membangkitkan data dalam makalah ini kami adopsi dari rancangan simulasi beberapa studi yang berkaitan dengan pendugaan model regresi nonparametrik. Fungsi-fungsi tersebut disajikan dalam Tabel 1 dengan catatan bahwa :

- a. $\varepsilon_i \sim N(0, 0.06)$
- b. $n = 200$
- c. untuk fungsi $f_1(x)$ dan fungsi $f_4(x)$: $x_i = \frac{i - 0.5}{n}$, $i = 1, \dots, 200$.
- d. untuk fungsi $f_2(x)$: $x_i \sim N(0,1)$
- e. untuk fungsi $f_3(x)$: $x_i \sim \text{uniform}(0,1)$

Tabel 1. Spesifikasi fungsi data simulasi

No.	Bentuk persamaan $y = f(x) + \epsilon$	Referensi
1.	$y_i = 1.5\phi\left(\frac{x_i-0.35}{0.15}\right) - \phi\left(\frac{x_i-0.8}{0.04}\right) + \epsilon_i, \quad \phi(u) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{u^2}{2}\right)$	Lee (2003)
2.	$y_i = \frac{\sin(\pi x / 2)}{1 + \{2x^2 (\text{sign}(x) + 1)\}} + \epsilon_i,$	Berry <i>et.al.</i> (2002)
3.	$y_i = \sin^3(2\pi x_i^3) + \epsilon_i,$	Faraway (2006)
4.	$y_i = 0.3 \exp\{-64(x_i - 0.25)^2\} + 0.7 \exp\{-256(x_i - 0.75)^2\} + \epsilon_i$	Aydin & Omay (2006)

2. Metode

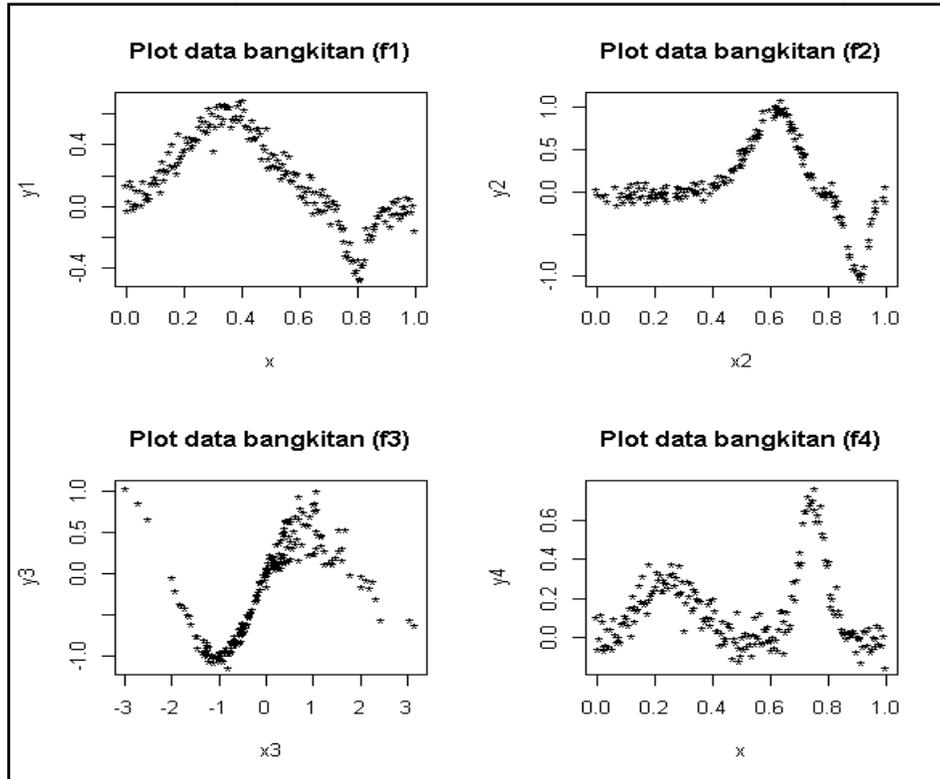
Untuk memperoleh dugaan kurva fungsi dari setiap metode regresi nonparametrik yang dibahas dalam makalah ini, kami melakukan langkah-langkah pendugaan sebagai berikut :

1. Pendugaan $f(x)$ dengan metode *kernel smoothing* dilakukan dengan menggunakan kernel Gaussian dan mencobakan beberapa nilai *bandwidth*. Selanjutnya memilih hasil terbaik secara subyektif yaitu dengan memilih bentuk kurva dugaan $f(x)$ yang mengikuti pola data secara optimal (tidak *oversmoothing* atau *undersmoothing*) dengan cara *trial and error*. Dalam makalah ini kami hanya menampilkan kurva dengan 3 nilai *bandwith* berbeda.
2. Pendugaan $f(x)$ menggunakan metode *smoothing spline* dilakukan dengan menggunakan parameter pemulus yang dihasilkan oleh metode GCV.
3. Pendugaan $f(x)$ dengan metode *loess* sama seperti pada metode *kernel smoothing*, yaitu dengan mencobakan beberapa nilai *span* dan memilih hasil terbaik secara subyektif yaitu secara *trial and error*. Dan kami hanya menampilkan kurva dengan 3 nilai *span* berbeda

HASIL DAN PEMBAHASAN

1. Plot pencar data bangkitan

Berikut ini kami sajikan gambar plot pencar data bangkitan dari ke empat fungsi :

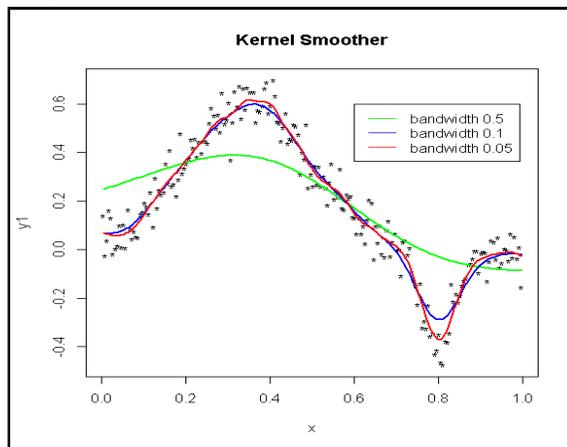


Gambar 1. Plot pencar (*scatterplot*) data bangkitan

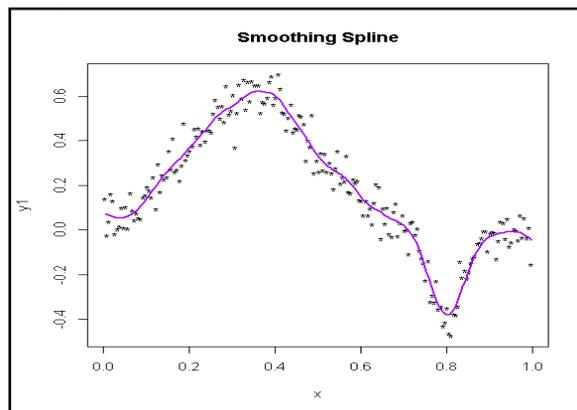
2. Hasil Analisis

a. Hasil analisis data bangkitan $f_1(x)$

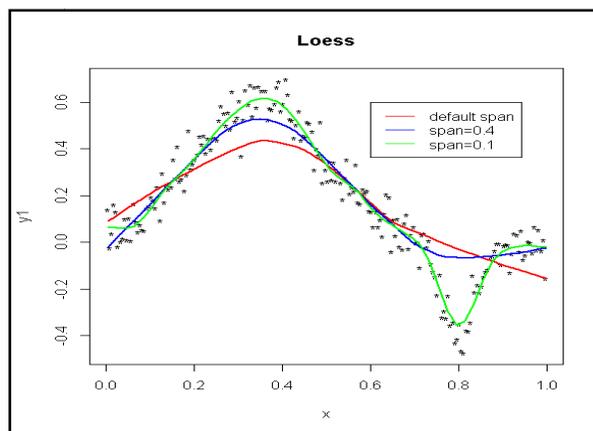
Gambar 2 - Gambar 5 di bawah ini berturut-turut menyajikan gambar dugaan kurva regresi untuk data bangkitan $f_1(x)$ menggunakan metode kernel *smoothing*, *smoothing spline* dan *loess*. Untuk data pertama ini, hasil dugaan kurva dengan metode kernel *smoothing* diperoleh hasil yang optimal pada *bandwith* 0.05 (warna hijau) paa Gambar 2. Sedangkan untuk kurva dugaan dengan metode *smoothing spline* disajikan pada Gambar 3, dan kurva dugaan *loess* pada Gambar 5. Untuk kurva *loess* diperoleh dugaan kurva terbaik pada *span*= 0.1 (warna hijau).



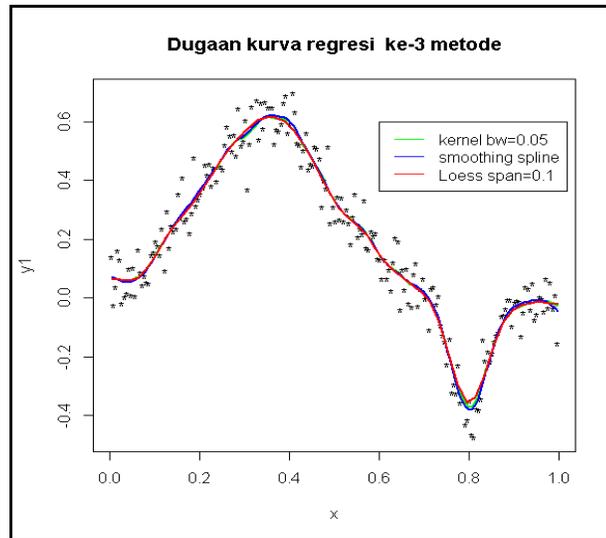
Gambar 2. Dugaan kurva fungsi $f_I(x)$ menggunakan *kernel smoothing*



Gambar 3. Dugaan kurva fungsi $f_I(x)$ menggunakan *smoothing spline*



Gambar 4. Dugaan kurva fungsi $f_I(x)$ menggunakan *loess*



Gambar 5. Dugaan kurva fungsi $f_1(x)$ dari ketiga metode

Berdasarkan nilai dugaan tersebut dapat dihitung kuadrat tengah galat masing-masing metode dan diperoleh sebagai berikut :

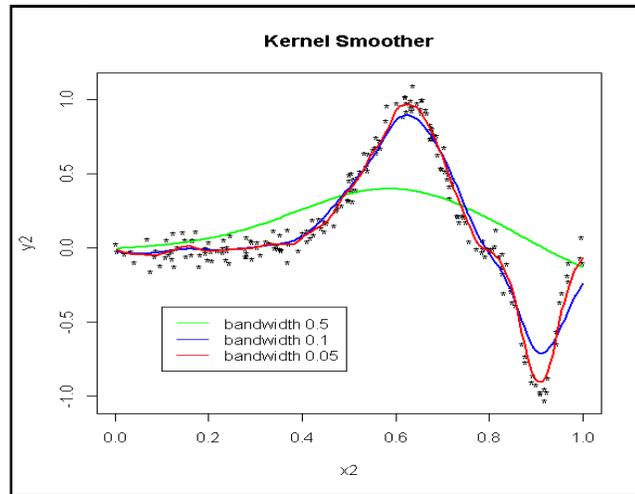
Tabel 2. Nilai Kuadrat Tengah Galat

Metode	Kuadrat Tengah Galat (KTG)
Kernel <i>Smoothing</i>	0.005085
<i>Smoothing Spline</i>	0.004745
Loess	0.004795

Berdasarkan hasil di atas dapat disimpulkan bahwa ketiga metode memberikan hasil yang hampir sama baiknya.

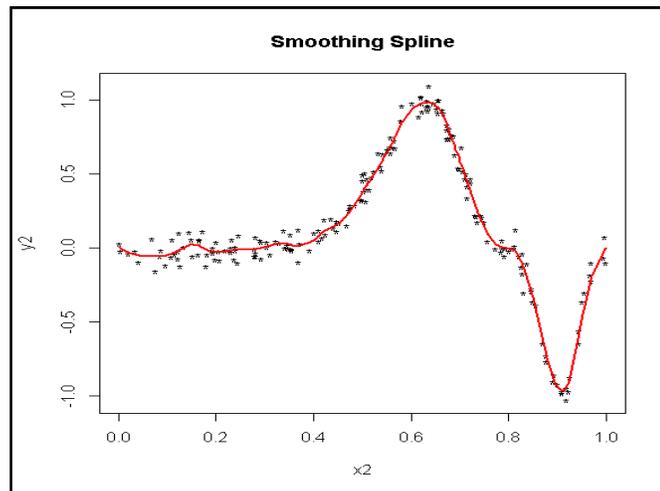
b. Hasil analisis data bangkitan $f_2(x)$

Gambar 6 - Gambar 8 di bawah ini berturut-turut menyajikan gambar dugaan kurva regresi untuk data bangkitan $f_2(x)$ menggunakan metode kernel *smoothing*, *smoothing spline* dan *loess*.

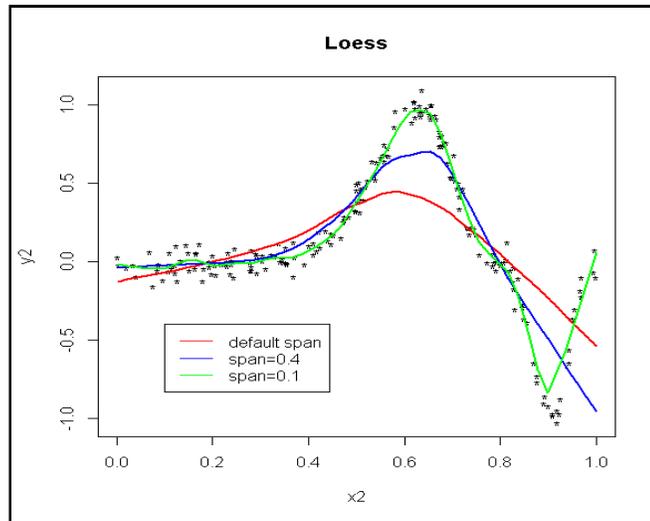


Gambar 6. Dugaan kurva fungsi $f_2(x)$ menggunakan *kernel smoothing*

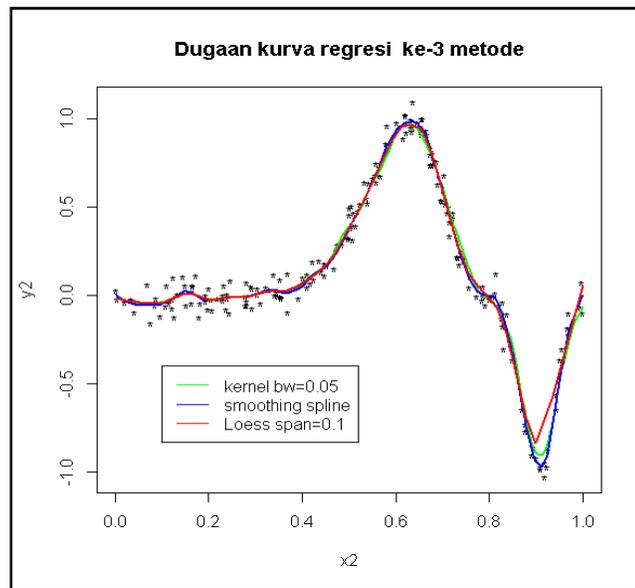
Untuk data kedua, hasil dugaan kurva dengan metode *kernel smoothing* diperoleh hasil yang optimal pada *bandwith* 0.05 (warna merah) pada Gambar 6. Sedangkan untuk kurva dugaan dengan metode *smoothing spline* disajikan pada Gambar 7 dan kurva dugaan *loess* pada Gambar 8. Untuk kurva *loess* diperoleh dugaan kurva terbaik pada *span*= 0.1 (warna hijau).



Gambar 7. Dugaan kurva fungsi $f_2(x)$ menggunakan *smoothing spline*



Gambar 8. Dugaan kurva fungsi $f_2(x)$ menggunakan *loess*

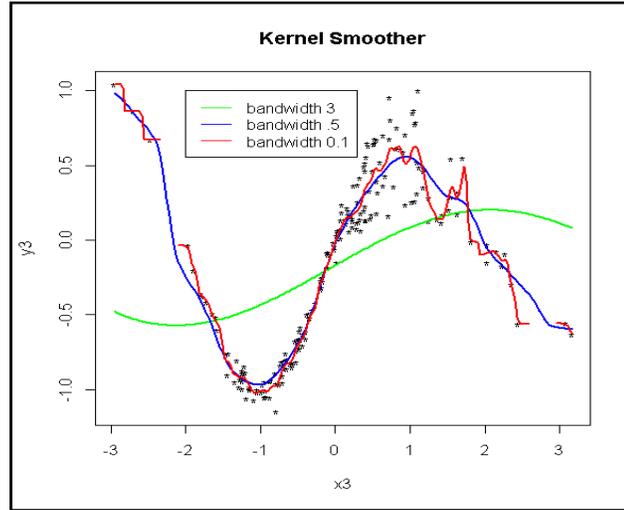


Gambar 9. Dugaan kurva fungsi $f_2(x)$ dari ketiga metode

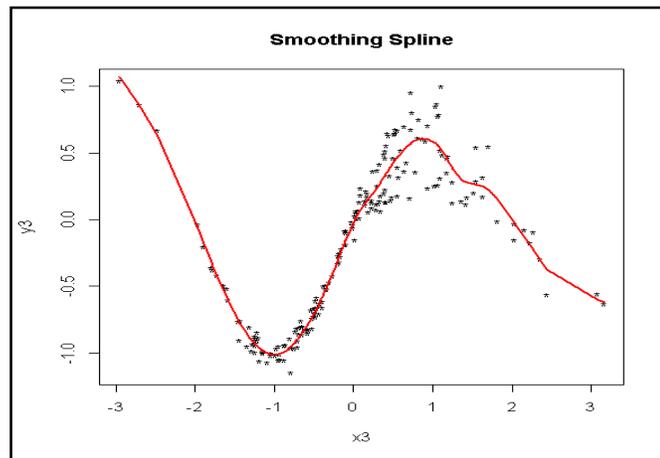
Berdasarkan gambar di atas dapat disimpulkan bahwa ketiga metode memberikan dugaan yang hampir sama baiknya yang diperlihatkan oleh kurva yang saling berhimpit. Namun secara subyektif kami memilih metode *smoothing spline* sebagai metode yang memberikan hasil terbaik pada data ini berdasarkan nilai-nilai dugaannya pada titik-titik disekitar $x = 0.9$ (sekitar titik minimum kurva).

c. Hasil analisis data bangkitan $f_3(x)$

Gambar 10 - Gambar 12 di bawah ini berturut-turut menyajikan gambar dugaan kurva regresi untuk data bangkitan $f_3(x)$.



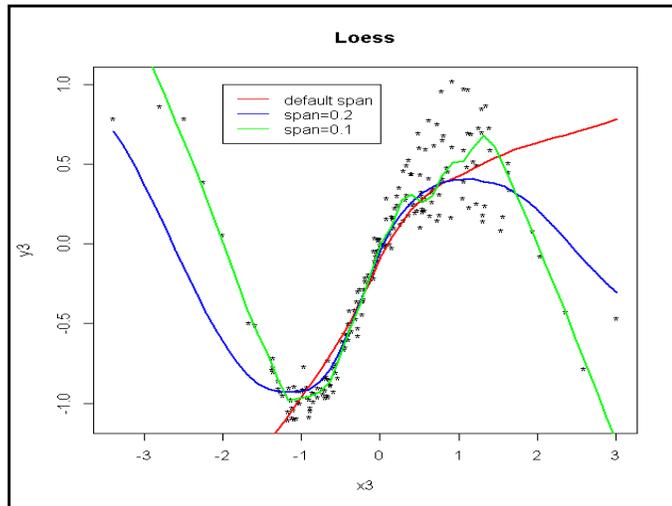
Gambar 10. Dugaan kurva fungsi $f_3(x)$ menggunakan *kernel smoothing*



Gambar 11. Dugaan kurva fungsi $f_3(x)$ menggunakan *smoothing spline*

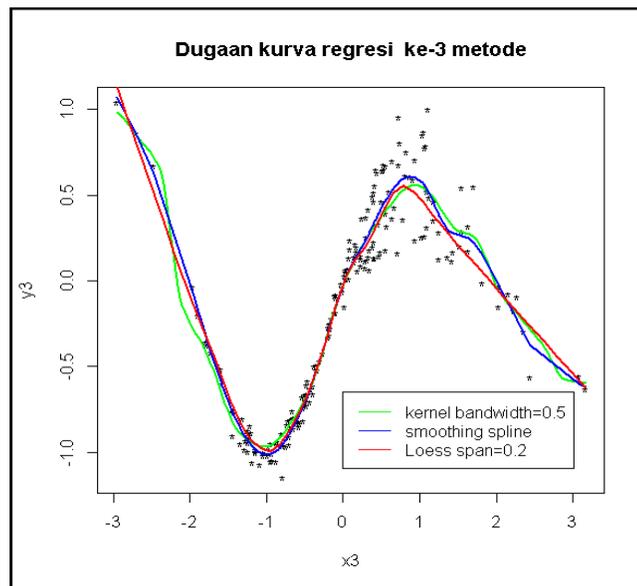
Untuk data ketiga, hasil dugaan kurva dengan metode *kernel smoothing* diperoleh hasil yang optimal pada *bandwith* 0.5 (warna biru) pada Gambar 10. Sedangkan untuk kurva dugaan dengan metode *smoothing spline* disajikan pada Gambar 11 dan

kurva dugaan *loess* pada Gambar 12. Untuk kurva *loess* diperoleh dugaan kurva terbaik pada *span*= 0.1 (warna hijau).



Gambar 12. Dugaan kurva fungsi $f_3(x)$ menggunakan *loess*

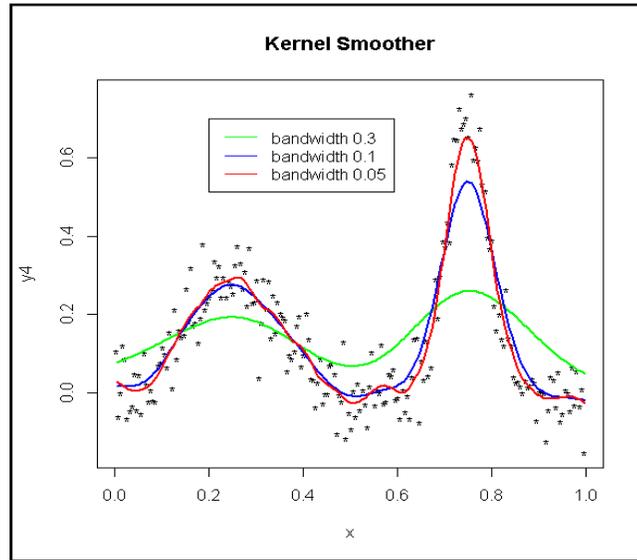
Hasil dugaan kurva regresi dari ketiga metode disajikan pada Gambar 13 dibawah. Dapat terlihat bahwa untuk kasus data ketiga ini ketiga metode memberikan hasil yang sedikit berbeda.



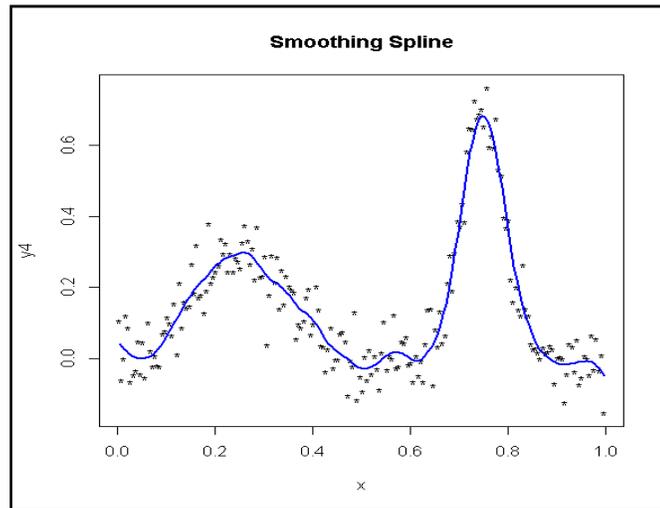
Gambar 13. Dugaan kurva fungsi $f_3(x)$ dari ketiga metode

d. Hasil analisis data bangkitan $f_4(x)$

Gambar 14 - Gambar 16 di bawah ini berturut-turut menyajikan gambar dugaan kurva regresi untuk data bangkitan $f_4(x)$.



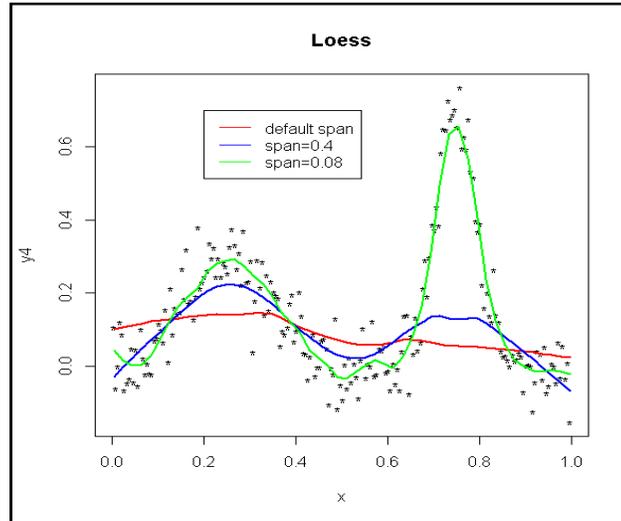
Gambar 14. Dugaan kurva fungsi $f_4(x)$ menggunakan *kernel smoothing*



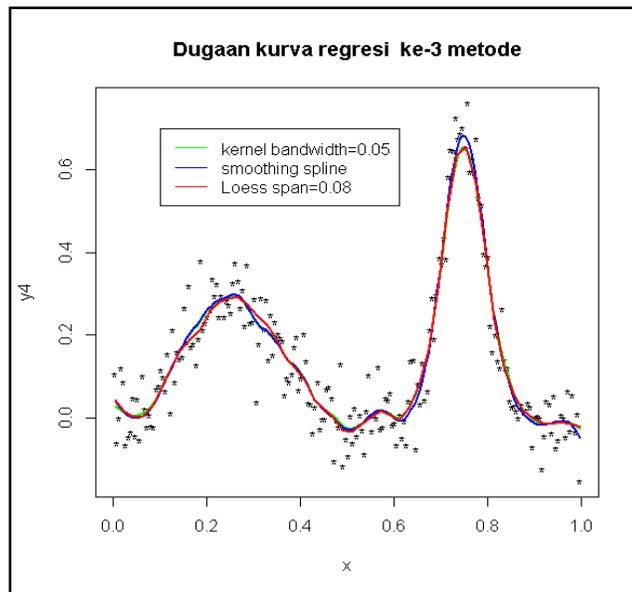
Gambar 15. Dugaan kurva fungsi $f_4(x)$ menggunakan *smoothing spline*

Untuk data keempat, hasil dugaan kurva dengan metode *kernel smoothing* diperoleh hasil yang optimal pada *bandwith* 0.05 (warna merah) pada Gambar 14. Sedangkan

untuk kurva dugaan dengan metode *smoothing spline* disajikan pada Gambar 15 dan kurva dugaan loess pada Gambar 16. Untuk kurva loess diperoleh dugaan kurva terbaik pada $\text{span} = 0.8$ (warna hijau).



Gambar 16. Dugaan kurva fungsi $f_4(x)$ menggunakan *loess*



Gambar 17. Dugaan kurva fungsi $f_4(x)$ dari ketiga metode

Berdasarkan gambar di atas dapat disimpulkan bahwa ketiga metode memberikan dugaan yang hampir sama baiknya yang diperlihatkan oleh kurva yang saling

berhimpit. Namun secara subyektif kami memilih metode *smoothing spline* sebagai metode yang memberikan hasil terbaik pada data ini berdasarkan nilai-nilai dugaannya pada titik-titik disekitar $x = 0.75$ (di sekitar puncak kurva).

KESIMPULAN

Berdasarkan uraian di atas dapat disimpulkan sebagai berikut :

1. Baik pendekatan *kernel smoothing*, *smoothing spline* maupun *loess* terhadap data bangkitan dalam makalah ini memberikan hasil yang cukup baik. Dan pendekatan terbaik dari ketiga metode ini adalah metode *smoothing spline*, hal ini sejalan dengan penelitian yang telah dilakukan oleh Aydin (2007) tentang perbandingan metode *kernel smoothing* dan *smoothing spline*.
2. Untuk metode *kernel smoothing* dan *loess* masing-masing memiliki keterbatasan, yaitu keduanya memerlukan tingkat subjektivitas pada penentuan *bandwidth* dan *span*. Subjektivitas dalam pemilihan *bandwidth* dan *span* ini sebenarnya dapat dihindari dengan berbagai cara misalnya dengan menggunakan kriteria kuadrat tengah galat (KTG) terkecil yang dihasilkan dari berbagai kemungkinan nilai *bandwidth* dan *span*.
3. Hasil analisis regresi menggunakan ketiga metode terhadap data dalam makalah ini memberikan model yang hampir sama yang ditunjukkan oleh bentuk kurva yang saling berhimpit.

Daftar Pustaka

1. Aydin, D. 2007. A Comparison of the Nonparametric Regression Models using Smoothing Spline and Kernel Regression. *World Academy of Science, Engineering and Technology* No. 36.
2. Aydin, D. Dan Omay, R.E. 2006. The Empirical Performances Of The Selection Criteria For Nonparametric Regression Using Smoothing Spline.

Proceedings Of The 5th WSEAS Int. Conf. On Computational Intelligence, Man-Machine Systems And Cybernetics, Venice, Italy, November 20-22.

3. Berry, S.M., Carroll, R.J., Ruppert, D. 2002. Bayesian Smoothing and Regression Splines for Measurement Error Problems. *Journal of the American Statistical Association*. March 2002, Vol. 97, No. 457, Theory and Methods.
4. Fox, J. 2002. *Nonparametric Regression*. Appendix to An R and S-PLUS Companion to Applied Regression
5. Gijbels, I., Pope, A., Wand, M. P. 1999. Understanding Exponential Smoothing via Kernel Regression. *Journal of the Royal Statistical Society. Series B (Statistical Methodology)*, Vol. 61, No. 1. pp. 39-50.
6. Hardle, W., & Marron, J.S., 1995. Fast and simple scatterplot smoothing. *Comput. Statist. Data Anal.* 20, 1–17.
7. Keele, L. 2008. *Semiparametric Regression for the Social Sciences*. John Wiley & Sons, Ltd.
8. Lee, T.C.M. 2003. Smoothing parameter selection for smoothing splines: a simulation study. *Computational Statistics & Data Analysis* 42 (2003) 139 – 148
9. Lee, T.C.M. 2004. Improved smoothing spline regression by combining estimates of different smoothness. *Statistics & Probability Letters* 67 (2004) 133–140
10. Rizzo, M.L. 2008. *Statistical Computing with R*. Chapman & Hall/CRC. Boca Raton.
11. Silverman, B.W. 1985. Some Aspects of the Spline Smoothing Approach to Non-parametric Regression Curve Fitting. *J. R. Statist. Soc. B* (1985) 47, No. 1, pp. 1-52
12. Zheng, Q. 2009. Local Adaptive Smoothing in Kernel Regression Estimation. A Master's Thesis Presented to the Graduate School of Clemson University.