Pemodelan Pola Hubungan Jumlah Penduduk Miskin Terhadap Rata-rata Lama Sekolah Wilayah Jawa-Bali Melalui Regresi Nonparametrik Metode *Smoothing Spline*

[1]Fanny Tania Safitri, [2]Suliadi, dan [3]Anneke Iswani A.

[1,2,3]*Statistika, Fakultas MIPA, Universitas Islam Bandung,*

e-mail: [1]funnytania@gmail.com, [2]suliadi@gmail.com

Abstrak. Pendekatan nonparametrik digunakan ketika bentuk fungsi data yang diperoleh tidak menunjukkan suatu pola hubungan yang mudah untuk digambarkan dengan fungsi tertentu. Bentuk fungsi yang tidak diketahui tersebut dapat diestimasi menggunakan metode *smoothing spline*. *Smoothing spline* mengestimasi fungsi regresi nonparametrik berdasarkan pada suatu persoalan optimasi, yaitu yang meminimumkan JKG dengan pembatasan bahwa fungsi cukup halus. Ini disebut dengan *Penalized Least Square Regression* (*PLS*). Metode *smoothing spline* diaplikasikan untuk mengestimasi rata-rata lama sekolah berdasarkan jumlah penduduk miskin Jawa-Bali tahun 2012. Kemudian estimator *smoothing spline* yang diperoleh dibandingkan dengan model linier dan kuadratik. Dalam skripsi ini diperoleh estimator *smoothing spline* pada model regresi nonparametrik yang memberikan estimasi fungsi regresi yang paling baik dengan menggunakan *MSE* sebagai kriteria pemilihan model paling andal.

Kata kunci : Regresi nonparametrik, *smoothing spline*, *penalized least square*, *generalized cross validation*

A. Pendahuluan

Untuk mengetahui hubungan fungsional antara dua variabel atau lebih, salah satu teknik statistika yang digunakan secara luas dalam ilmu terapan adalah analisis regresi. Dalam analisis regresi kita mengestimasi fungsi regresi. Estimasi fungsi regresi pada umumnya dilakukan dengan pendekatan parametrik, yang mulai diperkenalkan oleh Laplace sejak abad XVIII dan juga Boscovich pada tahun 1757 (Purnomo dkk, 2008). Pendekatan parametrik tersebut digunakan ketika bentuk fungsi data yang diperoleh menunjukkan suatu pola hubungan yang mudah digambarkan dengan fungsi tertentu seperti linier, kuadratik, kubik dan yang lainnya. Dimana pembuatan asumsi terhadap bentuk fungsi regresi didasarkan pada teori, pengalaman masa lalu atau tersedianya sumber-sumber lain yang dapat memberi pengetahuan atau informasi yang terperinci (Purnomo dkk, 2008).

Akan tetapi pada praktiknya asumsi tersebut mungkin tidak dapat dipenuhi. Terdapat banyak kasus dimana bentuk fungsi data yang diperoleh, terkadang tidak menunjukkan suatu pola hubungan yang mudah untuk digambarkan dengan fungsi tertentu (Diana dkk, 2012). Untuk mengatasi hal tersebut, estimasi fungsi regresi dapat dilakukan dengan pendekatan nonparametrik. Pendekatan nonparametrik tidak memberikan asumsi terhadap bentuk fungsi regresi, juga tidak menuntut terpenuhinya asumsi-asumsi regresi parametrik (Eubank, 1999). Sehingga pendekatan nonparametrik memiliki fleksibilitas yang tinggi untuk mengestimasi fungsi regresi (Mahler, 1995). Selain itu dalam pendekatan nonparametrik diasumsikan bahwa fungsi yang akan diduga berbentuk *smooth* (mulus), dalam arti termuat dalam suatu ruang fungsi tertentu (Adisantoso, 2010).

Beberapa penulis seperti Hardle (1990), Wahba (1990), Budiantara dan Subanar (1997) menyarankan penggunaan regresi nonparametrik sebagai pendekatan untuk pemodelan suatu data, agar mempunyai fleksibilitas yang baik. Diantara teknik estimasi yang cukup populer dalam regresi nonparametrik yaitu estimator *spline*, estimator kernel, estimator deret orthogonal, analisis wavelet dan lain-lain (Budiantara dkk, 2006).

Adapun Wahba (1990) menunjukkan bahwa *spline* memiliki sifat-sifat statistik yang berguna untuk menganalisis hubungan dalam regresi. Fungsi spline memiliki fleksibilitas yang tinggi dan mampu menangani data yang perilakunya berubah-ubah pada subinterval-subinterval tertentu (Eubank, 1999). Dalam pendekatan estimator *spline* dikenal beberapa metode seperti *spline* original, *spline type m*, *p-spline*, *smoothing spline*, *spline* terbobot dan lain-lain. Diantara banyak metode tersebut, yang akan digunakan untuk mengestimasi fungsi regresi adalah *smoothing spline*.

Smoothing spline mengestimasi fungsi regresi nonparametrik berdasarkan pada suatu persoalan optimasi, yaitu yang meminimumkan JKG dengan pembatasan bahwa fungsi cukup halus (*roughness penalty*). Ini disebut dengan *Penalized Least Square Regression* (*PLS*). Perhitungan dalam meminimumkan fungsi tujuan dari *PLS* tersebut didasarkan pada definisi *natural cubic spline* dengan teorema yang diberikan Green & Silverman (1994) mengenai bentuk matriks *roughness penalty*. Adapun untuk kasus yang lebih umum dimana memungkinkan terdapat nilai pengamatan yang sama, dapat dilakukan dengan sedikit modifikasi yaitu dengan membuat generalisasi titik desain menggunakan *incidence matrix*. Metode ini disebut dengan *smoothing spline* untuk *general design time point*.

Dalam penelitian ini, analisis regresi akan digunakan untuk mengestimasi rata-rata lama sekolah berdasarkan jumlah penduduk miskin Jawa-Bali tahun 2012. Dengan mengetahui berapa jumlah penduduk miskin di wilayah Jawa-Bali, kita dapat mengestimasi nilai rata-rata lama sekolah di wilayah tersebut. Yaitu penurunan/peningkatan jumlah penduduk miskin akan diikuti dengan penurunan/peningkatan rata-rata lama sekolah, atau peningkatan jumlah penduduk miskin akan diikuti dengan penurunan rata-rata lama sekolah dan sebaliknya.

Untuk mengetahui model pendekatan analisis regresi yang paling sesuai dalam menggambarkan hubungan antara rata-rata lama sekolah dan jumlah penduduk miskin tersebut, digunakan dua model pendekatan analisis regresi. Yaitu pendekatan parametrik menggunakan metode *OLS* yang akan membentuk model linier dan kuadratik, serta pendekatan nonparametrik menggunakan metode *smoothing spline*. Karena dalam data yang digunakan terdapat nilai pengamatan yang sama maka pendekatan nonparametrik menggunakan metode *smoothing spline* untuk *general design time point*.

B. Landasan Teori

Regresi Nonparametrik – *Smoothing Spline*

Dalam regresi nonparametrik, bentuk fungsi hubungan antara variabel bebas dan variabel tak bebas tidak diketahui. Fungsi hubungan tersebut adalah sembarang fungsi yang biasanya berasal dari kelas tertentu. Regresi nonparametrik ini digunakan ketika kita menduga bahwa fungsi regresi membentuk suatu fungsi, akan tetapi kita tidak memiliki model untuk fungsi tersebut.

Sama seperti dalam regresi parametrik, dalam regresi nonparametrik kita juga melakukan estimasi terhadap fungsi regresi. Hanya saja kita tidak dapat menggunakan metode *OLS* untuk mengestimasi fungsi regresi.

Misalkan kita memiliki model regresi

$$Y_i = f(t_i) + \varepsilon$$

Dalam notasi matriks dapat ditulis sebagai

$$\boldsymbol{Y} = \boldsymbol{f} + \boldsymbol{\varepsilon}$$

dimana

$$\boldsymbol{Y} = \begin{pmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{pmatrix};\ \boldsymbol{f} = \begin{pmatrix} f_1 \\ f_2 \\ \vdots \\ f_n \end{pmatrix} = \begin{pmatrix} f(t_1) \\ f(t_2) \\ \vdots \\ f(t_n) \end{pmatrix};\ \text{dan } \boldsymbol{\varepsilon} = \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{pmatrix}$$

Dalam regresi nonparametrik, $f(t_i)$ tidak diketahui. Jika kita mencoba mengestimasi $f(t_i)$ menggunakan *OLS* tanpa pembatasan, hasilnya akan tidak memuaskan. Salah satu cara untuk mendapatkan fungsi yang lebih halus tersebut adalah dengan memberikan *penalty* kemulusan fungsi $f(t_i)$. Tolak ukur kemulusan fungsi $f(t_i)$ adalah $\int (f''(t))^2\, dt$.

Sekarang $f(t_i)$ didasarkan pada meminimumkan JKG dengan pembatasan bahwa fungsi cukup halus. Ini disebut *Penalized Least Square Regression* (*PLS*). Maka fungsi tujuan dari *PLS* dapat dinyatakan sebagai berikut:

$$S(f) = \sum_{i=1}^{n} \underbrace{\left(Y_i - f(t_i)\right)^2}_{(a)} + \lambda \int_a^b \underbrace{(f''(t))^2\, dt}_{(b)} \qquad \ldots (2.1)$$

dimana bagian (a) merupakan JKG atau fungsi jarak antara data dan dugaan, bagian (b) merupakan *roughness penalty*, yaitu ukuran kemulusan kurva, dan λ merupakan parameter pemulus (Green & Silverman, 1994).

Smoothing **berdasarkan** ***Natural Cubic Spline***

Misalkan kita mempunyai *n* buah pengamatan $(t_1, y_1), (t_2, y_2), \ldots, (t_n, y_n)$ pada interval $[a, b]$ dengan $a < t_1 < t_2 < \cdots < t_n < b$. Fungsi *f* dikatakan *cubic spline* jika memenuhi 2 kondisi :

1) Setiap interval f adalah fungsi polinomial berordo 3
2) Polinomial-polinomial tersebut cocok (*fit*) secara bersama-sama pada titik t_i sedemikian sehingga f' dan f'' kontinu pada masing-masing t_i, sehingga fungsi f kontinu pada selang $[a, b]$

Nilai $t_1, t_2, \ldots t_n$ disebut sebagai titik knot. *Cubic spline* dikatakan *natural cubic spline* jika f'' dan f''' adalah nol pada titik *a* dan *b*, yang juga berimplikasi pada $f''(t_1) = 0$ dan $f''(t_n) = 0$.

Misalkan $f_i = f(t_i)$ dan $\gamma_i = f''(t_i)$ untuk $i = 1,2, \ldots, n$. Berdasarkan definisi *natural cubic spline,* maka $\gamma_1 = \gamma_n = 0$. Misalkan $\boldsymbol{f} = (f_1, f_2, \ldots f_n)^T$ adalah vektor n × 1, dan $\boldsymbol{\gamma} = (\gamma_2, \gamma_3, \ldots, \gamma_{n-1})^T$ adalah vektor berukuran $(n-2) \times 1$. Vektor $\boldsymbol{f}$ dan $\boldsymbol{\gamma}$ ini akan menentukan bentuk kurva f. Dua vektor ini ditentukan oleh dua matriks **Q** dan **R** yang didefinisikan di bawah ini (Green & Silverman, 1994).

Matriks **Q** adalah matriks *upper triagonal* berukuran $n \times (n-2)$ yang memiliki elemen q_{ij}, untuk $i = 1,2, \ldots, n$ dan $j = 2,3, \ldots, n-1$ yang diberikan oleh : $q_{j-1,j} = h_{j-1}^{-1}$, $q_{jj} = -h_{j-1}^{-1} - h_j^{-1}$, $q_{j+1,j} = h_j^{-1}$ dan $q_{ij} = 0$ untuk $|i-j| \geq 2$. Dengan h seperti pada persamaan (2.2). Karena $j = 2, 3, \ldots, n-1$ maka elemen *the top left* matriks $\boldsymbol{Q}$ adalah q_{12}.

Sedangkan **R** merupakan matriks simetris berukuran $(n-2) \times (n-2)$ dengan elemen r_{ij}, untuk i dan j bergerak dari 2 sampai dengan $n-1$ yang diberikan oleh : r_{ii}

$=\frac{1}{3}\left(h_{(i-1)}+h_i\right)$ untuk $i = 2,3,\dots,n-1$, $r_{i,i+1}=r_{i+1,i}=\frac{1}{6}h_i$ untuk $i = 2,3,\dots,n-2$, dan $r_{ij}=0$ untuk $|i-j|\geq 2$ dimana $|r_{ii}|>\sum_{i\neq j}|r_{ij}|$. Dengan

$$h_i = t_{i+1}-t_i \qquad \dots (2.2)$$

Untuk $i = 1, 2, \dots, n-1$. Kemudian matriks **K** diperoleh dari persamaan berikut :

$$\boldsymbol{K}=\boldsymbol{Q}\boldsymbol{R}^{-1}\boldsymbol{Q}^T \qquad \dots(2.3)$$

Dengan menggunakan teorema Green & Silverman (1994) mengenai bentuk matriks *roughness penalty,* fungsi tujuan dapat ditulis sebagai berikut :

$$S(f)=\boldsymbol{f}^T(\boldsymbol{I}+\lambda\boldsymbol{K})\boldsymbol{f}-2\boldsymbol{Y}^T\boldsymbol{f}+\boldsymbol{Y}^T\boldsymbol{Y} \qquad \dots (2.4)$$

Taksiran dari $\boldsymbol{f}$ adalah $\hat{\boldsymbol{f}}$ yang diperoleh dengan meminimumkan (2.11), sehingga diperoleh nilai taksiran $\boldsymbol{f}$

$$\hat{\boldsymbol{f}}=(\boldsymbol{I}+\lambda\boldsymbol{K})^{-1}\boldsymbol{Y} \qquad \dots (2.5)$$

Smoothing Spline* untuk *General Design Time Point

Penurunan $\boldsymbol{f}$ pada sub bab (2.1.1) adalah untuk kasus dimana tidak ada nilai pengamatan t yang sama. Untuk kasus yang lebih umum dimana memungkinkan ada nilai pengamatan t yang sama, maka perlu ada sedikit modifikasi. Generalisasi titik desain dapat dibuat dengan menggunakan *incidence matrix*. Misalkan kita memiliki *n* buah pengamatan $(t_1,y_1),(t_2,y_2),\dots,(t_n,y_n)$. Diantara nilai $(t_1,t_2,\dots,t_n)$ terdapat nilai t yang sama. Misalkan dari *n* nilai t tersebut terdapat $q\leq n$ nilai t yang berbeda. Nilai-nilai yang berbeda tersebut disimbolkan dengan $s_1,s_2,\dots,s_q$ dimana $s_1<s_2<\cdots<s_q$. Hubungan antara t_i dan s_j ditunjukkan oleh *incidence matrix* **X** yang berukuran $n\times q$. Elemen-elemen matriks **X** tersebut adalah

$$X_{ij}=\begin{cases}1, & jika\ t_i=s_j\\ 0, & jika\ t_i\neq s_j\end{cases} \qquad \dots (2.6)$$

dimana $i = 1,\dots,n$ dan $j=1,\dots,q$.

Untuk $\boldsymbol{f}=\left(f(s_1),\dots,f(s_q)\right)^T$ kita memiliki $\boldsymbol{Xf}=\left(f(t_1),\dots,f(t_n)\right)^T$, selanjutnya model regresi dapat ditulis dalam formulasi matriks

$$\boldsymbol{Y}=\boldsymbol{Xf}+\boldsymbol{\epsilon}$$

Fungsi tujuan sekarang dapat ditulis sebagai berikut:

$$S(f)=\boldsymbol{Y}^T\boldsymbol{Y}-2\boldsymbol{Y}^T\boldsymbol{Xf}+\boldsymbol{f}^T\boldsymbol{X}^T\boldsymbol{Xf}+\lambda\boldsymbol{f}^T\boldsymbol{Kf} \qquad \dots (2.7)$$

Dalam hal ini taksiran $\boldsymbol{f}$ dicari dengan meminimumkan persamaan (2.7) sehingga diperoleh nilai taksiran $\boldsymbol{f}$ sebagai berikut

$$\hat{\boldsymbol{f}}=(\boldsymbol{X}^T\boldsymbol{X}+\lambda\boldsymbol{K})^{-1}\boldsymbol{X}^T\boldsymbol{Y} \qquad \dots (2.8)$$

Matriks **K** diperoleh dengan perhitungan yang sama seperti pada sub bab 2.1.1. Hanya saja t_i diganti dengan s_j.

Plot Kurva

Dalam regresi nonparametrik, memplot kurva merupakan aspek penting karena inferensi dan analisis didasarkan pada kurva. Untuk *smoothing spline* menggunakan *natural cubic spline*, Green & Silverman (1994) memberikan $f(t)$ untuk seluruh rentang. Untuk $i = 1, 2, \dots, n-1$ didefinisikan $h_i = t_{i+1} - t_i$, kemudian

$$f(t) = \frac{(t - t_i)f_{i+1} + (t_{i+1} - t)f_i}{h_i} - \frac{1}{6}(t - t_i)(t_{i+1} - t)$$
$$\times \left\{\left(1 + \frac{t - t_i}{h_i}\right)\gamma_{i+1} + \left(1 + \frac{t_{i+1} - t}{h_i}\right)\gamma_i\right\},$$
$$\text{untuk } t_i \le t \le t_{i+1},\ i = 1, 2, \dots, n-1.$$

Untuk $t \le t_1$dan $t \ge t_n$, didapatkan $\gamma_1 = \gamma_n = 0$ sehingga f memiliki bentuk khusus dimana fungsinya adalah linier yang diberikan oleh

$$f'(t_1) = \frac{f(t_2) - f(t_1)}{t_2 - t_1} - \frac{1}{6}(t_2 - t_1)\gamma_2$$
$$f'(t_n) = \frac{f(t_n) - f(t_{n-1})}{t_n - t_{n-1}} + \frac{1}{6}(t_n - t_{n-1})\gamma_{n-1}$$

Danbentuk kurva diberikan oleh

$$f(t) = f_1 - (t_1 - t)f'(t_1), \quad \text{untuk } t \le t_1$$
$$f(t) = f_n + (t - t_n)f'(t_n), \quad \text{untuk } t \ge t_n$$

Pemilihan Parameter Pemulus

Estimator *smoothing spline* sangat tergantung pada parameter pemulus (*smoothing parameter*), sehingga pemilihan parameter pemulus merupakan hal yang penting dalam mencari estimator spline yang paling sesuai. Dalam penelitian ini akan digunakan kriteria pemilihan parameter pemulus melalui *GCV*. Fungsi *GCV* merupakan modifikasi dari *CV*. Fungsi GCV didefinisikan sebagai

$$GCV(\lambda) = \frac{1}{n}\frac{\sum_{i=1}^{n}\left(y_i - \hat{f}(t_i)\right)^2}{\left(1 - n^{-1}tr(\boldsymbol{A}(\lambda))\right)^2} \qquad \dots (2.9)$$

dimana $\boldsymbol{A}_{ii}$ adalah elemen diagonal ke-i dari matriks hat **A** dengan $\boldsymbol{A} = \boldsymbol{X}(\boldsymbol{X}^T\boldsymbol{X} + \lambda\boldsymbol{K})^{-1}\boldsymbol{X}^T$. Nilai λ yang optimal diperoleh dengan meminimumkan $GCV(\lambda)$.

Mean Square Error

Mean Square Error (*MSE*) merupakan fungsi risiko, yaitu perbedaan antara nilai dugaan dengan nilai sebenarnya. Dalam penelitian ini *MSE* digunakan untuk membandingkan seluruh model yang diperoleh, sehingga ditemukan model yang benar-benar andal. Adapun *MSE* diperoleh melalui perhitungan dibawah ini:

$$MSE = \frac{\sum_{i=1}^{n}\left(Y_i - \hat{Y}\right)^2}{db}$$

dimana db merupakan derajat bebas dari galat.

Green & Silverman (1994) mendefinisikan db pada *smoothing spline* sebagai

$$db = tr\{I - \boldsymbol{A}(\lambda)\} \qquad \dots (2.10)$$

dimana $\boldsymbol{A}(\lambda)$ merupakan matriks hat **A** terkait dengan *smoothing spline* dengan parameter pemulus λ.

Model Regresi Hubungan Jumlah Penduduk Miskin dengan Rata-rata Lama Sekolah Jawa-Bali

Bahan yang digunakan merupakan data sekunder yang diperoleh dari Sistem Informasi dan Manajemen Data Dasar Regional (SIMREG) Bappenas. Data yang diperoleh berskala rasio karena data tersebut dihasilkan melalui perhitungan matematis.

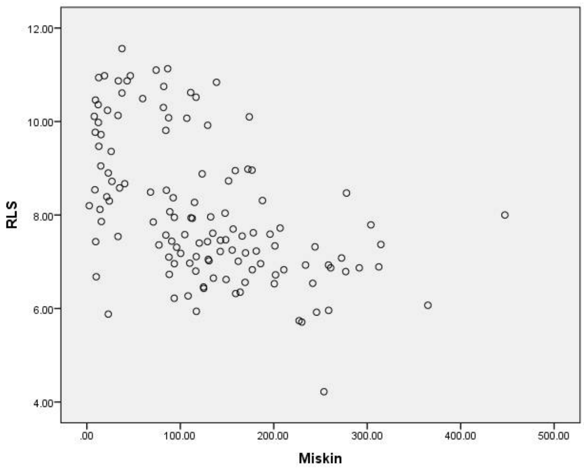
Data rata-rata lama sekolah dan jumlah penduduk miskin disajikan pada tabel 3.1. dengan versi lengkap terdapat pada lampiran.

Tabel 3.1. Data Rata-rata Lama Sekolah dan Jumlah Penduduk Miskin Kab/Kota Wilayah Jawa-Bali Tahun 2012

No.	Kabupaten/ Kota	Jumlah Penduduk Miskin (Ribu Jiwa) 2012	Rata-rata Lama Sekolah (Tahun) 2012
1	Kab. Kepulauan Seribu	2.60	8.20
2	Kota Jakarta Selatan	74.10	11.10
⋮	⋮	⋮	⋮
126	Kota Denpasar	12.70	10.94

Sumber : Data Statistik Indonesia (2012)

Berikut adalah s*catter plot* antara variabel jumlah penduduk miskin (X) dan variabel rata-rata lama sekolah (Y)



Gambar 3.1. S*catter plot* antara variabel jumlah penduduk miskin (X) dan variabel rata-rata lama sekolah (Y)

Dari gambar diatas terlihat bahwa data berpencar membentuk suatu fungsi data yang menunjukkan suatu pola hubungan tidak sederhana. Artinya kita tidak memiliki model untuk fungsi tersebut karena fungsi tersebut tidak dapat secara sederhana diwakili oleh suatu fungsi regresi parametrik seperti linier, kuadratik, dll.

a. Regresi Linier

Estimasi fungsi regresi dengan pendekatan parametrik untuk membentuk model linier menghasilkan :

$$\hat{\boldsymbol{\beta}} = \begin{pmatrix} 9.22915954 \\ -0.009100564 \end{pmatrix}$$

Sehingga diperoleh persamaan model regresi linier

$$\hat{y} = 9.2292 - 0.0091\, X$$

Selanjutnya melakukan pengujian menggunakan analisis varians dengan hipotesis

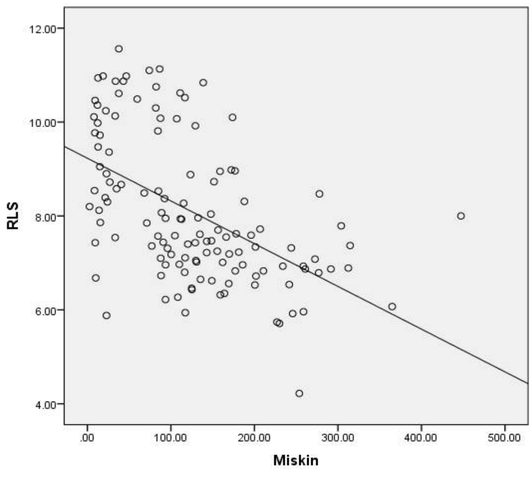
H_0 : $\beta_1 = 0$, koefisien regresi β_1 tidak signifikan
H_1 : $\beta_1 \neq 0$, koefisien regresi β_1 signifikan

dan nilai-nilai perhitungan yang tersaji pada tabel 3.1.

Tabel 3.2. Analisis Varians Regresi Linier

Sumber Variasi	**db**	**Jumlah Kuadrat**	**Kuadrat Tengah**	**F**	***P-value***
Regresi	1	80.292	80.292	46.199	0.000
Galat	124	215.505	1.738		
Total	125	295.797			

Karena *p-value* < α, maka tolak H_0 yang artinya koefisien regresi β_1 signifikan atau dengan kata lain variabel X berpengaruh signifikan terhadap variabel Y. Berikut ini adalah plot estimasi fungsi regresi linier



Gambar 3.2. Plot estimasi fungsi regresi linier

b. Regresi Kuadratik

Estimasi fungsi regresi dengan pendekatan regresi parametrik untuk membentuk model kuadratik, menghasilkan :

$$\hat{\boldsymbol{\beta}} = \begin{pmatrix} 9.68428392 \\ -0.017923189 \\ 0.0000277895 \end{pmatrix}$$

Sehingga diperoleh persamaan model regresi kuadratik

$$\hat{y} = 9.684 - 0.018\,X + 0.000028\,\mathrm{X}^2 \qquad \dots (3.2)$$

Selanjutnya melakukan pengujian menggunakan analisis varians dengan hipotesis

H_0 : $\beta_1 = \beta_2 = 0$, koefisien regresi β_1 dan β_2 tidak signifikan
H_1 : $\beta_1 \neq 0$ atau $\beta_2 \neq 0$, minimal ada satu koefisien regresi yang signifikan

dan nilai-nilai perhitungan yang tersaji pada tabel 3.3.

Tabel 3.3. Analisis Varians Regresi Kuadratik

Sumber Variasi	db	Jumlah Kuadrat	Kuadrat Tengah	F	*P-value*
Regresi	2	91.610	45.805	27.592	0.000
Galat	123	204.187	1.660		
Total	125	295.797			

Karena *p-value* < α, maka tolak H_0 yang artinya minimal ada satu koefisien regresi yang signifikan atau dengan kata lain kedua variabel X secara serempak berpengaruh signifikan terhadap variabel Y.

Selanjutnya melakukan pengujian koefisien regresi secara parsial melalui uji *t* untuk mengetahui variabel X mana yang berpengaruh signifikan terhadap variabel Y. Dengan hipotesis

H_0 : $\beta_i = 0$, koefisien regresi β_i tidak signifikan
H_1 : $\beta_i \neq 0$, koefisien regresi β_i signifikan

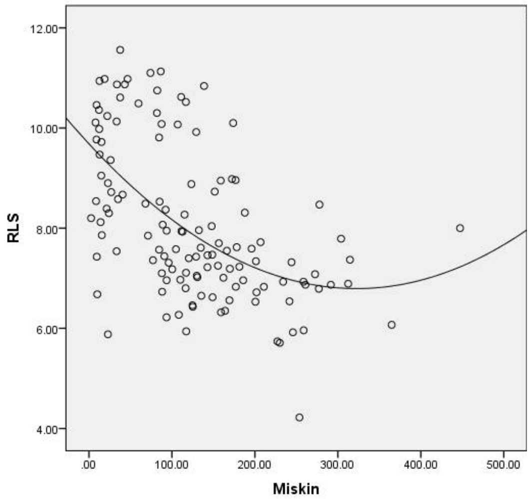
untuk i = 1 dan 2.

Adapun nilai-nilai perhitungan yang dihasilkan tersaji pada tabel 3.4.

Tabel 3.4. Taksiran Parameter Model

Variabel Bebas	Parameter	Taksiran	t	*P-value*
X	β_1	-0.018	-4.946	0.000
X^2	β_2	0.000028	2.611	0.010

Karena *p-value* dari kedua statistik uji t kurang dari α, maka tolak H_0 yang artinya koefisien regresi β_1 dan β_2 berpengaruh signifikan terhadap variabel Y. Berikut ini adalah plot estimasi fungsi regresi kuadratik



Gambar 3.3. Plot estimasi fungsi regresi kuadratik

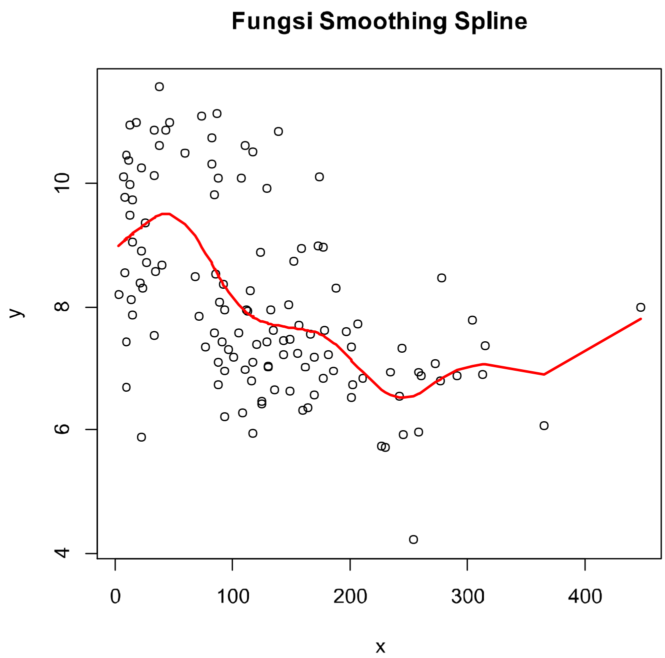
c. Regresi Nonparametrik

Estimasi fungsi regresi dengan pendekatan regresi nonparametrik menggunakan persamaan (2.15), menghasilkan :

$$\hat{\boldsymbol{f}} = \begin{pmatrix} 8.977156 \\ 8.963891 \\ 8.548305 \\ \vdots \\ 9.141203 \end{pmatrix}$$

dengan λ optimal yang dikeluarkan oleh program R adalah sebesar 0.0005068955. Sedangkan nilai *GCV* yang diperoleh adalah 1.675855 dan *MSE* sebesar 1.569789.

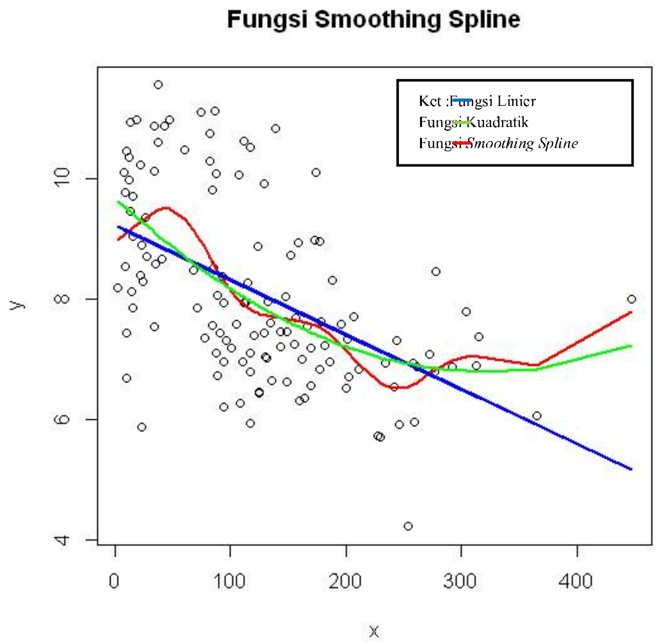
Selanjutnya membuat plot estimasi fungsi regresi nonparametrik. Adapun hasilnya adalah sebagai berikut :



Gambar 3.4. Plot estimasi fungsi regresi nonparametrik

d. Perbandingan Model

Hasil analisis secara keseluruhan dapat dilihat pada gambar di bawah ini :



Gambar 3.5. Plotestimasi fungsi regresi linier, kuadratik, dan nonparametrik

Dari gambar di atas dapat dilihat bahwa estimasi fungsi regresi menggunakan *smoothing spline* memberikan hasil yang sangat baik dengan fungsi regresi yang lebih fleksibel dalam mendekati pola data. Hal ini diperkuat dengan melihat *MSE* dari masing-masing model yang dapat dilihat pada tabel di bawah ini :

Tabel 3.5. *Mean Square Error*

No.	Model	MSE	Db Galat	Db Model
1	Linier	1.737945	124	2
2	Kuadratik	1.660059	123	3
3	Nonparametrik	1.569789	118.0254	7.974608

C. Kesimpulan

Dalam penelitian ini digunakan dua model pendekatan analisis regresi dalam mengestimasi pola hubungan rata-rata lama sekolah dan jumlah penduduk miskin wilayah Jawa-Bali. Yaitu pendekatan parametrik dan nonparametrik. Estimasi fungsi regresi dengan pendekatan parametrik menggunakan metode *OLS*. Prinsip dasar metode *OLS* adalah meminimumkan jumlah kuadrat dari galat (JKG). Adapun hasil yang diperoleh sebagai berikut :

1. Model Linier
   Persamaan model regresi linier yang diperoleh adalah

$$\hat{y} = 9.2292 - 0.0091\,X$$

   dengan nilai *MSE* sebesar 1.737945.

2. Model Kuadratik
   Persamaan model regresi kuadratik yang diperoleh adalah

$$\hat{y} = 9.684 - 0.018\,X + 0.000028\,\mathrm{X}^2$$

   dengan nilai *MSE* sebesar 1.660059.

Kemudian estimasi fungsi regresi dengan pendekatan nonparametrik menggunakan metode s*moothing spline* memberikan nilai *MSE* sebesar 1.569789.

Dari hasil analisis yang telah dilakukan, jelas terlihat bahwa nilai *MSE* untuk pendekatan parametrik lebih besar dibandingkan dengan pendekatan nonparametrik. Pendekatan nonparametrik menggunakan metode *smoothing spline* memberikan estimasi fungsi regresi yang paling baik dengan nilai *MSE* paling kecil yaitu sebesar 1.569789. Oleh karena itu, hubungan rata-rata lama sekolah dengan jumlah penduduk miskin di wilayah Jawa-Bali paling sesuai digambarkan dengan regresi nonparametrik. Hal ini sesuai dengan deskripsi data, yaitu pola data yang cenderung tidak linier ataupun kuadratik.

Maka berdasarkan pemodelan *smoothing spline,* seiring dengan bertambahnya jumlah penduduk miskin, angka rata-rata lama sekolah setiap kab/kota di wilayah Jawa-Bali cenderung mengalami penurunan dari titik tertinggi sekitar 9.5 (tahun) hingga *stagnant* di kisaran angka 7 (tahun).

Daftar Pustaka

Adisantoso, Julio. 2010. *Menentukan Parameter Pemulus pada Model Regresi Smoothing Spline.* [Online].http://julio.staff.ipb.ac.id/ (diakses tanggal 20 Mei 2014).

Budiantara, I.N., Suryadi, F., Otok, B.W., dan Guritno, S. 2006. *Pemodelan B-Spline dan MARS Pada Nilai Ujian Masuk Terhadap IPK Mahasiswa*. Jurnal Teknik Industri, 8 (1).

Budiantara, I.N. dan Subanar. 1997. *Pemilihan Parameter Penghalus dalam Regresi Spline Terbobot.* Surabaya : Jurnal Jurusan Statistika FMIPA-ITS.

Data Statistik Indonesia. 2012. *Jumlah Penduduk Miskin dan Rata-rata Lama Sekolah*. [Online]. http://simreg.bappenas.go.id/view/data/ (diakses tanggal 16 Juli 2014).

Diana, R.,Budiantara, I.N., Purhadi, dan Satwiko Darmesto. 2012. *Estimator Smoothing Spline dalam Model Regresi Nonparametrik Multivariabel*. ProsidingSeminar Nasional Matematika.

Eubank, R.L. 1999. *Nonparametric Regression and Spline Smoothing*, *2nd ed*. New York : Dekker.

Green, P.J. dan Silverman, B.W. 1994. *Nonparametrik Regression and Generalized Linear Models (a roughness penalty approach*). New York : Chapman & Hall.

Hardle, W. 1990. *Applied Non Paramertrik Regression.* New York : Cambridge University Press.

Mahler. 1995. *Variational Solution of Penalized likelihood problem and Smoth Curve*. Annal of Statistic, 23 : 1496-1517.

Purnomo, J.D.T., Budiantara, I.N., dan Fitriasari, Kartika. 2008. *Weight Estimation Using Generalized Moving Average*. IPTEK, The Journal for Technology and Science, 19 (4).

Wahba, G. 1990. *Spline Models for Observasional Data.* Pensylvania : SIAM.