

APLIKASI MODEL REGRESI SEMIPARAMETRIK *SPLINE TRUNCATED* (Studi Kasus: Pasien Demam Berdarah Dengue (DBD) di Rumah Sakit Puri Raharja)

Ni Wayan Merry Nirmala Yani¹, I Gusti Ayu Made Srinadi^{2§}, I Wayan Sumarjaya³

¹Jurusan Matematika, Fakultas MIPA – Universitas Udayana [Email: merrynirmala72@gmail.com]

²Jurusan Matematika, Fakultas MIPA – Universitas Udayana [Email: srinadi@unud.ac.id]

³Jurusan Matematika, Fakultas MIPA – Universitas Udayana [Email: sumarjaya@unud.ac.id]

ABSTRACT

Semiparametric regression is a regression model that includes parametric components and nonparametric components in a model. The regression model in this research is truncated spline semiparametric regression with case studies of patients with Dengue Hemorrhagic Fever (DHF) at Puri Raharja Hospital during the period of January to March 2015. The best regression model estimation is obtained from the selection of optimal knots which has minimum Generalized Cross Validation (GCV) is. Parametric components in this research include age (years), body temperature ($^{\circ}C$), platelets ($\times 10^3/\mu L$), and hematocrit (%) as a nonparametric component. The minimum value of GCV is 0.03552045 achieved at the point of 39.6 knots, MSE value of 0.0296922; and the value of coefficient determination is 98.91%, obtained from semiparametric regression model truncated linear spline (order 2) with a single point of knots.

Keywords: GCV, knots, semiparametric regression, spline

1. PENDAHULUAN

Analisis regresi merupakan suatu studi yang digunakan untuk melihat hubungan antara variabel respons pada satu atau lebih variabel prediktor. Terdapat dua pendekatan dalam analisis regresi untuk mengestimasi kurva regresi, yaitu pendekatan regresi parametrik dan regresi nonparametrik. Pada model regresi parametrik diasumsikan bahwa pola fungsi diketahui seperti linear, kuadratik, kubik, polinomial derajat- p , eksponensial, dan lain sebagainya. Asumsi pada pendekatan parametrik tersebut didasarkan pada teori atau tersedianya sumber-sumber lain yang dapat memberi suatu informasi yang rinci. Apabila dalam analisis regresi bentuk atau pola kurva tidak diketahui maka pendekatan model regresi tersebut disebut regresi nonparametrik. Regresi nonparametrik memiliki fleksibilitas yang tinggi di mana data diharapkan mencari sendiri bentuk estimasi kurva regresinya. Dalam regresi nonparametrik terdapat beberapa teknik estimasi seperti

pendekatan histogram, estimator *kernel*, estimator deret *orthogonal*, analisis *wavelet*, estimator MARS, estimator deret *Fourier*, estimator *spline* dan lain-lain. Pendekatan *spline* memiliki suatu basis fungsi, di mana basis fungsi yang biasa digunakan antara lain *spline truncated* dan *B-spline* (Lyche, *et al.*, 2008).

Spline adalah salah satu jenis *piecewise polynomial* (Eubank, 1988). Maksud *piecewise polynomial* adalah polinomial yang memiliki sifat tersegmen atau sifat terpotong-potong. Model polinomial dengan sifat tersegmen menyebabkan *spline* memiliki fleksibilitas yang lebih tinggi dari model polinomial biasa, sehingga menyebabkan regresi *spline* dapat menyesuaikan diri secara lebih efektif terhadap karakteristik lokal suatu fungsi data atau dengan kata lain regresi *spline* dapat menghasilkan suatu fungsi regresi yang sesuai dengan data.

Apabila pada sebuah model regresi terdapat komponen model yang diestimasi secara parametrik dan komponen lain menggunakan

pendekatan nonparametrik terbentuklah model regresi semiparametrik. Penelitian sebelumnya mengenai pemodelan menggunakan regresi semiparametrik *spline* pernah dilakukan oleh Marina & Budiantara (2013) yang memodelkan faktor-faktor yang memengaruhi persentase kriminalitas di Jawa Timur dengan pendekatan regresi semiparametrik *spline*. Laome (2009) memodelkan regresi semiparametrik *spline* untuk data longitudinal pada kadar CD4 penderita HIV. Sugiantari & Budiantara (2013) meneliti analisis faktor-faktor yang memengaruhi angka harapan hidup di Jawa Timur menggunakan regresi semiparametrik *spline*. Wibowo *et al.* (2013) yang meneliti estimasi parameter dalam regresi semiparametrik *spline*. Regresi semiparametrik *spline truncated* digunakan dalam penelitian ini dengan studi kasus pasien Demam Berdarah Dengue (DBD) di Rumah Sakit Puri Raharja yang diambil dari penelitian Fitriani (2015).

Penelitian ini bertujuan menentukan estimasi model regresi semiparametrik dengan *spline truncated* pada data pasien DBD di Rumah Sakit Puri Raharja dan mengetahui variabel-variabel apa saja yang berpengaruh signifikan.

2. KAJIAN PUSTAKA

2.1. Analisis Regresi Parametrik

Analisis regresi merupakan suatu studi yang digunakan untuk melihat ketergantungan atau hubungan antara suatu variabel respons (variabel terikat) pada satu atau lebih variabel prediktor (variabel bebas). Analisis regresi terdiri dari dua jenis variabel yaitu variabel tak bebas atau variabel respons yang disebut juga sebagai variabel dependen. Hubungan antara variabel respons (Y_i) dengan variabel prediktor (X_i) dapat dinyatakan dalam persamaan berikut (Kutner, *et al.*, 2004):

$$Y_i = f(X_i) + \varepsilon_i ; i = 1, 2, 3, \dots, n \quad (2.1)$$

dalam hal ini Y_i adalah nilai variabel respons dalam amatan ke- i , X_i adalah peubah bebas atau variabel prediktor dalam amatan ke- i , $f(X_i)$ adalah regresi yang telah diketahui bentuknya, ε_i adalah suku galat atau sisaan yang diasumsikan

independen dan bersifat acak dengan nilai tengah nol dan variansi σ^2 dalam amatan ke- i , dan n adalah banyaknya amatan Model antara dua atau lebih variabel independen atau variabel prediktor ($X_1, X_2, X_3, \dots, X_h$) dengan variabel dependen (Y_i) secara umum dapat ditulis sebagai berikut (Neter, *et al.*, 1990):

$$Y_i = \sum_{g=0}^h \gamma_g X_{ig} + \varepsilon_i \quad (2.2)$$

dengan $i = 1, 2, 3, \dots, n$; $\gamma_0, \gamma_1, \gamma_2, \dots, \gamma_h$ adalah parameter koefisien regresi, $X_{i1}, X_{i2}, \dots, X_{ih}$ adalah peubah bebas yang diketahui nilainya, dan ε_i adalah galat atau sisaan yang saling bebas atau independen yang berdistribusi normal dengan nilai tengah nol dan variansi σ^2 dalam amatan ke- i .

Adapun koefisien determinasi dalam regresi parametrik adalah sebagai berikut (Neter, *et al.*, 1990):

$$R^2 = \frac{JKR}{JKT} = 1 - \frac{JKG}{JKT} \quad (2.3)$$

2.2. Pengujian Parameter Model

Pengujian parameter model secara serentak (simultan) dilakukan dengan menggunakan uji F . Adapun hipotesis pada uji F ini adalah sebagai berikut (Neter, *et al.*, 1990):

$$H_0: \gamma_1 = \gamma_2 = \dots = \gamma_h = 0,$$

$$H_1: \text{tidak semua } \gamma_g \text{ sama dengan nol}$$

Adapun statistik uji yang digunakan pada uji F adalah sebagai berikut:

$$F_{hitung} = \frac{KTR}{KTG} = \frac{JKR/db}{JKG/ab} \quad (2.4)$$

dengan keputusan tolak H_0 jika $F_{hitung} \geq F_{\alpha; ((p+1)-1, n-(p+1))}$ atau $p\text{-value} < \alpha$.

Sedangkan Pengujian parameter model secara parsial (individu) yaitu menggunakan uji t . Adapun hipotesis yang digunakan dalam uji t adalah sebagai berikut (Neter, *et al.*, 1990):

$$H_0: \gamma_g = 0,$$

$$H_1: \gamma_g \neq 0 ; g = 1, 2, \dots, h.$$

Adapun statistik uji yang digunakan pada uji t adalah sebagai berikut:

$$t_{hitung} = \frac{\hat{\gamma}_g}{s(\hat{\gamma}_g)} \quad (2.5)$$

dan tolak H_0 jika $|t_{hitung}| > t_{\frac{\alpha}{2}, n-(p+1)}$; dengan

n merupakan jumlah pengamatan dan $(p + 1)$ merupakan banyaknya parameter.

2.3. Pengujian Asumsi Residual

Residual (*goodness of fit*) dari suatu model regresi harus memenuhi asumsi *IIDN*($0, \sigma^2$) yaitu identik, independen, dan berdistribusi normal. Pemeriksaan asumsi homogenitas atau uji asumsi identik dapat dilakukan dengan menggunakan uji Glejser (Gujarati, 2006). Pada uji Glejser mempertimbangkan regresi nilai $|e_i|$, terhadap variabel X yang dianggap berhubungan dekat dengan varians heteroskedastisitas σ_i^2 . Sementara itu asumsi residual independen dapat dilakukan dengan menggunakan uji d Durbin-Watson, adapun uji d Durbin-Watson dapat dirumuskan sebagai berikut (Gujarati, 2006):

$$d = \frac{\sum_{i=2}^n (e_i - e_{i-1})^2}{\sum_{i=1}^n e_i^2} \quad (2.6)$$

Sedangkan pengujian asumsi normalitas dapat dilakukan dengan melakukan uji Anderson-Darling. Adapun uji Anderson-Darling dapat dirumuskan sebagai berikut (Anderson & Darling, 1954):

$$A^2 = -n - S \quad (2.7)$$

dengan $S = \frac{1}{n} \sum_{i=1}^n [2i - 1] [\log(F(Z_i)) + \log(1 - F(Z_{n+1-i}))]$ adalah simpangan baku data, $Z_i = \frac{X_i - \bar{X}}{\sigma}$ adalah data X_i yang distandarisasi, X_i adalah data ke- i yang telah diurutkan, \bar{X} adalah rata-rata data, $F(Z_i)$ adalah nilai fungsi distribusi kumulatif normal baku di Z_i , A^2 adalah statistik uji untuk metode Anderson-Darling, n adalah ukuran sampel, dan $F_0(X)$ adalah fungsi distribusi kumulatif teoritis.

2.4. Regresi Nonparametrik Spline Truncated

Secara umum model regresi nonparametrik dapat dinyatakan sebagai berikut (Eubank, 1988):

$$Y_i = f(t_i) + \varepsilon_i ; i = 1, 2, 3, \dots, n \quad (2.8)$$

dengan Y_i adalah nilai peubah respons dalam amatan ke- i , t_i adalah variabel bebas atau variabel prediktor, ε_i adalah suku galat atau sisaan yang diasumsikan independen dan

bersifat acak dengan nilai tengah nol dan variansi σ^2 , dan $f(t_i)$ adalah regresi yang tidak diketahui bentuk atau polanya.

Regresi *spline* merupakan salah satu teknik estimasi dalam regresi nonparametrik dengan model polinomial yang memiliki sifat tersegmen atau terpotong-potong yang mulus. Apabila regresi $f(t_i)$ pada persamaan (2.8) dihampiri fungsi *spline*, maka untuk mengestimasi $f(t_i)$ dapat digunakan dengan model regresi *spline*. Secara umum model regresi *spline* pada suatu fungsi dengan orde p dapat dinyatakan sebagai berikut (Eubank, 1988):

$$f(t_i) = \sum_{j=0}^{p-1} \beta_j t_i^j + \sum_{l=1}^r \beta_{(p-1+l)} (t_i - K_l)_+^{p-1} \quad (2.9)$$

dengan $\beta_0, \beta_1, \beta_2, \dots, \beta_{p-1}, \beta_{(p-1+l)}$ adalah parameter koefisien regresi, $t_i^1, t_i^2, \dots, t_i^{p-1}$ adalah peubah bebas yang diketahui nilainya, K adalah banyaknya titik knot $K_1, K_2, K_3, \dots, K_r$, dan $(t_i - K_l)_+^{p-1}$ adalah fungsi potongan (*truncated*) yang dapat diuraikan sebagai berikut:

$$(t_i - K_l)_+^{p-1} = \begin{cases} (t_i - K_l)^{p-1}, & t_i \geq K_l \\ 0, & t_i < K_l. \end{cases} \quad (2.10)$$

Apabila persamaan (2.8) disubstitusikan ke dalam persamaan (2.9) maka akan diperoleh persamaan regresi nonparametrik *spline truncated* sebagai berikut:

$$Y_i = \sum_{j=0}^{p-1} \beta_j t_i^j + \sum_{l=1}^r \beta_{(p-1+l)} (t_i - K_l)_+^{p-1} + \varepsilon_i. \quad (2.11)$$

dengan $i = 1, 2, 3, \dots, n$.

2.5. Analisis Regresi Semiparametrik

Regresi semiparametrik merupakan gabungan antara regresi parametrik dan regresi nonparametrik. Secara umum model regresi semiparametrik dapat dinyatakan sebagai berikut (Ruppert, *et al.*, 2003):

$$Y_i = f(t_i) + X_i' \gamma + \varepsilon_i \quad (2.12)$$

dengan $i = 1, 2, 3, \dots, n$, Y_i adalah nilai variabel respons dalam amatan ke- i , $X_i = (X_{i1}, X_{i2}, X_{i3}, \dots, X_{ih})$ adalah peubah bebas atau variabel prediktor yang berhubungan secara

parametrik dengan variabel respons Y_i , $\gamma = (\gamma_0, \gamma_1, \gamma_2, \dots, \gamma_h)$ adalah parameter koefisien regresi, sementara itu t_i adalah variabel prediktor yang tidak diketahui bentuk polanya terhadap variabel respons (berhubungan secara nonparametrik dengan variabel respons Y_i). Apabila model persamaan (2.2) dan (2.9) disubstitusikan ke persamaan (2.12), sehingga diperoleh:

$$Y_i = \sum_{j=0}^{p-1} \beta_j t_i^j + \sum_{l=1}^r \beta_{(p-1+l)} (t_i - K_l)_+^{p-1} + \sum_{g=0}^h \gamma_g X_{ig} + \varepsilon_i \quad (2.13)$$

2.6. Pemilihan Titik Knot Optimal

Titik knot merupakan titik perpaduan bersama yang memperlihatkan terjadinya perubahan perilaku dari fungsi *spline* pada interval-interval yang berbeda sehingga kurva yang terbentuk tersegmentasi pada titik tersebut. Pada penentuan model regresi *spline* dapat dilakukan dengan melihat nilai *Generalized Cross Validation* (GCV) yang minimum. Adapun rumus untuk menghitung GCV adalah sebagai berikut (Eubank, 1988):

$$GCV(K) = \frac{MSE(K)}{(n^{-1} \text{tr}[I - A(K)])^2} \quad (2.14)$$

dengan $MSE(K) = n^{-1} \sum_{i=1}^n (Y_i - \hat{f}(t_i))^2$, K adalah titik knot $(K_1, K_2, K_3, \dots, K_n)$, $\hat{f}(t_i) = t(t't)^{-1}t'Y$, n adalah jumlah data, I adalah matriks identitas, $A(K) = t(t't)^{-1}t'$, dan $\hat{Y}_i = A(K)Y$.

2.7. Demam Berdarah Dengue (DBD)

Penyakit DBD yang disebabkan oleh virus dengue yang termasuk kelompok B *Arthropod* virus yang dikenal sebagai genus *Flavivirus* dengan family *Flaviviridae*. Virus dengue mempunyai empat jenis serotype yaitu DEN-1, DEN-2, DEN-3 dan DEN-4. Terdapat tiga faktor pemegang peran dalam penularan infeksi virus dengue yaitu manusia, virus, dan vektor perantara. Virus dengue yang menular ke manusia ditularkan oleh gigitan nyamuk *Aedes aegypti* (Departemen Kesehatan Republik Indonesia 2004).

3. METODE PENELITIAN

Jenis data yang digunakan adalah data sekunder yaitu pasien DBD yang menjalani rawat inap di Rumah Sakit Puri Raharja Denpasar dengan sampel data rekam medis pasien Demam Berdarah Dengue (DBD) sebanyak 70 sampel periode bulan Januari sampai bulan Maret 2015 yang diambil dari penelitian Fitriani (2015).

Variabel penelitian yang digunakan terdiri dari satu variabel respons dan tujuh variabel yang diduga berpengaruh. Adapun variabel respons yang digunakan yaitu lama kesembuhan pasien (hari), jenis kelamin, umur (tahun), suhu tubuh ($^{\circ}C$), nadi (\times / menit), lama demam (hari), jumlah trombosit ($\times 10^3 / \mu L$), dan kadar hematokrit (%). Teknik analisis data dalam model regresi semiparametrik *spline truncated* adalah sebagai berikut:

1. Menetapkan komponen parametrik dan komponen nonparametrik berdasarkan penelitian sebelumnya yang dilakukan oleh Fitriani (2015).
2. Memilih titik *knot* optimal dengan menggunakan *Generalized Cross Validation* (GCV) yang paling minimum.
3. Memodelkan data lama kesembuhan pasien DBD yang pernah menjalani rawat inap di Rumah Sakit Puri Raharja Denpasar (variabel respons) dengan variabel prediktor yang telah ditetapkan dengan regresi Semiparametrik *Spline Truncated* dengan titik *knot* optimal.
4. Menghitung nilai koefisien determinasi (R^2).
5. Menguji signifikansi parameter regresi Semiparametrik *Spline Truncated* secara serentak dengan uji F .
6. Melakukan uji parameter regresi Semiparametrik *Spline Truncated* secara parsial dengan uji t .
7. Menguji asumsi residual IIDN dari model regresi Semiparametrik *Spline Truncated*.
8. Menginterpretasikan model, membandingkan hasil analisis yang di-peroleh dengan menggunakan regresi semiparametrik

spline truncated terhadap estimasi regresi semiparametrik menggunakan estimator kernel uniform yang sebelumnya telah dianalisis oleh Fitriani (2015), dan menarik kesimpulan.

4. HASIL DAN PEMBAHASAN

4.1. Deskripsi Data

Gambaran umum data penelitian yaitu lama kesembuhan pasien DBD (hari) sebagai variabel respons dengan peubah bebas yaitu umur (tahun), suhu tubuh ($^{\circ}\text{C}$), nadi (kali/menit), kadar hematokrit (%), trombosit ($\times 10^3/\mu\text{L}$), dan lama demam (hari) diringkas dalam statistika deskriptif pada Tabel 1.

Tabel 1. Statistika Deskriptif Data Pasien DBD

Variabel	Rangkasan Statistik			
	Min	Maks	Mean	StDev
Lama Kesembuhan (Y)	3,0	10,0	7,329	1,666
Suhu (S)	36,0	40,3	37,464	1,174
Umur (U)	4,0	80,0	32,040	16,930
Lama Demam (LD)	1,0	9,0	3,700	1,397
Nadi (N)	60,0	112,0	84,830	9,950
Trombosit (PLT)	7,0	239,0	88,070	51,670
Kadar Hematokrit (HTC)	32,9	51,9	39,757	4,363

4.2. Penentuan Komponen Parametrik dan Komponen Nonparametrik

Penentuan komponen parametrik dan komponen nonparametrik ditentukan berdasarkan informasi sebelumnya yang dilakukan oleh Fitriani (2015). Pada penelitian tersebut variabel umur, suhu tubuh, dan trombosit diasumsikan sebagai komponen parametrik dan variabel kadar hematokrit diasumsikan sebagai komponen nonparametrik sedangkan variabel lain yaitu lama demam dan nadi diasumsi tidak berpengaruh signifikan terhadap lama kesembuhan pasien yang sudah diwakili oleh variabel bebas lain yang signifikan.

4.3. Pemilihan Titik Knot Optimal Regresi Semiparametrik *Spline Truncated*

Model yang dicobakan adalah regresi semiparametrik *spline truncated* dengan orde 2

dan orde 3, dengan satu knot, dua knot, dan tiga knot. Pemilihan titik knot optimal yang dicobakan adalah sebagai berikut.

- a. Titik knot optimal regresi semiparametrik *spline truncated* linear (orde 2) dengan satu titik knot, diperoleh model sebagai berikut:

$$Y_i = \beta_0 + \beta_1 t_{i1} + \beta_2 (t_{i1} - K_1)_+ + \gamma_1 X_{i1} + \gamma_2 X_{i2} + \gamma_3 X_{i3} + \varepsilon_i.$$

Nilai GCV minimum sebesar 0,03552 dicapai pada titik knot 39,6.

- b. Titik knot optimal regresi semiparametrik *spline truncated* linear (orde 2) dengan dua titik knot, diperoleh model sebagai berikut:

$$Y_i = \beta_0 + \beta_1 t_{i1} + \beta_2 (t_{i1} - K_1)_+ + \beta_3 (t_{i1} - K_2)_+ + \gamma_1 X_{i1} + \gamma_2 X_{i2} + \gamma_3 X_{i3} + \varepsilon_i.$$

Untuk model ini diperoleh nilai GCV minimum sebesar 0,03567 dicapai pada titik knot 39,7 dan titik knot 39,9.

- c. Pemilihan titik knot optimal regresi semiparametrik *spline truncated* linear (orde 2) dengan tiga titik knot, diperoleh model sebagai berikut:

$$Y_i = \beta_0 + \beta_1 t_{i1} + \beta_2 (t_{i1} - K_1)_+ + \beta_3 (t_{i1} - K_2)_+ + \beta_4 (t_{i1} - K_3)_+ + \gamma_1 X_{i1} + \gamma_2 X_{i2} + \gamma_3 X_{i3} + \varepsilon_i.$$

Dari model ini diperoleh nilai GCV minimum sebesar 0,03707 dicapai pada titik knot 34,1; titik knot 39,2; dan pada titik knot 47,7.

- d. Pemilihan titik knot optimal regresi semiparametrik *spline truncated* kuadratik (orde 3) dengan satu titik knot, diperoleh model sebagai berikut:

$$Y_i = \beta_0 + \beta_1 t_{i1} + \beta_2 t_{i1}^2 + \beta_3 (t_{i1} - K_1)_+^2 + \gamma_1 X_{i1} + \gamma_2 X_{i2} + \gamma_3 X_{i3} + \varepsilon_i.$$

Nilai GCV minimum sebesar 0,03632 dicapai pada titik knot 33,8.

- e. Pemilihan titik knot optimal regresi semiparametrik *spline truncated* kuadratik (orde 3) dengan dua titik knot, diperoleh model sebagai berikut:

$$Y_i = \beta_0 + \beta_1 t_{i1} + \beta_2 t_{i1}^2 + \beta_3 (t_{i1} - K_1)_+^2 + \beta_4 (t_{i1} - K_2)_+^2 + \gamma_1 X_{i1} + \gamma_2 X_{i2} + \gamma_3 X_{i3} + \varepsilon_i.$$

Untuk model ini diperoleh nilai GCV minimum sebesar 0,03682 dicapai pada titik knot 33,7 dan pada titik knot 40,5.

- f. Pemilihan titik knot optimal regresi semiparametrik *spline truncated* kuadrat (orde 3) dengan tiga titik knot, diperoleh model sebagai berikut:

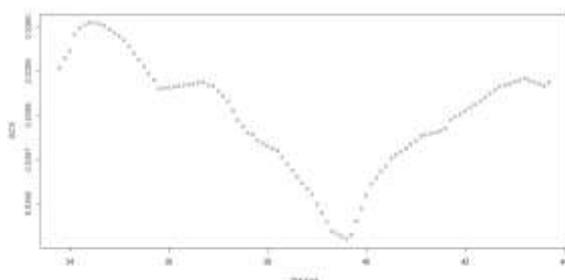
$$Y_i = \beta_0 + \beta_1 t_{i1} + \beta_2 t_{i1}^2 + \beta_3 (t_{i1} - K_1)_+^2 + \beta_4 (t_{i1} - K_2)_+^2 + \beta_5 (t_{i1} - K_3)_+^2 + \gamma_1 X_{i1} + \gamma_2 X_{i2} + \gamma_3 X_{i3} + \varepsilon_i.$$

Nilai GCV minimum sebesar 0,03795 dicapai pada titik knot 34,1; titik 35,2; dan pada titik 416,1.

Tabel 2. Nilai GCV dari Variasi Titik Knot dan Orde

Banyak Titik Knot	Orde	GCV
1 titik knot (39,6)	orde 2	0,03552
2 titik knot (39,7 dan 39,9)	orde 2	0,03567
3 titik knot (34,1; 39,2; dan 47,7)	orde 2	0,03707
1 titik knot (33,8)	orde 3	0,03632
2 titik knot (33,7 dan 40,5)	orde 3	0,03682
3 titik knot (34,1; 35,5; dan 46,1)	orde 3	0,03795

Dari enam nilai GCV diperoleh nilai GCV yang paling minimum adalah pada saat orde dua dengan satu titik knot. Sehingga pemilihan titik knot optimal terletak pada titik 39,6 dengan nilai GCV sebesar 0,03552. Plot titik knot dan nilai GCV secara kasar dapat dilihat pada Gambar 1.



Gambar 1. Nilai GCV regresi semiparametrik *spline truncated* linear satu titik knot

4.4. Penentuan Estimasi Model Regresi Semiparametrik *Spline Truncated*

Dari pemilihan titik knot optimal diperoleh estimasi model yang memiliki nilai GCV minimum yaitu pada saat orde 2 dengan satu titik knot yang terletak pada titik 39,6. Adapun estimasi model yang diperoleh adalah sebagai berikut:

$$\hat{Y}_i = -0,55021 - 0,01136HCT_i + 0,01648(HCT_i - 39,6)_+ + 0,28815S_i - 0,04751U_i - 0,01104PLT_i$$

dengan Y_i adalah lama kesembuhan pasien DBD yang menjalani rawat inap di Rumah Sakit Puri Raharja Denpasar, t_1 adalah kadar hematokrit (HCT), X_1 adalah suhu (S), X_2 adalah umur (U), dan X_3 adalah trombosit (PLT). Dengan nilai GCV sebesar 0,03552; MSE sebesar 0,02969; dan koefisien determinasi atau $R^2 = 0,98914$, hal ini menunjukkan bahwa dengan memuat enam parameter variabel kadar hematokrit, suhu, umur, dan trombosit mampu menerangkan 98,914% keragaman lama kesembuhan pasien DBD yang menjalani rawat inap di Rumah Sakit Puri Raharja Denpasar. Sisanya 1,086% dipengaruhi oleh variabel lain di luar model ataupun *error*.

4.5. Pengujian Parameter Model

Pengujian parameter model dilakukan secara serentak (simultan) kemudian dilanjutkan dengan pengujian secara parsial (individu).

Tabel 3. Analisis Variansi Regresi Semiparametrik *Spline Truncated* Linear

Sumber Keragaman	Df	JK	KT	F_{hitung}	F_{tabel} $\alpha = 1\%$
Regresi	5	189,3644	37,8729	1166,186	3,3177
Galat	64	2,0784	0,0325		
Total	69	191,4429			

Dari Tabel 3, diperoleh kesimpulan bahwa $F_{hitung} \geq F_{tabel}$ yaitu $1166,186 \geq 3,3177$ maka tolak H_0 yang mengindikasikan bahwa tidak semua koefisien regresi bernilai nol atau dengan kata lain terdapat pengaruh yang signifikan secara bersama-sama antara variabel bebas terhadap variabel terikat, sehingga model

signifikan. Selanjutnya dilakukan uji parameter secara parsial (secara individu) dengan uji t . Hasil pengujian tersebut disajikan dalam bentuk Tabel 4 berikut.

Tabel 4. Uji Individu Estimasi Model Regresi Semiparametrik *Spline Truncated Linear*

Parameter	Estimasi	Stdev	t-hitung
β_0	-0,55021	0,30671	-1,79393
β_1	-0,01136	0,00205	-5,53007
β_2	0,01648	0,00317	5,20711
γ_1	0,28815	0,00764	37,70848
γ_2	-0,04751	0,00088	-54,21042
γ_3	-0,01104	0,00023	-48,7238

Berdasarkan taraf nyata $\alpha = 1\%$ dan nilai $t - \text{tabel} = 2,654854$ yang mengindikasikan tolak H_0 . Dengan demikian dapat disimpulkan bahwa masing-masing parameter berpengaruh secara signifikan terhadap lama kesembuhan pasien.

4.6. Pengujian Residual Model

Residual (*goodness of fit*) dari suatu model regresi harus memenuhi asumsi $IIDN(0, \sigma^2)$ yaitu identik, independen dan berdistribusi normal dengan mean nol dan variansi σ^2 . Uji asumsi homogenitas dilakukan dengan menggunakan uji Glejser. Pada Tabel 5 diperoleh nilai F_{hit} sebesar 0,421 dan nilai signifikasinya sebesar 0,793 pada taraf nyata $\alpha = 1\%$, diperoleh bahwa nilai signifikasinya lebih besar dari 0,01. Maka dapat disimpulkan bahwa semua variabel tidak berpengaruh signifikan terhadap nilai mutlak residual. Hal tersebut membuktikan bahwa varians residual memenuhi asumsi homokedastisitas atau dengan kata lain tidak terjadi heteroskedastisitas.

Tabel 5. Analisis Variansi Uji Glejser

Sumber Keragaman	Db	JK	KT	F_{hit}	P - value
Regresi	4	0,019	0,005	0,421	0,793
Galat	65	0,724	0,011		
Total	69	0,743			

Selanjutnya dilakukan uji asumsi independen dengan menggunakan uji d Durbin-

Watson, diperoleh nilai d Durbin-Watson sebesar 1,579. Selanjutnya nilai d Durbin-Watson tersebut akan dibandingkan dengan nilai tabel signifikasi $\alpha = 1\%$, dengan jumlah sampel sebanyak 70 ($T = 70$), satu variabel dependen dan empat variabel independen ($k = 5$). Dari tabel d Durbin-Watson dengan $\alpha = 1\%$, $T = 70$, dan $k = 5$ diperoleh nilai d_L dan d_U secara berturut-turut yaitu 1,34263 dan 1,57778 dengan:

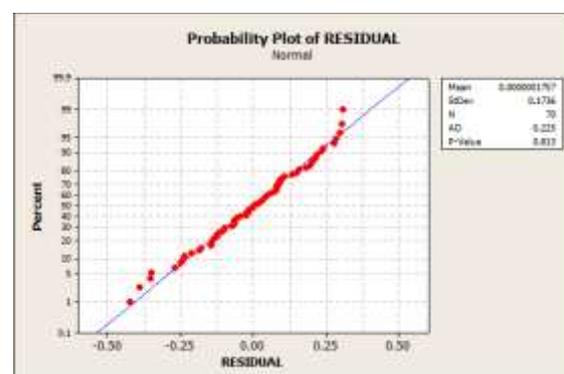
$$4 - d_L = 4 - 1,34263 = 2,65737 \text{ dan } 4 - d_U = 4 - 1,57778 = 2,42222.$$

Karena nilai dari d Durbin-Watson terletak di antara nilai d_U dan $4 - d_U$ maka terima H_0 yang mengindikasikan bahwa tidak terdapat autokorelasi positif ataupun negatif pada residual.

Tabel 6. Uji d Durbin-Watson

Model	R	R-Square	Adjusted R-Square	Std. Error of the Estimate	d Durbin-Watson
1	0,000	0,000	-0,062	0,7882	1,570

Sementara itu uji normalitas dilakukan dengan melakukan uji Anderson-Darling, diperoleh nilai Anderson-Darling sebesar 0,225 dan $p - \text{value}$ sebesar 0,813 pada taraf nyata $\alpha = 1\%$. Karena $p - \text{value}$ lebih besar dari 0,01, maka terima H_0 yang mengindikasikan bahwa residual model memenuhi asumsi distribusi normal.



Gambar 3. Uji Anderson-Darling

Dengan demikian dapat disimpulkan bahwa residual dari model regresi semiparametrik *spline truncated linear* dengan satu titik knot

memenuhi asumsi $IIDN(0, \sigma^2)$ yaitu identik, independen dan berdistribusi normal.

4.7. Interpretasi Model Regresi Semiparametrik *Spline Truncated* Linear dengan Satu Titik Knot

Model terbaik lama kesembuhan pasien DBD di Rumah Sakit Puri Raharja Denpasar menggunakan regresi semiparametrik *spline truncated* adalah model semiparametrik *spline truncated* linear (orde 2) dengan satu titik knot. Adapun interpretasi model regresi semiparametrik *spline truncated* linear (orde 2) dengan satu titik knot adalah sebagai berikut. Dengan mengasumsikan suhu (S), umur (U), dan trombosit (PLT), konstan maka pengaruh kadar hematokrit (HCT), terhadap lama kesembuhan pasien (Y), apabila kadar hematokrit naik sebesar 1% pada saat kadar hematokrit kurang dari 39,6%, maka Y akan mengalami penurunan sebesar 0,01136. Koefisien bernilai negatif artinya terjadi hubungan negatif antara kadar hematokrit dengan Y. Hal ini mengindikasikan bahwa apabila persentase kadar hematokrit menurun maka berakibat pada jenjang waktu yang lebih cepat pada kesembuhan pasien. Sedangkan ketika kadar hematokrit naik sebesar 1% pada saat kadar hematokrit lebih besar atau sama dengan 39,6% maka Y akan mengalami peningkatan sebesar 0,00512. Koefisien bernilai positif artinya terjadi hubungan positif antara kadar hematokrit dengan Y. Hal ini mengindikasikan bahwa apabila persentase kadar hematokrit meningkat maka berakibat pada jenjang waktu yang lebih lama pada kesembuhan pasien. Selanjutnya dengan mengasumsikan variabel yang lain konstan, maka interpretasi model berdasarkan masing-masing variabel (variabel suhu (S), umur (U), dan jumlah trombosit (PLT)) adalah sebagai berikut.

Apabila suhu (S) tubuh mengalami kenaikan 1°C maka Y akan mengalami kenaikan sebesar 0,28815. Koefisien bernilai positif artinya terjadi hubungan positif antara suhu (S) dengan Y. Hal ini mengindikasikan bahwa ketika suhu

pasien meningkat menyebabkan kesembuhan pasien cenderung menjadi lebih lama.

Pada saat umur (U) mengalami kenaikan 1 tahun dan variabel bebas lainnya tetap, maka akan menurunkan Y sebesar 0,04751. Koefisien bernilai negatif artinya terjadi hubungan negatif antara umur dengan Y. Hal ini mengindikasikan bahwa semakin bertambahnya umur (U) pasien menyebabkan kesembuhan pasien cenderung menjadi lebih cepat.

Sedangkan ketika jumlah trombosit (PLT) mengalami kenaikan $1(\times 10^3/\mu\text{L})$, dan variabel bebas lainnya tetap, maka akan menurunkan Y sebesar 0,01104. Koefisien bernilai negatif artinya terjadi hubungan negatif antara PLT dengan Y. Hal ini mengindikasikan bahwa semakin meningkat jumlah trombosit pasien, menyebabkan kesembuhannya pasien cenderung semakin cepat.

4.8. Hasil Regresi Semiparametrik *Spline Truncated* Linear dengan Estimasi Fungsi Regresi Semiparametrik dan Estimator *Kernel Uniform*

Berdasarkan penelitian sebelumnya yang dilakukan oleh Fitriani (2015), variabel umur, suhu tubuh, dan trombosit diasumsikan sebagai komponen parametrik dan variabel kadar hematokrit diasumsikan sebagai komponen nonparametrik sedangkan lama demam dan nadi diasumsikan tidak berpengaruh signifikan terhadap lama kesembuhan pasien yang sudah diwakili oleh variabel bebas lain yang signifikan.

Penduga fungsi regresi semiparametrik dengan *estimator kernel uniform* (\hat{y}_i), yang dilakukan oleh Fitriani (2015) diperoleh model regresi semiparametrik sebagai berikut:

$$\hat{y}_i = -0,005496 + 0,265939 S - 0,472707 U - 0,330655 \text{ PLT} + \hat{m}_h(t_i)$$

dengan *bandwidth* optimal sebesar 0,005, nilai GCV minimum sebesar $5,329214\text{e-}09$, nilai MSE minimum sebesar 0,00437 dan dengan nilai koefisien determinasi atau R^2 sebesar 98,90%.

Sedangkan pada regresi semiparametrik *spline truncated* linear diperoleh model sebagai

berikut:

$$\hat{Y}_i = -0,55021 - 0,01136HCT_i + 0,01648(HCT_i - 39,6)_+ + 0,28815S_i - 0,04751U_i - 0,01104PLT_i$$

dengan nilai GCV minimum sebesar 0,03552, nilai MSE sebesar 0,02969, nilai koefisien determinasi atau R^2 sebesar 98,91% yang terletak pada titik 39,6 dan dengan enam parameter.

5. KESIMPULAN DAN SARAN

Variabel yang signifikan berpengaruh terhadap model regresi semiparametrik *spline truncated* linear (orde 2) dengan satu titik knot yaitu umur (tahun), suhu tubuh ($^{\circ}\text{C}$), trombosit ($\times 10^3/\mu\text{L}$), dan kadar hematokrit (%). Nilai GCV minimum sebesar 0,03552045, nilai MSE sebesar 0,0296922, nilai R^2 sebesar 98,91% dicapai pada titik knot 39,6.

Dalam penelitian ini dibahas model regresi semiparametrik untuk satu variabel pada komponen nonparametriknya sehingga penelitian lebih lanjut dapat dilakukan untuk lebih dari satu variabel komponen nonparametrik.

DAFTAR PUSTAKA

- Anderson, T. W., and D. A. Darling. 1954. "A Test of Goodness of Fit." *Journal of the American Statistical Association* 49 (268): 765-769.
- Departemen Kesehatan Republik Indonesia . 2004. *Tata Laksana Demam Berdarah Dengue di Indonesia*. Ketiga. Jakarta: Departemen Kesehatan.
- Eubank, R. L. 1988. *Spline Smoothing and Nonparametric Regression*. New York: Marcel Dekker. inc.
- Fitriani, A. 2015. "Estimasi Model Regresi Semiparametrik Menggunakan Estimator Kernel Uniform (Studi Kasus: Pasien DBD di RS Puri Raharja)." *Skripsi*. Jurusan Matematika Universitas Udayana.
- Gujarati, D. N. 2006. *Essential of Econometrics*. Translated by Julius A. Mulyadi and Yelvi Andri. United States Military Academy, West Poit: McGraw-Hill Companies.
- Kutner, M. H., C. J. Nachtsheim, J. Neter, and W. Li. 2004. *Applied Linear Statistical Model*. 4. New York: McGraw-Hill/Irwin.
- Laome, L. 2009. "Model Regresi Semiparametrik Spline Untuk Data Longitudinal pada Kasus Kadar CD4 Penderita HIV." *Paradigma* 13 (2): 101-106.
- Lyche, T., and K. Morken. 2008. *Spline Methods Draft*. Norwegia: Department of Informatics Centre of Mathematics for Applications University of Oslo.
- Marina, S.M. T., and I N. Budiantara. 2013. "Pemodelan Faktor-Faktor yang Mempengaruhi Persentase Kriminalitas di Jawa Timur dengan Pendekatan Regresi Semiparametrik Spline." *Jurnal Sains dan Seni POMITS* 2 (2): D-147 - D-152.
- Neter, J., W. Wasserman, and M. H. Kutner. 1990. *Applied Linear Statistical Models*. Third. Translated by Bambang Sumantri. New York: McGraw-Hill College.
- Ruppert, D., M. P. Wand, and R. J. Carroll. 2003. *Semiparametric Regression*. New York: Cambridge University Press.
- Sugiantari, A. P., and I N. Budiantara. 2013. "Analisis Faktor-Faktor yang Mempengaruhi Angka Harapan Hidup di Jawa Timur Menggunakan Regresi Semiparametrik Spline." *Jurnal Sains dan Seni POMITS* 2 (1): D-37 - D-41.
- Wibowo, W., S. Haryatmi, and I N Budiantara. 2013. "Kajian Metode Estimasi Para-meter dalam Regresi Semi-parametrik Spline." *Berkala MIPA* 23 (1): 102-110.