

BAB I

PENDAHULUAN

1.1 Latar Belakang

Dalam berbagai bidang, seringkali diperoleh hasil pengamatan berupa data longitudinal. Karakteristik data longitudinal adalah jika suatu individu yang diamati lebih dari satu kali (pengamatan berulang) dalam periode waktu tertentu, sehingga nilai peubah respon berkorelasi dalam satu individu dan bebas (*independen*) antar individu (Wu dan Zhang, 2006: 2). Data longitudinal merupakan hasil pengamatan yang dilakukan berulang pada rentang waktu, sedangkan pada data *cross section* pengamatan dilakukan pada satu titik waktu. Data longitudinal memiliki keunggulan dibandingkan data *cross section* yaitu kekuatan uji statistik yang lebih besar, penduga parameter yang lebih efisien serta mampu menjelaskan pengaruh perubahan nilai peubah respon terhadap waktu.

Analisis statistika untuk mengetahui pengaruh hubungan peubah penjelas terhadap peubah respon adalah analisis regresi. Pendekatan regresi parametrik lebih sering digunakan dan didasarkan pada bentuk kecenderungan kurva yaitu polinomial derajat ke- k . Dalam perkembangannya, kurva yang terbentuk dari hubungan peubah penjelas terhadap peubah respon tidak selalu mengikuti bentuk standar. Menurut Budiantara (2006:77-78), pendekatan regresi parametrik untuk tujuan pemodelan dan prediksi, di mana bentuk kecenderungan kurva tidak standar akan menimbulkan galat yang relatif besar. Karena memodelkan peubah waktu (t_{ij}) terhadap peubah respon (y_{ij}) secara parametrik tidak dapat dilakukan, maka digunakan pendekatan nonparametrik. Permasalahan dalam regresi nonparametrik adalah bagaimana mendapatkan penduga model regresi.

Teknik pemulusan dalam regresi nonparametrik data longitudinal antara lain: histrogram (Green dan Silverman, 1994), penduga Kernel (Hardle, 1990), penduga K-nearest Neighbour (Hardle, 1990), deret Fourier (Eubank, 1988) dan penduga *Penalized Spline* (Wahba, 1990). Pendekatan yang digunakan dalam penelitian ini adalah penduga *Penalized Spline*, karena digunakan untuk menduga parameter model pada data yang mengalami perubahan perilaku dalam interval tertentu. *Penalized Spline* merupakan potongan-potongan polinomial yang digabungkan bersama titik-titik knot τ_r .

Wu dan Zhang (2006: 203) memperkenalkan pendekatan *Generalized Penalized Spline* dilengkapi dengan matriks pembobot sebagai penduga fungsi rata-rata populasi $\eta(t_{ij})$ tanpa didasari asumsi *homoskedastisitas*. Titik-titik knot τ_r pada pendekatan ini, terletak pada kuantil peubah penjelas yang ditetapkan secara pasti. Pendugaan model dilakukan dengan memilih derajat polinomial (k), jumlah knot optimal (K) dan nilai bobot (λ) optimal. Algoritma *full search* didasarkan nilai *Generalized Cross Validation* (GCV) terkecil untuk memilih K dan λ optimal, digunakan dalam penelitian ini (Ruppert, 2000: 5).

Penelitian ini akan membahas mengenai pendugaan model regresi nonparametrik pada data longitudinal menggunakan penduga *Generalized Penalized Spline* dan menerapkan metode tersebut pada data pertumbuhan berat badan bayi.

1.2 Rumusan masalah

Rumusan masalah dalam penelitian :

1. Bagaimana menduga parameter model data longitudinal satu peubah respon menggunakan pendekatan *Generalized Penalized Spline*?
2. Bagaimana menerapkan metode *Generalized Penalized Spline* pada data pertumbuhan berat badan bayi?

1.3 Tujuan

Berdasarkan rumusan masalah, tujuan penelitian ini adalah:

1. Menduga parameter model data longitudinal satu peubah respon dengan *Generalized Penalized Spline*.
2. Menerapkan metode *Generalized Penalized Spline* pada data pertumbuhan berat badan bayi.

1.4 Manfaat Penelitian

Manfaat yang diperoleh dari penelitian adalah:

1. Memberikan masukan kepada pembaca dalam memodelkan peubah penjelas terhadap peubah respon pada data longitudinal secara nonparametrik.
2. Mempelajari perilaku model serta perubahan nilai respon terhadap waktu.

3. Penggunaan regresi nonparametrik longitudinal berdasarkan metode *Generalized Penalized Spline* pada pertumbuhan berat badan bayi.

1.4 Batasan Masalah

Penelitian dibatasi pada masalah:

1. Data longitudinal seimbang dengan satu peubah respon.
2. Metode *Generalized Penalized Least Square* untuk menduga parameter model.
3. Derajat polinomial dibatasi pada $k = 1, 2, 3$ dan banyaknya titik knot $K=1, 2, 3$.
4. Algoritma *full search* dengan nilai GCV terkecil untuk menentukan λ optimal.



UNIVERSITAS BRAWIJAYA



BAB II TINJAUAN PUSTAKA

2.1 Data Longitudinal

Data merupakan hasil pengukuran atau pengamatan berupa angka, besaran, fakta maupun pernyataan untuk menggambarkan karakteristik obyek (Walpole, 1995:1). Menurut Weis (2005:2), data longitudinal adalah hasil pengamatan berulang berdasarkan waktu pada sejumlah individu. Data longitudinal merupakan gabungan antara data *time series* dan data *cross section*. Data *time series* diperoleh dari hasil pengukuran suatu peubah dalam runtun (berturut-turut) waktu tertentu, sedangkan data *cross section* dihasilkan dari individu berbeda pada suatu titik waktu tertentu.

Perbedaan data longitudinal pada *cross section* adalah pada data *cross section* pengamatan yang dilakukan pada satu titik waktu, sedangkan data longitudinal dihasilkan dari pengukuran berulang pada rentang waktu tertentu. Ada beberapa keuntungan mempelajari data longitudinal dibandingkan *cross section*. Pertama, data longitudinal lebih kuat dari data *cross section* pada ukuran contoh sama, dengan kata lain untuk memperoleh kekuatan uji statistik yang sama data longitudinal membutuhkan lebih kecil ukuran contoh. Kedua, dengan ukuran contoh sama, penduga parameter data longitudinal lebih efisien dibandingkan penduga parameter data *cross section*. Ketiga, data longitudinal mampu menyediakan informasi tentang perubahan nilai peubah respon (Laome, 2008: 1).

Tujuan analisis data longitudinal adalah untuk mempelajari bagaimana nilai peubah respon mengalami perubahan dari waktu ke waktu. Karena pengamatan yang menghasilkan data longitudinal dilakukan lebih dari satu unit waktu, maka terdapat korelasi antar nilai peubah respon. Diggle, et all (1995: 2) telah menunjukkan bahwa menggunakan analisis *cross section* pada data longitudinal akan menghasilkan kesimpulan yang tidak sah. Beberapa nama lain data longitudinal menurut Gujarati (2004: 636): panel data (bidang ekonomi), *pooled data*, *combination of time series and cross section data*, *micropanel data*, *event history analysis* dan *cohort analysis*. Terdapat dua macam data longitudinal, yaitu:

1. Data longitudinal seimbang dihasilkan dari banyaknya titik waktu pengamatan yang sama untuk setiap individu. Misal pada penelitian

di bidang pertanian, pengaruh pemberian kadmium (Cd) terhadap pertumbuhan lebar daun, sebanyak n tanaman kangkung (peubah respon) dengan peubah penjelas yaitu hari ke-14 sampai hari ke-37 sesudah tanam. Kerangka hasil pengukuran lebar daun tanaman kangkung menghasilkan data longitudinal seimbang yaitu :

Tabel 2.1. Kerangka Pengukuran Lebar Daun Kangkung

Tanaman kangkung ke-i	Lebar daun kangkung (cm) pada hari ke-j setelah tanam						
	14	15	16	17	18	...	37
1	y_{114}	y_{115}	y_{116}	y_{117}	y_{118}	...	y_{137}
2	y_{214}	y_{215}	y_{216}	y_{217}	y_{218}	...	y_{237}
3	y_{314}	y_{315}	y_{316}	y_{317}	y_{318}	...	y_{337}
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
n	y_{n14}	y_{n15}	y_{n16}	y_{n17}	y_{n18}	...	y_{n37}

Struktur data longitudinal disajikan pada Tabel 2.1:

Tabel 2.2. Struktur Data Longitudinal Seimbang

i	j	t_{ij}	y_{ij}	y_i
1	1	t_{11}	y_{11}	y_1
	2	t_{12}	y_{12}	
	⋮	⋮	⋮	
	T	t_{1T}	y_{1T}	
2	1	t_{21}	y_{21}	y_2
	2	t_{21}	y_{21}	
	⋮	⋮	⋮	
	T	t_{2T}	y_{2T}	
⋮	⋮	⋮	⋮	⋮
n	1	t_{n1}	y_{n1}	y_n
	2	t_{n2}	y_{n2}	
	⋮	⋮	⋮	
	T	t_{nT}	y_{nT}	

- $i = 1, \dots, n$; n = banyaknya tanaman kangkung
 $j = 1, \dots, T$; T = banyaknya titik waktu pengukuran setiap kangkung
 t_{ij} : titik waktu ke- j pada kangkung ke- i
 y_{ij} : lebar daun kangkung ke- i pada titik waktu ke- j

2. Data longitudinal tidak seimbang dihasilkan dari banyaknya titik waktu pengukuran berbeda untuk setiap individu. Misal, pengukuran kadar gula n pasien diabetes setiap hari hingga mengalami kesembuhan menghasilkan data longitudinal tidak seimbang. Karena pasien sembuh dalam rentang waktu berbeda, maka banyaknya titik waktu pengukuran kadar gula antar pasien tidak sama. Skema hasil pengukuran kadar gula pasien diabetes adalah :

Tabel 2.3. Kerangka Pengukuran Kadar Gula Pasien Diabetes

Pasien ke-	Kadar gula (mg/dl) pasien diabetes pada hari ke- j hingga sembuh						
	1	2	3	4	5	...	T_i
1	y_{11}	y_{12}	y_{13}				
			$T_1 = 3$				
2	y_{21}	y_{22}	y_{23}	y_{24}	y_{25}		
					$T_2 = 5$		
3	y_{31}	y_{32}					
		$T_3 = 2$					
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
n	y_{n1}	y_{n2}	y_{n3}	y_{n4}	y_{n5}	...	y_{nT_i}
							$T_n = T_i$

$i = 1, \dots, n$; n = banyaknya pasien diabetes

$j = 1, \dots, T_i$; T_i = banyaknya titik waktu pengukuran kadar gula pasien ke- i

t_{ij} : titik waktu ke- j pada pasien diabetes ke- i

y_{ij} : kadar gula pasien diabetes ke- i pada waktu ke- j

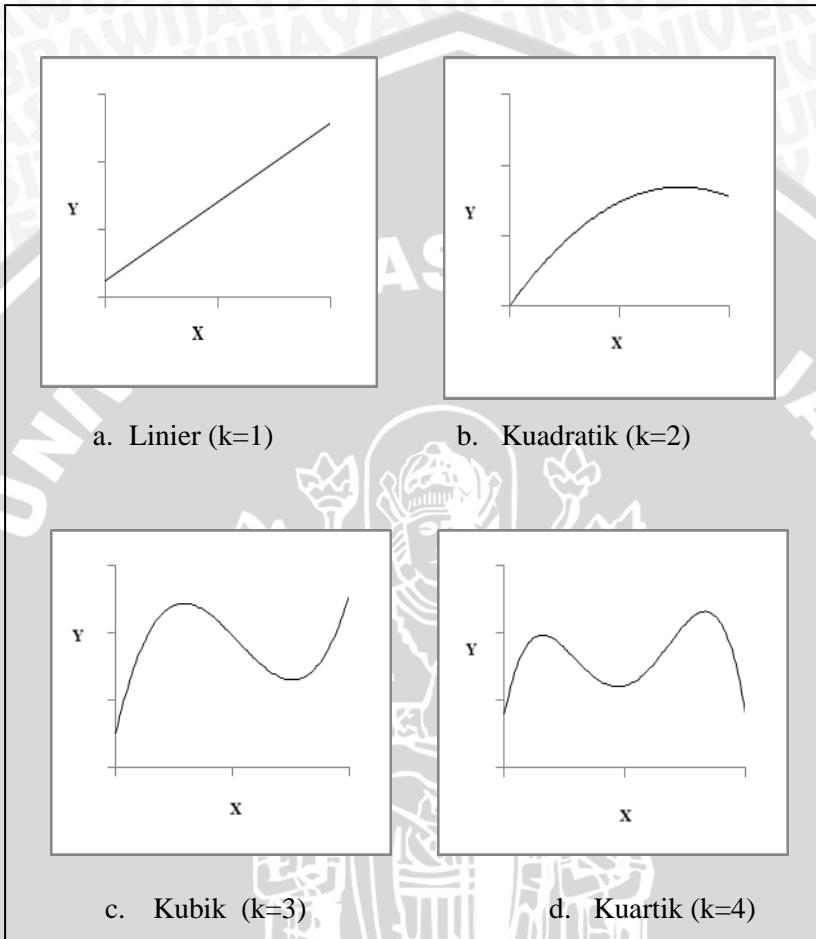
Berikut ini adalah struktur data longitudinal tidak seimbang :

Tabel 2.4. Struktur Data Longitudinal Tidak Seimbang

i	j	t_{ij}	y_{ij}	y_i
1	1	t_{11}	y_{11}	y_1
	2	t_{12}	y_{12}	
	\vdots	\vdots	\vdots	
	T_1	t_{1T_1}	y_{1T_1}	
2	1	t_{21}	y_{21}	y_2
	2	t_{22}	y_{22}	
	\vdots	\vdots	\vdots	
	T_2	t_{2T_2}	y_{2T_2}	
\vdots	\vdots	\vdots	\vdots	\vdots
n	1	t_{n1}	y_{n1}	y_n
	2	t_{n2}	y_{n2}	
	\vdots	\vdots	\vdots	
	T_n	t_{nT_n}	y_{nT_n}	

2.2 Regresi Nonparametrik

Analisis regresi digunakan untuk menduga pola hubungan antar peubah penjelas dengan peubah respon (terikat) (Eubank, 1988: 1). Terdapat dua pendekatan pendugaan model regresi yaitu pendekatan parametrik dan nonparametrik. Pendekatan parametrik digunakan jika kurva regresi dapat diwakili oleh suatu model parametrik. Bentuk kecenderungan kurva regresi parametrik mengikuti salah satu derajat polinomial k ($k=1,2,3, \dots$). Secara deskriptif diagram pencar digunakan untuk menentukan bentuk kecenderungan hubungan, sedangkan secara inferensia digunakan statistik uji F dan koefisien determinasi (R^2). Bentuk polinomial kurva regresi dijelaskan pada gambar berikut:



Gambar 2.1. Berbagai Bentuk Kurva Polinomial

Apabila bentuk kecenderungan tidak mengikuti kurva standar (polinomial derajat ke- k), maka digunakan regresi nonparametrik. Pendekatan regresi nonparametrik memiliki fleksibilitas tinggi karena data dapat mencari sendiri penduga model tanpa didasarkan pada asumsi bentuk kurva tertentu (misal kurva polinomial derajat ke- k) dan faktor rancangan penelitian (Eubank, 1988: 10). Model penduga pendekatan nonparametrik dapat disesuaikan berdasarkan bentuk kecenderungan

hubungan antar peubah respon dan peubah penjelas, sehingga lebih fleksibel.

Model parametrik data longitudinal mengasumsikan peubah respon mengikuti fungsi parametrik baik pengaruh tetap (*fixed effect*) maupun pengaruh acak (*random effect*) serta mengasumsikan hubungan yang terbentuk antara peubah respon dan peubah penjelas setiap individu adalah parametrik. Asumsi tersebut tidak selalu terpenuhi maka pendekatan yang digunakan adalah nonparametrik (Wu dan Zhang, 2006: 41).

2.3 Regresi Nonparametrik Data Longitudinal Seimbang

Pasangan data longitudinal (t_{ij}, y_{ij}) , dimodelkan sebagai:

$$y_{ij} = \eta(t_{ij}) + \varepsilon_{ij}, \quad i = 1, 2, \dots, n; \quad j = 1, 2, \dots, T_i \quad (2.2)$$

dalam bentuk matriks:

$$\begin{bmatrix} y_{11} \\ y_{12} \\ \vdots \\ y_{1T_1} \\ y_{21} \\ y_{22} \\ \vdots \\ y_{2T_2} \\ \vdots \\ y_{n1} \\ y_{n2} \\ \vdots \\ y_{nT_n} \end{bmatrix} = \begin{bmatrix} \eta(t_{11}) \\ \eta(t_{12}) \\ \vdots \\ \eta(t_{1T_1}) \\ \eta(t_{21}) \\ \eta(t_{22}) \\ \vdots \\ \eta(t_{2T_2}) \\ \vdots \\ \eta(t_{n1}) \\ \eta(t_{n2}) \\ \vdots \\ \eta(t_{nT_n}) \end{bmatrix} + \begin{bmatrix} \varepsilon_{11} \\ \varepsilon_{12} \\ \vdots \\ \varepsilon_{1T_1} \\ \varepsilon_{21} \\ \varepsilon_{22} \\ \vdots \\ \varepsilon_{2T_2} \\ \vdots \\ \varepsilon_{n1} \\ \varepsilon_{n2} \\ \vdots \\ \varepsilon_{nT_n} \end{bmatrix}$$

$$\sum_{i=1}^n T_i \times 1 \quad \sum_{i=1}^n T_i \times 1 \quad \sum_{i=1}^n T_i \times 1$$

t_{ij} = titik waktu pengukuran (*design time point*)

y_{ij} = peubah respon pada titik waktu pengukuran (*responses at the design time point*)

$\eta(t_{ij})$ = fungsi rata-rata populasi (*mean population function*)

Karena secara parametrik peubah waktu t_{ij} sulit dimodelkan terhadap respon y_{ij} , maka dilakukan pendekatan nonparametrik. Wu dan Zhang (2006: 43) menyarankan beberapa metode pendekatan nonparametrik seperti *local polynomial kernel smoother*, *regression splines*, *pemulusan splines*, *penalized spline* dan *linear smoother*.

2.4 Spline pada Data Longitudinal Seimbang

Spline merupakan potongan (*piecewise*) polinomial derajat ke- k yang memiliki sifat tersegmen kontinu sehingga dapat menjelaskan karakteristik lokal suatu model (Eubank, 1988: 227). Spline orde ke- k , dapat juga diartikan sebagai model polinomial orde ke k pada tiap interval segmennya, yakni $[\tau_r, \tau_{r+1}]$. Hal ini berarti, fungsi Spline merupakan suatu gabungan fungsi Polinomial, di mana penggabungan beberapa Polinomial tersebut pada titik knot dengan suatu cara yang menjamin sifat kontinuitas. Spline adalah potongan polinomial mulus yang masih memungkinkan memiliki sifat tersegmen. Terdapat tiga metode spline untuk menduga $\eta(t_{ij})$ yaitu regresi *spline*, pemulusan *spline* dan *penalized spline*.

2.4.1 Regresi Spline

Regresi Spline merupakan gabungan fungsi polinomial antar titik knot dengan suatu cara yang menjamin sifat kontinuitas kurva tersegmen (Didip, 2005: 8). Knot adalah suatu titik dalam fungsi spline, sehingga dihasilkan kurva tersegmen pada titik tersebut. Penduga model $\eta(t_{ij})$ didasarkan pada *truncated power basis* $\Phi_p(t_{ij})$:

$$\Phi_p(t_{ij}) = [1, t_{ij}, t_{ij}^2, \dots, t_{ij}^k, (t_{ij} - \tau_1)_+^k, (t_{ij} - \tau_2)_+^k, \dots, (t_{ij} - \tau_K)_+^k] \quad (2.3)$$

di mana

$p = K+k+1$ adalah banyaknya fungsi basis

$s = 0, 1, 2, \dots, k$; $k =$ derajat polinomial dari *truncated power basis*; $k = 0$ (konstan), 1 (linier), 2 (kuadratik) dan 3 (kubik)

$\tau_r =$ titik knot ke- r ($r = 1, 2, \dots, K$); $K =$ banyaknya titik knot

Pendugaan terhadap model $\eta(t_{ij})$ dilakukan dengan pendekatan regresi spline berdasarkan *truncated power basis*, sehingga diperoleh:

$$\eta(t_{ij}) = \sum_{s=0}^k \beta_s t_{ij}^s + \sum_{r=1}^K \beta_{k+r} (t_{ij} - \tau_r)_+^k \quad (2.4)$$

$$\beta_s = (\beta_0, \beta_1, \dots, \beta_k); \beta_{k+r} = \beta_{k+1}, \dots, \beta_{k+K}$$

$$(t_{ij} - \tau_r)_+^k = [(t_{ij} - \tau_r)_+]^k; (t_{ij} - \tau_r)_+ = \max[0, (t_{ij} - \tau_r)] \text{ atau}$$

$$(t_{ij} - \tau_r)_+^k = \begin{cases} (t_{ij} - \tau_r)^k & , t_{ij} \geq \tau_r \\ 0 & , t_{ij} < \tau_r \end{cases}$$

Anderson (2005: 31) menyajikan 3 kriteria untuk mendapatkan penduga model regresi spline yaitu menentukan derajat polinomial dengan melihat pola data, banyaknya knot dan lokasi penempatan knot. Pemilihan lokasi knot didasarkan pada metode seleksi yaitu mengambil semua titik pada peubah penjelas sebagai knot kemudian diseleksi menggunakan metode seleksi langkah maju, metode eliminasi langkah mundur atau regresi bertahap.

2.4.2 Pemulusan spline

Pemulusan *spline* menggunakan nilai ukuran kemulusan kurva untuk menduga nilai $\eta(t_{ij})$ pada interval $[a, b]$ didefinisikan sebagai integral dari kuadrat turunan ke- k (Wu dan Zhang, 2006: 54):

$$\int_a^b \{\eta^k(t_{ij})\}^2 dt; k \geq 1; \eta^k(t_{ij}) = \frac{\partial \eta(t_{ij})}{\partial^k t_{ij}} \quad (2.5)$$

Penduga nilai $\eta(t_{ij})$ dengan pendekatan *pemulusan spline* diperoleh dari meminimumkan persamaan :

$$\min_{\mathbf{f} \in \mathbf{W}_2^k} \left\{ \sum_{i=1}^n \sum_{j=1}^{T_i} (y_{ij} - \eta_{ij}(\mathbf{t}))^2 + \lambda \left(\int_a^b \{\eta^k(\mathbf{t}_{ij})\}^2 d\mathbf{t} \right) \right\} \quad (2.6)$$

λ = nilai bobot penalti yang mengukur kemulusan kurva

Kelemahan *pemulusan spline* terletak pada kesulitan pemilihan nilai λ yang menghasilkan kemungkinan nilai $\lambda > 0$ dengan ketelitian desimal 4 digit dan membutuhkan perhitungan integral suatu fungsi yang tidak diketahui.

2.4.3 Penalized Spline

Penduga *Penalized Spline* derajat ke- k dan titik-titik knot τ_r (Wu dan Zhang, 2006: 60) adalah:

$$\eta(t_{ij}) = \Phi'_p(t) \beta = \sum_{s=0}^k \beta_s t_{ij}^s + \sum_{r=1}^K \beta_{k+r} (t_{ij} - \tau_r)_+^k \quad (2.7)$$

$$\Phi'_p(t) = [\Phi_p(t_{11}), \Phi_p(t_{12}), \dots, \Phi_p(t_{1T_1}), \Phi_p(t_{21}), \dots, \Phi_p(t_{2T_2}), \dots, \Phi_p(t_{nT_n})]$$

$$\beta' = [\beta_s, \beta_{k+r}] \text{ atau } \beta' = [\beta_0, \beta_1, \beta_2, \dots, \beta_k, \beta_{k+1}, \dots, \beta_{k+K}]$$

Fungsi *Penalized Spline* untuk setiap pengamatan:

$$\begin{aligned} \eta(t_{11}) &= \beta_0 + \beta_1 t_{11} + \beta_2 t_{11}^2 + \dots + \beta_k t_{11}^k + \beta_{(k+1)} (t_{11} - \tau_1)_+^k + \dots + \beta_{(k+K)} (t_{11} - \tau_K)_+^k \\ \eta(t_{12}) &= \beta_0 + \beta_1 t_{12} + \beta_2 t_{12}^2 + \dots + \beta_k t_{12}^k + \beta_{(k+1)} (t_{12} - \tau_1)_+^k + \dots + \beta_{(k+K)} (t_{12} - \tau_K)_+^k \\ &\vdots \\ \eta(t_{1T_1}) &= \beta_0 + \beta_1 t_{1T_1} + \beta_2 t_{1T_1}^2 + \dots + \beta_k t_{1T_1}^k + \beta_{(k+1)} (t_{1T_1} - \tau_1)_+^k + \dots + \beta_{(k+K)} (t_{1T_1} - \tau_K)_+^k \\ \eta(t_{21}) &= \beta_0 + \beta_1 t_{21} + \beta_2 t_{21}^2 + \dots + \beta_k t_{21}^k + \beta_{(k+1)} (t_{21} - \tau_1)_+^k + \dots + \beta_{(k+K)} (t_{21} - \tau_K)_+^k \\ &\vdots \\ \eta(t_{2T_2}) &= \beta_0 + \beta_1 t_{2T_2} + \beta_2 t_{2T_2}^2 + \dots + \beta_k t_{2T_2}^k + \beta_{(k+1)} (t_{2T_2} - \tau_1)_+^k + \dots + \beta_{(k+K)} (t_{2T_2} - \tau_K)_+^k \\ &\vdots \\ \eta(t_{nT_n}) &= \beta_0 + \beta_1 t_{nT_n} + \beta_2 t_{nT_n}^2 + \dots + \beta_k t_{nT_n}^k + \beta_{(k+1)} (t_{nT_n} - \tau_1)_+^k + \dots + \beta_{(k+K)} (t_{nT_n} - \tau_K)_+^k \end{aligned}$$

Dalam bentuk matriks fungsi *penalized spline* ditulis sebagai:

$$\begin{bmatrix} \eta(t_{11}) \\ \eta(t_{12}) \\ \vdots \\ \eta(t_{1T_1}) \\ \eta(t_{21}) \\ \vdots \\ \eta(t_{2T_2}) \\ \vdots \\ \eta(t_{nT_n}) \end{bmatrix} = \begin{bmatrix} 1 & t_{11} & t_{11}^2 & \cdots & t_{11}^k & (t_{11} - \tau_1)_+^k & \cdots & (t_{11} - \tau_K)_+^k \\ 1 & t_{12} & t_{12}^2 & \cdots & t_{12}^k & (t_{12} - \tau_1)_+^k & \cdots & (t_{11} - \tau_K)_+^k \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & t_{1n_1} & t_{1n_1}^2 & \cdots & t_{1n_1}^k & (t_{1n_1} - \tau_1)_+^k & \cdots & (t_{1n_1} - \tau_K)_+^k \\ 1 & t_{21} & t_{21}^2 & \cdots & t_{21}^k & (t_{21} - \tau_1)_+^k & \cdots & (t_{21} - \tau_K)_+^k \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & t_{2n_2} & t_{2n_2}^2 & \cdots & t_{2n_2}^k & (t_{2n_2} - \tau_1)_+^k & \cdots & (t_{2n_2} - \tau_K)_+^k \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & t_{mn_n} & t_{mn_n}^2 & \cdots & t_{mn_n}^k & (t_{mn_n} - \tau_1)_+^k & \cdots & (t_{mn_n} - \tau_K)_+^k \end{bmatrix} \begin{bmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \\ \vdots \\ \beta_k \\ \beta_{k+1} \\ \vdots \\ \beta_{K+k} \end{bmatrix}$$

$$\sum_{i=1}^n T_i x1 \quad \sum_{i=1}^n T_i x(K+k+1) \quad (K+k+1)x1$$

atau $\eta(t) = \Phi_p(t)\beta = t\beta$

$$\eta(t) = \begin{bmatrix} \eta(t_{11}) \\ \eta(t_{12}) \\ \vdots \\ \eta(t_{1T_1}) \\ \eta(t_{21}) \\ \vdots \\ \eta(t_{2T_2}) \\ \vdots \\ \eta(t_{nT_n}) \end{bmatrix}; \beta = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \\ \vdots \\ \beta_k \\ \beta_{k+1} \\ \vdots \\ \beta_{K+k} \end{bmatrix}; t = \begin{bmatrix} 1 & t_{11} & t_{11}^2 & \cdots & t_{11}^k & (t_{11} - \tau_1)_+^k & \cdots & (t_{11} - \tau_K)_+^k \\ 1 & t_{12} & t_{12}^2 & \cdots & t_{12}^k & (t_{12} - \tau_1)_+^k & \cdots & (t_{11} - \tau_K)_+^k \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & t_{1n_1} & t_{1n_1}^2 & \cdots & t_{1n_1}^k & (t_{1n_1} - \tau_1)_+^k & \cdots & (t_{1n_1} - \tau_K)_+^k \\ 1 & t_{21} & t_{21}^2 & \cdots & t_{21}^k & (t_{21} - \tau_1)_+^k & \cdots & (t_{21} - \tau_K)_+^k \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & t_{2n_2} & t_{2n_2}^2 & \cdots & t_{2n_2}^k & (t_{2n_2} - \tau_1)_+^k & \cdots & (t_{2n_2} - \tau_K)_+^k \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & t_{nn_n} & t_{nn_n}^2 & \cdots & t_{nn_n}^k & (t_{nn_n} - \tau_1)_+^k & \cdots & (t_{nn_n} - \tau_K)_+^k \end{bmatrix}$$

$$\sum_{i=1}^n T_i x1 \quad (K+k+1)x1 \quad \sum_{i=1}^n T_i x(K+k+1)$$

Penduga bagi $\eta(t)$:

$$\hat{\eta}(t) = t\hat{\beta} \quad (2.8)$$

didapat melalui *Penalized Least Square* (PLS) sebagai penyelesaian optimasi persamaan (Eubank, 1988: 291):

$$\min_{f \in \mathbf{W}_2^k} \left\{ \sum_{i=1}^n \sum_{j=1}^{T_i} (y_{ij} - \eta(t_{ij}))^2 + \lambda \beta' G \beta \right\} \quad (2.9)$$

$$G_{K+k+1} = \begin{bmatrix} 0_{k+1} & k+1 \cdot 0_K \\ k \cdot 0_{k+1} & I_K \end{bmatrix}; I_K = \text{matriks identitas.}$$

$$\sum_{i=1}^n \sum_{j=1}^{T_i} (y_{ij} - \eta(t_{ij}))^2 = \text{fungsi kecocokan model (goodness of fit), } \lambda :$$

nilai bobot yang mengukur keseimbangan antara fungsi kesesuaian data dan kemulusan kurva, $\beta'G\beta$: ukuran kemulusan kurva.

2.4.4 Generalized Penalized Spline

Generalized Penalized Spline (GPS) adalah *penalized spline* yang menggunakan pembobot matriks. Wu dan Zhang (2006: 203) menjelaskan penduga GPS sebagai penduga fungsi rata-rata populasi $\eta(t_{ij})$ yang jauh lebih baik, karena dapat mengatasi *heteroskedastisitas* populasi. Statistik L adalah penduga *Penalized Generalized Least Square* (Wu dan Zhang, 2006: 203) :

$$L = (\mathbf{y} - \mathbf{t}\boldsymbol{\beta})' \mathbf{V}^{-1} (\mathbf{y} - \mathbf{t}\boldsymbol{\beta}) + \lambda \boldsymbol{\beta}' \mathbf{G} \boldsymbol{\beta} \quad (2.10)$$

$\mathbf{y}' = [\mathbf{y}'_1, \mathbf{y}'_2, \dots, \mathbf{y}'_n]$; $\mathbf{V}^{-1} = \text{diag}[\mathbf{V}_1, \mathbf{V}_2, \dots, \mathbf{V}_n]$,
 $\mathbf{V}_i = \text{diag}[W_{i1}, W_{i2}, \dots, W_{iT_i}]$ adalah matriks pembobot.

Steel dan Torie (1989) menyatakan bahwa, pembobot bergantung pada informasi dan ketepatan hasil pengamatan dan merupakan kebalikan ragam yakni $w_{ij} = \sigma_i^{-2}$ sebagai pembobot bagi respon ke-(i,j) karena $\hat{\sigma}_i^{-2} = s_i^{-2}$, maka :

$$w_{ij} = s_i^{-2} = \left(\frac{\sum_{j=1}^{T_i} (y_{ij} - \bar{y}_i)^2}{T_i - 1} \right)^{-1}$$

2.5 Pemilihan Lokasi Titik Knot

Pendugaan model Penalized Spline didasarkan pada lokasi titik knot, banyaknya titik knot dan derajat polinomial. Titik knot diartikan sebagai suatu titik fokus dalam fungsi spline, sehingga kurva yang dibentuk tersegmen pada titik tersebut. Pemilihan lokasi titik knot memiliki pengaruh yang lebih besar terhadap keakuratan model dibandingkan dengan derajat polinomial. Pada umumnya, Penalized Spline menggunakan derajat polinomial $k = 1, 2, 3$ untuk proses

komputasi sederhana. Wu dan Zhang (2006 :53-54) memaparkan tiga metode dalam pemilihan lokasi titik knot yaitu :

1. **Equally Spaced Method**

Pada interval [a,b] dipilih titik-titik knot sebanyak K dengan pemilihan lokasi titik-titik knot pada jarak sama.

$$\tau_r = a + (b - a)r / (K + 1), \quad r = 1, 2, 3, \dots, K$$

Metode ini digunakan pada t_{ij} yang tersebar seragam pada interval tertentu.

2. **Equally Spaced Samples Quantile**

Metode *equally spaced sample quantile* menggunakan titik-titik yang terletak pada kuantil peubah penjelas t_{ij} , $i = 1, 2, \dots, n$, $j = 1, 2, \dots, T$ sebagai titik knot. Statistik peringkat dari titik waktu pengukuran adalah $t_{(1)}, t_{(12)}, \dots, t_{(nT)}$. Knot ke-r didefinisikan :

$$\tau_r = t_{(1 + \lceil rn / (K + 1) \rceil)}, \quad r = 1, 2, \dots, K$$

$\lceil rn / (K + 1) \rceil$ adalah bilangan bulat dari $rn / (K + 1)$.

Pada t_{ij} yang tersebar seragam *equally spaced sample quantile* kira-kira ekuivalen dengan *equally spaced method*.

3. **Selection Based Method**

Metode ini menggunakan kombinasi dari seluruh titik waktu pengukuran yang mungkin berdasarkan banyaknya titik knot.

2.6 Pendugaan Parameter Regresi Nonparametrik dengan pendekatan Generalized Penalized Spline

Tujuan utama dalam analisis regresi nonparametrik longitudinal adalah menduga *mean population function* $\eta(t_{ij})$ dengan bentuk kurva yang tidak diketahui menggunakan metode *penalized generalized least square*. Penjabaran model (2.2) adalah:

$$y_{11} = \beta_0 + \beta_1 t_{11} + \beta_2 t_{11}^2 + \dots + \beta_k t_{11}^k + \beta_{(k+1)} (t_{11} - \tau_1)_+^k + \dots + \beta_{(k+K)} (t_{11} - \tau_K)_+^k + \varepsilon_{11}$$

$$y_{12} = \beta_0 + \beta_1 t_{12} + \beta_2 t_{12}^2 + \dots + \beta_k t_{12}^k + \beta_{(k+1)} (t_{12} - \tau_1)_+^k + \dots + \beta_{(k+K)} (t_{12} - \tau_K)_+^k + \varepsilon_{12}$$

⋮

$$y_{1T} = \beta_0 + \beta_1 t_{1T} + \beta_2 t_{1T}^2 + \dots + \beta_k t_{1T}^k + \beta_{(k+1)} (t_{1T} - \tau_1)_+^k + \dots + \beta_{(k+K)} (t_{1T} - \tau_K)_+^k + \varepsilon_{1T}$$

$$y_{21} = \beta_0 + \beta_1 t_{21} + \beta_2 t_{21}^2 + \dots + \beta_k t_{21}^k + \beta_{(k+1)} (t_{21} - \tau_1)_+^k + \dots + \beta_{(k+K)} (t_{21} - \tau_K)_+^k + \varepsilon_{21}$$

⋮

$$\begin{aligned}
 y_{2T_2} &= \beta_0 + \beta_1 t_{2T_2} + \beta_2 t_{2T_2}^2 + \dots + \beta_k t_{2T_2}^k + \beta_{(k+1)} (t_{2T_2} - \tau_1)_+^k + \dots + \beta_{(k+K)} (t_{2T_2} - \tau_K)_+^k + \varepsilon_{2T_2} \\
 &\vdots \\
 y_{nT_n} &= \beta_0 + \beta_1 t_{nT_n} + \beta_2 t_{nT_n}^2 + \dots + \beta_k t_{nT_n}^k + \beta_{(k+1)} (t_{nT_n} - \tau_1)_+^k + \dots + \beta_{(k+K)} (t_{nT_n} - \tau_K)_+^k + \varepsilon_{nT_n}
 \end{aligned}$$

Dalam bentuk matriks:

$$\begin{bmatrix} y_{11} \\ y_{12} \\ \vdots \\ y_{1T_1} \\ y_{21} \\ \vdots \\ y_{2T_2} \\ \vdots \\ y_{nT_n} \end{bmatrix} = \begin{bmatrix} 1 & t_{11} & t_{11}^2 & \dots & t_{11}^k & (t_{11} - \tau_1)_+^k & \dots & (t_{11} - \tau_K)_+^k \\ 1 & t_{12} & t_{12}^2 & \dots & t_{12}^k & (t_{12} - \tau_1)_+^k & \dots & (t_{11} - \tau_K)_+^k \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & t_{1n_1} & t_{1n_1}^2 & \dots & t_{1n_1}^k & (t_{1n_1} - \tau_1)_+^k & \dots & (t_{1n_1} - \tau_K)_+^k \\ 1 & t_{21} & t_{21}^2 & \dots & t_{21}^k & (t_{21} - \tau_1)_+^k & \dots & (t_{21} - \tau_K)_+^k \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & t_{2n_2} & t_{2n_2}^2 & \dots & t_{2n_2}^k & (t_{2n_2} - \tau_1)_+^k & \dots & (t_{2n_2} - \tau_K)_+^k \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & t_{nn} & t_{nn}^2 & \dots & t_{nn}^k & (t_{nn} - \tau_1)_+^k & \dots & (t_{nn} - \tau_K)_+^k \end{bmatrix} + \begin{bmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \\ \vdots \\ \beta_k \\ \beta_{k+1} \\ \vdots \\ \beta_{k+K} \end{bmatrix} + \begin{bmatrix} \varepsilon_{11} \\ \varepsilon_{12} \\ \vdots \\ \varepsilon_{1T_1} \\ \varepsilon_{21} \\ \vdots \\ \varepsilon_{2T_2} \\ \vdots \\ \varepsilon_{nT_n} \end{bmatrix}$$

$$\sum_{i=1}^n T_i x_i \quad \sum_{i=1}^n T_i x_i (K+k+1) \quad (K+k+1) x_i \quad \sum_{i=1}^n T_i x_i$$

$$\text{atau } \mathbf{y} = \mathbf{t}\boldsymbol{\beta} + \boldsymbol{\varepsilon} \quad (2.11)$$

di mana

$$\mathbf{y} = \begin{bmatrix} y_{11} \\ y_{12} \\ \vdots \\ y_{1T_1} \\ y_{21} \\ \vdots \\ y_{2T_2} \\ \vdots \\ y_{nT_n} \end{bmatrix}, \text{ karena } \mathbf{y}_1 = \begin{bmatrix} y_{11} \\ y_{12} \\ \vdots \\ y_{1T_1} \end{bmatrix}, \mathbf{y}_2 = \begin{bmatrix} y_{21} \\ y_{22} \\ \vdots \\ y_{2T_2} \end{bmatrix}, \dots, \mathbf{y}_n = \begin{bmatrix} y_{n1} \\ y_{n2} \\ \vdots \\ y_{nT_n} \end{bmatrix} \text{ maka } \mathbf{y} = \begin{bmatrix} \mathbf{y}_1 \\ \mathbf{y}_2 \\ \vdots \\ \mathbf{y}_n \end{bmatrix}$$

$$\sum_{i=1}^n T_i x_i \quad T_1 x_i \quad T_2 x_i \quad T_n x_i \quad nx_i$$

$$\boldsymbol{\varepsilon} = \begin{bmatrix} \varepsilon_{11} \\ \varepsilon_{12} \\ \vdots \\ \varepsilon_{1T_1} \\ \varepsilon_{21} \\ \vdots \\ \varepsilon_{2T_2} \\ \vdots \\ \varepsilon_{iT_n} \end{bmatrix}$$

$$\sum_{i=1}^n T_i x_i$$

Statistik $\eta(t_{ij})$ didapat melalui kriteria *Penalized Generalized Least Square* berdasarkan model 2.10:

$$\begin{aligned} L &= (\mathbf{y} - \mathbf{t}\boldsymbol{\beta})' \mathbf{V}^{-1} (\mathbf{y} - \mathbf{t}\boldsymbol{\beta}) + \lambda \boldsymbol{\beta}' \mathbf{G} \boldsymbol{\beta} \\ &= (\mathbf{y}' - \boldsymbol{\beta}' \mathbf{t}') \mathbf{V}^{-1} (\mathbf{y} - \mathbf{t}\boldsymbol{\beta}) + \lambda \boldsymbol{\beta}' \mathbf{G} \boldsymbol{\beta} \\ &= (\mathbf{y}' - \boldsymbol{\beta}' \mathbf{t}') (\mathbf{V}^{-1} \mathbf{y} - \mathbf{V}^{-1} \mathbf{t}\boldsymbol{\beta}) + \lambda \boldsymbol{\beta}' \mathbf{G} \boldsymbol{\beta} \\ &= \mathbf{y}' \mathbf{V}^{-1} \mathbf{y} - \mathbf{y}' \mathbf{V}^{-1} \mathbf{t}\boldsymbol{\beta} - \boldsymbol{\beta}' \mathbf{t}' \mathbf{V}^{-1} \mathbf{y} + \boldsymbol{\beta}' \mathbf{t}' \mathbf{V}^{-1} \mathbf{t}\boldsymbol{\beta} + \lambda \boldsymbol{\beta}' \mathbf{G} \boldsymbol{\beta} \\ &= \mathbf{y}' \mathbf{V}^{-1} \mathbf{y} - 2\boldsymbol{\beta}' \mathbf{t}' \mathbf{V}^{-1} \mathbf{y} + \boldsymbol{\beta}' \mathbf{t}' \mathbf{V}^{-1} \mathbf{t}\boldsymbol{\beta} + \lambda \boldsymbol{\beta}' \mathbf{G} \boldsymbol{\beta} \end{aligned}$$

Syarat perlu untuk meminimumkan nilai PGLS adalah menghitung turunan pertama L terhadap parameter $\boldsymbol{\beta}$ dan menyamakan dengan nol:

$$\begin{aligned} \frac{\partial L}{\partial \hat{\boldsymbol{\beta}}} &= 0 \\ -2\mathbf{t}' \mathbf{V}^{-1} \mathbf{y} + 2\mathbf{t}' \mathbf{V}^{-1} \mathbf{t} \hat{\boldsymbol{\beta}} + \lambda \hat{\boldsymbol{\beta}}' \mathbf{G} \hat{\boldsymbol{\beta}} &= 0 \\ (\mathbf{t}' \mathbf{V}^{-1} \mathbf{t} + \lambda \mathbf{G}) \hat{\boldsymbol{\beta}} &= \mathbf{V}^{-1} \mathbf{t}' \mathbf{y} \\ \hat{\boldsymbol{\beta}} &= (\mathbf{t}' \mathbf{V}^{-1} \mathbf{t} + \lambda \mathbf{G})^{-1} \mathbf{t}' \mathbf{V}^{-1} \mathbf{y} \end{aligned} \quad (2.12)$$

Penduga *mean population function* adalah:

$$\begin{aligned} \hat{\eta}(\mathbf{t}) &= \Phi_p(\mathbf{t})' \hat{\boldsymbol{\beta}} \\ &= \Phi_p(\mathbf{t})' (\mathbf{t}' \mathbf{V}^{-1} \mathbf{t} + \lambda \mathbf{G})^{-1} \mathbf{t}' \mathbf{V}^{-1} \mathbf{y} \end{aligned} \quad (2.13)$$

Penduga model y dengan menggunakan GPS:

$$\begin{aligned}\hat{y}_{\text{gps}} &= \mathbf{t}\hat{\boldsymbol{\beta}} \\ &= \mathbf{t}(\mathbf{t}'\mathbf{V}^{-1}\mathbf{t} + \lambda\mathbf{G})^{-1} \mathbf{t}'\mathbf{V}^{-1}\mathbf{y} \\ &= \mathbf{A}_{\text{gps}} \mathbf{y}\end{aligned}\quad (2.14)$$

$\mathbf{A}_{\text{gps}} = (\mathbf{t}'\mathbf{V}^{-1}\mathbf{t} + \lambda\mathbf{G})^{-1} \mathbf{t}'\mathbf{V}^{-1}$ adalah matriks penghalus (*smoother matrix*).

2.7 Pemilihan Parameter Penghalus (λ) Optimal

Menurut Eubank (1988: 239), pemilihan nilai λ optimal merupakan hal penting dalam pendugaan model regresi dengan bentuk kecenderungan kurva tidak standar. Nilai bobot λ adalah pengontrol keseimbangan yang berpengaruh terhadap kemulusan kurva dan kemampuan model dalam memetakan data. Semakin besar λ bentuk kurva akan semakin mulus, tetapi kemampuan model dalam memetakan data kurang baik sebaliknya nilai λ kecil akan menghasilkan penduga model dengan bentuk kurva kasar.

Metode yang digunakan untuk pemilihan nilai bobot λ adalah *Generalized Cross Validation* (GCV). Wu dan Zhang (2006: 66) menyatakan bahwa GCV merupakan modifikasi dari *Cross-Validation* (CV) didapat dengan meminimumkan fungsi CV:

$$\text{GCV}_{\text{gps}}(\lambda) = \frac{(\mathbf{y} - \hat{\mathbf{y}}_{\text{gps}})' \mathbf{V}^{-1} (\mathbf{y} - \hat{\mathbf{y}}_{\text{gps}})}{(1 - \text{tr}(\mathbf{A}_{\text{gps}}) / n)^2} \quad (2.15)$$

2.8 Koefisien Determinasi

Koefisien determinasi (R^2) adalah besaran untuk mengukur kekuatan model regresi yang mampu menjelaskan hubungan antar peubah penjelas dan peubah respon (Draper dan Smith, 1992: 197):

$$R^2 = \frac{\sum_{i=1}^n \sum_{j=1}^{T_i} (\hat{y}_{ij} - \bar{y})^2}{\sum_{i=1}^n \sum_{j=1}^{T_i} (y_{ij} - \bar{y})^2} = \frac{\text{SSR}}{\text{SST}} \quad (2.16)$$

\hat{y}_{ij} = penduga nilai peubah respon ke- i pada waktu ke- j

\bar{y} = rata-rata peubah respon

y_{ij} = nilai peubah respon ke- i pada waktu ke- j

SSR = jumlah kuadrat regresi

SST = jumlah kuadrat total

Menurut Draper dan Smith (1992: 286), statistik lain pengukur proporsi keragaman, yakni koefisien determinasi disesuaikan R^2_{adjusted} , yang melibatkan derajat bebas galat dan total.

$$R^2_{\text{adjusted}} = \frac{\text{SSE}/(n - m)}{\text{SST}/(n - 1)} = 1 - (1 - R^2) \left(\frac{n - 1}{n - m} \right) \quad (2.17)$$

SSE = SST-SSR= jumlah Kuadrat Galat

m = banyaknya parameter dalam model

n = ukuran contoh

2.9 Uji Signifikansi Model Generalized Penalized Spline (GPS)

Pengujian signifikansi koefisien vektor β dilakukan untuk mendapatkan nilai-nilai parameter yang signifikan terhadap model dan untuk mengetahui apakah model GPS yang terbentuk dapat menggambar data yang sebenarnya, sehingga layak digunakan. Pengujian GPS menggunakan uji parsial dengan statistik uji t . Uji ini digunakan untuk mengetahui apakah setiap parameter dalam model mempunyai pengaruh yang signifikan atau tidak, dilakukan dengan cara menguji setiap koefisien.

Hipotesis yang melandasi pengujian signifikansi parsial:

$$H_0 : \beta_s = 0$$

$$H_1 : \beta_s \neq 0 \quad s = 0, 1, \dots, k$$

$$H_0 : \beta_{k+r} = 0$$

$$H_1 : \beta_{k+r} \neq 0 \quad r = 0, 1, \dots, K$$

Penduga bagi β_s atau β_{k+r} adalah $\hat{\beta}_s$ atau $\hat{\beta}_{k+r}$ dan jika H_0 benar $\hat{\beta}_s$ atau $\hat{\beta}_{k+r} \sim N(0, \sigma^2(\hat{\beta}_s \text{ atau } \hat{\beta}_{k+r}))$ karena nilai $\hat{\sigma}^2 = s^2$, maka statistik uji $\frac{\hat{\beta}_s \text{ atau } \hat{\beta}_{k+r}}{S_{\hat{\beta}_s \text{ atau } \hat{\beta}_{k+r}}} \sim t_{n-1}$. Matrik ragam peragam (var-cov($\hat{\beta}_s$ atau $\hat{\beta}_{k+r}$)) = $(t'V^{-1}t + \lambda G)^{-1}$.

$$\text{Var-cov}(\hat{\beta}_s \text{ atau } \hat{\beta}_{k+r}) = \begin{bmatrix} s_{\hat{\beta}_0}^2 & 0 & \dots & 0 \\ & s_{\hat{\beta}_1}^2 & & \\ & & \ddots & \\ \vdots & & & s_{\hat{\beta}_k}^2 & \vdots \\ & & & & s_{\hat{\beta}_{k+1}}^2 \\ 0 & \dots & & & s_{\hat{\beta}_{k+k}}^2 \end{bmatrix}$$

(Wu dan Zhang, 2006 : 217)

Kriteria penolakan H_0 jika p-value $< \alpha$.

2.10 Pengujian Asumsi Galat

2.10.1 Asumsi $E(\varepsilon_{ij}) = 0$

Penduga diharapkan mendekati nilai parameter sebenarnya. Asumsi nilai harapan peubah galat sama dengan nol, menunjukkan bahwa secara keseluruhan diharapkan kurva regresi tepat menghubungkan titik-titik. Hipotesis yang melandasi pengujian nilai harapan galat (Walpole, 1995 : 302):

$$H_0 : \mu_{\varepsilon_{ij}} = 0$$

$$H_1 : \mu_{\varepsilon_{ij}} \neq 0$$

Penduga bagi $\hat{\mu}_{\varepsilon_{ij}} = \bar{x}_{\varepsilon_{ij}}$ dan jika H_0 benar $\bar{x}_{\varepsilon_{ij}} \sim N(0, \sigma^2/n)$ karena nilai $\hat{\sigma}^2 = s^2$, maka statistik uji $\frac{|\bar{X}_{\varepsilon_{ij}} - \mu_{\varepsilon_{ij}}|}{s/\sqrt{n}} \sim t_{n-1}$. Kriteria penolakan H_0 jika p-value $< \alpha$.

2.10.2 Homogenitas Ragam Galat

Asumsi *homoskedastisitas* terpenuhi jika ragam populasi σ^2 konstan untuk setiap individu ke- i , pelanggaran terhadap asumsi ini disebut dengan *heteroskedastisitas* (Gujarati, 2004:341). Hipotesis yang melandasi pengujian kehomogenan ragam adalah:

$$H_0 : \sigma_i^2 = \sigma^2$$

$$H_1 : \text{paling tidak ada satu } i \text{ di mana } \sigma_i^2 \neq \sigma^2$$

Statistik $\hat{\sigma}_i^2 = s_i^2$, apabila H_0 benar $\sigma_1^2 = \sigma_2^2 = \dots = \sigma_n^2$ dan statistik uji W :

$$\frac{(nT - n) \sum_{i=1}^n T (\bar{Z}_{i.} - \bar{Z}_{..})^2}{(n - 1) \sum_{i=1}^n \sum_{j=1}^T (Z_{ij} - \bar{Z}_{i.})^2} \sim F_{(\alpha, n-1, nT-n)} \quad (2.19)$$

nT = banyaknya nilai peubah respon (y_{ij})

n = banyaknya individu

T = banyaknya titik waktu pengukuran setiap individu

$$Z_{ij} = |Y_{ij} - \bar{Y}_{i.}|; \bar{Y}_{i.} = \text{rata-rata nilai peubah respon ke-}i$$

BAB III METODE PENELITIAN

3.1 Sumber Data

Data sekunder yang digunakan berasal dari Aristya Rachmadani (2011) tentang pertumbuhan berat badan bayi. Pada 9 bayi dilakukan pengukuran berat badan secara berulang selama 13 kali, yang menghasilkan data pada Lampiran 1.

3.2 Prosedur Analisis Data

3.3.1 Membuat algoritma untuk menduga parameter model $\eta(t)$

i) Algoritma untuk menentukan derajat polinomial (k), jumlah knot (K), dan nilai bobot lambda (λ) optimal:

a. Membuat matriks

$$t = \begin{bmatrix} 1 & t_{11} & t_{11}^2 & \dots & t_{11}^k & (t_{11} - \tau_1)_+^k & \dots & (t_{11} - \tau_K)_+^k \\ 1 & t_{12} & t_{12}^2 & \dots & t_{12}^k & (t_{12} - \tau_1)_+^k & \dots & (t_{12} - \tau_K)_+^k \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & t_{1n_1} & t_{1n_1}^2 & \dots & t_{1n_1}^k & (t_{1n_1} - \tau_1)_+^k & \dots & (t_{1n_1} - \tau_K)_+^k \\ 1 & t_{21} & t_{21}^2 & \dots & t_{21}^k & (t_{21} - \tau_1)_+^k & \dots & (t_{21} - \tau_K)_+^k \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & t_{2n_2} & t_{2n_2}^2 & \dots & t_{2n_2}^k & (t_{2n_2} - \tau_1)_+^k & \dots & (t_{2n_2} - \tau_K)_+^k \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & t_{nn_n} & t_{nn_n}^2 & \dots & t_{nn_n}^k & (t_{nn_n} - \tau_1)_+^k & \dots & (t_{nn_n} - \tau_K)_+^k \end{bmatrix}$$

b. Membuat matriks $G_{K+k+1} = \begin{bmatrix} \mathbf{0}_{k+1} & \mathbf{0}_{k+1} \\ \mathbf{0}_{k+1} & \mathbf{I}_K \end{bmatrix}$

c. Membuat matriks pembobot

$$V^{-1} = \text{diag}[V_1, V_2, \dots, V_n];$$

$$V_i = \text{diag}[W_{i1}, W_{i2}, \dots, W_{iT_i}]$$

$$W_{ij} = S_i^{-2} = \left(\frac{\sum_{j=1}^{T_i} (y_{ij} - \bar{y}_i)^2}{T_i - 1} \right)^{-1}$$

- d. Meminimumkan GCV dengan memilih j λ umlah knot optimal:

$$\text{GCV}_{\text{gps}}(\lambda) = \frac{(\mathbf{y} - \hat{\mathbf{y}}_{\text{gps}})' \mathbf{V}^{-1} (\mathbf{y} - \hat{\mathbf{y}}_{\text{gps}})}{(1 - \text{tr}(\mathbf{A}_{\text{gps}}) / n)^2}$$

- ii) Menduga parameter model $\eta(\mathbf{t})$ dengan substitusi nilai k , K , dan λ optimal berdasarkan GCV yang menghasilkan nilai terkecil.

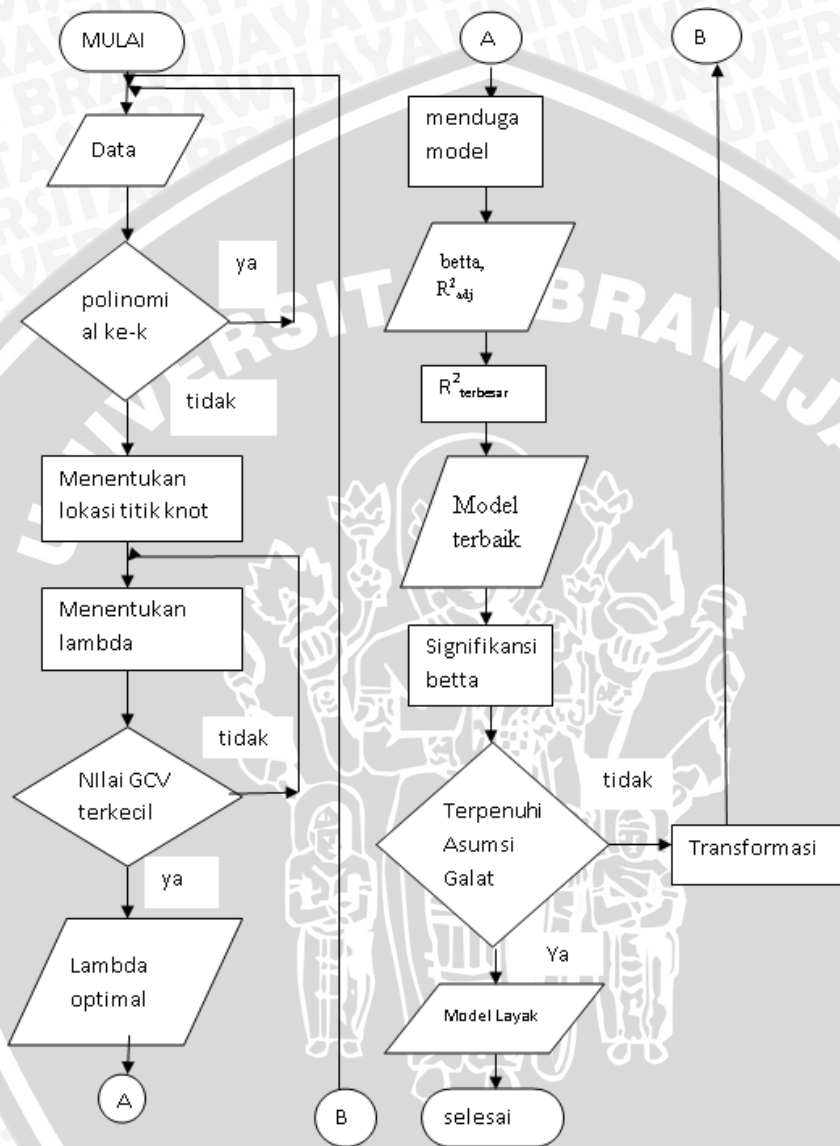
3.3.2 Membuat program untuk menduga parameter model regresi nonparametrik longitudinal ($y_{ij} = \eta(t_{ij}) + \varepsilon_{ij}$) pada Lampiran.2-4 berdasarkan penduga *Generalized Penalized Splines* menggunakan bantuan Minitab 15.

3.3.3 Menganalisis data

- Membuat plot y_{ij} terhadap t_{ij}
- Menduga parameter model regresi nonparametrik longitudinal \hat{y}_{gps} dan galat (ε_{ij})
- Melakukan pengujian terhadap asumsi $E(\varepsilon_{ij}) = 0$
- Menguji kehomogenan ragam galat menggunakan uji Levene dengan statistik uji W .

Langkah-langkah pendugaan parameter model hingga menghasilkan model terbaik dan sesuai (layak) disajikan pada Gambar

3.1



Gambar 3.1 Diagram Alir Pendekatan Regresi Nonparametrik

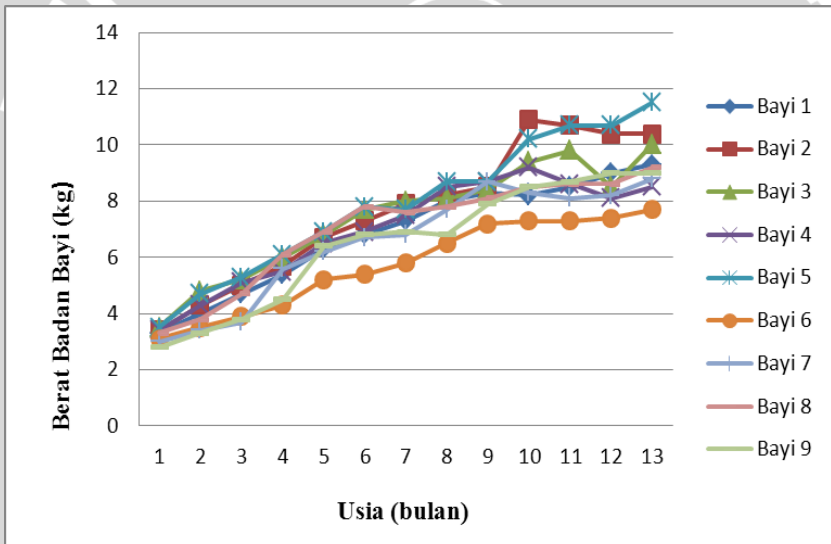
UNIVERSITAS BRAWIJAYA



BAB IV PEMBAHASAN

4.1 Deskripsi Umum Bayi 0-12 Bulan di Kecamatan Bandulan Kota Malang

Berdasarkan Kartu Menuju Sehat (KMS), secara umum pola pertumbuhan berat badan bayi (subyek) berbeda untuk setiap usia, ragam terkecil terletak pada usia 0 bulan. Semakin usia bertambah, maka keragaman akan semakin besar. Untuk mengetahui pola kecenderungan pertumbuhan bayi di kelurahan Bandulan, maka dibuat diagram pencar antar berat badan bayi dan usia 0-12 bulan.



Gambar 4.1. Plot berat badan (kg) terhadap usia bayi (bulan)

Sebanyak 9 garis bertitik yang terbentuk pada Gambar 4.1 menunjukkan perubahan berat badan bayi pada usia 0-12 bulan. Berdasarkan diagram pencar diperoleh informasi berupa keragaman peubah respon antar umur dan antar subyek. Bayi berusia 0 bulan memiliki berat antara 2.8-3.5 kg dengan ragam antar subyek 0.0578 dan semakin bertambahnya usia maka keragaman berat badan bayi akan semakin besar, mendekati usia 12 bulan berat badan bayi berkisar antara

7.7-11.5 kg. Keragaman berat badan antar usia dalam satu subyek lebih besar dibandingkan keragaman antar subyek, hal ini mengingat terdapat peubah lain yang mempengaruhi pertumbuhan berat badan bayi, yakni asupan nutrisi bayi. Keragaman berat badan antar usia dalam subyek dan antar subyek disajikan dalam tabel berikut:

Tabel 4.1. Keragaman Berat Badan Antar Usia dalam Subyek (a) dan Antar Subyek (b)

Bulan	Keragaman Berat Badan Antar Subyek	Bayi	Keragaman Berat Badan Dalam Subyek
0	0.0578	1	3.8794
1	0.3061	2	6.4344
2	0.4025	3	4.0667
3	0.4325	4	3.5514
4	0.2800	5	6.3214
5	0.5744	6	2.6692
6	0.4744	7	4.2460
7	0.5361	8	3.7583
8	0.2419	9	4.9058
9	1.2178		
10	1.3475		
11	1.1419		
12	1.2544		

(a)
(b)

4.2 Eksplorasi Profil Individu

Eksplorasi ini menggambarkan bagaimana perubahan nilai peubah respon terhadap waktu pada setiap subyek yang diamati, secara visual digambarkan melalui diagram pencar antar peubah respon dan waktu yang disajikan pada Lampiran 5.

Profil individu yang terbentuk pada bayi ke-1 di mana pertumbuhan berat badan bayi terus meningkat dari waktu ke waktu. Pada usia 0-8 bulan terjadi peningkatan berat badan bayi. Penurunan berat badan bayi

terjadi pada usia 9 bulan, kemudian mengalami kenaikan hingga bulan ke-12.

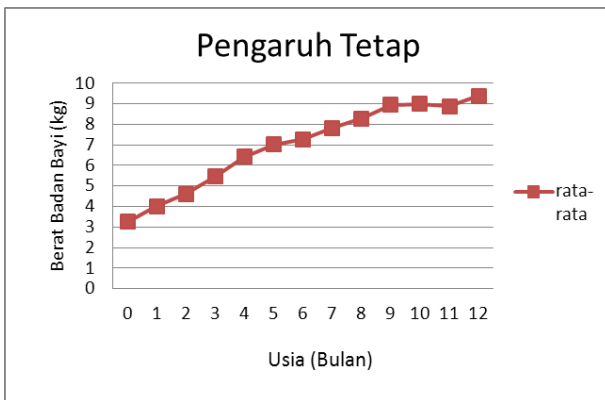
Pada bayi ke-2, terjadi kenaikan berat badan yang tinggi dari bulan ke-8 hingga bulan ke-9, setelah itu terjadi penurunan berat terus menerus hingga bulan ke-12. Perubahan perilaku pertumbuhan terjadi pada usia 8-12 bulan. Berat badan bayi ke-3 mengalami penurunan berat yang besar dari 9.8 kg menjadi 8.5 kg dengan perubahan perilaku terjadi pada usia 10 bulan

Terjadi penurunan berat bayi pada bulan ke 9,10,11 pada bayi ke-4. Berat badan bayi ke-5 mengalami peningkatan cukup stabil, serta tidak mengalami penambahan berat bayi pada bulan 6-7, 8-9 dan 11-12, sedangkan berat bayi ke-6 terus mengalami peningkatan setiap bulan. Bayi ke-7 mengalami pertumbuhan yang cepat pada usia 2-3 bulan dan 6-7 bulan, mengalami penurunan pada usia 8-10 bulan serta cenderung konstan pada usia 10-11 bulan,.

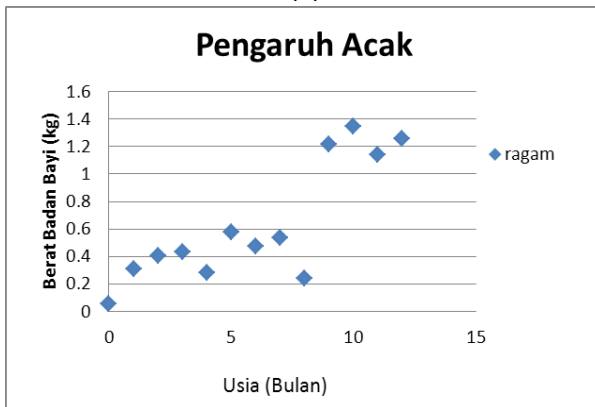
Pada bayi ke-8 terjadi penurunan berat pada saat berusia 6 bulan. Peningkatan berat badan bayi cenderung tidak konstan pada bayi ke-9. Secara umum, bentuk kecenderungan kurva pertumbuhan masing-masing bayi tidak mengikuti bentuk kecenderungan kurva parametrik serta terjadi perubahan perilaku pada usia-usia tertentu. Untuk memodelkan data longitudinal dengan perubahan perilaku tertentu, maka digunakan metode *Spline*. Eksplorasi data tidak hanya dilakukan untuk mendapatkan profil individu, tetapi juga pada pengaruh tetap dan pengaruh acak.

4.3 Eksplorasi Pengaruh Tetap dan Pengaruh Acak

Salah satu asumsi dasar model nonparametrik longitudinal adalah peubah respon tidak mengikuti pola kecenderungan fungsi nonparametrik baik *fixed-effect* (pengaruh tetap) maupun *random-effect* (pengaruh acak). Hal ini dilakukan melalui eksplorasi terhadap pola rata-rata dan pola ragam. Kesimpulan pengaruh tetap pada model diperoleh dari eksplorasi rata-rata, sedangkan ragam memberikan kesimpulan terhadap pengaruh acak.



(a)



(b)

Gambar 4.2. (a) Pengaruh Tetap dan (b) Pengaruh Acak

Gambar 4.2 (a) memperlihatkan adanya pengaruh perubahan waktu (usia) terhadap berat badan bayi. Pertumbuhan berat badan bayi mengalami peningkatan serta perubahan perilaku pada usia 5, 9 dan 10 bulan, serta bentuk kecenderungan kurva tidak mewakili bentuk parametrik. Akibat adanya perubahan pola pertumbuhan berat badan bayi menyebabkan kurva pertumbuhan tidak membentuk suatu fungsi yang diketahui (polinomial derajat ke-k), sehingga tidak dapat dimodelkan secara parametrik.

Pola kecenderungan ragam yang terbentuk pada Gambar 4.2 (b) menunjukkan adanya perubahan keragaman berat bayi terhadap waktu. Perilaku yang digambarkan memberikan bentuk kecenderungan kurva yang tidak mengikuti bentuk parametrik (polinomial derajat ke-k).

Perubahan pola pertumbuhan berat badan bayi di Kelurahan Bandulan baik pengaruh tetap maupun pengaruh acak akan membentuk suatu fungsi yang tidak diketahui. Model penduga didekati dengan metode Generalized Penalized Spline (GPS), yang memberikan informasi kapan atau usia berapa pertumbuhan berat badan bayi mengalami perubahan pola.

4.4 Penentuan Lokasi Titik Knot Pada GPS

Tiga model GPS sebagai pendekatan terhadap suatu bentuk kurva adalah linier (derajat 1), kuadratik (derajat 2) dan kubik (derajat 3). Berdasarkan bentuk kurva pertumbuhan berat badan bayi, banyaknya titik knot yang mungkin adalah 1, 2 dan 3. Metode untuk penentuan lokasi titik knot didasarkan pada *Equally Spaced* (ES), *Equally Spaced Samples Quantiles* (ESQ) dan *Selection Based* (SB). Hasil kombinasi yang diperoleh berdasarkan ketiga metode:

Tabel 4.2. Lokasi Titik Knot

K	Metode	letak titik knot		
		t1	t2	t3
1	ESP	6	-	-
	ESQ	6	-	-
	SB	3, 4, 5, 6, 7, 8, 9, 10, 11	-	-
2	ESP	4	8	-
	ESQ	4	9	-
	SB	3, 3, 3, 3, 4, 4, 4, 5, 5, 6	6, 7, 8, 9, 7, 8, 9, 8, 9, 9	-
3	ESP	3	6	9
	ESQ	3	6	9
	SB	3,4	6,7	9,10

4.5 Penentuan Parameter Penghalus Optimal

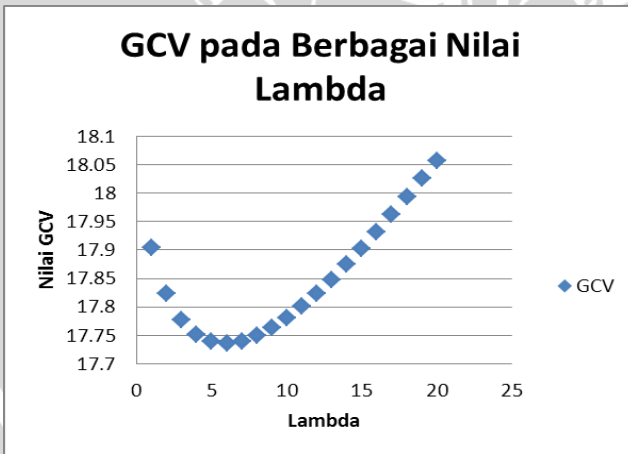
Bentuk penduga GPS dipengaruhi oleh parameter penghalus λ , sehingga sesuai atau tidaknya pendekatan ini bergantung pada parameter penghalus λ optimal. Pemilihan λ didekati dengan pendekatan *Generalized Cross Validation*. Berdasarkan output program GPS2

(Lampiran 3), nilai GCV pada berbagai lambda (λ) disajikan pada Tabel 4.3 menunjukkan nilai λ optimal = 6, dan nilai GCV sebesar 17.73697. Nilai λ tersebut digunakan untuk menghitung \hat{y} , berdasarkan titik knot optimal.

Tabel 4.3. Nilai GCV dan Lambda (λ)

Lambda	GCV	Lambda	GCV
1	17.90386	11	17.80068
2	17.82346	12	17.82356
3	17.77786	13	17.8485
4	17.75206	14	17.87516
5	17.73932	15	17.90323
6	17.73597	16	17.93249
7	17.73973	17	17.96273
8	17.74905	18	17.99376
9	17.76284	19	18.02544
10	17.78026	20	18.05764

Gambar 4.3 menunjukkan grafik GCV pada berbagai nilai λ , di mana pada $0 < \lambda < 6$ nilai GCV mengalami penurunan, lambda optimal belum tercapai karena belum didapatkan nilai GCV minimum. Pada nilai $6 < \lambda \leq 20$ terjadi peningkatan nilai GCV, sehingga nilai GCV minimum terjadi pada saat $\lambda=6$.



Gambar 4.3. Plot nilai λ terhadap GCV

4.6 Pemodelan Data Longitudinal dengan Pendekatan Generalized Penalized Spline

Setelah nilai λ optimal, dilakukan pemilihan titik knot untuk mendapatkan 9 model tentatif yaitu model dengan pendekatan GPS linier ($K = 1, 2, 3$), kuadratik ($K=1, 2, 3$) dan kubik ($K= 1, 2, 3$). Pemilihan model terbaik menggunakan kriteria R^2_{adjusted} terbesar.

4.6.1 Pendekatan GPS linier

Hasil eksekusi program GPS1 dengan $K=1$, $\lambda = 6$, nilai GCV pada berbagai kombinasi letak titik knot disajikan pada Tabel 4.3. Pemilihan titik knot optimal didasarkan pada nilai GCV terkecil. Mengacu hasil pada Tabel 4.4, didapatkan titik knot optimal yaitu pada $\tau_1 = 5$ dengan nilai $GCV = 18.04204$.

Tabel 4.4. Nilai GCV dengan pendekatan GPS linier satu titik Knot

No	k	K	t1	GCV
1	1	1	3	19.77097
2	1	1	4	18.56205
3	1	1	5	18.04204
4	1	1	6	18.04422
5	1	1	7	18.05205
6	1	1	8	18.2234
7	1	1	9	18.76123
8	1	1	10	20.30154
9	1	1	11	22.11055

Berdasarkan eksekusi program GPS3 didapatkan penduga parameter GPS linier satu titik knot yang disajikan pada Lampiran 6, sehingga model penduga adalah

$$\hat{y}_{ij} = 3.2975052 + 0.716647 t_{ij} - 0.35907(t_{ij} - \tau_1)_+$$

$R^2_{\text{adjusted}} 79.4544 \%$.

Selanjutnya dilakukan pendekatan GPS linier untuk $K=2$, $\lambda = 6$. Hasil pemilihan titik knot optimal dari hasil eksekusi program GPS2 berdasarkan kriteria nilai GCV disajikan pada Tabel 4.5, diperoleh nilai GCV terkecil yaitu 17.50674 dengan dua titik knot $\tau_1 = 5$ dan $\tau_2 = 9$. Hasil pendugaan parameter fungsi rata-rata populasi dengan pendekatan GPS ditampilkan pada Lampiran 7. Model penduga \hat{y}_{ij} yang terbentuk memiliki $R^2_{\text{adjusted}} = 79.857\%$:

$$\hat{y}_{ij} = 3.301371 + 0.701069t_{ij} - 0.26945(t_{ij} - \tau_1)_+ - 0.22507(t_{ij} - \tau_2)_+$$

Tabel 4.5. Nilai GCV dengan pendekatan GPS linier dua titik knot

No	k	K	t1	t2	GCV
1	1	2	3	6	18.11358
2	1	2	3	7	17.97485
3	1	2	3	8	17.89554
4	1	2	3	9	18.01526
5	1	2	3	10	18.77291
6	1	2	4	7	17.77144
7	1	2	4	8	17.60496
8	1	2	4	9	17.57177
9	1	2	5	8	17.59061
10	1	2	5	9	17.50674

Nilai GCV dari program GPS2 (Lampiran 8) $K=3$, $\lambda = 6$, dengan titik knot $\tau_1 = 3$, $\tau_2 = 6$ dan $\tau_3 = 9$ adalah 17.76411. Nilai R^2_{adj} yang dihasilkan sebesar 79.8101%, untuk model:

$$\hat{y}_{ij} = 3.264363 + 0.733285t_{ij} - 0.11778(t_{ij} - \tau_1)_+ - 0.21104(t_{ij} - \tau_2)_+ - 0.19913(t_{ij} - \tau_3)_+$$

4.6.2 Pendekatan GPS kuadratik

Nilai GCV pada satu titik knot, $\lambda = 6$ pada berbagai lokasi menggunakan pendekatan GPS kuadratik disajikan pada Tabel 4.6. Diperoleh nilai GCV minimum sebesar 17.60056 yang terletak pada titik knot $\tau_1 = 11$.

Tabel 4.6. Nilai GCV dengan pendekatan GPS kuadratik satu titik knot

No	K	K	t1	GCV
1	2	1	3	17.71466
2	2	1	4	17.77763
3	2	1	5	17.81913
4	2	1	6	17.83184
5	2	1	7	17.83134
6	2	1	8	17.83261
7	2	1	9	17.83372
8	2	1	10	17.77086
9	2	1	11	17.60056

Penduga parameter beta diperoleh dari hasil eksekusi program GPS3 yang disajikan pada Lampiran 9. Proporsi keragaman yang mampu dijelaskan oleh model sebesar $R^2_{adj} = 80.9453\%$:

$$\hat{y}_{ij} = 3.094699 + 0.90837t_{ij} - 0.03447t_{ij}^2 + 0.030231(t_{ij} - \tau_1)_+^2$$

Nilai GCV $K=2$, $\lambda = 6$ dengan berbagai lokasi menggunakan pendekatan GPS kuadratik disajikan pada Tabel 4.7. Titik knot optimal pada nilai GCV terkecil yaitu 17.91448 terletak pada $\tau_1 = 3$, $\tau_2 = 6$.

Koefisien $\hat{\beta}$ berdasarkan output program GPS3 disajikan di Lampiran 10, sebesar 80.8619% keragaman data mampu dijelaskan oleh model ($R^2_{adj} = 0.808619$):

$$\hat{y}_{ij} = 3.335605 + 0.692655t_{ij} + 0.013353t_{ij}^2 - 0.05673(t_{ij} - \tau_1)_+^2 + 0.046918(t_{ij} - \tau_2)_+^2$$

Tabel 4.7. Nilai GCV dengan pendekatan GPS kuadratik dua titik knot

No	k	K	t1	t2	GCV
1	2	2	3	6	17.91448
2	2	2	3	7	17.95713
3	2	2	3	8	17.97015
4	2	2	3	9	17.94715
5	2	2	3	10	17.84492
6	2	2	4	7	18.00837
7	2	2	4	8	18.02698
8	2	2	4	9	18.0034
9	2	2	5	8	18.05338
10	2	2	5	9	18.03658

Nilai GCV $K=3$, $\lambda = 6$ menggunakan pendekatan GPS kuadratik dengan titik knot $\tau_1 = 3$, $\tau_2 = 6$ dan $\tau_3 = 9$ adalah 18.10045. Proporsi keragaman yang mampu dijelaskan oleh model $R^2_{\text{adj}} = 80.6781\%$. Hasil eksekusi program GPS 3 disajikan pada Lampiran 11. Penduga model :

$$\hat{y}_{ij} = 3.225325 + 0.686135t_{ij} + 0.016168t_{ij}^2 - 0.06418(t_{ij} - \tau_1)_+^2 + 0.0102722(t_{ij} - \tau_2)_+^2 + 0.007449(t_{ij} - \tau_3)_+^2$$

4.6.3 Pendekatan GPS Kubik

Melalui pendekatan GPS kubik $K = 1$, $\lambda = 6$, nilai GCV yang diperoleh dengan berbagai lokasi titik knot disajikan pada Tabel 4.7. Titik knot optimal dengan kriteria GCV terkecil yaitu 17.82179 terletak pada titik knot $\tau_1 = 11$.

Tabel 4.8. Nilai GCV dengan pendekatan GPS kubik satu titik knot

No	k	K	t1	GCV
1	3	1	3	18.03393
2	3	1	4	18.03843
3	3	1	5	18.06394
4	3	1	6	18.08723
5	3	1	7	18.09081
6	3	1	8	18.06938
7	3	1	9	18.01066
8	3	1	10	17.9012
9	3	1	11	17.82179

Model penduga yang mampu menjelaskan hubungan antara peubah respon dan peubah penjelas dengan $R^2_{adj} = 80.8114\%$. Hasil program GPS tersaji pada Lampiran 12. Model yang diduga adalah

$$\hat{y}_{ij} = 3.168291 + 0.814568t_{ij} - 0.01388t_{ij}^2 - 0.00116t_{ij}^3 + 0.053846(t_{ij} - \tau_1)_+^3$$

Nilai GCV dengan banyaknya titik knot $K=2$, $\lambda = 6$ dengan pendekatan GPS kubik, pada berbagai lokasi titik knot tersaji pada Tabel 4.9. terlihat bahwa titik knot optimal terjadi pada saat $\tau_1 = 3$ dan $\tau_2 = 10$, dengan nilai GCV = 18.18014.

Penduga parameter betta dari hasil eksekusi program GPS3 disajikan pada Lampiran 13. Kekuatan model regresi yang terbentuk mampu menjelaskan hubungan peubah penjelas dan peubah respon sebesar 80.764% untuk model:

$$\hat{y}_{ij} = 3.255378 + 0.567742t_{ij} + 0.093362t_{ij}^2 - 0.0139t_{ij}^3 + 0.012597(t_{ij} - \tau_1)_+^3 + 0.054115(t_{ij} - \tau_2)_+^3$$

Tabel 4.9. Nilai GCV dengan pendekatan GPS kubik dua titik knot

No	k	K	t1	t2	GCV
1	3	2	3	6	18.32435
2	3	2	3	7	18.33933
3	3	2	3	8	18.3379
4	3	2	3	9	18.29467
5	3	2	3	10	18.18014
6	3	2	4	7	18.33724
7	3	2	4	8	18.34791
8	3	2	4	9	18.30704
9	3	2	5	8	18.36654
10	3	2	5	9	18.31558

Nilai GCV banyaknya titik knot $K = 3$, $\lambda = 6$ menggunakan pendekatan GPS kubik pada titik knot $\tau_1 = 3$, $\tau_2 = 6$ dan $\tau_3 = 9$ adalah 18.3294. Hasil output eksekusi program seperti terlihat pada Lampiran 14. Kekuatan model yang mampu dijelaskan peubah penjelas dan peubah terikat sebesar $R^2_{adj} = 80.6251\%$ untuk model:

$$\hat{y}_{ij} = 3.2691945 + 0.372947t_{ij} + 0.237365t_{ij}^2 - 0.03782t_{ij}^3 + 0.051457(t_{ij} - \tau_1)_+^3 - 0.03021(t_{ij} - \tau_2)_+^3 + 0.057139(t_{ij} - \tau_3)_+^3$$

4.7 Model Terbaik

Model terbaik dipilih berdasarkan kriteria R^2_{adj} dan MSE untuk masing-masing titik knot yang optimal pada model linier, kuadratik dan kubik dengan $K = 1, 2, 3$ disajikan pada Tabel 4.10. Kelemahan pendekatan nonparametrik adalah bersifat tidak unik, di mana akan didapatkan model penduga yang lebih dari satu dan bersifat bias. Keunggulan dari pendekatan ini, memiliki fleksibilitas yang tinggi, yaitu model penduga menyesuaikan bentuk kecenderungan kurva hubungan antar peubah respon dan peubah penjelas.

Tabel 4.10. Nilai R^2_{adj} untuk setiap derajat polinomial

Derajat Polinomial	Banyaknya Titik Knot	GCV	R^2_{adj}	MSE
Linier	1	18.04204	0.794544	0.64328
	2	17.50674	0.79857	0.64635
	3	17.76411	0.79801	0.64829
Kuadratik	1	17.60056	0.809453	0.63283
	2	17.91448	0.808619	0.63422
	3	18.10045	0.806781	0.64032
Kubik	1	17.82179	0.808114	0.63653
	2	18.18014	0.80764	0.63390
	3	18.3294	0.806251	0.64092

Pemilihan model terbaik berdasarkan metode semua kemungkinan regresi (*all possible regression*). Pada model linier 1 titik knot nilai $R^2_{adj} = 79.4544\%$, dengan menambahkan 1 titik knot (GPS linier 2 titik knot), R^2_{adj} bertambah sangat kecil yaitu 0.40%. Besarnya R^2_{adj} akan semakin kecil, jika ditambahkan 1 titik knot menjadi model GPS linier 3 titik knot.

Apabila model linier 1 titik knot, ditambahkan peubah baru yaitu waktu dikuadratkan (t^2_{ij}) maka R^2_{adj} akan bertambah sebesar 1.5%, dan akan semakin kecil jika ditambahkan 1 titik knot dan 2 titik knot. Peubah baru (t^3_{ij}) dimasukkan dalam GPS kuadratik 1 titik knot akan menurunkan R^2_{adj} sebesar 0.13%.

Nilai MSE terkecil yaitu sebesar 0.63283, sehingga model terbaik yaitu model GPS kuadratik 1 titik knot ($\tau_1 = 11$), dengan nilai $R^2_{adj} = 80.9453\%$, dengan model penduga yang terbentuk :

$$\hat{y}_{ij} = 3.094699 + 0.90837t_{ij} - 0.03447t_{ij}^2 + 0.030231(t_{ij} - 11)_+$$

di mana:

$$(t_{ij} - 11)_+ = \begin{cases} (t_{ij} - 11), & \text{jika } t_{ij} > 11 \\ 0, & \text{jika } t_{ij} \leq 11 \end{cases}$$

Nilai \hat{y}_{ij} adalah piecewises polinomial derajat 2, sehingga penduga model berat badan bayi di kelurahan Bandulan dapat ditulis dalam 2 model dengan 1 titik knot, seperti terlihat pada tabel berikut:

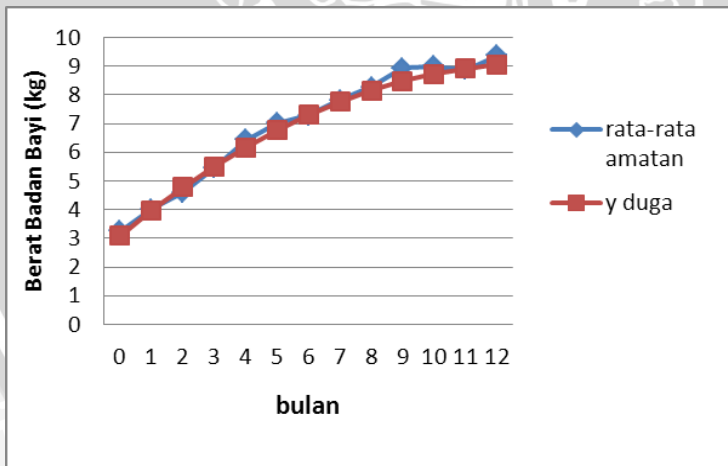
Tabel 4.11. Penduga model y_{ij} dalam interval waktu (t_{ij})

Bulan ke-	Model penduga (\hat{y}_{ij})
$t_{ij} \leq 11$	$3.094699 + 0.90837t_{ij} - 0.03447t_{ij}^2$
$t_{ij} > 11$	$6.75265 + 0.243218t_{ij} - 4.169 \cdot 10^{-3}t_{ij}^2$

Pendugaan rata-rata berat badan bayi di Kelurahan Bandulan :

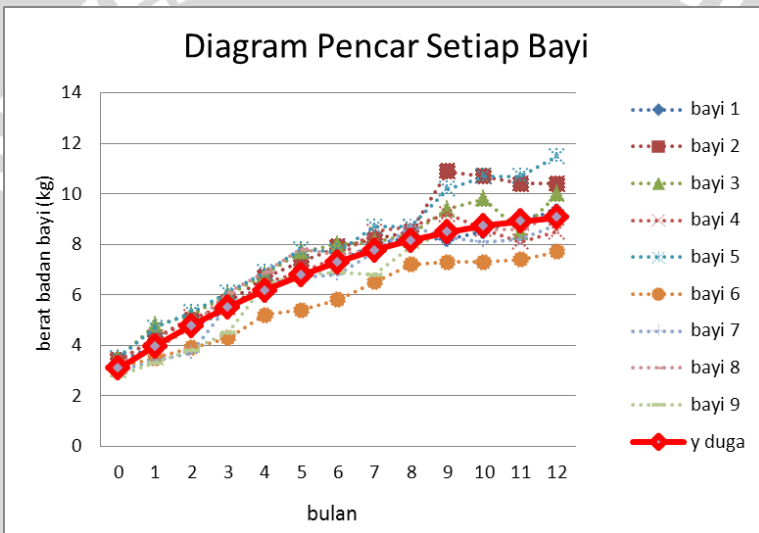
- Rata-rata berat badan bayi pada minggu ke-19 (4.75 bulan) dalam interval $t_{ij} \leq 11$, $\hat{y}(t) = 6.631395 \approx 6.63$ kg.
- Penduga berat badan bayi pada minggu ke-45 (11.25 bulan) pada interval $11 < t_{ij} \leq 12$, $\hat{y}(t) = 8.9612134 \approx 8.96$ kg.

Secara visual Gambar 4.4 memperlihatkan seberapa dekat model mampu menjelaskan berat badan bayi sebagai fungsi dari usia di kelurahan Bandulan. Diagram memperlihatkan bahwa garis \bar{y}_{ij} dengan \hat{y}_{ij} berhimpitan, yang menunjukkan bahwa berat badan bayi yang ditimbang dan penduga hampir sama.



Gambar 4.4. Plot antara \bar{y}_{ij} dengan \hat{y}_{ij}

Gambar 4.5 menyajikan pola perubahan berat badan bayi berdasarkan model penduga dan hasil pengukuran berat badan setiap bayi. Gambar tersebut menunjukkan berat badan bayi normal berada disekitar garis y duga. Apabila garis bayi ke- i di bawah garis y duga maka bayi dikatakan kurus (*underweight*), tetapi bila garis bayi ke- i di atas garis y duga maka bayi mengalami kegemukan (*overweight*). Berat badan bayi ke-6 dibawah garis y duga pada bulan ke-1 hingga bulan ke-12, hal ini perlu diwaspadai karena beresiko mengalami gizi buruk. Pengaruh acak mencerminkan seberapa besar penyimpangan berat badan bayi dari model penduga.



Gambar 4.5. Plot antara berat badan hasil pengukuran setiap bayi dan \hat{y}_{ij}

4.8 Pengujian Signifikansi Koefisien Vektor GPS

Pengujian koefisien β dilakukan untuk mengetahui apakah model yang terbentuk dapat dengan sebenarnya menggambarkan data. Pengujian parameter secara parsial untuk mengetahui apakah setiap parameter mempunyai pengaruh signifikan terhadap model berdasarkan statistik uji t .

Hipotesis yang melandasi pengujian signifikansi model secara parsial:

$$H_0 : \beta_0 = 0 \quad \text{lawan} \quad H_1 : \beta_0 \neq 0$$

$$H_0 : \beta_1 = 0 \quad \text{lawan} \quad H_1 : \beta_1 \neq 0$$

$$H_0 : \beta_2 = 0 \quad \text{lawan} \quad H_1 : \beta_2 \neq 0$$

$$H_0 : \beta_3 = 0 \quad \text{lawan} \quad H_1 : \beta_3 \neq 0$$

Berdasarkan hasil program GPS3 (Lampiran 8), hasil pengujian untuk setiap parameter dapat disajikan pada Tabel 4.8:

Tabel 4.12. Hasil Pengujian Parsial Terhadap Koefisien Parameter GPS

Beta	covbeta	thitung	pvalue	Keputusan
3.094699	0.000239	200.1348	0	Tolak H_0
0.90837	3.69E-05	149.4987	0	Tolak H_0
-0.03447	2.46E-07	-69.4551	0	Tolak H_0
0.030231	0.000142	2.538955	0.011118	Tolak H_0

Setiap parameter memberikan keputusan tolak H_0 dapat dikatakan bahwa β_0 , β_1 , β_2 dan β_3 berpengaruh signifikan terhadap model.

4.9 Uji asumsi Galat

Untuk menguji ketepatan model dilakukan pengujian asumsi galat. Model nonparametrik longitudinal akan dikatakan layak jika memenuhi asumsi kehomogenan ragam galat dan nilai harapan galat sama dengan nol.

4.9.1 Asumsi $E(\varepsilon) = 0$

Pengujian asumsi $E(\varepsilon) = 0$, menggunakan statistik uji t, berdasarkan hipotesis:

$$H_0 : \mu_\varepsilon = 0$$

$$H_1 : \mu_\varepsilon \neq 0$$

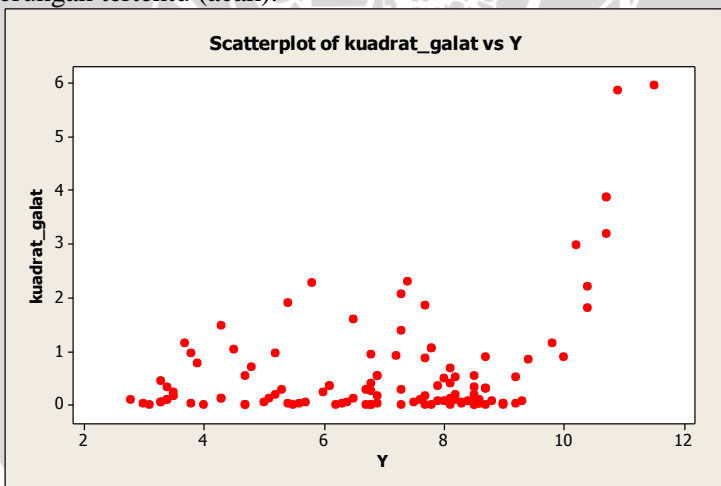
Tabel 4.9 memperlihatkan bahwa $p (0.79) > \alpha (0.05)$, $t (1.771) < t_{(0.025,116)} (2.27089)$ sehingga H_0 diterima, dapat dikatakan bahwa $E(\varepsilon) = 0$.

Tabel 4.13. Hasil Uji t Satu Contoh

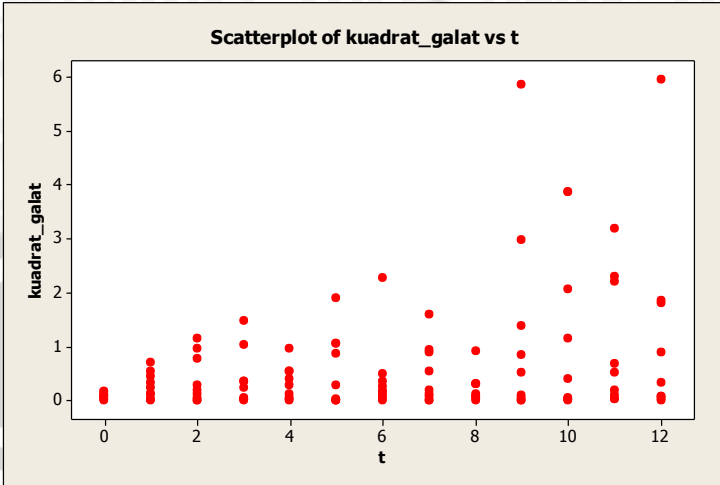
$\mu_0 = 0$					
t	Db	Sig.(2-arah)	Rata-Rata	Limit-limit kepercayaan 95% bagi μ	
				Batas Bawah	Batas Atas
1.771	116	.079	.12688	-.0150	.2687

4.9.2 Asumsi Kehomogenan Ragam Galat

Ragam galat dikatakan homogen jika tidak terdapat pola hubungan tertentu antara kuadrat galat dengan peubah respon dan peubah penjelas. Gambar 4.5 dan Gambar 4.6 plot hubungan kuadrat galat dengan peubah respon dan kuadrat galat dengan peubah penjelas. Kedua grafik memperlihatkan galat tidak membentuk pola kecenderungan tertentu (acak).



Gambar 4.6. Diagram Pencar antar Kuadarat galat dan peubah respon



Gambar 4.7. Diagram Pencar antar kuadrat galat dan peubah penjelas

Secara inferensia, digunakan uji Levene yang dilandasi pada hipotesis :

$$H_0 : \sigma_{a_i}^2 = \sigma^2$$

$$H_1 : \text{paling tidak ada satu } i \text{ di mana } \sigma_{a_i}^2 \neq \sigma^2$$

Tabel 4.14. Hasil Uji Levene

	Uji Levene	df1	df2	Sig.
Berdasarkan rata-rata	0.528	8	108	0.833
Berdasarkan nilai median	0.466	8	108	0.877

Berdasarkan nilai rata-rata dan median peubah respon pengujian kehomogenan ragam galat diperoleh keputusan yang sama. Karena nilai $p (0.833) > \alpha (0.05)$ dan $p (0.877) > \alpha (0.05)$ maka H_0 diterima, sehingga dapat dikatakan bahwa ragam galat berat badan bayi homogen.

4.10 Kelayakan Model

Model GPS kuadratik dengan satu titik knot adalah model terbaik. Uji parsial betta menghasilkan penduga parameter secara signifikan berpengaruh terhadap model. Keragaman berat badan bayi yang dapat dijelaskan oleh hubungan peubah penjelas (waktu) sebesar 80.945%. Usia merupakan faktor penentu terbesar terhadap berat badan bayi, dapat dikatakan bahwa model yang digunakan bisa menunjukkan keadaan yang sebenarnya (model sah).



UNIVERSITAS BRAWIJAYA



BAB V

KESIMPULAN DAN SARAN

5.1 Kesimpulan

Berdasarkan atas hasil analisis, dapat diambil beberapa kesimpulan:

- 1) Model GPS longitudinal nonparametrik pada data pertumbuhan berat badan bayi didapatkan dengan menduga $\hat{\beta} = (\hat{\beta}_0, \hat{\beta}_1, \hat{\beta}_1, \hat{\beta}_2)$. Nilai lambda optimal yang diperoleh $\lambda = 6$ dengan nilai GCV = 17.73697. Model GPS paling baik pada saat $K=1$ ($\tau_1 = 11$), $k = 2$ dengan nilai GCV 17.60056 dan proporsi keragaman 80.9453% berat badan bayi dipengaruhi oleh waktu.
- 2) Model GPS yang dihasilkan:

Bulan ke-	Penduga model (\hat{y}_{ij})
$t_{ij} \leq 11$	$3.094699 + 0.90837t_{ij} - 0.03447t_{ij}^2$
$t_{ij} > 11$	$6.75265 + 0.243218t_{ij} - 4.169 \cdot 10^{-3}t_{ij}^2$

5.2 Saran

Pemodelan data longitudinal nonparametrik pada penelitian ini dibatasi pada data longitudinal seimbang satu peubah respon (uniresponse) dengan $K = 3$, $k = 3$, sehingga pada penelitian selanjutnya disarankan untuk:

1. Memodelkan data longitudinal tidak seimbang
2. Memodelkan data longitudinal dengan nilai K dan k yang lebih besar, serta menggunakan peubah respon lebih dari satu baik biresponse maupun multirespon.
3. Menggunakan pendekatan nonparametrik longitudinal selain GPS, salah satunya dengan metode *polynomial smoothers kernel*.

UNIVERSITAS BRAWIJAYA



DAFTAR PUSTAKA

- Anderson, B. 2005. *Regression Spline*. <http://socserv.mcmaster.ca>. Tanggal akses: 5 Oktober 2011.
- Budiantara, I.N. 2006. *Model Spline dengan Knots Optimal*. Jurnal Natural FMIPA Universitas Jember.
- Didip, S. 2005. *Kajian Penggunaan Knot dan Orde pada Regresi Spline*. Skripsi Statistika. Universitas Brawijaya. Tidak dipublikasikan.
- Diggle, P.J., K.Y. Liang dan S.L. Zeger. 1994. *Analysis Of Longitudinal Data*. Oxford University Press. Oxford.
- Draper, N.R. dan H. Smith. 1992. *Analisis Regresi Terapan, Ed. 2*. Penerjemah: Bambang Sumantri. Gramedia. Jakarta.
- Eubank, R. 1988. *Spline Smoothing and Nonparametric Regression*. Marcel Dekker. New York.
- Fox, J. 2002. *Nonparametric Regression*. <http://www.math.itb.ac.id> . Tanggal akses : 5 Oktober 2011.
- Green, P. and B.W. Silverman.1994. *Nonparametric Regression and Generalized Linier Models*. Chapman and Hall. London.
- Gujarati, D. 2004. *Basic Econometrics*. McGraw-Hill, Inc. New York.
- Hardle, W. 1990. *Applied Nonparametric Regression*. Cambridge University Press. Boston.
- Laome, L. 2008. *Estimasi Parameter Pada Regresi Semiparametrik Untuk Data Longitudinal*. <http://jurnal.unhalu.ac.id>. Tanggal akses : 5 Oktober 2011.
- Myers, R.H. 1990. *Classical and Modern Regression With Applications*. Duxbury Press. California.

- Rachmadani, A. 2011. *Penerapan Model Gompertz Dan Logistic Pada Data Pertumbuhan Berat Badan Balita*. Skripsi Statistika. Universitas Brawijaya. Tidak dipublikasikan.
- Ruppert, D., M.P. Wand and R.J. Carroll. 2003. *Semiparametric Regression*. Cambridge University Press. New York.
- Steel, R.G.D. dan J.H. Torrie 1991. *Prinsip dan Prosedur Statistika*. Penerjemah: Bambang Soemantri. PT. Gramedia Pustaka Utama. Jakarta.
- Walpole, R.E. 1992. *Pengantar Statistika*. Penerjemah Bambang Soemantri. PT Gramedia. Jakarta.
- Wahba, G. 1990. *Spline Models for Observational Data*. SIAM. Philadelphia.
- Weiss, R. 2005. *Modelling Longitudinal Data*. Springer Text in Statistics. New York.
- Wu, H. and J.T. Zhang. 2006. *Nonparametric regression Methods for Longitudinal Data Analysis*. Wiley-Interscience. Singapore.

Lampiran 1. Data Berat Badan Bayi

BULAN	BAYI1	BAYI2	BAYI3	BAYI4	BAYI5	BAYI6	BAYI7	BAYI8	BAYI9
0	3.3	3.4	3.5	3.4	3.5	3.1	3	3.3	2.8
1	4	4.3	4.8	4.3	4.7	3.5	3.4	3.8	3.3
2	4.7	5	5.2	5.1	5.3	3.9	3.7	4.7	3.8
3	5.4	5.7	6	5.5	6.1	4.3	5.6	6.1	4.5
4	6.3	6.7	6.8	6.5	6.9	5.2	6.2	6.9	6.4
5	6.8	7.3	7.7	6.9	7.8	5.4	6.7	7.8	6.8
6	7.3	7.9	8	7.5	7.7	5.8	6.8	7.6	6.9
7	8	8.2	8.1	8.5	8.7	6.5	7.7	7.8	6.8
8	8.3	8.5	8.4	8.7	8.7	7.2	8.7	8.1	7.9
9	8.2	10.9	9.4	9.2	10.2	7.3	8.3	8.5	8.5
10	8.5	10.7	9.8	8.6	10.7	7.3	8.1	8.6	8.7
11	9	10.4	8.5	8.1	10.7	7.4	8.2	8.6	9
12	9.3	10.4	10	8.5	11.5	7.7	8.8	9.2	9

Sumber data: Rachmadani (2011)

Lampiran 2. Program GPS1

MACRO

GPS1 Y t p knots m lambda beta yfit error cR2adj cGCV

```
#####  
# PROGRAM GPS UNIRESPON PARAMETER ESTIMATE  
#  
# Y : Kolom Variabel Respon #  
# t : Kolom variabel prediktor time #  
# p : Kolom variabel identitas subjek #  
# m : Konstanta ordo polinomial (linier m=2, kuadrat m=3, kubik m=4)  
#  
# knots : Kolom titik knot dan ukuran knot #  
#####  
  
# Simpan di Drive Komputer D:\  
# Cara model ordo polinomial kubik dg memanggil %D:\GPS1.txt c1-c4  
4 6 c5-c9  
# Deklarasi variabel  
MCONSTANT m nknot npar i j n nt m1 np lambda  
MCONSTANT ybar SSR SST R2 R2adj db GCV  
MCOLUMN Y t p knots ct.1-ct.20 cv hc.1-hc.20 h1cv.1-h1cv.120  
MCOLUMN hcg.1-hcg.20 hclg.1-hclg.20 yfit cA error cgev cR2adj beta  
MMATRIX mt mv1 mg mlg myfit mtb1 mA  
MMATRIX my mtt mttv1 mttv1t mb1 mb2 mbeta  
MMATRIX me met metv1 metv1e  
  
# Membentuk matriks t  
# nknot adalah banyaknya knot  
# npar adalah banyaknya parameter  
# nt adalah banyaknya time repeated  
# n adalah banyaknya observasi keseluruhan  
# ct adalah kolom t  
# mt adalah matriks t  
# np adalah banyaknya subjek  
let nknot=count(knots)  
let npar=m+nknot
```

```

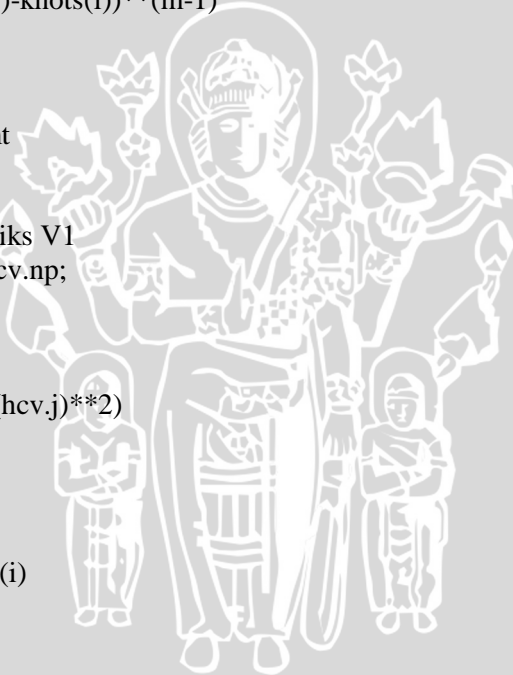
let n=count(Y)
let nt=max(t)
let np=max(p)
do i=1:m
  let ct.i=t**(i-1)
enddo
do i=1:nknot
  let m1=m+i
  do j=1:n
    if t(j)<knots(i)
      let ct.m1(j)=0
    else
      let ct.m1(j)=(t(j)-knots(i))**(m-1)
    endif
  enddo
enddo
copy ct.1-ct.npar mt
print mt

```

```

# Membentuk matriks V1
Unstack Y hcv.1-hcv.np;
subscripts p.
do i=1:n
  let j=p(i)
  let cv(i)=1/(stdev(hcv.j)**2)
enddo
do i=1:n
  do j=1:n
    if i=j
      let h1cv.i(j)=cv(i)
    else
      let h1cv.i(j)=0
    endif
  enddo
enddo
copy h1cv.1-h1cv.n mv1
print mv1

```



```
# Membentuk matriks G dan lambda G
```

```
do i=1:npar  
  do j=1:npar  
    if i=j  
      if i>m  
        let hcg.i(j)=1  
        let hclg.i(j)=lambda  
      else  
        let hcg.i(j)=0  
        let hclg.i(j)=0  
      endif  
    else  
      let hcg.i(j)=0  
      let hclg.i(j)=0  
    endif  
  enddo  
enddo  
copy hcg.1-hcg.npar mg  
copy hclg.1-hclg.npar mlg  
print mlg
```

```
# Menduga parameter Beta
```

```
copy Y my  
transpose mt mtt  
multiply mtt mv1 mttv1  
multiply mttv1 mt mttv1t  
add mttv1t mlg mb1  
invert mb1 mb1  
multiply mttv1 my mb2  
multiply mb1 mb2 mbeta  
copy mbeta beta
```

```
# Memperoleh nilai fit dan Rsquare adjusted
```

```
multiply mt mbeta myfit  
print myfit  
copy myfit yfit  
let ybar=mean(Y)  
let SSR=0
```

```

let SST=0
do i=1:n
  let SSR=SSR+(yfit(i)-ybar)**2
  let SST=SST+(y(i)-ybar)**2
enddo
let R2=SSR/SST
let R2adj=1-(1-R2)*((n-1)/(n-npar))
let cR2adj(1)=R2adj

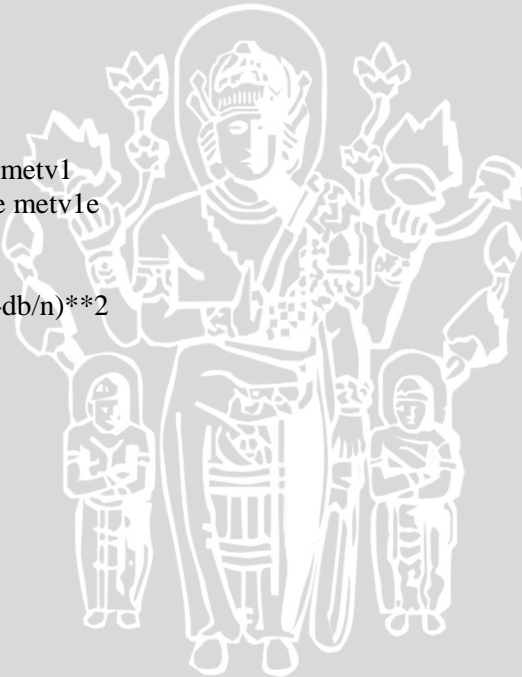
```

```

# Memperoleh nilai GCV
multiply mt mb1 mtb1
multiply mtb1 mttv1 mA
diagonal mA cA
let db=sum(cA)
let error=Y-yfit
copy error me
transpose me met
multiply met mv1 metv1
multiply metv1 me metvle
copy metvle cgcv
let GCV=cgcv(1)
let GCV=GCV/(1-db/n)**2
print GCV
let cgcv(1)=GCV

```

ENDMACRO



Lampiran 3. Program GPS2

MACRO

GPS2 Y t p knots m clambda cGCV cR2adj

```
#####  
# PROGRAM GPS UNIRESPON PARAMETER ESTIMATE DAN  
OPTIMALISASI LAMBDA #  
# Y : Kolom Variabel Respon #  
# t : Kolom variabel prediktor time #  
# p : Kolom variabel identitas subjek #  
# m : Konstanta ordo polinomial (linier m=2, kuadrat m=3, kubik m=4  
#  
# knots : Kolom titik knot dan ukuran knot #  
#####  
  
# Simpan di Drive Komputer D:\  
# Cara model ordo polinomial kubik dg memanggil %D:\GPS2.txt c1-c4  
4 C5-C7  
# Deklarasi variabel  
MCONSTANT m nknot npar i j n nt m1 np lambda loop mloop  
MCONSTANT ybar SSR SST R2 R2adj db GCV  
MCOLUMN Y t p knots ct.1-ct.20 cv hcv.1-hcv.20 h1cv.1-h1cv.120  
MCOLUMN hcg.1-hcg.20 hclg.1-hclg.20 yfit cA error cgcvcgcv1  
cR2adj clambda  
MMATRIX mt mv1 mg mlg myfit mtb1 mA  
MMATRIX my mtt mttv1 mttvlt mb1 mb2 mbeta  
MMATRIX me met metv1 metvle  
  
# Membentuk matriks t  
# nknot adalah banyaknya knot  
# npar adalah banyaknya parameter  
# nt adalah banyaknya time repeated  
# n adalah banyaknya observasi keseluruhan  
# ct adalah kolom t  
# mt adalah matriks t  
# np adalah banyaknya subjek  
let nknot=count(knots)
```



```

let npar=m+nknot
let n=count(Y)
let nt=max(t)
let np=max(p)
do i=1:m
  let ct.i=t**(i-1)
enddo
do i=1:nknot
  let m1=m-i
  do j=1:n
    if t(j)<knots(i)
      let ct.m1(j)=0
    else
      let ct.m1(j)=(t(j)-knots(i))**(m-1)
    endif
  enddo
enddo
copy ct.1-ct.npar mt

```

```

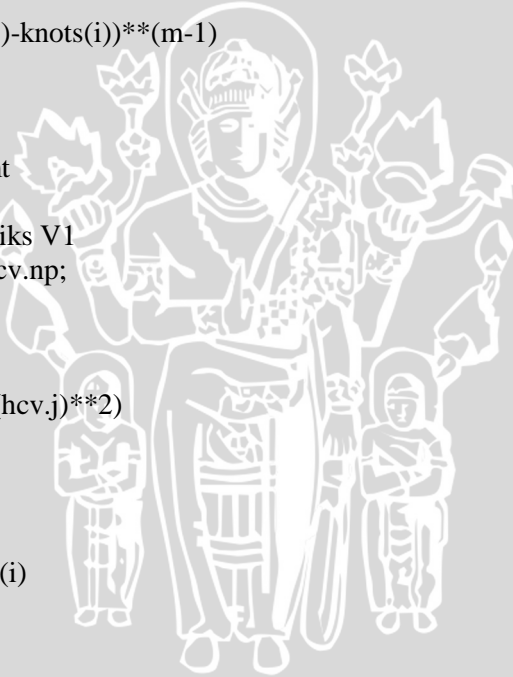
# Membentuk matriks V1
Unstack Y hcv.1-hcv.np;
subscripts p.
do i=1:n
  let j=p(i)
  let cv(i)=1/(stdev(hcv.j)**2)
enddo
do i=1:n
  do j=1:n
    if i=j
      let h1cv.i(j)=cv(i)
    else
      let h1cv.i(j)=0
    endif
  enddo
enddo
copy h1cv.1-h1cv.n mv1

```

```

# LOOPING TIAP NILAI LAMBDA

```



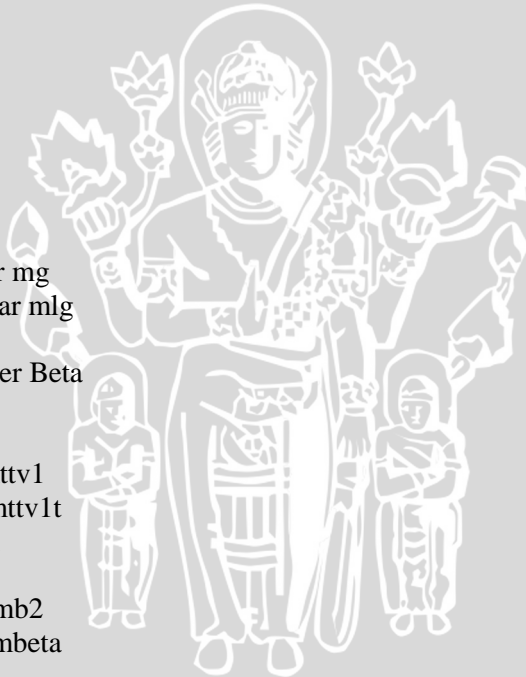
```

let mloop=count(clambda)
do loop=1:mloop
  let lambda=clambda(loop)
  # Membentuk matriks G dan lambda G
  do i=1:npar
    do j=1:npar
      if i=j
        if i>m
          let hcg.i(j)=1
          let hclg.i(j)=lambda
        else
          let hcg.i(j)=0
          let hclg.i(j)=0
        endif
      else
        let hcg.i(j)=0
        let hclg.i(j)=0
      endif
    enddo
  enddo
  copy hcg.1-hcg.npar mg
  copy hclg.1-hclg.npar mlg

  # Menduga parameter Beta
  copy Y my
  transpose mt mtt
  multiply mtt mv1 mttv1
  multiply mttv1 mt mttv1t
  add mttv1t mlg mb1
  invert mb1 mb1
  multiply mttv1 my mb2
  multiply mb1 mb2 mbeta

  # Memperoleh nilai fit dan Rsquare adjusted
  multiply mt mbeta myfit
  copy myfit yfit
  let ybar=mean(Y)
  let SSR=0

```



```

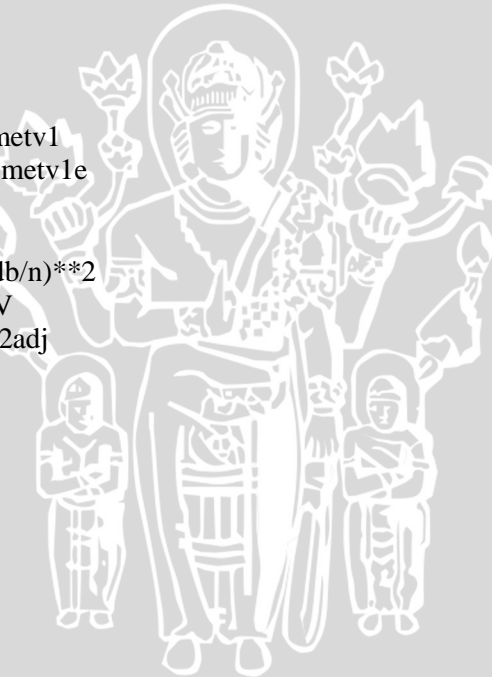
let SST=0
do i=1:n
  let SSR=SSR+(yfit(i)-ybar)**2
  let SST=SST+(y(i)-ybar)**2
enddo
let R2=SSR/SST
let R2adj=1-(1-R2)*((n-1)/(n-npar))

```

```

# Memperoleh nilai GCV
multiply mt mb1 mtb1
multiply mtb1 mttv1 mA
diagonal mA cA
let db=sum(cA)
let error=Y-yfit
copy error me
transpose me met
multiply met mv1 metv1
multiply metv1 me metv1e
copy metv1e cgcv1
let GCV=cgcv1(1)
let GCV=GCV/(1-db/n)**2
let cgcv(loop)=GCV
let cR2adj(loop)=R2adj
enddo
ENDMACRO

```



Lampiran 4. Program GPS3

MACRO

GPS1 Y t p knots m lambda beta covbeta thitung pvalue yfit error
cR2adj cGCV

```
#####  
# PROGRAM GPS UNIRESPON PARAMETER ESTIMATE  
#  
# Y : Kolom Variabel Respon #  
# t : Kolom variabel prediktor time #  
# p : Kolom variabel identitas subjek #  
# m : Konstanta ordo polinomial (linier m=2, kuadratik m=3, kubik  
m=4)#  
# knots : Kolom titik knot dan ukuran knot #  
#####  
  
# Simpan di Drive Komputer D:\  
# Cara model ordo polinomial kubik dg memanggil %D:\GPS3.txt c1-c4  
4 6 c5-c12  
# Deklarasi variabel  
MCONSTANT m nknot npar i j n nt m1 np lambda  
MCONSTANT ybar SSR SST R2 R2adj db GCV  
MCOLUMN Y t p knots ct.1-ct.20 cv hcv.1-hev.20 h1cv.1-h1cv.300  
MCOLUMN hcg.1-hcg.20 hclg.1-hclg.20 yfit cA error cgcvc cR2adj beta  
MCOLUMN covbeta thitung pvalue cb.1-cb.20 thitung1  
MMATRIX mt mv1 mg mlg myfit mtb1 mA  
MMATRIX my mtt mttv1 mttv1t mb1 mb2 mbeta  
MMATRIX me met metv1 metv1e  
  
# Membentuk matriks t  
# nknot adalah banyaknya knot  
# npar adalah banyaknya parameter  
# nt adalah banyaknya time repeated  
# n adalah banyaknya observasi keseluruhan  
# ct adalah kolom t  
# mt adalah matriks t  
# np adalah banyaknya subjek
```

```
let nknot=count(knots)
```

```
let npar=m+nknot
```

```
let n=count(Y)
```

```
let nt=max(t)
```

```
let np=max(p)
```

```
do i=1:m
```

```
  let ct.i=t**(i-1)
```

```
enddo
```

```
do i=1:nknot
```

```
  let m1=m-i
```

```
  do j=1:n
```

```
    if t(j)<knots(i)
```

```
      let ct.m1(j)=0
```

```
    else
```

```
      let ct.m1(j)=(t(j)-knots(i))**(m-1)
```

```
    endif
```

```
  enddo
```

```
enddo
```

```
copy ct.1-ct.npar mt
```

```
print mt
```

```
# Membentuk matriks V1
```

```
Unstack Y hcv.1-hcv.np;
```

```
  subscripts p.
```

```
do i=1:n
```

```
  let j=p(i)
```

```
  let cv(i)=1/(stdev(hcv.j)**2)
```

```
enddo
```

```
do i=1:n
```

```
  do j=1:n
```

```
    if i=j
```

```
      let h1cv.i(j)=cv(i)
```

```
    else
```

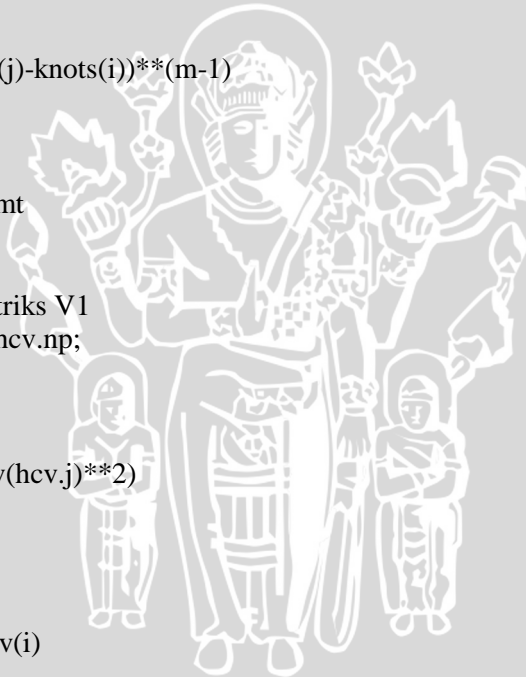
```
      let h1cv.i(j)=0
```

```
    endif
```

```
  enddo
```

```
enddo
```

```
copy h1cv.1-h1cv.n mv1
```



print mv1

Membentuk matriks G dan lambda G

do i=1:npar

do j=1:npar

if i=j

if i>m

let hcg.i(j)=1

let hclg.i(j)=lambda

else

let hcg.i(j)=0

let hclg.i(j)=0

endif

else

let hcg.i(j)=0

let hclg.i(j)=0

endif

enddo

enddo

copy hcg.1-hcg.npar mg

copy hclg.1-hclg.npar mlg

print mlg

Menduga parameter Beta dan Cov Beta

copy Y my

transpose mt mtt

multiply mtt mv1 mttv1

multiply mttv1 mt mttv1t

add mttv1t mlg mb1

invert mb1 mb1

multiply mttv1 my mb2

multiply mb1 mb2 mbeta

copy mbeta beta

copy mb1 cb.1-cb.npar

do i=1:npar

let covbeta(i)=cb.i(i)*.001

let thitung(i)=beta(i)/covbeta(i)**0.5

enddo

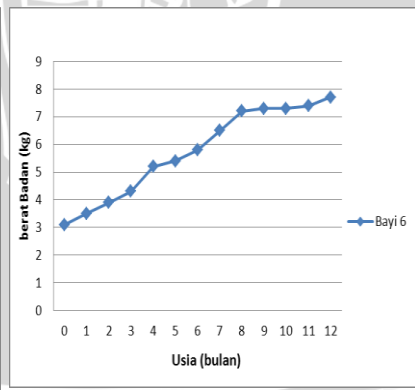
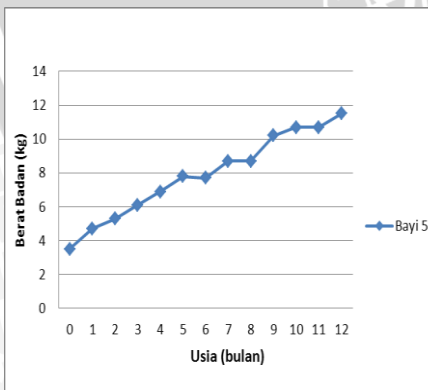
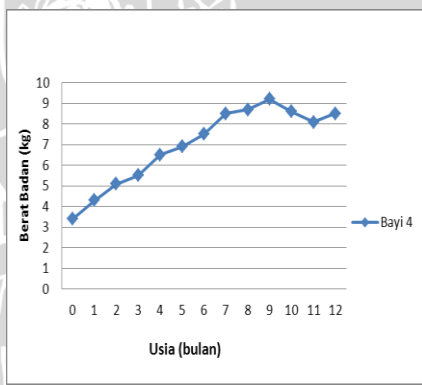
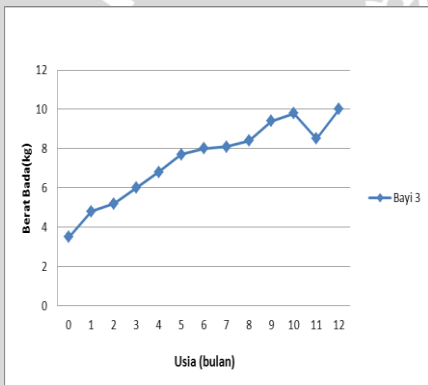
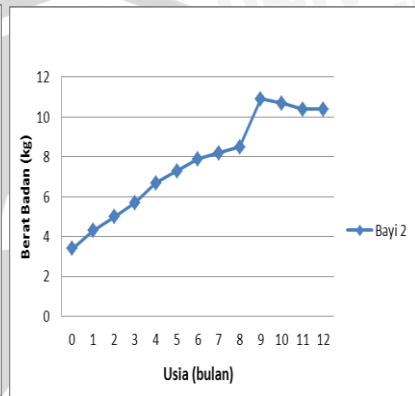
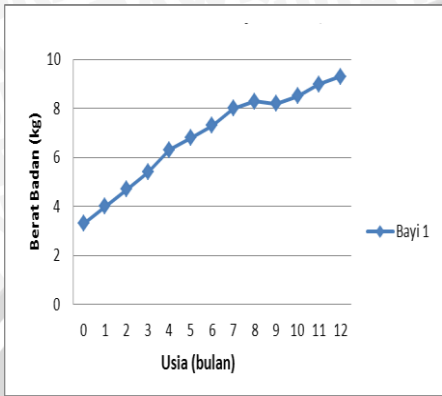
```
let thitung1=abso(thitung)
CDF thitung1 pvalue;
normal 0 1.
let pvalue=2*(1-pvalue)
```

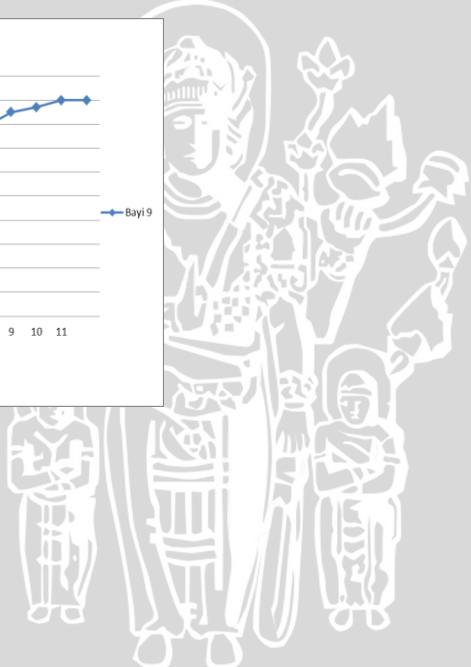
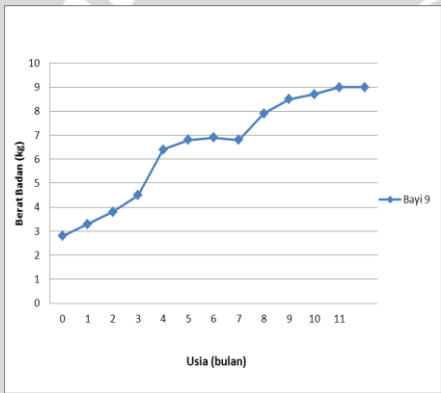
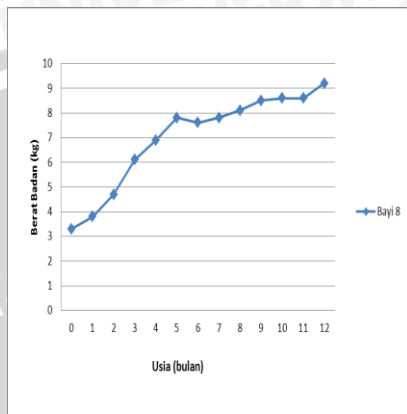
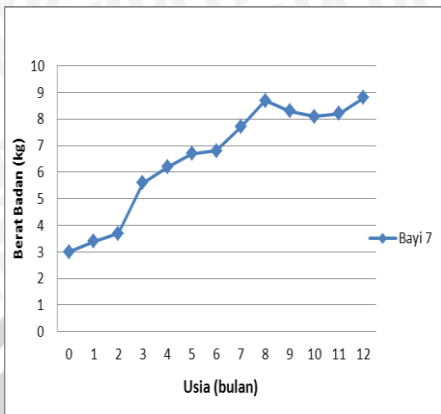
```
# Memperoleh nilai fit dan Rsquare adjusted
multiply mt mbeta myfit
print myfit
copy myfit yfit
let ybar=mean(Y)
let SSR=0
let SST=0
do i=1:n
  let SSR=SSR+(yfit(i)-ybar)**2
  let SST=SST+(y(i)-ybar)**2
enddo
let R2=SSR/SST
let R2adj=1-(1-R2)*((n-1)/(n-npar))
let cR2adj(1)=R2adj
```

```
# Memperoleh nilai GCV
multiply mt mb1 mtb1
multiply mtb1 mttv1 mA
diagonal mA cA
let db=sum(cA)
let error=Y-yfit
copy error me
transpose me met
multiply met mv1 metv1
multiply metv1 me metv1e
copy metv1e cgcv
let GCV=cgcv(1)
let GCV=GCV/(1-db/n)**2
print GCV
let cgcv(1)=GCV
```

```
ENDMACRO
```

Lampiran 5. Diagram Pencar Berat Badan Bayi dan Usia





Lampiran 6. Hasil pendugaan parameter pendekatan GPS linier satu titik knot ($\tau_1 = 5$)

knots	beta	covbeta	thitung	pvalue	GCV	R^2_{adj}
5	3.297052	0.000206	229.4887	0	18.04204	0.794544
	0.716647	1.54E-05	182.7734	0		
	-0.35907	3.39E-05	-61.6364	0		

yfit	error
3.297052	0.002948
4.0137	-0.0137
4.730347	-0.03035
5.446994	-0.04699
6.163641	0.136359
6.880288	-0.08029
7.237864	0.062136
7.59544	0.40456
7.953016	0.346984
8.310591	-0.11059
8.668167	-0.16817
9.025743	-0.02574
9.383319	-0.08332
3.297052	0.102948
4.0137	0.2863
4.730347	0.269653
5.446994	0.253006
:	:
7.953016	-0.05302
8.310591	0.189409
8.668167	0.031833
9.025743	-0.02574
9.383319	-0.38332

Lampiran 7. Hasil pendugaan parameter pendekatan GPS linier dua titik knot ($\tau_1 = 5$ dan $\tau_2 = 9$)

knots	beta	covbeta	thitung	pvalue	R^2_{adj}
5	3.301371	0.000206	229.7741	0	
	0.701069	1.57E-05	176.7778	0	
9	-0.26945	4.56E-05	-39.8813	0	
	-0.22507	7.39E-05	-26.1893	0	

yfit	error
3.301371	-0.00137
4.00244	-0.00244
4.703509	-0.00351
5.404579	-0.00458
6.105648	0.194352
6.806717	-0.00672
7.238335	0.061665
7.669953	0.330047
8.10157	0.19843
8.533188	-0.33319
8.739736	-0.23974
8.946283	0.053717
9.152831	0.147169
3.301371	0.098629
4.00244	0.29756
⋮	⋮
8.10157	-0.20157
8.533188	-0.03319
8.739736	-0.03974
8.946283	0.053717
9.152831	-0.15283

Lampiran 8. Hasil pendugaan parameter pendekatan GPS linier tiga titik knot ($\tau_1 = 3$, $\tau_2 = 6$ dan $\tau_3 = 9$)

knots	beta	covbeta	thitung	pvalue	R^2_{adj}
3	3.264363	0.00025	206.3057	0	
6	0.733285	4.08E-05	114.8497	0	
9	-0.11778	8.22E-05	-12.9875	0	
	-0.21104	6.34E-05	-26.4988	0	
	-0.19913	8.22E-05	-21.958	0	

yfit	error
3.264363	0.035637
3.997648	0.002352
4.730932	-0.03093
5.464217	-0.06422
6.079722	0.220278
6.695228	0.104772
7.310733	-0.01073
7.715195	0.284805
8.119656	0.180344
8.524117	-0.32412
8.72945	-0.22945
8.934783	0.065217
9.140115	0.159885
3.264363	0.135637
⋮	⋮
8.119656	-0.21966
8.524117	-0.02412
8.72945	-0.02945
8.934783	0.065217
9.140115	-0.14012

Lampiran 9. Hasil pendugaan parameter pendekatan GPS kuadratik satu titik knot ($\tau_1 = 11$)

knots	beta	covbeta	thitung	pvalue	R^2_{adj}
11	3.094699	0.000239	200.1348	0	
	0.90837	3.69E-05	149.4987	0	
	-0.03447	2.46E-07	-69.4551	0	
	0.030231	0.000142	2.538955	0.011118	

yfit	error
3.094699	0.205301
3.968595	0.031405
4.773544	-0.07354
5.509544	-0.10954
6.176597	0.123403
6.774702	0.025298
7.30386	-0.00386
7.764069	0.235931
8.155331	0.144669
8.477645	-0.27765
8.731012	-0.23101
8.91543	0.08457
9.061132	0.238868
3.094699	0.305301
⋮	⋮
8.155331	-0.25533
8.477645	0.022355
8.731012	-0.03101
8.91543	0.08457
9.061132	-0.06113

Lampiran 10. Hasil pendugaan parameter pendekatan GPS kuadratik dua titik knot ($\tau_1 = 3, \tau_2 = 6$)

knots	beta	covbeta	thitung	pvalue	R^2_{adj}
3	3.225605	0.000383	164.8598	0	
	0.692655	0.000395	34.86847	0	
	0.013353	1.73E-05	3.208525	0.001334	
10	-0.05673	2.36E-05	-11.6863	0	0.808619
	0.046918	6.38E-05	5.87399	4.25E-09	

yfit	error
3.225605	0.074395
3.931612	0.068388
4.664325	0.035675
5.423743	-0.02374
6.153138	0.146862
6.795781	0.004219
7.351674	-0.05167
7.820814	0.179186
8.203204	0.096796
8.498841	-0.29884
8.707728	-0.20773
8.87678	0.12322
9.052917	0.247083
⋮	⋮
8.203204	-0.3032
8.498841	0.001159
8.707728	-0.00773
8.87678	0.12322
9.052917	-0.05292

Lampiran 11. Hasil pendugaan parameter pendekatan GPS kuadratik tiga titik knot ($\tau_1 = 3$, $\tau_2 = 6$ dan $\tau_3 = 9$)

knots	beta	covbeta	thitung	pvalue	R^2_{adj}
3	3.225325	0.000383	164.7445	0	
	0.686135	0.000449	32.3976	0	
6	0.016168	2.41E-05	3.295912	0.000981	
	-0.06418	5.3E-05	-8.81236	0	
9	0.010272	3.06E-05	1.855424	0.063536	
	0.007449	5.3E-05	1.022888	0.306361	

yfit	error
3.225325	0.074675
3.927628	0.072372
4.662268	0.037732
5.429243	-0.02924
6.16438	0.13562
6.803503	-0.0035
7.346612	-0.04661
7.803979	0.196021
8.185876	0.114124
8.492302	-0.2923
8.730706	-0.23071
8.908539	0.091461
9.025799	0.274201
⋮	⋮
8.185876	-0.28588
8.492302	0.007698
8.730706	-0.03071
8.908539	0.091461
9.025799	-0.0258

Lampiran 12. Hasil pendugaan parameter pendekatan GPS kubik satu titik knot ($\tau_1 = 11$)

knots	beta	covbeta	thitung	pvalue	R^2_{adj}
11	3.168291	0.000335	173.008	0	
	0.814568	0.000193	58.58747	0	
	-0.01388	7.78E-06	-4.97741	6.44E-07	
	-0.00116	2.38E-08	-7.50084	6.35E-14	
	0.053846	0.000152	4.37204	1.23E-05	
				0.808114	

yfit	Error
3.168291	0.131709
3.967817	0.032183
4.732624	-0.03262
5.455763	-0.05576
6.130285	0.169715
6.749239	0.050761
7.305676	-0.00568
7.792647	0.207353
8.2032	0.0968
8.530388	-0.33039
8.76726	-0.26726
8.906867	0.093133
⋮	⋮
8.2032	-0.3032
8.530388	-0.03039
8.76726	-0.06726
8.906867	0.093133
8.996104	0.003896

Lampiran 13. Hasil pendugaan parameter pendekatan GPS kubik dua titik knot ($\tau_1 = 3$ dan $\tau_2 = 10$)

knots	beta	covbeta	thitung	pvalue	R^2_{adj}
3	3.255378	0.00044	155.2763	0	
	0.567742	0.001494	14.68651	0	
	0.093362	0.000331	5.130766	2.89E-07	
10	-0.0139	5.59E-06	-5.88223	4.05E-09	
	0.012597	7.03E-06	4.751584	2.02E-06	
	0.054115	3.83E-05	8.746318	0	

yfit	error
3.255378	0.044622
3.90258	0.09742
4.653097	0.046903
5.423518	-0.02352
6.143028	0.156972
6.791201	0.008799
7.360205	-0.0602
7.842208	0.157792
8.229381	0.070619
8.513891	-0.31389
8.687907	-0.18791
8.797713	0.202287
9.106053	0.193947
⋮	⋮
8.229381	-0.32938
8.513891	-0.01389
8.687907	0.012093
8.797713	0.202287
9.106053	-0.10605

Lampiran 14. Hasil pendugaan parameter pendekatan GPS kubik tiga titik knot ($\tau_1 = 3$, $\tau_2 = 16$ dan $\tau_3 = 9$)

knots	beta	covbeta	thitung	pvalue	R^2_{adj}
3	3.269145	0.000442	155.5136	0	
	0.372947	0.001932	8.484941	0	
6	0.237365	0.000565	9.989376	0	
	-0.03782	1.19E-05	-10.9546	0	
9	0.051457	2.32E-05	10.6863	0	
	-0.03021	8.86E-06	-10.145	0	
	0.057139	2.32E-05	11.86622	0	

yfit	Error
3.269145	0.030855
3.841635	0.158365
4.661925	0.038075
5.503085	-0.10309
6.189644	0.110356
6.751957	0.048043
7.271838	0.028162
7.800893	0.199107
8.269906	0.030094
8.579455	-0.37946
8.687257	-0.18726
8.779581	0.220419
⋮	⋮
8.269906	-0.36991
8.579455	-0.07946
8.687257	0.012743
8.779581	0.220419
9.099838	-0.09984