

**PEMILIHAN PARAMETER PENGHALUS PADA ESTIMATOR DERET  
FOURIER DALAM REGRESI NONPARAMETRIK**

**Agustini Tripena Br.Sb.**

Fakultas Sains dan Teknik, Universitas Jenderal Soedirman  
Purwokerto, Indonesia

**ABSTRACT.** Given regression function  $y_i = g(t_i) + e_i$ ,  $i = 1, 2, \dots, n$  with regression curve  $g(t_i)$  and  $e_i \sim N(0, \sigma^2)$ . Regression  $g(t_i)$  is assumed to be bounded in function space  $C(0, p)$ . Fourier Series estimator in nonparametric regression is obtained by mimimizing PLS  $n^{-1} \sum_{i=1}^n (y_i - g(t_i))^2 + I \int_0^p \frac{2}{p} (g^{(2)}(t))^2 dt$ . The solution of this optimization yields fourier series estimator is given by

$$\hat{g}_1(t) = \hat{b}(I)t + \frac{1}{2} \hat{a}_0(I) + \sum_{k=1}^K \hat{a}_k(I) \cos kt$$

with  $\hat{a}_k(I) = \left( \hat{b}(I), \frac{1}{2} \hat{a}_0(I), \hat{a}_1(I), \mathbf{K}, \hat{a}_K(I) \right)' = (n^{-1} \mathbf{X}' \mathbf{X} + I \mathbf{D})^{-1} n^{-1} \mathbf{X}' y$ ,

$D = \text{diag}(0, 0, 1^4, 2^4, \mathbf{K}, K^4)$  and  $\mathbf{X}$  is coefficient matrix.

Forier series estimator  $\hat{g}_1(t)$  is class of linear estimator in observations  $y$  and it has bias property for regression curve  $g(t)$ . Beside that, fourier series estimator  $\hat{g}_1(t)$  is normally distributed if error model has normal distribution.

Choice of smothing parameter in fourier series estimator using UBR is given by:

$$UBR(I) = n^{-1} y' (\mathbf{S}(I) - \mathbf{I})' (\mathbf{S}(I) - \mathbf{I}) y + n^{-1} \mathbf{S}^2 \text{trace}(\mathbf{S}'(I) \mathbf{S}(I)) + \\ - n^{-1} \mathbf{S}^2 \text{trace}(\mathbf{S}(I) - \mathbf{I})' (\mathbf{S}(I) - \mathbf{I})$$

Using trigonometrics and exponential function with  $n = 50, n = 100, n = 200$ ,  $\mathbf{S}^2 = 0,1$  dan,  $K = 5, K = 20$ , the simulation study yields that MSE for UBR is less than the CV's and GCV's

**Keywords :** Nonparametric Regression, Fourier Series, UBR

## 1. Pendahuluan

Analisa regresi merupakan metode yang banyak digunakan untuk mengetahui hubungan antara sepasang variabel atau lebih. Misalkan  $y$  adalah variabel respon dan  $t$  adalah variabel prediktor, maka hubungan variabel  $t$  dan  $y$  dapat dinyatakan sebagai berikut

$$y_i = g(t_i) + e_i, \quad i = 1, 2, \dots, n \quad (1)$$

$e_i$  adalah error random yang diasumsikan independen dengan mean nol dan variansi  $\sigma^2$  dan  $g(t_i)$  merupakan kurva regresi. Untuk mengestimasi  $g(t_i)$  ada dua pendekatan yang dapat digunakan yaitu pendekatan regresi parametrik dan regresi nonparametrik (Hardle, 1990).

Pendekatan regresi parametrik digunakan bila bentuk fungsi  $g(t_i)$  diketahui dari informasi sebelumnya berdasarkan teori ataupun pengalaman masa lalu. Sedangkan pendekatan regresi nonparametrik tidak memberikan asumsi terhadap bentuk kurva regresi sehingga memiliki fleksibilitas yang tinggi untuk mengestimasi kurva regresi  $g(t_i)$  (Mahler, 1995). Fungsi regresi  $g(t_i)$  hanya diasumsikan termuat dalam suatu ruang fungsi tertentu, dimana pemilihan ruang fungsi tersebut biasanya dimotivasi oleh sifat kemulusan (*smoothness*) yang dimiliki oleh fungsi  $g(t_i)$  tersebut.

Estimator-estimator kurva regresi ini mempunyai latar belakang dan motivasi tersendiri, sebagai suatu pendekatan untuk model data. Dalam tulisan ini akan dikaji pendekatan regresi nonparametrik pada Deret Fourier. Apabila fungsi  $g \in C(0, p) = \{g ; g \text{ kontinu pada } (0, p)\}$  maka ukuran kesesuaian kurva terhadap

data adalah  $n^{-1} \sum_{i=1}^n (y_i - g(t_i))^2$  dan ukuran kekasaran kurva adalah

$$\int_0^p \frac{2}{p} \left( g^{(2)}(t) \right)^2 dt.$$

Estimator  $g$  diperoleh dengan meminimumkan *Penalized Least Square*

$$n^{-1} \sum_{i=1}^n (y_i - g(t_i))^2 + I \int_0^P \frac{2}{p} (g^{(2)}(t))^2 dt \quad (2)$$

$I$  adalah parameter penghalus dan  $I > 0$ .

Untuk mendapatkan estimasi kurva regresi yang baik, diperlukan pemilihan  $I$  yang optimal dan merupakan hal yang sangat penting. Apabila digunakan Deret Fourier untuk mengestimasi kurva regresi pada (1) maka perlu dipilih suatu nilai  $I$  yang optimal. Beberapa metode untuk memilih  $I$  yaitu *Unbiased Risk (UBR)* (Wang, 1998), *Cross Validation (CV)* dan *Generalized Cross Validation (GCV)* (Craven dan Wahba, 1979). Untuk nilai  $I$  yang sangat besar akan menghasilkan estimator kurva regresi yang sangat halus. Sebaliknya untuk nilai  $I$  yang kecil akan memberikan estimator kurva regresi yang sangat kasar (Wahba, 1990; Eubank, 1988; Budiantara, 1998). Akibatnya pemilihan parameter penghalus optimal merupakan hal yang sangat penting dalam regresi nonparametrik.

Penelitian ini bertujuan untuk menentukan bentuk estimator Deret Fourier dalam model regresi nonparametric dan memilih parameter penghalus  $I$  optimal. Dengan metode UBR Selanjutnya akan dibandingkan metode UBR, CV dan GCV untuk memilih parameter penghalus  $I$  optimal dalam estimator Deret Fourier dengan menggunakan data simulasi.

## 2. Metode Penelitian

Penyelesaian penelitian ini dilakukan langkah-langkah sebagai berikut:

1. Mengkaji Estimator Deret Fourier dalam regresi nonparametrik
2. Menyelidiki sifat-sifat estimator Deret Fourier dalam regresi nonparametrik
3. Mengkaji pemilihan parameter penghalus dalam estimator Deret Fourier
4. Membandingkan metode UBR, CV dan GCV untuk memilih parameter penghalus optimal dalam estimator Deret Fourier dengan menggunakan data simulasi berdasarkan nilai MSE.

Untuk menyelesaikan langkah-langkah tersebut, perlu diketahui beberapa hal berikut ini.

## 2.1 Regresi Nonparametrik

Regresi nonparametrik merupakan suatu metoda Statistika yang digunakan untuk mengetahui hubungan antara variabel respon dan prediktor, jika bentuk hubungan antara variabel respon dan variabel prediktor tersebut tidak diketahui atau tidak didapat informasi sebelumnya. Jika diberikan pasangan data  $(t_i, y_i)$ ,  $i = 1, 2, \dots, n$  dan hubungan antara variabel respon  $y_i$  dengan independen  $t_i$  mengikuti model

$$y_i = g(t_i) + e_i, \quad t_i \in [a, b], \quad i = 1, 2, \dots, n \quad (3)$$

dimana  $g(t_i)$  kurva regresi yang tidak diketahui bentuknya dan  $e_i$  error random yang independen berdistribusi normal dengan mean nol variansi  $\sigma^2$ .

Dalam regresi nonparametrik fleksibilitas sangat dipertahankan, fungsi  $g(t_i)$  diasumsikan mulus dalam arti kontinu dan diffrensiabel (Eubank, 1988). Pendekatan regresi nonparametrik Deret Fourier diperoleh dengan meminimumkan *Penalized Least Squares*, yaitu kriteria pendugaan yang menggabungkan *goodness-of fit* dengan kemulusan kurva, dimana diantara keduanya dikontrol oleh suatu parameter pemulusan. Misalkan kurva regresi diasumsikan termuat didalam ruang  $C(0, p)$  yaitu  $g \in C(0, p)$ , maka estimasi untuk  $g$  diperoleh dengan meminimumkan *Penalized Leas Square (PLS)*

$$n^{-1} \sum_{i=1}^n (y_i - g(t_i))^2 + I \int_0^p \frac{2}{p} (g^{(m)}(t))^2 dt \quad (4)$$

dengan  $I$  merupakan parameter penghalus.

## 2.2 Fungsi Kerugian dan Fungsi Resiko

Berikut diuraikan definisi fungsi kerugian dan fungsi resiko yang direferensi dari Eubank, (1988) sebagai berikut :

**Definisi 1. Fungsi Kerugian**

Jika  $\hat{g}_{\mathbf{0}'} = (\hat{g}_{11}, \hat{g}_{12}, \dots, \hat{g}_{1n})'$  adalah estimator untuk  $g_{\mathbf{0}'} = (g_1, g_2, \dots, g_n)'$  maka fungsi kerugian (*loss function*) kuadrat didefinisikan sebagai

$$L(I) = n^{-1} \sum_{i=1}^n (g_i - \hat{g}_{1i})^2 . \quad (5)$$

**Definisi 2 .Fungsi Resiko**

Jika  $\hat{g}_{\mathbf{0}'} = (\hat{g}_{11}, \hat{g}_{12}, \dots, \hat{g}_{1n})'$  adalah estimator untuk  $g_{\mathbf{0}'} = (g_1, g_2, \dots, g_n)'$  maka ekspektasi dari fungsi resiko (*risk function*) kuadrat didefinisikan sebagai :

$$\begin{aligned} R(I) &= E[L(I)] \\ &= n^{-1} \sum_{i=1}^n E(g_i - \hat{g}_{1i})^2 . \end{aligned} \quad (6)$$

$L(I)$  dan  $R(I)$  merupakan ukuran dari kinerja suatu estimator

**2.3 Estimator Deret Fourier.**

Deret Fourier merupakan polinomial trigonometri yang mempunyai fleksibilitas, sehingga dapat menyesuaikan diri secara efektif terhadap sifat lokal data. Deret Fourier baik digunakan untuk menjelaskan kurva yang menunjukkan gelombang sinus dan cosinus. Diberikan data  $(t_i, y_i)$ ,  $i = 1, 2, \dots, n$  dan hubungan antara  $t_i$  dan  $y_i$  diasumsikan mengikuti model regresi

$$y_i = g(t_i) + e_i \quad (7)$$

dengan  $g$  adalah kurva regresi yang diasumsikan termuat dalam ruang  $C(0, p)$ . Bilangan  $e_i$  adalah *error* random yang diasumsikan berdistribusi independen dengan mean nol dan variansi  $\sigma^2$ . Dalam analisis regresi untuk estimasi kurva regresi  $g$  dapat digunakan metode *least square*, yaitu meminimumkan jumlah kuadrat *error*.

Dengan kata lain penduga untuk  $g$  dapat diperoleh dari

$$\text{Min}_{g \in C(0,p)} \left\{ \sum_{i=1}^n e_i^2 \right\} = \text{Min}_{g \in C(0,p)} \left\{ \sum_{i=1}^n (y_i - g(t_i))^2 \right\}. \quad (8)$$

Disamping meminimumkan (8) diberikan pula suatu *penalized* untuk ukuran kemulusan fungsi  $g$  sebagai berikut

$$\int_0^p \frac{2}{p} (g^{(2)}(t))^2 dt. \quad (9)$$

Dengan demikian estimator untuk kurva regresi  $g$  dapat diperoleh dari menyelesaikan optimasi PLS

$$\text{Min}_{g \in C(0,p)} \left\{ n^{-1} \sum_{i=1}^n (y_i - g(t_i))^2 + I \int_0^p \frac{2}{p} (g^{(2)}(t))^2 dt \right\} \quad (10)$$

dengan  $I$  merupakan parameter penghalus yang mengontrol antara *goodnes-of-fit* dan kemulusan fungsi. Untuk  $I$  yang sangat besar akan diperoleh fungsi penyelesaian yang sangat mulus (*smooth*). Sedangkan untuk  $I$  yang sangat kecil akan diperoleh fungsi penyelesaian yang sangat kasar. Karena  $g$  adalah fungsi yang kontinu maka  $g$  dapat dihamperi dengan fungsi  $T$ , dengan

$$T(t) = bt + \frac{1}{2}a_0 + \sum_{k=1}^K a_k \cos kt, \quad (11)$$

Untuk menyelesaikan (10) terlebih dahulu dicari nilai  $P(a)$  dengan

$$\begin{aligned} P(a) &= \int_0^p \frac{2}{p} \{g^{(2)}(t)\}^2 dt \\ &= \frac{2}{p} \int_0^p \sum_{k=1}^K (k^2 a_k \cos kt)^2 dt + \frac{2}{p} \int_0^p 2 \sum_{k < j} (k^2 a_k \cos kt)(j^2 a_j \cos jt) dt \\ &= \sum_{k=1}^K k^4 a_k^2. \end{aligned} \quad (12)$$

Berdasarkan (12) diperoleh PLS

$$\begin{aligned}
 & \underset{g \in C(0,p)}{\text{Min}} \left\{ n^{-1} \sum_{i=1}^n [y_i - g(t_i)]^2 + I \int_0^p \frac{2}{p} [g^{(2)}(t)]^2 dt \right\} \\
 & = \underset{b, a_0, a_1, \mathbf{K}, a_k \in R}{\text{Min}} \left\{ n^{-1} \sum_{i=1}^n \left( y_i - b t_i - \frac{1}{2} a_0 - \sum_{k=1}^K a_k \cos k t_i \right)^2 + I \sum_{k=1}^K k^4 a_k^2 \right\} \\
 & \quad n^{-1} \left( \underset{\%}{y} - \underset{\%}{\mathbf{X}} \underset{\%}{a} \right)' \left( \underset{\%}{y} - \underset{\%}{\mathbf{X}} \underset{\%}{a} \right) + I \underset{\%}{a}' \underset{\%}{\mathbf{D}} \underset{\%}{a}. \tag{13}
 \end{aligned}$$

Akibatnya persamaan (13) dapat ditulis

$$\begin{aligned}
 & \underset{g \in C(0,p)}{\text{Min}} \left\{ n^{-1} \sum_{i=1}^n [y_i - g(t_i)]^2 + I \int_0^p \frac{2}{p} [g^{(2)}(t)]^2 dt \right\} \\
 & = n^{-1} \underset{\%}{y}' \underset{\%}{y} - 2n^{-1} \underset{\%}{a}' \underset{\%}{\mathbf{X}}' \underset{\%}{y} + n^{-1} \left( \underset{\%}{a}' \underset{\%}{\mathbf{X}}' \underset{\%}{y} \right)' + \underset{\%}{a}' \left( n^{-1} \underset{\%}{\mathbf{X}}' \underset{\%}{\mathbf{X}} + I \underset{\%}{\mathbf{D}} \right) \underset{\%}{a}.
 \end{aligned}$$

Misalkan  $n^{-1} \underset{\%}{y}' \underset{\%}{y} - 2n^{-1} \underset{\%}{a}' \underset{\%}{\mathbf{X}}' \underset{\%}{y} + \underset{\%}{a}' \left( n^{-1} \underset{\%}{\mathbf{X}}' \underset{\%}{\mathbf{X}} + I \underset{\%}{\mathbf{D}} \right) \underset{\%}{a} = Q(\underset{\%}{a})$ .

Dengan menurunkan secara parsial  $Q(\underset{\%}{a})$  terhadap  $\underset{\%}{a}$  dan hasilnya disamakan dengan nol didapat

$$\frac{\partial Q(\underset{\%}{a})}{\partial \underset{\%}{a}} = 0 - 2n^{-1} \underset{\%}{\mathbf{X}}' \underset{\%}{y} + 2 \left( n^{-1} \underset{\%}{\mathbf{X}}' \underset{\%}{\mathbf{X}} + I \underset{\%}{\mathbf{D}} \right) \underset{\%}{a}.$$

Akibatnya, diperoleh estimator untuk  $\underset{\%}{a}$

$$\hat{\underset{\%}{a}}(I) = \left( n^{-1} \underset{\%}{\mathbf{X}}' \underset{\%}{\mathbf{X}} + I \underset{\%}{\mathbf{D}} \right)^{-1} n^{-1} \underset{\%}{\mathbf{X}}' \underset{\%}{y}. \tag{14}$$

Karenanya, estimator untuk kurva regresi  $g$  diberikan oleh

$$\hat{g}_I(t_i) = \hat{b}(I) t_i + \frac{1}{2} \hat{a}_0(I) + \sum_{k=1}^K \hat{a}_k(I) \cos k t_i \quad (\text{Bilodeau, 1992}). \tag{15}$$

## 2.4 Sifat-sifat Estimator Deret Fourier.

Berikut ini diberikan sifat-sifat yang dimiliki oleh estimator Deret Fourier.

Misalkan diberikan model regresi

$$y_i = g(t_i) + e_i ; i = 1, 2, \dots, n, \tag{16}$$

Persamaan (16) dapat ditulis dalam bentuk

$$\begin{pmatrix} y_1 \\ y_2 \\ \mathbf{M} \\ y_n \end{pmatrix} = \begin{pmatrix} t_1 & 1 & \cos t_1 & \cos 2t_1 & \mathbf{L} & \cos K t_1 \\ t_2 & 1 & \cos t_2 & \cos 2t_2 & \mathbf{L} & \cos K t_2 \\ t_3 & 1 & \cos t_3 & \cos 2t_3 & \mathbf{L} & \cos K t_3 \\ \mathbf{M} & \mathbf{M} & \mathbf{M} & \mathbf{M} & \mathbf{L} & \mathbf{M} \\ t_n & 1 & \cos t_n & \cos 2t_n & \mathbf{L} & \cos K t_n \end{pmatrix} \begin{pmatrix} b \\ \frac{1}{2} a_0 \\ a_1 \\ \mathbf{M} \\ a_K \end{pmatrix} + \begin{pmatrix} e_1 \\ e_2 \\ \mathbf{M} \\ e_n \end{pmatrix}. \quad (17)$$

Persamaan (17) dapat ditulis dalam bentuk matriks  $y = \mathbf{X}a + e$ .

Estimasi persamaan regresi  $g$  diberikan oleh

$$\hat{g}_I(t) = \mathbf{X} \hat{a}(I) = \mathbf{S}(I) y. \quad (18)$$

Berdasarkan (18) terlihat bahwa estimator Deret Fourier  $\hat{g}_I(t)$  merupakan kelas estimator linear dalam observasi  $y$ . Berikut ini diperlihatkan estimator Deret

Fourier adalah bias untuk kurva regresi  $g$ . Dicari ekspektasi dari  $\hat{g}_I(t)$  adalah

$$E\left(\hat{g}_I(t)\right) = E\left(\mathbf{S}(I) y\right) = \mathbf{S}(I) E(y). \quad (19)$$

Jadi  $E\left(\hat{g}_I(t)\right) \neq g(t)$ , artinya estimator  $\hat{g}_I(t)$  bias untuk kurva regresi  $g(t)$ .

Jika dalam penelitian ini *error* random dari model regresi Deret Fourier diasumsikan berdistribusi  $N(0, \mathbf{S}^2)$  maka berdasarkan persamaan (16) diperoleh Moment Generating Functions (MGF) untuk variabel respon  $y_i$  diberikan oleh

$$M_{y_i}(q) = M_{g(t_i)+e_i}(q) = e^{g(t_i)q + \frac{1}{2}q^2 \mathbf{S}^2}. \quad (20)$$

Selanjutnya, berdasarkan (4.12) dicari MGF untuk  $y$

$$M_y(q) = M_{\mathbf{X}a+e}(q). \quad (21)$$

Selanjutnya akan dicari distribusi dari estimator  $\hat{a}(I)$

$$\hat{a}(I) = \left(n^{-1} \mathbf{X}' \mathbf{X} + I \mathbf{D}\right)^{-1} n^{-1} \mathbf{X}' y.$$



MGF dari  $\hat{a}(I)$  adalah

$$\begin{aligned} M_{\hat{a}(I)}(q) &= M_{(n^{-1}\mathbf{X}'\mathbf{X}+I\mathbf{D})^{-1}n^{-1}\mathbf{X}'y} (q) \\ &= e^{q' \left[ n^{-1}\mathbf{X}(n^{-1}\mathbf{X}'\mathbf{X}+I\mathbf{D})^{-1}\mathbf{X}a \right] + \frac{1}{2}q' \left[ n^{-2}\mathbf{X}(n^{-1}\mathbf{X}'\mathbf{X}+I\mathbf{D})^{-2}\mathbf{X}'s^2 \right]} q. \end{aligned} \quad (22)$$

MGF dari distribusi normal dengan mean  $n^{-1}\mathbf{X}(n^{-1}\mathbf{X}'\mathbf{X}+I\mathbf{D})^{-1}\mathbf{X}a$  dan variansi  $n^{-2}\mathbf{X}(n^{-1}\mathbf{X}'\mathbf{X}+I\mathbf{D})^{-2}\mathbf{X}'s^2$ . Selanjutnya akan dicari distribusi dari estimator

$\hat{g}_I(t) = \mathbf{S}(I)y$ . MGF dari  $\hat{g}_I(t)$  diberikan oleh

$$\begin{aligned} M_{\hat{g}_I(t)}(q) &= M_{\mathbf{S}(I)y} (q) \\ &= e^{q'[\mathbf{S}'(I)\mathbf{X}a] + \frac{1}{2}q'[\mathbf{S}'(I)s^2\mathbf{S}(I)]} q. \end{aligned} \quad (23)$$

MGF dari distribusi normal dengan mean  $\mathbf{S}'(I)\mathbf{X}a$  dan variansi  $\mathbf{S}'(I)\mathbf{S}(I)s^2$ .

Karena  $\mathbf{S}(I) = \mathbf{X}(n^{-1}\mathbf{X}'\mathbf{X}+I\mathbf{D})^{-1}n^{-1}\mathbf{X}'$ , maka distribusi dari  $\hat{g}_I(t)$  adalah

normal dengan mean  $\mathbf{S}'(I)\mathbf{X}a = n^{-1}\mathbf{X}(n^{-1}\mathbf{X}'\mathbf{X}+I\mathbf{D})^{-1}\mathbf{X}'\mathbf{X}a$  dan variansi

$$\mathbf{S}'(I)\mathbf{S}(I)s^2 = n^{-2}\mathbf{X}(n^{-1}\mathbf{X}'\mathbf{X}+I\mathbf{D})^{-1}\mathbf{X}'\mathbf{X}(n^{-1}\mathbf{X}'\mathbf{X}+I\mathbf{D})^{-1}\mathbf{X}'s^2.$$

Jadi didapat

$$\hat{g}_I(t) \sim N_n \left( n^{-1}\mathbf{X}(n^{-1}\mathbf{X}'\mathbf{X} + \lambda\mathbf{D})^{-1}\mathbf{X}'\mathbf{X} a, n^{-2}\mathbf{X}(n^{-1}\mathbf{X}'\mathbf{X} + \lambda\mathbf{D})^{-1}\mathbf{X}'\mathbf{X}(n^{-1}\mathbf{X}'\mathbf{X} + \lambda\mathbf{D})^{-1}\mathbf{X}'\sigma^2 \right). \quad (24)$$

### 3. Hasil dan Pembahasan

#### 3.1 Pemilihan Parameter Penghalus dalam Estimator Deret Fourier

Diperoleh estimator Deret Fourier berbentuk  $\hat{g}_I(t) = \mathbf{S}(I)y$ . Estimator Deret Fourier ini sangat tergantung kepada  $I$  (parameter penghalus) Wahba, (1990) dan Eubank (1988) menyatakan parameter penghalus  $I$  mempunyai peran sangat penting dalam regresi nonparametrik.

Nilai  $I$  sangat kecil ( $I \rightarrow 0$ ) akan memberikan estimator yang sangat kasar, sebaliknya untuk nilai  $I$  sangat besar ( $I \rightarrow \infty$ ) akan menghasilkan estimator yang sangat mulus. Terdapat berbagai metode untuk memilih parameter penghalus optimal, salah satunya adalah metode UBR. Dalam penelitian ini akan diturunkan metode UBR untuk memilih parameter penghalus dalam estimator Deret Fourier. Didefinisikan fungsi Kerugian (*Loss*) kuadrat

$$L(I) = n^{-1} \sum_{i=1}^n (\hat{g}_I(t_i) - g(t_i))^2. \quad (25)$$

Berkaitan dengan fungsi Loss didefinisikan fungsi Resiko

$$R(I) = n^{-1} \sum_{i=1}^n E(\hat{g}_I(t_i) - g(t_i))^2. \quad (26)$$

Fungsi Resiko  $R(I)$  diatas dapat disajikan dalam bentuk vektor

$$R(I) = n^{-1} \mathbf{g}'(t) (\mathbf{S}(I) - \mathbf{I})' (\mathbf{S}(I) - \mathbf{I}) \mathbf{g}(t) + n^{-1} \mathbf{S}^2 \text{trace}(\mathbf{S}'(I) \mathbf{S}(I)). \quad (27)$$

Berikut akan dicari suatu kuantitas  $UBR(I)$  sedemikian sehingga

$$UBR(I) = n^{-1} \mathbf{y}' (\mathbf{S}(I) - \mathbf{I})' (\mathbf{S}(I) - \mathbf{I}) \mathbf{y} + n^{-1} \mathbf{S}^2 \text{trace}(\mathbf{S}'(I) \mathbf{S}(I)) + \\ - n^{-1} \mathbf{S}^2 \text{trace} \left( (\mathbf{S}(I) - \mathbf{I})' (\mathbf{S}(I) - \mathbf{I}) \right).$$

Selanjutnya akan dibuktikan  $UBR(I)$  tak bias untuk  $R(\lambda)$ . Perhatikan bahwa

$$E[UBR(I)] = E \left( \left[ n^{-1} \mathbf{y}' (\mathbf{S}(I) - \mathbf{I})' (\mathbf{S}(I) - \mathbf{I}) \mathbf{y} \right] + n^{-1} \mathbf{S}^2 \text{trace}[\mathbf{S}'(I) \mathbf{S}(I)] + \right. \\ \left. - n^{-1} \mathbf{S}^2 \text{trace} \left[ (\mathbf{S}(I) - \mathbf{I})' (\mathbf{S}(I) - \mathbf{I}) \right] \right) \\ = R(I). \quad (28)$$

Jadi diperoleh  $UBR(I)$  tak bias untuk  $R(I)$ .

Dengan demikian metode UBR untuk memilih parameter penghalus dalam estimator Deret Fourier diberikan oleh:

$$UBR(I) = n^{-1} \mathbf{y}' (\mathbf{S}(I) - \mathbf{I})' (\mathbf{S}(I) - \mathbf{I}) \mathbf{y} + n^{-1} \mathbf{S}^2 \text{trace}(\mathbf{S}'(I) \mathbf{S}(I)) + \\ n^{-1} \mathbf{S}^2 \text{trace}(\mathbf{S}(I) - \mathbf{I})' (\mathbf{S}(I) - \mathbf{I}). \quad (29)$$

Bilangan  $\lambda$  optimal diperoleh dari nilai  $I$  yang meminimumkan  $UBR(I)$ .

Apabila  $s^2$  tidak diketahui,  $s^2$  dapat diestimasi dengan

$$\hat{s}^2 = \frac{y'(\mathbf{I}-\mathbf{S}(I))'(\mathbf{I}-\mathbf{S}(I))y}{\text{trace}[(\mathbf{I}-\mathbf{S}(I))]} \quad (30)$$

### 3.2 Metode UBR, CV dan GCV pada Estimator Deret Fourier dengan Menggunakan Data Simulasi

Dalam penelitian ini dilakukan simulasi untuk memberikan gambaran tentang model regresi Deret Fourier. Simulasi dalam penelitian ini dilakukan untuk mengevaluasi kebaikan metode UBR, CV dan GCV kemudian membandingkan kebaikan antara ketiga metode tersebut. Realibilitas pengukuran berdasarkan nilai MSE terkecil yang diperoleh pada metode UBR, CV dan GCV. Simulasi ini menggunakan Software MATLAB version 6.5. Diberikan dua kurva percobaan yaitu:

1. Fungsi Trigonometri :  $g(t) = \sin(2\pi t^3)$
2. Fungsi Eksponen :  $g(t) = 4,26(e^{-t} - 4e^{-2t} + 3e^{-3t})$ .

Cara simulasi diuraikan sebagai berikut. Membangun model regresi  $y_i = g(t_i) + e_i$ ,  $i = 1, 2, \dots, n$ , dan ukuran sampel  $n = 50, n = 100, n = 200$  dengan nilai  $K = 5, K = 20$ . Diberikan  $t_i = \frac{i}{n}$ , dengan  $i = 1, 2, 3, \dots, n$ . Membangkitkan  $e_i \sim N(0, s^2)$ , dengan  $s^2 = 0,1$ . Setelah membangkitkan data, dihitung matriks  $\mathbf{X}$ , matriks  $\mathbf{D}$  dan matriks hat  $\mathbf{S}(I) = \mathbf{X}(n^{-1}\mathbf{X}'\mathbf{X} + I \mathbf{D})^{-1} n^{-1} \mathbf{X}'$ .

Selanjutnya dihitung estimator  $\hat{\mathbf{a}}_{\%}(I) = (n^{-1}\mathbf{X}'\mathbf{X} + I \mathbf{D})^{-1} n^{-1} \mathbf{X}' y_{\%}$ , dan menentukan estimator Deret Fourier dengan  $\hat{g}_{\%}(t) = \mathbf{S}(I) y_{\%}$ , sehingga didapat

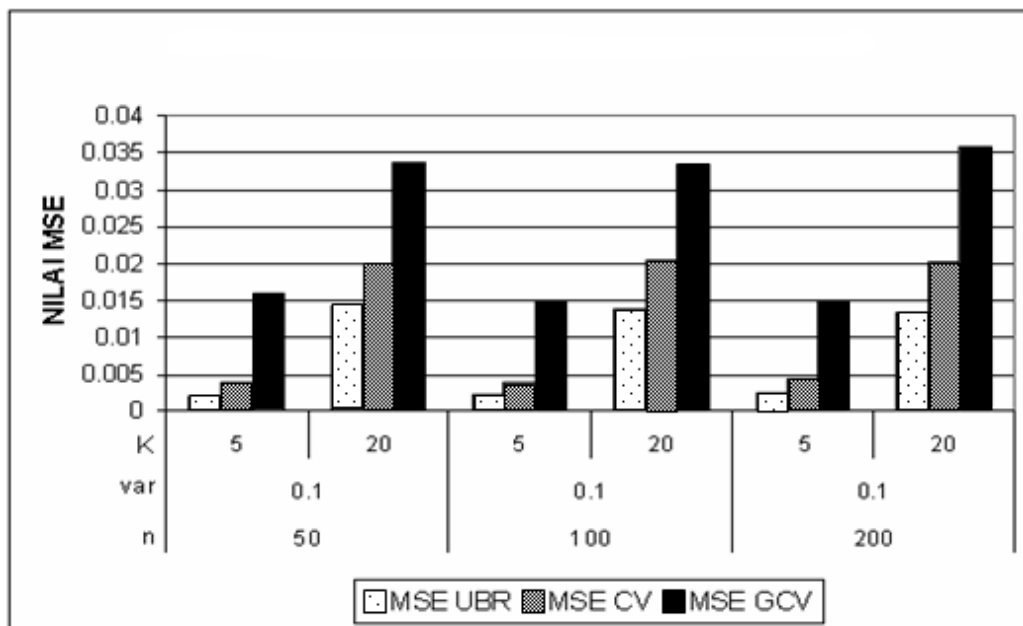
$$\hat{g}_I(t) = \hat{b}(I)t + \frac{1}{2}\hat{a}_0(I) + \sum_{k=1}^K \hat{a}_k(I) \cos kt.$$

### 3.3 Simulasi Fungsi Trigonometri $g(t) = \sin(2p t^3)$

Tabel 1 Hasil Simulasi Nilai  $I$  Optimal dan MSE Metode UBR, CV dan GCV pada Estimator Deret Fourie dengan Fungsi  $g(t) = \sin(2p t^3)$ , dengan  $n = 50, n = 100, n = 200, s^2 = 0,1$  dan  $K = 5, K = 20$

N	Var	k	Metode UBR		Metode CV		Metode GCV	
			$I_{optimal}$	MSE UBR	$I_{optimal}$	MSE CV	$I_{optimal}$	MSE GCV
50	0.1	5	0,2067	0,0019053	0,107620	0,004077	0,015056	0,021495
		20	0,08617	0,014198	0,045473	0,019947	0,003230	0,0359330
100	0.1	5	0,20015	0,0021103	0,102020	0,0040874	0,014231	0,0179270
		20	0,094627	0,013752	0,048525	0,020311	0,003442	0,0359630
200	0.1	5	0,1808	0,0023361	0,091257	0,0044296	0,012714	0,0179280
		20	0,10285	0,013313	0,052058	0,020157	0,003691	0,0375740

Secara visual nilai MSE pada Tabel 1 dapat dilihat pada diagram batang Gambar 1



Gambar 1: Diagram Batang Nilai MSE untuk Fungsi Trigonometri  $g(t)$ , untuk  $n = 50, n = 100, n = 200, s^2 = 0,1$  dan  $K = 5, K = 20$

**3.4 Simulasi Fungsi Eksponen**  $g(t) = 4,26(e^{-t} - 4e^{-2t} + 3e^{-3t})$

Berikut diuraikan hasil simulasi untuk fungsi eksponen.

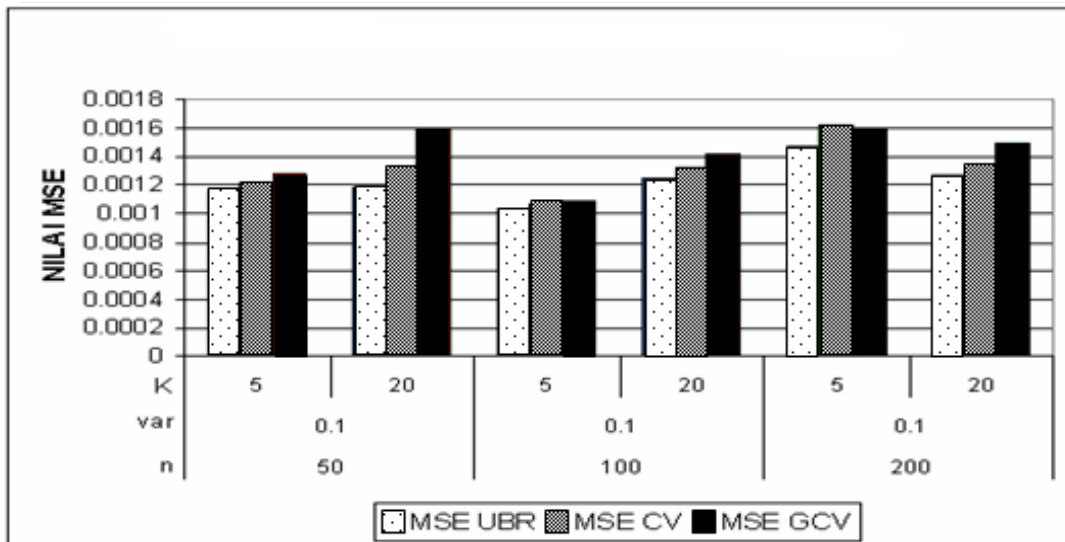
Tabel 2 Hasil Simulasi nilai  $I$  Optimal dan MSE Metode UBR, CV dan GCV

pada Estimator Deret Fourie dengan Fungsi  $g(t) = 4,26(e^{-t} - 4e^{-2t} + 3e^{-3t})$ ,

$n = 50, n = 100, n = 200, S^2 = 0,1$  dan  $K = 5, K = 20$

N	Var	k	Metode UBR		Metode CV		Metode GCV	
			$I_{optimal}$	MSE UBR	$I_{optimal}$	MSE CV	$I_{optimal}$	MSE GCV
50	0.1	5	0,208130	0,0029387	0,108360	0,0056457	0,015160	0,02517
		20	0,023513	0,0011852	0,012408	0,0013142	0,000881	0,0015975
100	0.1	5	0,018596	0,0010388	0,009479	0,0010671	0,001322	0,0010937
		20	0,017993	0,0012463	0,009227	0,001326	0,000654	0,0014255
200	0.1	5	0,025844	0,0014638	0,013044	0,0016284	0,001807	0,0018173
		20	0,019389	0,0012701	0,009814	0,0013433	0,000696	0,0014924

Secara visual nilai MSE pada Tabel 2 dapat dilihat pada diagram batang Gambar 2.



Gambar 2: Diagram Batang Nilai MSE Fungsi Eksponen  $g(t) = 4,26(e^{-t} - 4e^{-2t} + 3e^{-3t})$ ,  $n = 50, n = 100, n = 200, S^2 = 0,1$  dan  $K = 5, K = 20$

#### 4. Kesimpulan

Berdasarkan analisis dan pembahasan yang telah dilakukan maka dapat diambil beberapa kesimpulan sebagai berikut:

1. Berdasarkan hasil kajian estimator Deret Fourier dalam regresi nonparametrik

diperoleh dari meminimumkan PLS  $n^{-1} \sum_{i=1}^n (y_i - g(t_i))^2 + I \int_0^P \frac{2}{P} (g^{(2)}(t))^2 dt$ .

Optimasi ini berupa estimator Deret Fourier

$$\hat{g}_I(t_i) = \hat{b}(I)t_i + \frac{1}{2}\hat{a}_0(I) + \sum_{k=1}^K \hat{a}_k(I) \cos k t_i$$

dengan  $\hat{a}_I(I) = \left( \hat{b}(I), \frac{1}{2}\hat{a}_0(I), \hat{a}_1(I), \mathbf{K}, \hat{a}_K(I) \right)'$ .

2. Estimator Deret Fourier  $\hat{g}_I(t)$  merupakan estimator linear dalam observasi  $y$  dan mempunyai sifat yang bias untuk kurva regresi  $g(t)$ . Estimator Deret Fourier  $\hat{g}_I(t)$  berdistribusi normal, apabila error model juga berdistribusi normal.
3. Pemilihan parameter penghalus dalam estimator Deret Fourier dengan metode UBR diberikan oleh

$$UBR(I) = n^{-1} y'(S(I) - I)'(S(I) - I)y + n^{-1} S^2 \text{trace}(S'(I)S(I)) + n^{-1} S^2 \text{trace}(S(I) - I)'(S(I) - I).$$

4. Berdasarkan hasil simulasi dengan fungsi trigonometri dan fungsi eksponen untuk  $n = 50, n = 100, n = 200, S^2 = 0,1$ , dan  $k = 5, k = 20$  diperoleh bahwa nilai MSE metode UBR cenderung lebih kecil dari pada nilai MSE metode CV dan nilai MSE GCV pada setiap model simulasi. Untuk nilai MSE metode CV lebih kecil dari pada nilai MSE metode GCV pada setiap model simulasi. Makin besar nilai  $k$  maka memberikan nilai MSE makin besar baik metode UBR, CV dan GCV. Dapat disimpulkan bahwa pemilihan parameter penghalus  $I$  optimal dengan metode UBR cenderung lebih baik dibanding dengan metode CV dan GCV.

## 5. Daftar Pustaka

Budiantara, I N., dan Subanar, *Pemilihan Parameter Penghalus Dalam Regresi Spline*, Majalah Ilmiah Matematika dan Ilmu Pengetahuan Alam, **7**(1997), 37 – 49.

Budiantara, I N., *Parameter Penghalus Dalam Regresi Nonparametrik*, Prosiding Seminar Alumni S-2 Matematika PPS UGM, 1998..

Bilodeau, M, *Fourier Smoother and Additive Models*, The Canadian Journal of Statistics, **3**(1992), 257 – 269.

Craven, P., and Wahba, G., *Smoothing Noisy Data with Spline Functions: Estimating the Correct Degree of Smoothing by the Method of Generalized Cross-Validation*, Numer Math, **31**(1997), 377- 403.

Draper, N., and Smith, H., *Applied Regression Analysis*, John Wiley & Sons, New York, 1996.

Eubank, R.L., *Spline Smoothing and Nonparametric Regression*, Marcel Dekker, New York, 1998.

Green, P.J., dan Silverman, B.W., *Nonparametrik Regression and Generalized Linear Model*, Chapman & Hall, London, 1994.

Hardle, G., *Applied Nonparametric Regression*, Cambridge University Press, New York, 1990.

Kohn. R, *The performance of cross validation and maximum likelihood estimators of spline smoothing parameters*, Journal of the American Statistical Association, **86**(1991), 1042 -1050.

Lehman, R, *Teory of Point Estimation*, John Wiley and Sons, New York, 1983.

Li, K. C., Asymtotic optimality of  $C_L$  and Generalized Cross Validation in Ridge Regression with application to spline smoothing, *Ann. Statist.*, **14**(1986)), 1101 - 1112.

---

Mahler, *Variational Solution Of Penalized Likelihood Problem and Smooth Curve*, Annal of Statistic, **23**(1985), 1496-1517.

Montgomery, D.C. and Peck, EA, *Introduction to linear Regression Analysis*, New York, John Wiley & Sons, 1982.

Seber, G.A.F., *Linear Regression Analysis*, John Wiley & Sons, New York, 1977.

————— dan Lee, A.J., *Linear Regression Analysis*, Second Edition, John Wiley & Sons Inc., New Jersey Canada, 2003.

Shao, *Linear Model Selection by Cross Validation*, *Journal of the American Statistical Association*, **88**(1993), 486-494.

Simonoff, *Smoothing Methods in Statistics*, Springer-Verlag, 1996.

Wahba, G., *A Comparison of GCV dan GML for Chosing the Smoting Parameter in Generalized Spline Smoothing Problem*, Annal of Statistic, **13**(1985), 1378-1402.

Wahba, G., *Spline Models Opservatione Data*, SIAM, Pennsylvania, 1990.

Wang, Y., *Smoothing Spline Models with Corelated Errorrs*, Journal of the American Statistic Association, **93**(1998), 343-348.