

# PEMILIHAN PARAMETER THRESHOLD OPTIMAL DALAM ESTIMATOR REGRESI WAVELET THRESHOLDING DENGAN PROSEDUR *FALSE DISCOVERY RATE* (FDR)

Suparti<sup>1</sup>, Tarno<sup>1</sup> dan Yon Haryono<sup>2</sup>  
<sup>1</sup> Staf Jurusan Matematika FMIPA UNDIP  
<sup>2</sup> Alumni Jurusan Matematika FMIPA UNDIP

## Abstract

If  $X$  is predictor variable and  $Y$  is response variable of following model  $Y = f(X) + e$  with function  $f$  is regression which not yet been known and  $e$  is independent random variable with mean 0 and variant  $\sigma^2$ , hence function of  $f$  can estimate with parametric and nonparametric approach. At this paper estimate  $f$  with nonparametric approach. Nonparametric approach that used is wavelet shrinkage or wavelet thresholding method. At function estimation with method of wavelet thresholding, what most dominant determine level of smoothing estimator is value of threshold. The small threshold give function estimation very no smoothly, while the big value of threshold give function estimation very smoothly. Therefore require to be selected value of optimal threshold to determine optimal function estimation.

One of the method to determine the value of optimal threshold is with procedure of False Discovery Rate (FDR). In procedure of FDR, the optimal threshold determined by selection of level of significance. Smaller mount used significance progressively smoothly its.

**Keywords:** Nonparametric regression, wavelet thresholding estimator, procedure of False Discovery Rate.

## 1. Pendahuluan

Model regresi non parametrik standar dari data observasi berpasangan  $\{(X_i, Y_i)\}_{i=1}^n$  adalah  $Y_i = f(X_i) + \varepsilon_i$  (1)

dengan  $X_i$  variabel prediktor,  $Y_i$  variabel respon,  $f$  fungsi regresi yang tidak diketahui, dan  $\varepsilon_i$  variabel random independen dengan mean 0 dan varian  $\sigma^2$ . Karena  $f$  fungsi regresi yang tidak diketahui maka  $f$  perlu diestimasi. Fungsi  $f$  dapat diestimasi, salah satunya dengan pendekatan non parametrik. Pendekatan non parametrik yang sudah populer adalah metode kernel dan metode Fourier. Pada pendekatan ini mengasumsikan bahwa fungsi  $f$  termuat dalam kelas fungsi mulus, artinya mempunyai turunan yang kontinu. Sehingga jika fungsinya tidak mulus metode ini kurang baik untuk digunakan. Pendekatan non parametrik yang dapat mengatasi kekurangan metode kernel dan deret Fourier adalah metode wavelet. Dalam metode wavelet diasumsikan fungsi yang akan diestimasi dapat diintegrasikan secara kuadrat. Jadi dengan metode wavelet fungsi yang akan diestimasi dapat berupa fungsi mulus maupun tidak mulus.

Estimator wavelet dari regresi non parametrik adalah pengembangan dari estimator regresi deret Fourier dan estimator kernel. Estimator wavelet sendiri dibedakan menjadi dua macam, yaitu estimator wavelet linier dan estimator wavelet non linier. Estimator wavelet non linier memiliki nilai Error Kuadrat Rata-rata Terintegrasi / Integrated Mean Square Error (IMSE) yang merupakan salah satu ukuran kebaikan dari estimator, lebih cepat menuju nol daripada IMSE wavelet linier<sup>[8]</sup>.

Prinsip dari estimator wavelet thresholding, mempertahankan koefisien wavelet yang nilainya lebih besar dari suatu nilai threshold tertentu dan mengabaikan koefisien wavelet yang kecil. Selanjutnya koefisien yang besar ini digunakan untuk merekonstruksi estimator fungsi yang dicari.

Pada estimasi fungsi dengan metode wavelet thresholding, tingkat kemulusan estimator ditentukan oleh pemilihan fungsi wavelet, level resolusi, fungsi thresholding, dan parameter threshold. Namun yang paling dominan menentukan tingkat kemulusan estimator adalah parameter threshold. Nilai threshold yang kecil memberikan estimasi fungsi yang sangat tidak mulus (under smooth), sedangkan nilai threshold yang besar memberikan estimasi yang sangat mulus (over smooth). Oleh karena itu perlu dipilih nilai threshold yang optimal.

Makalah ini membahas penentuan nilai threshold optimal pada estimator wavelet thresholding untuk fungsi regresi non parametrik dengan prosedur FDR.

## 2. Bahan dan Metode

Penelitian ini merupakan kajian literatur yang kemudian dikembangkan dengan simulasi menggunakan software S-Plus For Windows. Dalam tulisan ini dilakukan pembahasan tentang estimator fungsi regresi non parametrik dengan wavelet thresholding dan cara mendapatkan parameter threshold optimal dengan prosedur False Discovery Rate (FDR) serta contoh aplikasinya.

## 3. Hasil dan Pembahasan

### Estimator Deret Orthogonal

Diasumsikan bahwa  $f \in L^2(\mathbb{R})$  dengan  $L^2(\mathbb{R}) = \{f : \int_{-\infty}^{\infty} f(x)^2 dx < \infty\}$ . Didefinisikan sebuah hasil kali dalam pada ruang  $L^2(\mathbb{R})$  adalah fungsi yang mengasosiasikan bilangan riil  $\langle f, g \rangle$ , dengan masing-masing pasangan fungsi  $f(x)$  dan  $g(x)$  pada  $L^2(\mathbb{R})$ . Hasil kali dalam  $L^2(\mathbb{R})$  dari dua fungsi dan norma sebuah fungsi didefinisikan  $\langle f, g \rangle = \int_{-\infty}^{\infty} f(x)g(x)dx$  dan  $\|f\| = \sqrt{\langle f, f \rangle} = \sqrt{\int_{-\infty}^{\infty} (f(x))^2 dx}$ . Menurut Vetterli dan Kovacevic<sup>[9]</sup>,  $L^2(\mathbb{R})$  merupakan ruang Hilbert.

Jika  $\{\varphi_j\}_{j=1,2,\dots}$  sistem ortonormal lengkap (CONS) dari  $L^2(\mathbb{R})$ , maka sembarang  $f \in L^2(\mathbb{R})$  dapat dinyatakan sebagai  $f = \sum_{j=1}^{\infty} \alpha_j \varphi_j$  dengan  $\alpha_j = \langle f, \varphi_j \rangle$  dan memenuhi identitas Parseval  $\|f\|^2 = \sum_{j=1}^{\infty} \alpha_j^2$ . Karena  $\int_{-\infty}^{\infty} f(x)^2 dx < \infty$  maka  $\sum_{j=1}^n \alpha_j^2 < \infty$  sehingga  $\alpha_j \rightarrow 0$ , untuk  $j \rightarrow \infty$ . Oleh karena itu,  $f$  dapat didekati oleh  $f = \sum_{j=1}^J \alpha_j \varphi_j$  untuk bilangan bulat  $J$  cukup besar. Khususnya jika  $f \in L^2[0, 2\pi]$ , maka  $f$  dapat didekati dengan deret Fourier,

$$f_j(x) = \frac{1}{2} a_0 + \sum_{j=1}^J (a_j \cos(jx) + b_j \sin(jx)) \quad (2)$$

dengan koefisien Fourier  $a_j = \frac{1}{\pi} \langle f, \cos(j \cdot) \rangle = \frac{1}{\pi} \int_0^{2\pi} f(x) \cos(jx) dx$  dengan  $j = 0, 1, \dots, J$  dan  $b_j = \frac{1}{\pi} \langle f, \sin(j \cdot) \rangle = \frac{1}{\pi} \int_0^{2\pi} f(x) \sin(jx) dx$  dengan  $j=1, 2, \dots, J$ .

Jika  $\{(X_i, Y_i)\}_{i=1}^n$  merupakan data observasi independen mempunyai model (1) dengan  $X_i = 2\pi i/n$  dan  $X_i \in [0, 2\pi]$ , maka estimator deret Fourier dari regresi  $f$  adalah

$$\hat{f}_J(x) = \frac{1}{2} \hat{a}_0 + \sum_{j=1}^J (\hat{a}_j \cos(jx) + \hat{b}_j \sin(jx)) \quad (3)$$

dengan  $\hat{a}_j = \frac{1}{\pi} \langle f^{\sim}(X_i), \cos(j \cdot) \rangle, j = 0, 1, \dots, J$  dan  $\hat{b}_j = \frac{1}{\pi} \langle f^{\sim}(X_i), \sin(j \cdot) \rangle, j = 1, 2, 3, \dots, J$ .

Dalam hal ini  $\hat{a}_j$  dan  $\hat{b}_j$  merupakan estimator tak bias dari  $a_j$  dan  $b_j$ .

### Fungsi Wavelet

Fungsi wavelet adalah suatu fungsi matematika yang mempunyai sifat- sifat tertentu diantaranya beresilasi di sekitar nol (seperti fungsi sinus dan cosinus) dan terlokalisasi dalam domain waktu artinya pada saat nilai domain relatif besar, fungsi wavelet berharga nol. Fungsi wavelet dibedakan atas dua jenis, yaitu wavelet ayah ( $\phi$ ) dan wavelet ibu ( $\psi$ )

yang mempunyai sifat  $\int_{-\infty}^{\infty} \phi(x) dx = 1$  dan  $\int_{-\infty}^{\infty} \psi(x) dx = 0$ . Dengan dilatasi diadik dan translasi integer, wavelet ayah dan wavelet ibu melahirkan keluarga wavelet yaitu  $\phi_{j,k}(x) = (p2^j)^{1/2} \phi(p2^j x - k)$  dan  $\psi_{j,k}(x) = (p2^j)^{1/2} \psi(p2^j x - k)$  untuk suatu skalar  $p > 0$ , dan tanpa mengurangi keumuman dapat diambil  $p=1$ , sehingga  $\phi_{j,k}(x) = 2^{j/2} \phi(2^j x - k)$  dan  $\psi_{j,k}(x) = 2^{j/2} \psi(2^j x - k)$ . Fungsi  $\phi_{j,k}(x)$  dan  $\psi_{j,k}(x)$

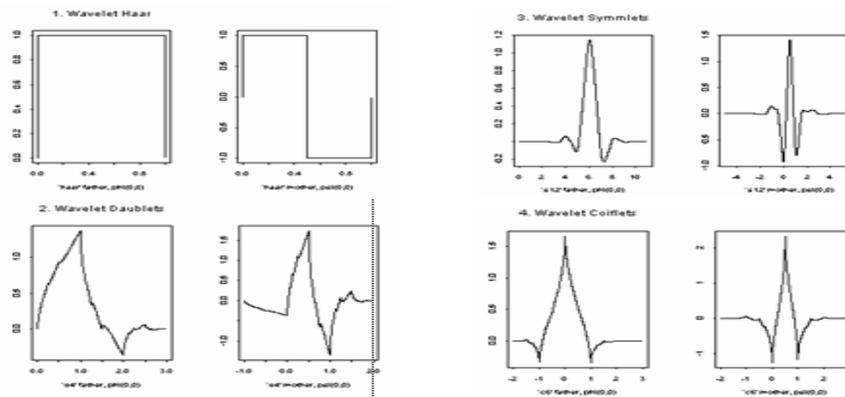
mempunyai sifat  $\int_{-\infty}^{\infty} \phi_{j,k}(x) \phi_{j,k'}(x) dx = \delta_{k,k'}$ ,  $\int_{-\infty}^{\infty} \psi_{j,k}(x) \phi_{j,k'}(x) dx = 0$ , dan

$$\int_{-\infty}^{\infty} \psi_{j,k}(x) \psi_{j,k'}(x) dx = \delta_{j,j'} \delta_{k,k'} \text{ dengan } \delta_{i,j} = \begin{cases} 1 & \text{jika } i = j \\ 0 & \text{jika } i \neq j. \end{cases}$$

Contoh wavelet paling sederhana adalah wavelet Haar yang mempunyai rumus

$$\psi(x) = \begin{cases} 1 & , 0 \leq x < 1/2 \\ -1 & , 1/2 \leq x < 1 \\ 0 & , x \text{ yang lain} \end{cases} \text{ dan } \phi(x) = \begin{cases} 1 & , 0 \leq x < 1 \\ 0 & , x \text{ yang lain.} \end{cases} \quad (4)$$

Beberapa contoh wavelet selain wavelet haar diantaranya adalah wavelet Daubechies (Daublet), symmetris (Symmlet), dan Coifman (Coiflet)<sup>[2]</sup>. Visualisasi beberapa wavelet dapat ditunjukkan pada gambar 1 berikut:



Gambar 1. Visualisasi beberapa Wavelet

### Analisis Multiresolusi

Analisis multiresolusi  $L^2(\mathbb{R})$  adalah ruang bagian tertutup  $\{V_j, j \in \mathbb{Z}\}$  yang memenuhi :

- i)  $\dots \subset V_{-2} \subset V_{-1} \subset V_0 \subset V_1 \subset V_2 \subset \dots$

- ii)  $\bigcap_{j \in \mathbb{Z}} V_j = \{0\}$ ,  $\bigcup_{j \in \mathbb{Z}} V_j = L^2(\mathbb{R})$
- iii)  $f \in V_j \Leftrightarrow f(2 \cdot) \in V_{j+1}$
- iv)  $f \in V_0 \Rightarrow f(-k) \in V_0, \forall k \in \mathbb{Z}$
- v) Terdapat sebuah fungsi  $\phi \in V_0$  sehingga  $\phi_{0,k} = \phi(\cdot - k), k \in \mathbb{Z}$  membentuk basis ortonormal untuk  $V_0$  dimana untuk semua  $j, k \in \mathbb{Z}$ ,  $\phi_{j,k}(x) = 2^{j/2} \phi(2^j x - k)$ .

Jika  $\{V_j, j \in \mathbb{Z}\}$  analisis multiresolusi dari  $L^2(\mathbb{R})$ , maka ada basis ortonormal  $\{\psi_{j,k}; j, k \in \mathbb{Z}\}$  untuk  $L^2(\mathbb{R})$ :  $\psi_{j,k} = 2^{j/2} \psi(2^j x - k)$ , sehingga untuk sembarang  $f \in L^2(\mathbb{R})$ ,  $P^j f = P^{j-1} f + \sum_{k \in \mathbb{Z}} \langle f, \psi_{j-1,k} \rangle \psi_{j-1,k}$ , yaitu  $\psi(x)$  yang diturunkan dari  $\psi(x) = \sum_{k \in \mathbb{Z}} (-1)^k c_{(-k+1)} \phi_{1,k}(x)$ .

**Akibat.** Bila  $\phi$  adalah fungsi skala yang membangun analisis multiresolusi dan  $\psi(x) = \sum_{k \in \mathbb{Z}} (-1)^k c_{(-k+1)} \phi_{1,k}(x)$  maka dekomposisi ke dalam wavelet ortonormal untuk sembarang  $f \in L^2(\mathbb{R})$  dapat dilakukan menjadi

$$f(x) = \sum_{k \in \mathbb{Z}} c_{j_0,k} \phi_{j_0,k}(x) + \sum_{j \geq j_0} \sum_{k \in \mathbb{Z}} d_{j,k} \psi_{j,k}(x) \tag{5}$$

dengan  $c_{j_0,k} = \langle f, \phi_{j_0,k} \rangle$  dan  $d_{j,k} = \langle f, \psi_{j,k} \rangle$ .

**Estimator Wavelet Linier**

Jika terdapat sekumpulan data berpasangan independen  $\{(X_i, Y_i)\}_{i=1}^n$  yang mempunyai model (1) dan  $n = 2^m$  dengan  $m$  bilangan bulat positif. Jika  $X_i$  rancangan titik reguler pada interval  $[0,1]$  dengan  $X_i = i/n$ , maka proyeksi  $f$  pada ruang  $V_j$  dapat ditulis menjadi  $(P^j f)(x) = \sum_{k \in \mathbb{Z}} c_{j,k} \phi_{j,k}(x)$  atau  $f_j(x) = \sum_{k \in \mathbb{Z}} c_{j,k} \phi_{j,k}(x)$  dengan

$c_{j,k} = \langle f, \phi_{j,k} \rangle = \int_0^1 f(x) \phi_{j,k}(x) dx$ . Berdasarkan dekomposisi fungsi ke dalam wavelet ortonormal (5) untuk sembarang fungsi  $f \in L^2(\mathbb{R})$  diperoleh

$$f_j(x) = \sum_{k \in \mathbb{Z}} c_{j_0,k} \phi_{j_0,k}(x) + \sum_{j \geq j_0} \sum_{k \in \mathbb{Z}} d_{j,k} \psi_{j,k}(x) \text{ dengan } c_{j_0,k} = \langle f, \phi_{j_0,k} \rangle = \int_0^1 f(x) \phi_{j_0,k}(x) dx \text{ dan}$$

$d_{j,k} = \langle f, \psi_{j,k} \rangle = \int_0^1 f(x) \psi_{j,k}(x) dx$ . Karena fungsi regresi  $f$  tidak diketahui maka estimator

$f$  pada ruang  $V_j$  dapat ditulis sebagai  $\hat{f}_j(x) = \sum_{k \in \mathbb{Z}} \hat{c}_{j,k} \phi_{j,k}(x)$  dengan  $\hat{c}_{j,k} = \frac{1}{n} \sum_{i=1}^n Y_i \phi_{j,k}(X_i)$ ,

atau

$$\hat{f}_j(x) = \sum_{k \in \mathbb{Z}} \hat{c}_{j_0,k} \phi_{j_0,k}(x) + \sum_{j \geq j_0} \sum_{k \in \mathbb{Z}} \hat{d}_{j,k} \psi_{j,k}(x) \tag{6}$$

dengan  $\hat{c}_{j_0,k} = \frac{1}{n} \sum_{i=1}^n Y_i \phi_{j_0,k}(X_i)$  dan  $\hat{d}_{j,k} = \frac{1}{n} \sum_{i=1}^n Y_i \psi_{j,k}(X_i)$ , yang merupakan estimator tak bias dari  $c_{j_0,k}$  dan  $d_{j,k}$ . Estimator wavelet (6) dinamakan estimator wavelet linier.

### Estimator Wavelet Shrinkage

Jika diberikan data  $\{(X_i, Y_i)\}_{i=1}^n$  dengan model (1),  $n = 2^m$  dan  $X_i = i/n$ , maka  $Y_i \sim N(g(i/n), \sigma^2)$ . Mean dan varian dari  $\hat{d}_{j,k}$  adalah  $E[\hat{d}_{j,k}] = d_{j,k}$  dan  $\text{var}(\hat{d}_{j,k}) = \frac{\sigma^2}{n}$  sehingga  $\hat{d}_{j,k} \sim N(d_{j,k}, \frac{\sigma^2}{n})$ . Jadi koefisien wavelet empiris  $\hat{d}_{j,k}$  memuat sejumlah noise dan hanya

relatif sedikit yang memuat sinyal signifikan sehingga dapat direkonstruksi estimator wavelet dengan menggunakan sejumlah koefisien terbesar. Oleh karena itu Hall&Patil<sup>[4]</sup> dan Ogden<sup>[7]</sup> memberikan metode yang menekankan rekonstruksi wavelet dengan menggunakan sejumlah koefisien wavelet terbesar, yakni hanya koefisien yang lebih besar dari suatu nilai tertentu yang diambil, sedangkan koefisien selebihnya diabaikan, karena dianggap 0. Nilai tertentu tersebut dinamakan nilai threshold (nilai ambang) dan estimatornya menghasilkan

$$\hat{f}_\lambda(x) = \sum_k \hat{c}_{j_0,k} \phi_{j_0,k}(x) + \sum_{j \geq j_0} \sum_{k=0}^{2^j-1} \partial_\lambda(\hat{d}_{j,k}) \psi_{j,k}(x) \quad (7)$$

dengan  $\partial_\lambda$  menyatakan fungsi thresholding atau fungsi ambang dengan nilai ambang atau threshold  $\lambda$ . Estimator (7) dinamakan estimator wavelet non linier, estimator wavelet shrinkage, atau estimator wavelet thresholding.

Prinsip dari estimator wavelet thresholding adalah mempertahankan koefisien wavelet yang nilainya lebih besar dari suatu nilai ambang atau nilai threshold tertentu dan mengabaikan koefisien wavelet yang kecil. Selanjutnya koefisien yang besar ini digunakan untuk merekonstruksi fungsi (estimator) yang dicari. Karena thresholding ini dirancang untuk membedakan antara koefisien wavelet empiris yang masuk dan yang keluar dari rekonstruksi wavelet, sedangkan untuk membuat keputusan ada 2 faktor yang mempengaruhi ketepatan estimator, yaitu ukuran sampel  $n$  dan tingkat noise  $\sigma^2$ , maka setiap koefisien merupakan calon kuat masuk didalam rekonstruksi wavelet jika ukuran sampel besar atau tingkat noise kecil. Karena  $\sqrt{n} \hat{d}_{j,k} / \sigma$  berdistribusi normal dengan varian 1 untuk seluruh  $n$  dan  $\sigma$ , maka estimator thresholding dari  $d_{j,k}$  adalah

$$\tilde{d}_{j,k} = \frac{\sigma}{\sqrt{n}} \partial_\lambda \left( \frac{\sqrt{n} \hat{d}_{j,k}}{\sigma} \right) \text{ sehingga estimator wavelet thresholding adalah}$$

$$\hat{f}_\lambda(x) = \sum_k \hat{c}_{j_0,k} \phi_{j_0,k}(x) + \sum_{j \geq j_0} \sum_{k=0}^{2^j-1} \frac{\sigma}{\sqrt{n}} \partial_\lambda \left( \frac{\sqrt{n} \hat{d}_{j,k}}{\sigma} \right) \psi_{j,k}(x) \quad (8)$$

dengan  $\hat{c}_{j_0,k}$  : penduga koefisien fungsi skala,

$\hat{d}_{j,k}$  : penduga koefisien wavelet  $d_{j,k}$

$\lambda$  : parameter nilai threshold

$\partial_\lambda$  : fungsi threshold

### Langkah-langkah Thresholding

Langkah-langkah thresholding terdiri dari :

#### 1. Pemilihan Fungsi Thresholding

Ada dua jenis fungsi thresholding  $\partial_\lambda$ , yaitu Hard Thresholding,  $\partial_\lambda^H(x) = \begin{cases} x, & |x| > \lambda \\ 0, & x \text{ yang lain} \end{cases}$  dan Soft Thresholding,  $\partial_\lambda^S(x) = \begin{cases} x - \lambda, & x > \lambda \\ 0, & x \leq \lambda \\ x + \lambda, & x < -\lambda \end{cases}$  dengan  $\lambda$

merupakan parameter thresholding.

Fungsi Hard thresholding lebih dikenal karena terdapat diskontinyu dalam fungsi thresholding sehingga nilai  $x$  yang berada diatas threshold  $\lambda$  tidak disentuh. Sebaliknya, fungsi soft thresholding kontinyu yaitu sejak nilai  $x$  berada diatas threshold  $\lambda$ . Motivasi penggunaan soft thresholding berasal dari prinsip bahwa noise mempengaruhi seluruh koefisien wavelet. Juga kekontinyuan dari fungsi soft shrinkage membuat kondisi yang lebih baik untuk alasan statistik.

## 2. Estimasi $\sigma$

Dalam merekonstruksi fungsi wavelet biasanya nilai  $\sigma$  tidak diketahui. Oleh karena itu,  $\sigma$  harus diestimasi dari data. Ogden<sup>[7]</sup> memberikan estimasi  $\sigma$  berdasarkan koefisien wavelet empiris pada level resolusi tertinggi dengan fungsi Median Deviasi Absolut (MAD), yaitu:

$$\hat{\sigma} = \frac{\text{median}\left(\left|\hat{d}_{J-1,k} - \text{median}\left(\hat{d}_{J-1,k}\right)\right|\right)}{0,6745}.$$

## 3. Pemilihan Parameter Thresholding

Pada estimasi wavelet thresholding, tingkat kemulusan estimator ditentukan oleh level resolusi  $J$ , fungsi thresholding  $\partial_\lambda$  dan parameter threshold  $\lambda$ . Namun pemilihan  $J$  dan  $\partial_\lambda$  tidak sedominan  $\lambda$ . Nilai  $\lambda$  yang terlalu kecil memberikan estimasi fungsi yang sangat tidak mulus (under smooth) sedangkan nilai  $\lambda$  yang terlalu besar memberikan estimasi yang sangat mulus (over smooth). Oleh karena itu perlu dipilih parameter threshold yang optimal untuk mendapatkan fungsi yang optimal. Untuk memilih nilai threshold optimal, ada dua kategori pemilihan yaitu memilih satu harga threshold untuk seluruh level resolusi (pemilihan secara global) dan pemilihan threshold yang tergantung pada level resolusi (dependent level thresholding).

Untuk pemilihan threshold global, Ogden<sup>[7]</sup> memberikan 2 pemilihan threshold yang hanya bergantung pada banyaknya data pengamatan  $n$  yaitu threshold universal ( $\lambda_j = \sqrt{2 \log n}$ ) dan threshold minimax yang telah ditabelkan oleh Donoho dan Johnstone<sup>[3]</sup>. Nilai-nilai threshold minimax selalu lebih kecil dibandingkan dengan nilai threshold universal untuk ukuran sampel yang sama. Pemilihan threshold yang tergantung pada level resolusi berarti memilih  $\lambda_j$  bergantung level resolusi  $j$ . Dengan demikian ada kemungkinan perbedaan nilai threshold  $\lambda_j$  yang dipilih untuk tiap level wavelet  $j$ . Ada beberapa cara level-dependent thresholding diantaranya yaitu threshold Adapt dan threshold Top. Threshold adapt didasarkan pada prinsip untuk meminimalkan *Stein Unbiased Risk Estimator* (SURE) pada suatu level resolusi. Threshold adapt untuk himpunan koefisien detail  $d_j$  yang beranggotakan  $K$  koefisien didefinisikan sebagai,  $\lambda_j = \arg \min_{t \geq 0} \text{SURE}(d_j, t)$ , dengan

$$\text{SURE}(d_j, t) = K - 2 \sum_{k=1}^K 1_{\{|d_{j,k}| \leq t\sigma_j\}} + \sum_{k=1}^K \min\left\{\left(d_{j,k} / \sigma_j\right)^2, t^2\right\}.$$

Sedangkan nilai threshold top ditentukan berdasarkan besar prosentase koefisien yang akan digunakan dari keseluruhan koefisien wavelet yang ada.

Pemilihan parameter threshold optimal dengan universal, minimax, adapt, dan top merupakan pemilihan parameter threshold optimal standar dalam estimasi dengan wavelet thresholding. Pemilihan dengan cara tersebut telah tersedia dalam software S+Wavelet. Selain cara tersebut ada beberapa cara atau prosedur lain untuk mendapatkan threshold optimal dalam estimasi fungsi wavelet thresholding yaitu dengan prosedur False Discovery rate (FDR). Pemilihan threshold optimal dengan prosedur FDR merupakan pengembangan dari prosedur uji hipotesis multipel.

a. Penentuan parameter threshold optimal dengan prosedur uji hipotesis multipel.

Penentuan parameter threshold optimal dengan prosedur uji hipotesis multipel pada hakekatnya mengacu pada prosedur pengujian suatu hipotesis tunggal. Neter, *et al.* [6] menggunakan uji hipotesis untuk mengetahui signifikan atau tidaknya setiap koefisien regresi untuk memutuskan apakah koefisien-koefisien dalam regresi signifikan atau tidak. Dengan cara seperti ini, Abramovich dan Benjamini [1] menguji semua koefisien wavelet. Jika dalam uji hipotesis menyimpulkan bahwa koefisien wavelet signifikan maka koefisien ini akan dipertahankan dalam merekonstruksi fungsi tetapi kalau dalam uji hipotesis koefisien wavelet memutuskan sama dengan nol maka koefisien ini akan diabaikan. Misalkan ada sejumlah  $n$  koefisien wavelet, maka untuk setiap koefisien wavelet diuji dengan uji hipotesis :

$$H_0 : d_{jk} = 0 \text{ ( koefisien wavelet tidak signifikan )},$$

$$H_1 : d_{jk} \neq 0 \text{ ( koefisien wavelet signifikan)}$$

Karena  $\hat{d}_{j,k} \sim N(d_{j,k}, \frac{\sigma^2}{n})$  maka statistik uji yang digunakan

$$Z = \frac{\hat{d}_{j,k} - d_{j,k}}{\sigma/\sqrt{n}}. \text{ Dibawah } H_0 \text{ benar maka statistik uji } Z \sim \frac{\hat{d}_{j,k}}{\sigma/\sqrt{n}} N(0,1) \text{ sehingga } H_0 \text{ ditolak}$$

jika  $|Z| > Z_{\alpha/2}$  dengan  $Z_{\alpha/2}$  adalah quantil normal standar. Jika  $H_0$  ditolak maka  $\hat{d}_{jk}$  masuk dalam rekonstruksi estimator regresi wavelet thresholding, tetapi jika  $H_0$  diterima maka  $\hat{d}_{jk}$  dihilangkan. Threshold optimal dengan pendekatan uji hipotesis tunggal adalah  $\hat{d}_{jk}$  terkecil yang memenuhi  $|Z| > Z_{\alpha/2}$ . Tingkat kemulusan estimator wavelet yang diperoleh dengan prosedur ini dipengaruhi oleh besar kecilnya tingkat signifikansi  $\alpha$  yang diambil. Semakin kecil  $\alpha$  maka harga  $Z_{\alpha/2}$  akan semakin besar sehingga nilai threshold optimal juga semakin besar sehingga estimator yang diperoleh semakin mulus. Menurut Ogden [7], dalam pendekatan uji hipotesis tunggal koefisien yang masuk dalam rekonstruksi terlalu banyak, sehingga estimasi fungsi yang dihasilkan cenderung kurang mulus sehingga uji hipotesis tunggal diperbaiki dengan uji rekursif dan prosedur FDR.

b. Pemilihan threshold dengan prosedur FDR( False Discovery Rate)

FDR (tingkat kesalahan yang ditemukan) adalah proporsi koefisien wavelet yang tidak signifikan masuk dalam rekonstruksi terhadap banyaknya semua koefisien yang dimasukkan dalam rekonstruksi. Prosedur ini mengontrol agar ekspektasi dari tingkat kesalahan ini sekecil mungkin..

Misalkan  $R$  menyatakan banyaknya koefisien wavelet yang tidak diabaikan dalam prosedur estimasi wavelet thresholding. Berarti rekonstruksi fungsi menggunakan  $R$  koefisien wavelet. Dari  $R$  ini ada sebanyak  $S$  koefisien dalam rekonstruksi yang benar dan sebanyak  $V$  koefisien yang salah tetapi masuk dalam rekonstruksi.  $Q=V/R$  menyatakan proporsi banyaknya koefisien wavelet yang seharusnya dikeluarkan dari

rekonstruksi. FDR dari koefisien wavelet didefinisikan sebagai ekspektasi dari Q. Tujuan prosedur ini memasukkan koefisien wavelet sebanyak mungkin sehingga ekspektasi dari nilai Q di bawah nilai yang ditentukan (tingkat signifikansi).

Prosedur FDR untuk menentukan threshold optimal sbb :

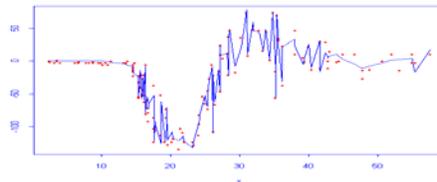
1. Untuk setiap koefisien empiris  $\hat{d}_{jk}$ , dihitung p-value dua sisi ( $p_{jk}$ ) untuk n hipotesis  $H_{jk} : d_{jk} = 0$ , yaitu  $p_{jk} = 2(1 - \phi(|\hat{d}_{jk}|/\sigma))$  dengan  $\phi$  suatu nilai yang berkenaan dengan distribusi dari  $\hat{d}_{jk}$ .
2.  $p_{jk}$  diurutkan yaitu  $p_{(1)} \leq p_{(2)} \leq \dots \leq p_{(m)}$  dengan  $p_{(i)}$  bersesuaian dengan beberapa  $d_{jk}$ .
3. Mulai dengan  $i=1$ , dan misalkan  $k =$  bilangan terbesar  $i$  yang memenuhi  $p_{(i)} \leq (i/m)\alpha$  dengan  $\alpha$  merupakan tingkat signifikansi biasanya  $\alpha = 0,01$  atau  $0,05$  maka  $\lambda$  optimal adalah  $\lambda_k = \sigma \phi^{-1}(1 - p_{(k)}/2)$ .

**Penerapan pada Data Simulasi Tubrukan**

Data diambil dari buku *Applied Nonparametric Regresión* [5], yakni data simulasi tubrukan sepeda motor pada suatu PTMO (*post mortem human test object* / obyek uji pemeriksaan mayat manusia). Dalam hal ini variabel-variabelnya adalah sebagai berikut:

- Sebagai variabel respon, Y (percepatan dalam g) menyatakan percepatan setelah tubrukan yang disimulasikan.
- Sebagai variabel prediktor, X (waktu dalam milisekon) menyatakan waktu setelah simulasi tubrukan.

Dari data tersebut, dicari estimasi kurva yang optimal menggunakan prosedur uji hipotesis dan prosedur FDR dengan  $\alpha = 0,01$ . Dengan menggunakan pendekatan uji hipotesis tunggal diperoleh  $\lambda$  optimal = 2.575 dan dengan prosedur FDR  $\lambda$  optimal = 25.014 Visualisasi dari estimasi fungsi optimalnya sebagai berikut.

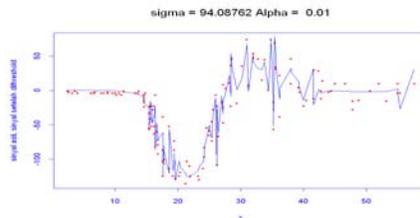


**Gambar 2.** Kurva Estimasi dengan Pendekatan Uji Hipotesis Tunggal

Keterangan :

..... : Sebaran data

———— : Estimasi dengan uji hipotesis tunggal



**Gambar 3.** Kurva Estimasi dengan prosedur FDR

Keterangan :

..... : Sebaran data

———— : Estimasi dengan prosedur FDR

#### 4. Penutup

Estimasi fungsi regresi dengan wavelet merupakan estimasi fungsi dengan pendekatan non parametrik yang tidak mengasumsikan bahwa fungsi yang akan diestimasi harus mulus seperti asumsi pada metode kernel dan deret Fourier. Dalam estimasi fungsi wavelet, tingkat kemulusan yang dominan ditentukan oleh parameter threshold. Semakin besar nilai threshold, akan semakin mulus hasil estimasi fungsinya. Parameter threshold optimal yang diperoleh dengan prosedur FDR dipengaruhi oleh besarnya tingkat signifikansi yang digunakan. Semakin kecil tingkat signifikansi yang digunakan semakin besar threshold optimalnya sehingga semakin mulus fungsi estimasinya.

#### DAFTAR PUSTAKA

1. Abramovich, F. and Benjamini, Y., *Thresholding of Wavelet Coefficients as Multiple Hypothesis Testing Procedure*. In *Wavelets and Statistics*. Antoniadis, A., and Oppenheim, G. (eds.). Springer -Verlag, New York, 1995: 5-14.
2. Daubechies, I. , *Ten Lectures on Wavelets*, Capital City Press, Philadelphia, 1992.
3. Donoho, D.L and Johnstone, I.M., Ideal Spatial Adaptation by Wavelet Shrinkage, *Biometrika*, 1994, Vol. 81, No. 3: 425-455.
4. Hall, P. and Patil, P., On Wavelet Methods for Estimating Smooth Functions, *Bernoulli*, 1995, Vol. 1, No. 2: 041-058.
5. Hardle, W. , *Applied Nonparametric Regression*, Cambridge University Press, New York, 1993.
6. Neter, J., Wasserman, W. and Kutner, M.H. *Applied Linear Statistical Model*, Irwin Inc., Boston, 1990.
7. Ogden, R.T., *Essential Wavelets for Statistical Applications and Data Analysis*, Birkhauser, Boston, 1997.
8. Suparti dan Subanar, H., Estimasi Regresi dengan Metode Wavelet Shrinkage, *Jurnal Sains & Matematika*, 2000, Vol. 8, No. 3: 105-113.
9. Vetterli, M. and Kovacevic, J., *Wavelets And Subband Coding*, Prentice Hall PTR, New Jersey, 1995.