Pemilihan Threshold Optimal pada Estimator Regresi Wavelet thresholding dengan Prosedur Uji Hipotesis Multipel

Suparti[1] , Sidharta Wahyu Putra[2] dan Rukun Santoso[3]
[1,3] Staf Jurusan Matematika FMIPA UNDIP
[2] Alumni Jurusan Matematika FMIPA UNDIP

ABSTRAK-- Misalkan $\{(X_i, Y_i)\}_{i=1}^{n}$ data pengamatan independen yang mengikuti model $Y_i = f(X_i)+\epsilon_i$, i = 1,2,…..,n dengan f fungsi regresi yang belum diketahui dan ε_i variabel random independen dengan mean 0 dan varian σ^2. Fungsi f dapat diestimasi dengan pendekatan parametrik dan non parametrik. Pada makalah ini dilakukan estimasi f dengan pendekatan non parametrik. Pendekatan non parametrik yang digunakan adalah dengan metode wavelet shrinkage atau metode wavelet thresholding. Pada estimasi fungsi dengan metode wavelet thersholding, yang paling dominan menentukan tingkat kemulusan estimator adalah nilai threshold. Nilai threshold yang kecil memberikan estimasi fungsi yang sangat tidak mulus, sedangkan nilai threshold yang besar memberikan estimasi fungsi yang sangat mulus. Oleh karena itu perlu dipilih nilai threshold optimal untuk menentukan estimasi fungsi yang optimal. Salah satu cara untuk menentukan nilai threshold optimal dengan uji hipotesis multipel. Pada uji hipotesis multipel, besarnya nilai threshold dipengaruhi oleh nilai signifikansi alpha. Nilai alpha yang kecil menghasilkan nilai threshold optimal yang besar, sehingga dihasilkan fungsi yang cenderung lebih mulus dan sebaliknya.

Kata kunci: estimator wavelet thresholding, uji hipotesis multipel

PENDAHULUAN

Bentuk model regresi non parametrik standar dari data observasi $\{(X_i, Y_i)\}_{i=1}^{n}$ adalah :

$$Y_i = f(X_i) + \varepsilon_i \quad (1)$$

dengan X_i variabel prediktor, Y_i variabel respon, f fungsi regresi yang tidak diketahui, dan ε_i variabel random independen dengan mean 0 dan varian σ^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 dalah metode wavelet. Dalam metode wavelet diasumsikan fungsi yang akan diestimasi dapat diintegralkan 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. Menurut Suparti dan Subanar[9] estimator wavelet nonlinier 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. Selain itu bila dibandingkan dengan metode Fourier, estimator wavelet non linier mampu mengestimasi fungsi baik mulus maupun tidak mulus.

Estimator wavelet non linier disebut juga dengan estimator wavelet shrinkage atau estimator wavelet thresholding. Prinsip dari estimator wavelet thresholding, mempertahankan koefisien wavelet yang nilainya lebih besar dari suatu nilai threshold tertentu dan mengabaikan koefisien wavelet yang kecil. Selanjutnya menurut Ogden[8] koefisien yang besar ini digunakam untuk merekonstruksi estimator fungsiyang 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.

Tulisan ini membahas penentuan nilai threshold optimal pada estimator wavelet thresholding untuk fungsi regresi non parametrik dengan prosedur uji hipotesis multipel, berikut sifat-sifat dan contoh simulasinya dengan menggunakan program S+Wavelets for Windows.

Estimator Deret Fourier. Diasumsikan bahwa f$\in$ $L^2(R)$ dengan $L^2(R)$ $=\left\{f : \int_{-\infty}^{\infty} f(x)^2\,dx < \infty\right\}$. Didefinisikan sebuah hasil kali dalam pada ruang $L^2(R)$ adalah fungsi yang mengasosiasikan bilangan riil $\langle f, g\rangle$, dengan masing-masing pasangan fungsi f(x) dan g(x) pada $L^2(R)$. Hasil kali dalam $L^2(R)$ 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[10], $L^2(R)$ merupakan ruang Hilbert.

Andaikan $\{\varphi_j\}_{j=1,2,...}$ sistem ortonormal lengkap (CONS) dari $L^2(R)$, maka sembarang f $\in$ $L^2(R)$ 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 $\alpha_j \to 0$, untuk $j \to \infty$. 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,

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

dengan koefisien Fourier

$$a_j = \frac{1}{\pi}\langle f, \cos(j.)\rangle = \frac{1}{\pi}\int_{-\pi}^{\pi} f(x)\cos(jx)dx \quad ,$$
j=0,1,…,J dan

$$b_j = \frac{1}{\pi}\langle f, \sin(j.)\rangle = \frac{1}{\pi}\int_{-\pi}^{\pi} f(x)\sin(jx)dx \quad ,$$
j=1,2,…,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 g adalah

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

dengan

$\hat{a}_j = \frac{1}{\pi}\left\langle f^{\sim}(X_i), \cos(j.)\right\rangle$, dengan j = 0,1,…,J dan

$\hat{b}_j = \frac{1}{\pi}\left\langle f^{\sim}(X_i), \sin(j.)\right\rangle$, dengan j=1,2,3,…,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 berosilasi 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 (ϕ) dan wavelet ibu (ψ) yang mempunyai sifat:

$$\int_{-\infty}^{\infty} \phi(x)dx = 1 \quad \text{dan} \quad \int_{-\infty}^{\infty} \psi(x)dx = 0 .$$

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

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

dengan $\delta_{i,j} = \begin{cases} 1 & jika\ i = j \\ 0 & jika\ i \neq j. \end{cases}$

Contoh wavelet paling sederhana adalah wavelet Haar yang mempunyai rumus

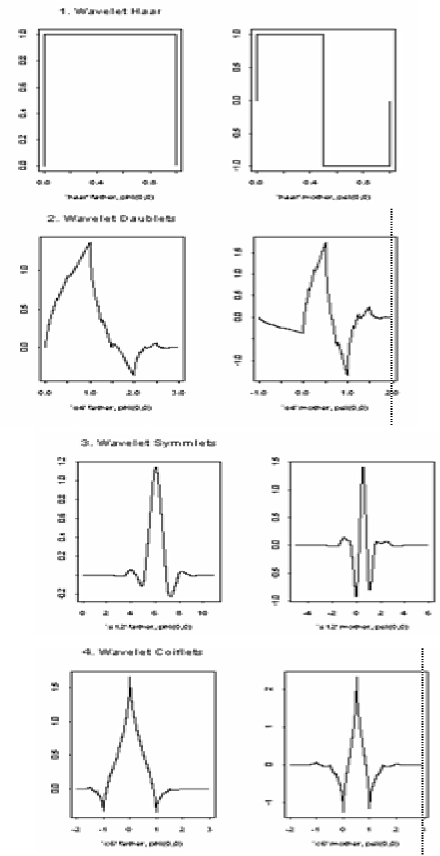
$$\psi(x) = \begin{cases} 1 & ,0 \le x < 1/2 \\ -1 & ,1/2 \le x < 1 \\ 0 & ,x \quad \text{yang} \quad \text{lain} \end{cases}$$

dan

$$\phi(x) = \begin{cases} 1 & ,0 \le x < 1 \\ 0 & ,x \quad \text{yang} \quad \text{lain.} \end{cases}$$

(3)

Beberapa contoh wavelet selain wavelet haar diantaranya adalah wavelet Daubechies (Daublet), symmetris (Symmlet), dan Coifman (Coiflet). Visualisasi beberapa wavelet dapat ditunjukkan pada gambar 3 berikut:



Gambar 3. Visualisasi beberapa Wavelet

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

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

ii) $\cap_{j\in Z} V_j = \{0\}$, $\overline{\cup}_{j\in Z} V_j = L^2(R)$

iii) $f \in V_j \Leftrightarrow f(2.) \in V_{j+1}$

iv) $f \in V_0 \Rightarrow f(.-k) \in V_0, \forall k \in Z$

v) Terdapat sebuah fungsi $\phi \in V_0$ sehingga $\phi_{0,k} = \phi(.-k), k \in Z$ membentuk basis ortonormal untuk V_0 dimana untuk semua $j,k \in Z$, $\phi_{j,k}(x) = 2^{j/2}\phi(2^j x - k)$.

Jika $\{V_j, j \in Z\}$ analisis multiresolusi dari $L^2(R)$, maka ada basis ortonormal $\{\psi_{j,k}; j,k \in Z\}$ untuk

$L^2(R)$: $\psi_{j,k} = 2^{j/2}\psi(2^j x - k)$,

sehingga untuk semua f pada $L^2(R)$,

$$P^j f = P^{j-1} f + \sum_{k\in Z} < f, \psi_{j-1,k} > \psi_{j-1,k},$$

yaitu $\psi(x)$ yang diturunkan dari

$$\psi(x) = \sum_{k\in Z} (-1)^k c_{(-k+1)} \phi_{1,k}(x).$$

Akibat. Bila ϕ adalah fungsi skala yang membangun analisis multiresolusi dan

$$\psi(x) = \sum_{k\in Z} (-1)^k c_{(-k+1)} \phi_{1,k}(x)$$

maka dekomposisi ke dalam wavelet ortonormal untuk sembarang $f \in L^2(R)$ dapat dilakukan menjadi

$$f(x) = \sum_{k\in Z} c_{jo,k}\phi_{jo,k}(x) + \sum_{j\ge jo}\sum_{k\in Z} d_{j,k}\psi_{j,k}(x)$$

(4)

dengan $c_{jo,k} =< f, \phi_{jo,k} >$ dan $d_{j,k} =< f, \psi_{j,k} >$.

Estimator Wavelet Linier. Jika terdapat sekumpulan data independen

$$\{((Xi, Yi))\}_{i=1}^{n}$$

yang mempunyai model (1) dan $n = 2^m$ dengan m bilangan bulat positip. 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

$$(P^J f)(x) = \sum_{k\in Z} c_{J,k}\phi_{J,k}(x)$$

$$f_J(x) = \sum_{k\in Z} c_{J,k}\phi_{J,k}(x)$$

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 (4) untuk sembarang fungsi $f \in L^2(R)$ diperoleh

$$f_J(x) = \sum_{k\in Z} c_{jo,k}\phi_{jo,k}(x) + \sum_{j\ge jo}^{J-1}\sum_{k\in Z} d_{j,k}\psi_{j,k}(x)$$

dengan

$$c_{jo,k} =< f, \phi_{jo,k} > = \int_0^1 f(x)\phi_{jo,k}(x)dx \text{ dan}$$

$$d_{j,k} =< f, \psi_{j,k} > = \int_0^1 f(x)\psi_{j,k}(x)dx.$$

Karena fungsi regresi f tidak diketahui maka estimator f pada ruang V_J dapat ditulis sebagai

$$\hat{f}_J(x) = \sum_{k\in Z} \hat{c}_{J,k}\phi_{J,k}(x)$$

dengan $\hat{c}_{J,k} = \frac{1}{n}\sum_{i=1}^{n} Y_i \phi_{J,k}(X_i)$, atau

$$\hat{f}_J(x) = \sum_{k\in Z} \hat{c}_{jo,k}\phi_{jo,k}(x) + \sum_{j\geq jo}^{J-1}\sum_{k\in Z} \hat{d}_{j,k}\psi_{j,k}(x)$$

(5)

dengan $\hat{c}_{jo,k} = \frac{1}{n}\sum_{i=1}^{n} Y_i \phi_{jo,k}(X_i)$ 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 $c_{jo,k}$ dan $d_{j,k}$. Estimator wavelet (5) 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 $Y_i \sim N\left(g\left(i/n\right), \sigma^2\right)$. Mean dan varian dari $\hat{d}_{j,k}$ adalah $E\left[\hat{d}_{j,k}\right] = d_{j,k}$ dan $\mathrm{Var}\left(\hat{d}_{j,k}\right) = \frac{\sigma^2}{n}$. Jadi $\hat{d}_{j,k} \sim N(d_{j,k}, \frac{\sigma^2}{n})$. Jadi koefisien wavelet empiris $\hat{d}_{j,k}$ memuat sejumlah noise dan hanya relatif sedikit yang memuat sinyal signifikan. Karena itu, dapat direkonstruksi wavelet dengan menggunakan sejumlah koefisien terbesar. Oleh karena itu Hall&Patil[4] dan Ogden[8] 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

$$\hat{f}_\lambda(x) = \sum_k \hat{c}_{jo,k}\phi_{jo,k}(x) + \sum_{j\geq jo}^{J-1}\sum_{k=o}^{2^j-1} \partial_\lambda\left(\hat{d}_{j,k}\right)\psi_{j,k}(x)$$

(6)

dengan ∂_λ menyatakan fungsi thresholding atau fungsi ambang dengan nilai ambang atau threshold λ. Estimator (6) 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 σ^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 σ, maka estimator thresholding dari $d_{j,k}$ adalah

$$\tilde{d}_{j,k} = \frac{\sigma}{\sqrt{n}}\partial_\lambda\left(\frac{\sqrt{n}\,\hat{d}_{j,k}}{\sigma}\right)$$

sehingga estimator wavelet thresholding adalah

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

(7)

dengan

$\hat{c}_{jo,k}$: penduga koefisien fungsi skala $c_{jo,k}$

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

λ : parameter nilai threshold

∂_λ : fungsi threshold

Langkah-langkah Thresholding

Langkah-langkah thresholding terdiri dari :

1. **Pemilihan Fungsi Thresholding**

   Ada dua jenis fungsi thresholding ∂_λ, yaitu:

   Hard Thresholding,

$$\partial_\lambda^{H}(x) = \begin{cases} x, & |x| > \lambda \\ 0, & x \text{ yang lain} \end{cases}$$

   Soft Thresholding,

$$\partial_\lambda^{S}(x) = \begin{cases} x - \lambda, & x > \lambda \\ 0, & x \leq \lambda \\ x + \lambda, & x < -\lambda \end{cases}$$

   dengan λ merupakan parameter thresholding.

   Fungsi Hard thresholding lebih dikenal karena terdapat diskontinyu dalam fungsi thresholding sehingga nilai x yang berada diatas threshold λ tidak disentuh. Sebaliknya, fungsi soft thresholding kontinyu yaitu sejak nilai x berada diatas threshold λ. 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 σ**

Dalam merekontruksi fungsi wavelet biasanya nilai σ tidak diketahui. Oleh karena itu, σ harus diestimasi dari data. Ogden (1997) memberikan estimasi σ 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 fungsi dengan metode wavelet thresholding, tingkat kemulusan estimator ditentukan oleh level resolusi J, fungsi thresholding ∂_λ dan parameter threshold λ. Namun pemilihan J dan ∂_λ tidak sedominan λ. Nilai λ yang terlalu kecil memberikan estimasi fungsi yang sangat tidak mulus (under smooth) sedangkan nilai λ 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.

Untuk pemilihan global threshold, Ogden (1997) 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 (1994). Nilai-nilai threshold minimax selalu lebih kecil dibandingkan dengan nilai threshold universal untuk ukuran sampel yang sama.

Level-dependent thresholding berarti memilih λ_j bergantung level resolui j. Dengan demikian ada kemungkinan perbedaan nilai threshold λ_j yang dipilh 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

$$\lambda_j = \arg\min_{t\geq 0} \text{SURE}\left(d_j, t\right)$$

dengan

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

Sedangkan nilai threshold Top ditentukan berdasarkan besar prosentase koefisien yang akan digunakan dari keseluruhan koefisien wavelet dalam merekonstruksi fungsi.

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 uji hipotesis multiple.

i. Pendekatan Uji Hipotesis Tunggal

Seperti halnya pada regresi linier, uji hipotesis koefisien regresi digunakan untuk mengetahui apakah variabel-varabel prediktor dalam regresi linier berpengaruh atau tidak. Dengan cara seperti ini, Abramovich dan Benjamini[1] menguji semua koefisien wavelet. Misalkan ada sejumlah n koefisien wavelet, maka untuk setiap koefisien wavelet diuji dengan uji hipótesis :

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

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

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

Dibawah H_0 benar maka statistik uji Z ~ N(0,1) sehingga H_0 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 uji hipotesa 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 α yang diambil. Semakin kecil α maka harga $Z_{\alpha/2}$ akan semakin besar sehingga nilai threshold optimal juga semakin besat sehingga estimator yang diperoleh semakin mulus. Menurut Ogden[10] dalam pendekatan uji hipotesis tunggal koefisien yang masuk dalam rekonstruksi terlalu banyak, sehingga estimasi fungsi yang dihasilkan cenderung kurang mulus sehingga uji hipotesa tunggal diperbaiki dengan uji rekursif.

ii. Uji Rekursif

Misalkan $X_1, X_2, \ldots, X_d$ adalah koefisien wavelet empiris pada level j = ${}^2\log d$, dengan $X_k \sim N(\mu_k, 1)$. Masing-masing koefisien memiliki rata-rata $\mu_1, \mu_2, \ldots, \mu_d$. Misalkan I_d adalah himpunan bagian tak kosong dari indeks $\{1, 2, \ldots, d\}$, akan diuji hipotesis sebagai berikut :

$H_0 : \mu_1 = \ldots = \mu_d = 0$

$H_1 : \mu_i \neq 0 \ , \forall i \in I_d \ ; \mu_i = 0 \ , \forall i \notin I_d$

Jika banyaknya anggota dari I_d tidak diketahui, maka untuk menguji hipotesis didasarkan pada statistik uji $\sum_{i=1}^{d} X_i^2$, yang memiliki distribusi χ^2 dengan derajat bebas d ketika H_0 benar. Statistik uji ini juga bisa digunakan jika $I_d = \{1, 2, \ldots, d\}$. Tetapi statistik uji bukanlah satatistik uji yang paling tepat untuk situasi ini karena diyakini bahwa banyaknya μ_i yang tidak nol dalam estimasi wavelet shrinkage sangat sedikit. Hasil dari penggunaan statistik uji ini kurang dapat mendeteksi ketika I_d hanya terdiri dari sedikit koefisien, karena noise dari koefisien nol akan cenderung untuk meliputi sinyal koefisien yang tidak nol. Oleh karena itu, Ogden dan Parzen (1996) cit Ogden dan Parzen (1997) memberikan pendekatan alternatif. Jika banyaknya anggota dari I_d diketahui, katakan m, maka digunakan statistik uji jumlah kuadrat dari m X_i terbesar. Pada prakteknya m tidak diketahui, sehingga pendekatan dari Ogden dan Parzen terdiri dari prosedur uji rekursif untuk I_d yang hanya berisikan satu elemen setiap diuji. Statistik uji yang digunakan untuk uji rekursif ini adalah kuadrat dari X_i terbesar. Titik kritis dalam uji ini adalah :

$$x_d^\alpha = \left\{ \Phi^{-1}\left[\frac{1}{2}((1-\alpha)^{1/d} + 1) \right] \right\}^2$$

Langkah-langkah dalam metode rekursif untuk memilih nilai threshold optimal adalah sebagai berikut :

1) Urutkan X_i^2 dari yang terbesar hingga terkecil ,dengan i=1,2,3,...,d.
2) Bandingkan X_i^2 terbesar dengan titik kritis x_d^α.
3) Jika X_i^2 lebih besar dari x_d^α maka X_i adalah koefisien signifikan, dan masuk dalam rekonstruksi fungsi. Selanjutnya keluarkan X_i dan ganti d dengan d -1 dan kembali ke langkah 2)
4) Jika $X_i^2 < x_d^\alpha$ maka X_i merupakan koefisien yang tidak signifikan dan nilai mutlak X_i merupakan nilai threshold optimal dan proses terhenti.

Setelah didapatkan nilai threshold optimal, koefisien yang lebih kecil dari nilai threshold akan dithreshold menjadi nol. Sedangkan koefisien yang lebih besar (dalam nilai mutlak) dari nilai threshold akan dipakai dalam rekonstruksi fungsi. Tingkat kemulusan estimator wavelet yang diperoleh dengan prosedur ini juga dipengaruhi oleh besar kecilnya tingkat signifikansi α yang diambil. Semakin kecil α maka estimator yang diperoleh semakin mulus.

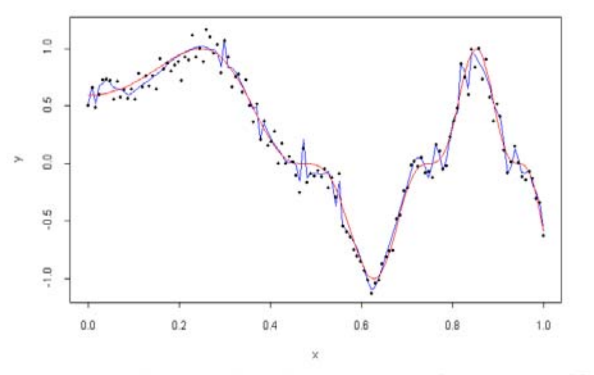
Contoh penerapan pada Sebuah Fungsi

Untuk menerapkan uji hipotesis multipel, digunakan sekumpulan data yang dibangun dari sebuah fungsi yakni

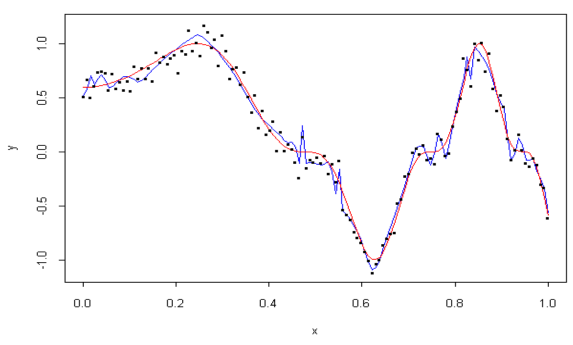
$f_i = \sin^3\left(3\pi X_i^2 + 1\right), (i = 1,2,\ldots,128)$ dengan X_i data pada interval [0,1], dengan $X_i = i/n$. Kemudian data tersebut akan diberi error(noise) yang berdistribusi normal dengan nilai ekspektasi 0 dan variansi $\sigma^2 = 0{,}01$, sehingga menjadi

$$Y_i = \sin^3\left(3\pi X_i^2 + 1\right) + \varepsilon_i$$

dengan menggunakan program S+Wavelets diperolehhasil sebagai berikut.



Gambar 5.5 Kurva Estimasi Wavelet Thresholding dengan Uji rekursif alpha= 0.05, threshold = 3.070008



Gambar 5.6 Kurva Estimasi Wavelet Thresholding dengan Uji rekursif alpha= 0.01, threshold = 3.855581

Keterangan :

: Fungsi asli

: Fungsi asli ditambah noise

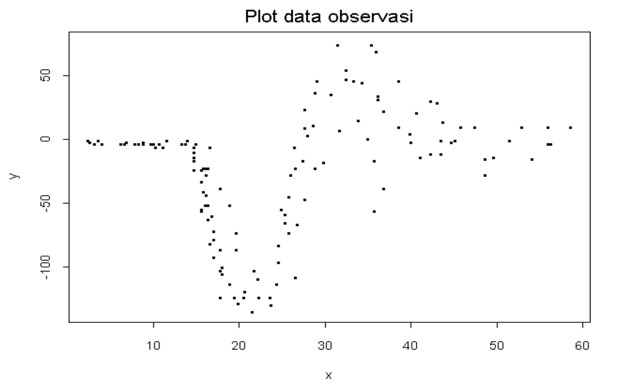
: Estimasi dengan uji rekursif

Penerapan pada Data Simulasi Tubrukan

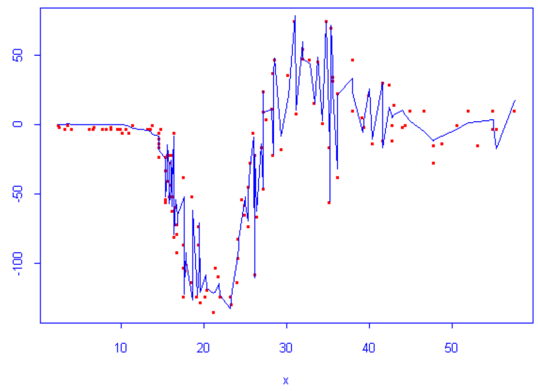
Data diambil dari buku *Applied Nonparametric Regression* (W. Hardle,1993), yakni data simulasi tubrukan sepeda motor pada suatu PTMO (*post mortem human test object*/ obyek uji pemeriksaan mayat manusia). Dalam hal ini variabel-variabelnya adalah sebagai berikut:

- 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 uji hipotesis multipel sebagai berikut



a. Dengan Pendekatan Uji Hipotesis Tunggal



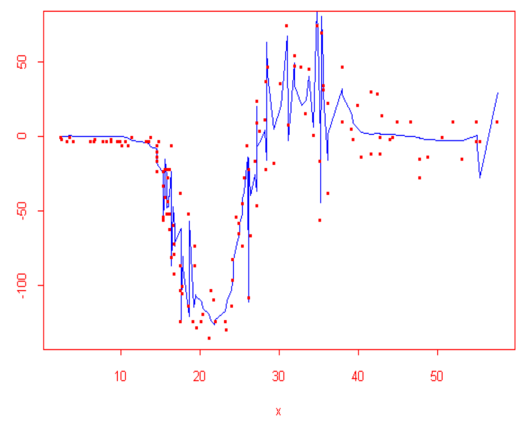
Gambar 6 Kurva Estimasi Wavelet Thresholding dengan Pendekatan Uji Hipotesis Tunggal, alpha= 0.01, threshold = 2.575829

Keterangan :

: Sebaran data

: Estimasi uji hipotesis tunggal

b. Dengan Uji Rekursif



Gambar 7 Kurva Estimasi Wavelet Thresholding dengan Uji Rekursif, alpha = 0.01, threshold = 3.820991

Keterangan :
. . . . : Sebaran data
——— : Estimasi dengan uji rekursif

Dengan pemilihan nilai alpha 0.01, pendekatan uji hipotesis tunggal dan uji rekursif menghasilkan estimasi fungsi yang berbeda. Dari hasil visualisasi gambar terlihat bahwa uji rekursif lebih baik dibandingkan dengan pendekatan uji hipotesis tunggal. Hal ini juga dapat ditunjukkan dari banyaknya koefisien wavelet yang masuk rekonstruksi. Pada uji rekursif koefisien yang masuk rekonstruksi fungsi ($d_{j,k} \neq 0$) lebih sedikit dibandingkan dengan pendekatan uji hipotesis tunggal. Akibatnya hasil visualisasi gambar dengan uji rekursif lebih baik daripada uji hipotesis tunggal. Dengan pendekatan uji hipotesa tunggal banyaknya koefisien wavelet empiris $d_{j,k} \neq 0)$ yang masuk dalam rekonstruksi ada 43 buah sedangkan dengan uji Rekursif koefisien wavelet empiris yang masuk dalam rekonstruksi ada 23 buah. Jadi dalam simulasi ini tingkat efisiensinya uji rekursi dibanding uji hipotesa tunggal hampir sekitar 50%.

KESIMPULAN

1. Wavelet mampu mengestimasi fungsi, baik itu fungsi mulus maupun fungsi yang tidak mulus.
2. Tingkat kemulusan estimator wavelet thresholding dipengaruhi oleh besarnya nilai threshold. Semakin besar nilai threshold, akan semakin mulus hasil estimasi fungsinya.
3. Penentuan nilai threshold optimal dapat dilakukan dengan menggunakan uji hipotesis multipel yang dipengaruhi oleh nilai signifikansi alpha (α). Semakin kecil alpha (α), nilai threshold akan semakin besar sehingga akan menghasilkan estimasi fungsi yang semakin mulus.

DAFTAR PUSTAKA

1. Abramovich and Benjamini. 1995. *Wavelets and Statistics*. Springer-Verlag. New York.
2. Anton, Howard. 1995. *Aljabar Linear Elementer*. Edisi kelima. Erlangga. Jakarta.
3. Bruce, A. and Gao, H Y. 1996. *Applied Wavelet Analysis with S-PLUS*. Springer-Verlag. New York..
4. Hall, P. and Patil, P. 1995. *On Wavelet Methods for Estimating Smooth Functions*. Bernoulli 1(1/2). 041-058.
5. Hardle, W. 1993. *Applied Nonparametric Regression*. Cambridge University Press. New York.
6. Martono, K. 1999. *Kalkulus*. Erlangga. Bandung.
7. Mustafid. 2003. *Statistika Elementer.* Jurusan Matematika UNDIP. Semarang.
8. Ogden, R.T. 1997. *Essential Wavelets for Statistical Applications and Data Analysis*. Birkhauser. Boston.
9. Suparti dan Subanar, H. 2000. *Estimasi Regresi dengan Metode Wavelet Shrinkage*. Jurnal Sains & Matematika. Volume 8. Nomor 3. 105-113.
10. Vetterli, M. and Kovacevic, J. 1995. *Wavelets And Subband Coding*. Prentice Hall PTR. New Jersey.