

PEMILIHAN THRESHOLD OPTIMAL PADA ESTIMATOR REGRESI WAVELET SHRINKAGE

Suparti

Staf Pengajar PS Statistika Jurusan Matematika FMIPA UNDIP
Jl. Prof. Soedarto, Kampus UNDIP Tembalang, Semarang

Abstrak: Misalkan $\{(X_i, Y_i)\}_{i=1}^n$ data observasi independen yang mempunyai model $Y_i = g(X_i) + \varepsilon_i$ dengan g fungsi regresi yang tidak diketahui dan ε_i variabel random independen dengan mean 0 dan varian σ^2 maka untuk mengestimasi fungsi g dapat dilakukan dengan pendekatan parametrik atau non parametrik.

Salah satu pendekatan non parametrik yang telah populer adalah metode wavelet. Metode wavelet merupakan pengembangan dari metode Fourier. Estimator wavelet dibedakan dalam dua jenis yaitu estimator wavelet linier dan estimator wavelet non linier atau wavelet shrinkage. Estimator wavelet shrinkage lebih unggul dibandingkan dari estimator yang lain, karena mampu mengestimasi fungsi mulus maupun tidak mulus. Tingkat kemulusan estimator wavelet shrinkage ditentukan oleh pemilihan parameter threshold. Pengambilan nilai threshold yang terlalu kecil memberikan hasil estimasi yang sangat tidak mulus (under smooth) sedangkan nilai threshold yang terlalu besar memberikan hasil estimasi yang sangat mulus (over smooth). Oleh karena itu perlu dipilih nilai threshold optimal untuk memberikan hasil estimasi yang optimal.

Ada beberapa cara pemilihan parameter threshold optimal diantaranya dengan prosedur threshold universal, adapt, minimax, top, wp, uji hipotesa multipel, FDR, cross validasi. Pemilihan threshold optimal universal memberikan hasil estimasi yang relatif lebih mulus. Sedangkan pemilihan threshold optimal dengan prosedur uji hipotesa cenderung memberikan hasil estimasi yang kurang mulus.

Kata Kunci: Regresi non parametrik, estimator wavelet shrinkage, parameter threshold.

PENDAHULUAN

Representasi fungsi dalam deret wavelet merupakan generalisasi dari representasi fungsi dalam deret Fourier. Representasi fungsi dengan wavelet lebih efisien dari representasi deret Fourier, karena koefisien wavelet yang tidak nol dalam rekonstruksi fungsi dengan wavelet relatif lebih sedikit dibandingkan dengan banyaknya koefisien Fourier yang tidak nol dalam rekonstruksi fungsi pada level resolusi yang sama.

Selain efisien, wavelet mampu merepresentasikan fungsi-fungsi yang bersifat tidak mulus, yang mana hal ini tidak dapat dilakukan oleh deret Fourier. Ini dikarenakan basis dalam wavelet ditentukan oleh letak dan skala (translasi dan dilatasi). Pada bagian fungsi yang tidak mulus, representasi wavelet akan menggunakan panjang support yang sempit dan pada bagian fungsi yang mulus akan menggunakan support yang lebih lebar. Dengan demikian fungsi wavelet mempunyai panjang support yang bersifat adaptif secara lokal.

Dalam statistika, ada banyak aplikasi wavelet. Salah satu diantaranya untuk mengestimasi fungsi regresi non parametrik. Metode paling sahih dalam estimasi fungsi regresi dengan wavelet adalah metode wavelet shrinkage atau metode wavelet thresholding. Pada estimasi fungsi dengan metode wavelet thresholding, tingkat kemulusan estimator ditentukan oleh pemilihan jenis fungsi wavelet, level resolusi, jenis fungsi thresholding dan parameter thresholdnya. Namun pemilihan jenis fungsi wavelet, level resolusi dan jenis fungsi thresholdingnya tidak sedominan pada pemilihan parameter thresholdnya. 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 parameter threshold optimal untuk mendapatkan estimasi fungsi yang optimal. Untuk memilih nilai threshold optimal, ada dua kategori pemilihan yaitu pemilihan secara global artinya memilih satu harga threshold untuk seluruh level resolusi dan pemilihan threshold tergantung pada setiap level resolusi.

Dalam paper ini dibahas tentang penentuan nilai threshold optimal pada estimator wavelet thresholding untuk fungsi regresi non parametrik.

PEMBAHASAN

Representasi Fungsi Dalam Wavelet

Fungsi wavelet adalah suatu fungsi yang berosilasi di sekitar nol dengan sifat-sifat tertentu. Fungsi wavelet dibedakan atas dua jenis, yaitu wavelet ayah (ϕ) dan wavelet ibu (ψ) yang mempunyai sifat:

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

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

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

Beberapa contoh wavelet selain wavelet Haar diantaranya adalah wavelet Daubechies (Daublet), symmetris (Symmlet), dan Coifman (Coiflet)[3].

Jika $g \in L^2(\mathbb{R})$ maka g dapat direpresentasikan dalam deret wavelet ortonormal

$$g(x) = \sum_{k \in \mathbb{Z}} c_{j_0,k} \phi_{j_0,k}(x) + \sum_{j \geq j_0} \sum_{k \in \mathbb{Z}} d_{j,k} \psi_{j,k}(x) \quad (3)$$

dengan $c_{j_0,k} = \langle g, \phi_{j_0,k} \rangle = \int_{-\infty}^{\infty} g(x) \phi_{j_0,k}(x) dx$ dan $d_{j,k} = \langle g, \psi_{j,k} \rangle = \int_{-\infty}^{\infty} g(x) \psi_{j,k}(x) dx$.

Dengan mengambil J cukup besar maka deret wavelet (3) dapat didekati oleh

$$g_J(x) = \sum_{k \in \mathbb{Z}} c_{j_0,k} \phi_{j_0,k}(x) + \sum_{j \geq j_0}^{J-1} \sum_{k \in \mathbb{Z}} d_{j,k} \psi_{j,k}(x) \quad (4)$$

g_J merupakan pendekatan g pada level ke J .

Estimator Wavelet Linier

Jika terdapat sekumpulan data independen $\{(X_i, Y_i)\}_{i=1}^n$ yang mempunyai model

$$Y_i = g(X_i) + \varepsilon_i, \quad (5)$$

dengan ε_i variabel random independen berdistribusi Normal $(0, \sigma^2)$, $n = 2^m$ dengan m bilangan bulat positif dan jika X_i rancangan titik reguler pada interval $[0,1]$ dengan $X_i = i/n$, maka estimator g pada level J adalah

$$\hat{g}_J(x) = \sum_{k \in \mathbb{Z}} \hat{c}_{j_0,k} \phi_{j_0,k}(x) + \sum_{j \geq j_0}^{J-1} \sum_{k \in \mathbb{Z}} \hat{d}_{j,k} \psi_{j,k}(x) \quad (6)$$

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 $c_{j_0,k}$, $d_{j,k}$ [10].

Estimator Wavelet Shrinkage

Jika diberikan data independen $\{(X_i, Y_i)\}_{i=1}^n$ dengan model (5) maka $Y_i \sim N(g(X_i), \sigma^2)$.

Selanjutnya, $\sqrt{n}(\hat{d}_{j,k} - d_{j,k}) \sim N(0, \sigma^2)$ atau $\hat{d}_{j,k} = d_{j,k} + \left(\frac{1}{\sqrt{n}}\right) z_{j,k}$ dengan $\hat{d}_{j,k}$ adalah estimator dari koefisien $d_{j,k}$ dan $z_{j,k}$ adalah n himpunan yang tidak teramati berdistribusi $N(0, \sigma^2)$. 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. Dengan ide demikian, berdasarkan referensi [5], [6], [7] dan [9] 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 (dianggap 0). Nilai tertentu tersebut dinamakan nilai threshold dan estimator waveletnya dinamakan estimator wavelet thresholding (shrinkage). Misalkan tersedia nilai threshold λ , maka estimator wavelet shrinkage dari fungsi regresi g dapat dituliskan sebagai

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

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

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

λ : parameter nilai threshold

∂_λ : fungsi threshold

Thresholding ini merupakan operator non linier pada vektor koefisien wavelet d yang diestimasi dengan $\hat{d}$. Dengan demikian, estimator thresholding (shrinkage) tersebut dinamakan juga estimator wavelet non linier. Karena thresholding ini dirancang untuk membedakan antara koefisien wavelet empiris mana yang masuk dan mana 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 di dalam 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 shrinkage 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 shrinkage adalah

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

Metode wavelet shrinkage mampu mengestimasi fungsi baik mulus maupun tidak mulus karena wavelet shrinkage mempunyai panjang support yang adaptif secara lokal.

Algoritma Penentuan Estimator Wavelet Shrinkage

Algoritma penentuan estimator wavelet shrinkage terdiri atas tiga tahap, yaitu:

1. Menentukan transformasi wavelet diskrit pada level J terhadap sebuah sinyal Y sehingga diperoleh koefisien detail $d_1, d_2, \dots, d_J$, dan koefisien penghalus s_J .
2. Melakukan thresholding yaitu menyusutkan koefisien detail pada skala terbaik j untuk mendapatkan koefisien detail baru $\tilde{d}_1 = \partial_{\lambda, \sigma_1}(d_1), \dots, \tilde{d}_j = \partial_{\lambda, \sigma_j}(d_j)$. Fungsi $\partial_{\lambda, \sigma}$ menyusutkan nilai x menuju nol, dengan λ merupakan nilai threshold dan σ adalah estimasi skala noise.
3. Menentukan transformasi wavelet invers pada koefisien detail (yang baru) untuk mendapatkan estimasi wavelet shrinkage dari fungsi g , yaitu $\hat{g}_\lambda$.

Langkah-langkah Thresholding

Pada dasarnya, proses thresholding koefisien wavelet dapat dibagi kedalam empat langkah, yaitu:

1. Pemilihan Fungsi Thresholding

Ada dua jenis fungsi thresholding ∂_λ , yaitu thresholding kuat, $\partial_\lambda^H(x) = \begin{cases} x, & |x| > \lambda \\ 0, & \text{yang lain} \end{cases}$ dan

thresholding lemah $\partial_\lambda^S(x) = \begin{cases} x - \lambda, & x > \lambda \\ 0, & x \leq \lambda \\ x + \lambda, & x < -\lambda \end{cases}$ dengan λ merupakan parameter threshold.

Fungsi thresholding kuat lebih dikenal karena terdapat diskontinyu sehingga nilai x yang berada diatas threshold λ tidak disentuh. Sebaliknya, fungsi shrinkage lemah bersifat kontinyu sejak nilai x berada diatas threshold λ .

2. Pemilihan Parameter Threshold

Tingkat kemulusan estimator wavelet shrinkage ditentukan oleh nilai threshold λ . Nilai λ kecil memberikan fungsi estimasi yang berayun dan λ besar memberikan fungsi estimasi yang sangat mulus. Oleh karena itu perlu dipilih nilai λ optimal.

3. Estimasi σ dari Noise

Dalam merekonstruksi estimator regresi wavelet shrinkage biasanya nilai σ tidak diketahui. Oleh karena itu, σ harus diestimasi dari data. Referensi [9] memberikan estimasi σ berdasarkan koefisien wavelet empiris pada level resolusi tertinggi dengan fungsi Median Deviasi Absolut (MAD), yaitu

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

Beberapa cara untuk menentukan nilai threshold λ_j optimal pada fungsi threshold (shrinkage), yaitu:

1. Threshold Universal

Threshold universal ditentukan dengan memilih satu nilai threshold λ untuk seluruh level resolusi j , dan didefinisikan dengan $\lambda_j = \sqrt{2 \log n}$ dengan n adalah banyaknya observasi [2].

2. Threshold wp

Threshold wp ditentukan dengan memilih satu nilai threshold λ untuk seluruh level resolusi j , dan didefinisikan $\lambda_j = \sqrt{2 \log(n \cdot \log(n))}$ dengan n adalah banyaknya observasi [2].

3. Threshold Minimax

Nilai-nilai threshold minimax ditentukan berdasarkan ukuran sampel n , dan sudah ditabelkan menurut referensi [4], yaitu:

Tabel 1. Nilai threshold minimax berdasarkan ukuran sampel

n	λ	n	λ
2	0	512	2,074
4	0	1024	2,232
8	0	2048	2,414
16	1,200	4096	2,594
32	1,270	8192	2,773
64	1,474	16384	2,952
128	1,669	32768	3,131
256	1,860	65536	3,310

Nilai-nilai threshold minimax selalu lebih kecil dibandingkan dengan nilai threshold universal untuk ukuran sampel yang sama.

4. Threshold Adapt

Threshold adapt adalah nilai threshold yang ditentukan berdasarkan pada level resolusi j . Pemilihan threshold ini 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 $\lambda_j = \text{argmin}_{t \geq 0} \text{SURE}(d_j, t)$

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

Threshold adapt akan memberikan hasil yang kurang baik jika koefisien-koefisiennya sangat jarang (sebagian besar koefisien pada level tersebut mendekati nol). Oleh karena itu, himpunan koefisien ini diuji dengan persamaan berikut:

$$\frac{1}{K} \sum_{k=1}^K \frac{d_{j,k}}{\sigma_j} - 1 \leq \frac{(2 \log K)^{3/2}}{\sqrt{K}}$$

Jika persamaan tersebut terpenuhi maka threshold yang digunakan pada level resolusi j adalah threshold universal, sedangkan jika tidak maka threshold adaptlah yang digunakan [2].

5. Threshold Top

Nilai threshold ini ditentukan dengan menentukan besarnya prosentase koefisien yang akan digunakan dari keseluruhan koefisien wavelet dalam merekonstruksi fungsi sehingga dapat ditentukan berapa prosen koefisien wavelet yang akan digunakan untuk mengestimasi suatu fungsi. Besar threshold adalah min dari besar koefisien wavelet yang digunakan.

6. Threshold dengan uji hipotesa multipel

Penentuan parameter threshold optimal dengan prosedur uji hipotesa menguji apakah koefisien-koefisien wavelet signifikan atau tidak [1]. Jika dalam uji hipotesa menyimpulkan bahwa koefisien wavelet signifikan maka koefisien ini akan dipertahankan dalam merekonstruksi fungsi, tetapi jika dalam uji hipotesa koefisien wavelet memutuskan sama dengan nol maka koefisien ini akan diabaikan. Hipotesa yang akan diuji (ada n hipotesa yaitu sebanyak koefisien detail) adalah :

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

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

Karena $\hat{d}_{j,k}$ berdistribusi $N(d_{j,k}, \sigma^2/n)$, dengan menggunakan tingkat signifikansi α , prosedur pengujian tunggal $H_0 : d_{jk} = 0$ melawan $H_1 : d_{jk} \neq 0$, untuk sembarang j dan k yang telah ditetapkan, akan menolak H_0

jika statistik uji $\left| \frac{\hat{d}_{j,k}}{\sigma/\sqrt{n}} \right| > z_{\alpha/2}$. Jadi threshold optimal dengan prosedur uji hipotesa multipel adalah nilai

terkecil $\hat{d}_{j,k}$ yang memenuhi $\left| \frac{\hat{d}_{j,k}}{\sigma/\sqrt{n}} \right| > z_{\alpha/2}$. Dengan uji multipel ini diperkirakan ada sebanyak $n\alpha$ koefisien

wavelet yang masuk dalam rekonstruksi fungsi. Jumlah ini terlalu banyak sehingga menghasilkan fungsi estimasi yang sangat tidak mulus. Untuk mengatasinya dipilih estimasi σ pada level terendah.

7. Threshold dengan prosedur FDR(False Discovery Rate)

Berdasarkan referensi [1] dan [9], prosedur False Discovery Rate (FDR)) merupakan pengembangan dari uji hipotesa multipel. Misalkan R menyatakan banyaknya koefisien wavelet yang tidak diabaikan dalam prosedur estimasi wavelet shrinkage. 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 :

i. Untuk setiap koefisien empiris $\hat{d}_{jk}$, dihitung p-value dua sisi (p_{jk}) untuk n hipotesa $H_{jk} : d_{jk} = 0$, yaitu p_{jk}

$= 2(1 - \phi(\hat{d}_{jk}/\sigma))$ dengan ϕ suatu nilai yang berkenaan dengan distribusi dari $\hat{d}_{jk}$.

ii. p_{jk} diurutkan yaitu $p_{(1)} \leq p_{(2)} \leq \dots \leq p_{(m)}$ dengan $p_{(i)}$ bersesuaian dengan beberapa d_{jk} .

iii. Mulai dengan $i = 1$, dan misalkan $k =$ bilangan terbesar i yang memenuhi $p_{(i)} \leq (i/m)q$ dengan q merupakan tingkat signifikansi biasanya $q = 0,01$ atau $0,05$ maka λ optimal adalah $\lambda_k = \sigma \phi^{-1}(1 - p_{(k)}/2)$.

8. Threshold dengan cross validasi (CV)

Mencari threshold optimal dengan metode cross validasi adalah mencari λ yang meminimumkan MISE antara estimator wavelet thresholding $\hat{g}_\lambda$ dan fungsi sebenarnya g dengan $MISE(\hat{g}_\lambda) = M(\lambda) = E \int \{\hat{g}_\lambda(x) - g(x)\}^2 dx$. Dalam prakteknya fungsi sebenarnya g tidak diketahui (di sini yang akan dicari estimatornya) sehingga M harus diestimasi katakanlah $\hat{M}$. Metode cross validasi klasik meminimumkan besarnya $\hat{M}$ (cross validasi) dengan meninggalkan satu data yang dikeluarkan (Wang, 1996). Tetapi karena transformasi koefisien diskrit cepat pada wavelet menggunakan 2^m data maka metode cross validasi klasik tidak dapat langsung diterapkan untuk memilih parameter optimal dalam estimator wavelet shrinkage. Metode cross validasi klasik kemudian dimodifikasi dengan mengeluarkan 2^{m-v} data (Nason, 1995). Tanpa mengurangi keumuman dapat diambil $v=1$ dan metodenya dinamakan metode cross validasi two-fold (cross validasi 2 lipatan). Prosedur Cross validasi untuk menentukan threshold optimal dengan metode cross validasi dua lipatan sbb :

i. Data semula $Y_1, Y_2, \dots, Y_n$ dengan $n = 2^m$. 2^{m-1} data berindeks ganjil dikeluarkan dan 2^{m-1} data berindeks genap disisakan. Data berindeks genap ini diindeks ulang dari $j = 1, 2, \dots, 2^{m-1}$.

ii. Fungsi estimasi indeks genap $\hat{g}_{\lambda}^E$ dikonstruksi menggunakan threshold λ tertentu berdasarkan data yang diindeks ulang Y_j .

iii. Dibentuk fungsi interpolasi dari $\hat{g}_{\lambda}^E$ dengan

$$\tilde{g}_{\lambda,j}^E = \begin{cases} \frac{1}{2}(\hat{g}_{\lambda,j-1}^E + \hat{g}_{\lambda,j}^E), & j = 2, \dots, n/2 \\ \frac{1}{2}(\hat{g}_{\lambda,1}^E + \hat{g}_{\lambda,n/2}^E), & j = 1 \end{cases}$$

iv. Dihitung interpolasi $\tilde{g}_{\lambda,j}^O$ untuk titik-titik yang berasal indeks ganjil.

v. Dihitung nilai cross validasi berdasarkan $n/2$ data yaitu

$$\hat{M}(\lambda) = \sum_{j=1}^{n/2} \left\{ \left(\tilde{g}_{\lambda,j}^E - Y_{2j-1} \right)^2 + \left(\tilde{g}_{\lambda,j}^O - Y_{2j} \right)^2 \right\}$$

vi. Mengulangi langkah ii s.d v untuk berbagai nilai λ .

vii. Nilai λ yang meminumkan cross validasi (5) merupakan threshold optimal untuk $n/2$ data yaitu $\lambda(n/2)$.

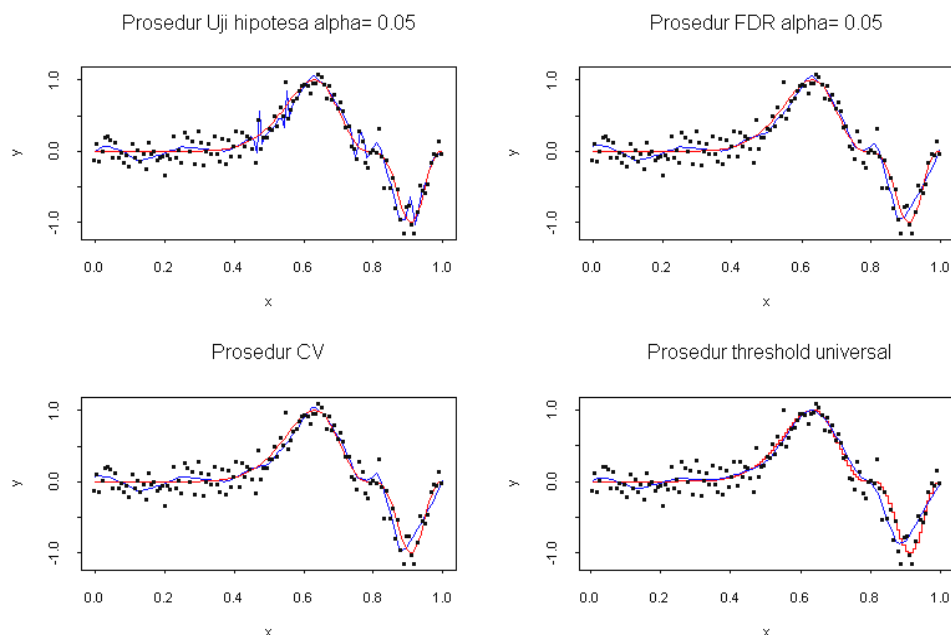
viii. Dihitung threshold optimal untuk n data dengan $\lambda(n) \approx \left(1 - \frac{\log 2}{\log n} \right)^{-1/2} \lambda(n/2)$.

Contoh Simulasi

Diberikan suatu fungsi $y = (\sin(2\pi x^3))^3 + \varepsilon$ dengan ε adalah 128 bilangan random normal dengan mean = 0, sd = 0,15 dan $x = i/128, i=1,2,\dots,128$. Dilakukan estimasi fungsi wavelet shrinkage dengan pemilihan threshold optimal menggunakan prosedur uji hipotesa, FDR, CV dan threshold universal yang hasilnya ditunjukkan pada gambar 1. Pada hasil estimasi fungsi, prosedur uji hipotesa memberikan estimasi yang paling berayun. Ini menunjukkan bahwa banyak koefisien yang masuk dalam rekonstruksi atau dengan kata lain lamda optimal yang diperoleh paling kecil. Sedangkan prosedur FDR, CV, universal hampir memberikan hasil estimasi yang mendekati, namun jika diperhatikan lebih cermat estimasi dengan threshold universal terlihat paling mulus. Pada prosedur FDR, CV dan universal, lamda optimal yang diperoleh masing-masing adalah 0,813; 0,540 dan 3,115. Sedangkan pada prosedur uji hipotesa diperoleh lamda optimalnya adalah 0,335.

KESIMPULAN

Prosedur uji hipotesa memberikan hasil estimasi yang sangat tidak mulus. Ini berarti lamda optimal kecil sehingga terlalu banyak koefisien wavelet yang masuk dalam rekonstruksi. Sedangkan prosedur threshold universal memberikan hasil estimasi yang paling mulus.



Gambar 1. Fungsi estimasi regresi wavelet shrinkage dengan berbagai pemilihan lamda optimal

..... : diagram pencar
 _____ : fungsi estimasi
 ----- : fungsi sesungguhnya

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, pp. 5-14, 1995.
- [2]. Bruce, A. and Gao, H Y., *Applied Wavelet Analysis with S-PLUS*, Springer-Verlag, New York, 1996
- [3]. Daubechies, I., *Ten Lectures on Wavelet*, Capital City Press, Philadelphia, 1992.
- [4]. Donoho, D.L and Johnstone, I.M., Ideal Spatial Adaptation by Wavelet Shrinkage. *Biometrika*, Vol. 81, No. 3, pp. 425-455, 1994.
- [5]. Donoho, D.L and Johnstone, I.M., Wavelet Shrinkage: Asymptopia? deal Spatial Adaptation by Wavelet Shrinkage, *J.R.Statist.Soc.B*, Vol. 57, No.2 , pp. 301-369, 1995.
- [6]. Hall, P. and Patil, P., On Wavelet Methods for Estimating Smooth Functions, *Bernoulli*, Vol.1, No.1/2, pp. 041-058, 1995.
- [7]. Hall, P. and Patil, P., On the Choice of Smoothing Parameter, Threshold and Truncation in Nonparametric Regression by non-linier wavelet Methods, *J.R.Statist.Soc.B*, Vol.58, pp. 361-377, 1996.
- [8]. Nason, G.P, Choice of Threshold Parameter in wavelet Function Estimation, In *Wavelets and Statistics* . Antoniadis, A., and Oppenheim, G.(eds.). Springer -Verlag, New York , pp. 261-280, 1995.
- [9]. Ogden, R.T., *Essential Wavelets for Statistical Applications and Data Analysis*, Birkhauser, Boston, 1997.
- [10]. Suparti dan Subanar, Estimasi Regresi dengan Metode Wavelet Shrinkage, *Jurnal Sains & Matematika*, Vol. 8, No. 3, 2000.
- [11]. Wang, Y., Function Estimation Via Wavelet Shrinkage For Long Memory Data, *The Annals of Statistics*, Vol. 24, No.2, pp. 466-484, 1996.