APLIKASI *SPLINE ESTIMATOR* TERBOBOT

I Nyoman Budiantara
Dosen Fakultas Matematika dan Ilmu Pengetahuan Alam, Jurusan Statistika
Institut Teknologi 10 November Surabaya

ABSTRAK

Diberikan model regresi nonparametrik : $Z_j = X(t_j) + e_j$, $j = 1,2,\ldots,n$, dengan $X(t_j)$ kurva regresi dan e_j sesatan *random* yang diasumsikan berdistribusi normal dengan *mean* nol dan variansi σ^2/b_j, $b_j > 0$. Estimasi kurva regresi X yang meminimumkan suatu *Penalized Least Square Terbobot*, merupakan estimator Polinomial Spline Natural Terbobot. Selanjutnya diberikan suatu aplikasi estimator *spline* terbobot dalam regresi nonparametrik.

Kata kunci: *Spline* terbobot, Regresi nonparametrik, *Penalized Least Square*.

ABSTRACT

We considered the nonparametric regression model : $Z_j = X(t_j) + e_j$, $j = 1,2,\ldots,n$, where $X(t_j)$ is the regression curve. The random error e_j are independently distributed normal with a zero mean and a variance s^2/b_j, $b_j > 0$. The estimation of X obtained by minimizing a Weighted Least Square. The solution of this optimation is a Weighted Spline Polynomial. Further, we give an application of weigted spline estimator in nonparametric regression.

Keywords: Weighted Spline, Nonparametric Regression, Penalized Least Square.

1. PENDAHULUAN

Diberikan data (t_j, Z_j) dengan hubungan antara t_j dan Z_j diasumsikan mengikuti model regresi nonparametrik :

$$Z_j = X(t_j) + e_j,\ j = 1,2,\ldots,n,$$

dengan $X(t_j)$ merupakan kurva regresi yang diasumsikan licin (*smooth*), dalam arti merupakan anggota ruang fungsi tertentu (*Sobolev order dua*), khususnya $X \in W[a,b]$ dengan :

$$W[a,b] = \left\{ X : \int_a^b [X^{(m)}(s)]^2\, ds < \infty \right\},$$ untuk suatu m bilangan bulat positif.

Sesatan random e_j berdistribusi normal dengan *mean* nol dan variansi σ^2/b_j, $b_j > 0$. Untuk mendapatkan bentuk estimasi kurva regresi X, digunakan pendekatan *Penalized Least Square* Terbobot (Budiantara, et al, 1997) :

$$\underset{X \in W}{Arg\ Min} \left\{ n^{-1} \sum_{j=1}^{n} b_j (Z_j - X(t_j))^2 \right\},\ b_j > 0.$$

dengan syarat :

$$\int_a^b [X^{(m)}(s)]^2\, ds < \Phi,\ \Phi > 0.$$

Parameter Φ merupakan kontrol antara kemulusan kurva dan *Goodness of fit*. Penyelesaian optimasi dengan kendala ini berupa Polinomial Spline Natural Terbobot berorde 2*m*-1, yang dapat disajikan menjadi (Budiantara dan Subanar,1998) :

$$\hat{X}(t) = \sum_{j=1}^{n} B(t, t_j)\, Z_j$$

dengan $B(t, t_j)$ koefisien bobot yang bebas dari observasi Z_j. Salah satu sifat baik dalam regresi parametrik yang masih dapat dipertahankan dalam estimator spline terbobot adalah sifat estimator linier dalam observasi (Budiantara,1999). Bentuk estimator linier ini dapat memberikan suatu keuntungan dalam mengkonstruksi inferensi dalam analisis regresi nonparametrik, khususnya *spline* terbobot (Budiantara, 2000). Meskipun estimator *spline* terbobot bersifat bias untuk $X(t_j)$, yaitu :

$$E[\hat{X}(t)] = \sum_{j=1}^{n} B(t, t_j)\, X(t_j) \quad \neq \quad X(t_j),$$

tetapi estimator ini tak bias asimtotik, sehingga inferensinya secara matematik cukup baik berdasarkan sampel berukuran besar (Budiantara dan Subanar, 1997).

Estimator *spline* terbobot dalam regresi nonparametrik berdasarkan sudut pandang teoritis telah diberikan dan dikembangkan oleh Budiantara (2000). Dalam tulisan ini diberikan suatu aplikasi model *spline* terbobot, dengan data diambil dari penelitian yang dilakukan oleh Schmidt, Mattern dan Schuler pada tahun 1981 (Hardle ,1990).

2. APLIKASI *SPLINE ESTIMATOR* TERBOBOT

Pada bagian ini, disajikan contoh aplikasi model *spline* terbobot, dengan data diambil dari penelitian yang dilakukan oleh Schmidt, Mattern dan Schuler pada tahun 1981 (Hardle ,1990). Diberikan data t_j, j=1,2,…,133 yang menyatakan waktu (dalam mili detik) setelah tabrakan simulasi dengan sepeda motor. Sedangkan variabel respon Z_j menyatakan percepatan yang tercatat pada suatu PMTO (*Post Mortem Human Test Object*). Ingin diselidiki pengaruh waktu setelah tabrakan terhadap percepatan dalam PMTO.

Untuk menyelidiki keinginan di atas dan menyelesaikan inferensi-inferensinya dengan model spline terbobot dapat menggunakan program pada lampiran 1 berikut, dengan memanfaatkan fasilitas dalan Program S-PLUS for Windows.

Berdasarkan plot antara (t_j,Z_j), j=1,2,…,133, terlihat bahwa ada indikasi ketidak konstanan variansi data. Untuk nilai-nilai t yang makin besar, perilaku variansi data cendrung makin besar dan kemudian mengecil. Selanjutnya, model data didekati dengan *spline* kubik terbobot dengan bobot $1/t_j^2$ dan titik-titik knots pada 15, 22, 28 dan 36. Model *spline* ini dapat ditulis menjadi :

$$X(t) = \text{¡}_0 + \text{¡}_1 t + \text{¡}_2 t^2 + \text{¡}_3 t^3 + \text{¡}_4 (t-15)_+^3 + \text{¡}_5 (t-22)_+^3 + \text{¡}_6 (t-28)_+^3 + \text{¡}_7 (t-36)_+^3,$$

dengan :

$$(t-a)_+^k = \begin{cases} 0, & , t < a \\ (t-a)^k & , t \geq a \end{cases}.$$

Ringkasan nilai GCV untuk model-model *spline* kubik terbobot dengan berbagai titik-titik *knots*, diberikan dalam Tabel 1.

Tabel 1. Ringkasan nilai GCV model *spline* kubik terbobot dengan berbagai knots

Titik-titik knots	Nilai GCV
10, 22, 28, 36	1,465313
14, 22, 28, 36	1,018173
15, 22, 28, 36*	**0,984761***
16, 22, 28, 36	0,992416
17, 22, 28, 36	1,032384
21, 22, 28, 36	1,305274
15, 16, 28, 36	1,375524
15, 20, 28, 36	1,028140
15, 21, 28, 36	0,997346
15, 25, 28, 36	1,078732
15, 27, 28, 36	1,223866
15, 22, 23, 36	1,105363
15, 22, 25, 36	1,034937
15, 22, 27, 36	0,993632
15, 22, 29, 36	0,985012
15, 22, 32, 36	1,038564
15, 22, 28, 30	1,008589
15, 22, 28, 33	0,992419
15, 22, 28, 35	0,985711
15, 16, 28, 37	0,985560
15, 16, 28, 40	0,996436
15, 16, 28, 45	1,029835
15, 16, 28, 50	1,061733
15, 16, 28, 55	1,086187

Terlihat dari Tabel 1 bahwa model *spline* terbobot dengan titik-titik *knots* 15, 22, 28 dan 36 mempunyai nilai GCV yang terkecil, yaitu 0,984761. Berdasarkan analisis *residual* diperoleh kesimpulan model *spline* terbobot dengan titik-titik knots 15, 22, 28 dan 36 mempunyai variansi data yang konstan dan residualnya tidak ada indikasi penyimpangan terhadap distribusi normal.

Selanjutnya, dilakukan uji hipotesis koefisien-koefisien regresi $\beta_0, \beta_1, ..., \beta_7$. Pertama dilakukan uji hipotesis :

H_0 : $\beta_0 = \beta_1 = \ldots = \beta_7 = 0$,

H_1 : Tidak semua β_i, $i = 0,1,2,\ldots,7$ sama dengan nol.

Ringkasan analisis variansi model *spline* terbobot diberikan dalam Tabel 2.

Tabel 2. Analisis Variansi Model *Spline* Terbobot

Sumber Variasi	Derajat bebas	Jumlah Kuadrat (JK)	Rata-rata JK	F
Regresi	7	890,03	127,15	137,38
Sesatan	125	115,69	0,93	
Total	132	1.005,72		

Dengan tingkat signifikansi α = 5 % diperoleh kesimpulan bahwa tidak semua β_i sama dengan nol, i = 0,1,2,…,7.

Lebih lanjut diuji koefisien-koefisien regresi, khususnya koefisien-koefisien fungsi *truncated* (β_4, β_5, β_6 dan β_7) yang memberikan pengaruh signifikan terhadap model. Untuk uji hipotesis :

$H_0 : \beta_4 = 0$, $H_1 : \beta_4 \neq 0$,
$H_0 : \beta_5 = 0$, $H_1 : \beta_5 \neq 0$,
$H_0 : \beta_6 = 0$, $H_1 : \beta_6 \neq 0$,
$H_0 : \beta_7 = 0$, $H_1 : \beta_7 \neq 0$.

Ringkasan inferensi untuk uji ini diberikan dalam Tabel 3.

Tabel 3. Ringkasan Estimasi Parameter Model *Spline* Terbobot

Parameter	Estimasi	Sesatan Standar	Nilai t
β_0	26,627	6,473	4,11
β_1	-15,852	3,043	-5,21
β_2	2,411	0,359	6,72
β_3	-0,106	0,012	-8,83
β_4	0,476	0,035	13,60
β_5	-0,851	0,059	-14,42
β_6	0,632	0,061	10,36
β_7	-0,158	0,039	-4,05

Berdasarkan tingkat signifikansi α = 5 % diperoleh kesimpulan bahwa parameter β_4, β_5, β_6 dan β_7 signifikan dalam model. Jadi dapat disimpulkan bahwa model *spline* kubik terbobot dengan titik-tiktik *knots* 15, 22, 28 dan 36 cukup memadai sebagai model pendekatan untuk data model di atas.

DAFTAR PUSTAKA

Budiantara, I N., Subanar, and Zoejoeti, Z., 1997. Weighted Spline Estimator, *Bulletin of the International Statistical Institute*, 51, 333-334.

Budiantara, I N. dan Subanar, 1997. Sifat Asimtotik Estimator Spline Terbobot, *Majalah Ilmiah Matematika dan Pengetahuan Alam*, 7, 23-36.

Budiantara, I N. dan Subanar, 1998. Estimator Spline Terbobot, *Majalah Ilmiah Himpunan Matematika Indonesia*, 4, 35-45.

Budiantara, I N., 1999. Estimator Spline Terbobot Dalam Regresi Semiparametrik, *Majalah Ilmu Pengetahuan dan Teknologi*, 10, 103-109.

Budiantara, I N., 2000. *Estimator Spline Dalam Regresi Nonparametrik dan Semiparametrik*, Disertasi Doktor pada Universitas Gadjah Mada, Yogyakarta.

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

Lampiran 1. Program S-PLUS *Spline Estimator* Terbobot

```
spline<- function(x, y, w1, k1, k2, k3, k4)
        {
          n <- length(y)
          trun <- function(data, knots, power)
              {
              ((data - knots)^power) * (data >= knots)
               }
             m <- matrix(0, ncol = 8, nrow = n)
              m[, 1] <- 1
              m[, 2] <- x
              m[, 3] <- x^2
              m[, 4] <- x^3
              m[, 5] <- trun(x, k1, 3)
              m[, 6] <- trun(x, k2, 3)
              m[, 7] <- trun(x, k3, 3)
              m[, 8] <- trun(x, k4, 3)
              w <- diag(w1)
              wbeta <- solve(t(m) %*% w %*% m) %*% t(m) %*% w %*%y
              flamda <- m %*% wbeta
              alamda <- m %*% solve(t(m) %*% w %*% m) %*% t(m) %*% w
              ia <- diag(n) - alamda
              bgcv <- (sum(diag(ia))/n)^2
              agcv <- t(ia %*% y) %*% w %*% (ia %*% y)/n
              gcv <- agcv/bgcv
              i <- seq(min(x), max(x), length = n)
              fest <- wbeta[1] + wbeta[2] * i + wbeta[3] * i^2 + wbeta[4] * i^3 +
                      wbeta[5] * trun(i,k1,3)+wbeta[6]* trun(i, k2,3)+
                      wbeta[7]*  trun(i,k3,3)+wbeta[8] * trun(i, k4, 3)
              residual <- y - flamda
             ybar <- sum(y)/n
              sse <- t(y - flamda) %*% w %*% (y - flamda)
              syy <- t(y - ybar) %*% w %*% (y - ybar)
              ssr <- t(flamda - ybar) %*% w %*% (flamda - ybar)
              koef.det. <- ssr/syy
              mse <- as.vector(sse)/(n - 8)
              rmse <- 1/sqrt(mse)
              d <- residual %*% rmse
              msr <- as.vector(ssr)/7
              Fh <- msr/mse
              covbeta <- solve(t(m) %*% w %*% m) * mse
              seb0 <- sqrt(covbeta[1, 1] * mse)
              tb0 <- wbeta[1]/seb0
              seb1 <- sqrt(covbeta[2, 2] * mse)
              tb1 <- wbeta[2]/seb1
              seb2 <- sqrt(covbeta[3, 3] * mse)
              tb2 <- wbeta[3]/seb2
              seb3 <- sqrt(covbeta[4, 4] * mse)
```

```
        tb3 <- wbeta[4]/seb3
        seb4 <- sqrt(covbeta[5, 5] * mse)
        tb4 <- wbeta[5]/seb4
        seb5 <- sqrt(covbeta[6, 6] * mse)
        tb5 <- wbeta[6]/seb5
        seb6 <- sqrt(covbeta[7, 7] * mse)
        tb6 <- wbeta[7]/seb6
        seb7 <- sqrt(covbeta[8, 8] * mse)
        tb7 <- wbeta[8]/seb7
win.graph()
 par(mfrow = c(2, 2))
 plot(x, y, xlab = "t", ylab = "Data y")
 plot(i, fest, type = "l", xlab = "Data t", ylab = "Spline terbobot")
 plot(flamda, residual, type = "p", xlab = "flamda", ylab = "Residual")
 qqnorm(residual)
 return(wbeta, gcv, syy, ssr, sse, mse, msr, koef.det., covbeta, Fh, seb0, tb0,
       seb1, tb1, seb2, tb2, seb3, tb3, seb4, tb4, seb5, tb5)
 }.
```