Menentukan Parameter Pemulus pada Model Regresi Smoothing Spline

Julio Adisantoso

Agustus 2010

Abstrak/*Abstract*

Makalah ini membahas dan melakukan percobaan pendugaan kurva regresi non parametrik dengan menggunakan metode *smoothing spline*. Masalah yang dihadapi dalam menduga kurva tersebut adalah memilih parameter pemulus. Beberapa teknik tersedia untuk menentukan parameter pemulus yaitu rataan kuadrat sisaan (MSE), fungsi *loss* dan fungsi resiko, serta *Generalized Cross-Validation* (GCV). Ketiga teknik tersebut tersedia dalam program R.

Nilai parameter pemulus optimal untuk menduga kurva regresi spline dapat ditentukan melalui nilai *spar* dan derajat bebas (df). Hasil percobaan untuk menduga kurva regresi dari data simulasi dengan menggunakan fungsi **smooth.spline** menunjukkan hasil yang baik pada nilai parameter pemulus optimal. Nilai parameter pemulus menggunakan *df* memiliki hubungan terbalik dengan *spar* dan λ. Semakin besar nilai *df* maka semakin kecil nilai *spar* maupun λ yang mengakibatkan kurva dugaan semakin tidak mulus atau mendekati nilai data.

1 Pendahuluan

Tujuan analisis regresi adalah mempelajari bagaimana respon sebuah variabel Y akibat perubahan pada peubah lain yaitu X. Hubungan antara X dan Y dapat dituliskan sebagai

$$Y_i = f(x_i) + \epsilon_i,\ i = 1, ..., n \tag{1}$$

sedangkan f adalah fungsi regresi dan ϵ_i adalah *error* dari setiap individu ke-i. Pada prakteknya, kita memperoleh pasangan data (x_1,Y_1),...,(x_n,Y_n) yang berisi informasi tentang fungsi f. Dari setiap pasangan data ini kita dapat menduga fungsi f tersebut.

Pendugaan fungsi f dapat dilakukan dengan pendekatan parametrik atau non parametrik. Jika pengetahuan tentang fungsi f minim maka pendugaan terhadap fungsi f menggunakan regresi non parametrik. Agar pendekatan non-parametrik ini menghasilkan pendugaan f yang baik, maka seperti halnya dengan pendugaan fungsi kepekatan, sangat dipenagaruhi oleh derajat pemulusan (*bandwidth*). Biasanya kemulusan dari f merupakan syarat yang cukup untuk menjamin sebuah penduga akan konvergen pada f yang sesungguhnya bila jumlah data bertambah banyak tanpa batas. Pendekatan non parametrik lebih fleksibel karena tidak dibatasi oleh asumsi-asumsi seperti halnya pada regresi parametrik. Kurva regresi non parametrik hanya diasumsikan *smooth* (mulus), artinya termuat di dalam suatu ruang fungsi tertentu.

Sama dengan regresi parametrik, jumlah terboboti dari pengamatan y digunakan untuk memperoleh nilai fit. Pada kssus regresi dengan satu variabel penjelas, pengamatan dengan informasi di sekitar $f(x_0)$ pada lokasi x_i akan dekat ke x_0. Oleh karena itu, fungsi turun pada selang $(x_0,\ x_i)$ menentukan bobot y_i. Titik yang dekat ke x_0 memiliki bobot lebih tinggi dibanding yang jauh dari x_0.

Salah satu pendekatan regresi non parametrik untuk memperoleh dugaan kurva regresi f adalah *smoothing spline*. Masalah utama pada saat menduga fungsi regresi menggunakan *smoothing spline* adalah memilih dan menentukan parameter pemulus (Cantoni & Hastie, 2000). Beberapa pendekatan telah dilakukan untuk memilih parameter pemulus yang optimal ini, yaitu subyektif dan otomatis. Makalah ini menyajikan pendekatan otomatis yang menentukan parameter pemulus pada *smoothing spline* berdasarkan data hasil simulasi seperti yang dilakukan oleh Faraway(2006) yaitu $y = f(x) = sin^3(2\pi x^3) + \epsilon$ dimana $\epsilon \sim$N(0,0.2).

2 Regresi Spline

Pemulusan merupakan salah satu metode yang digunakan dalam analisis data non parametrik. Tujuan dari pemulusan adalah untuk memperkecil keragaman dari data yang tidak memiliki pengaruh sehingga ciri-ciri dari data akan tampak lebih jelas. Pemulusan telah menjadi teknik umum di dalam metode-metode non parametrik yang digunakan untuk menduga fungsi.

Misalnya diberikan data pengamatan Y_1, ..., Y_n pada titik-titik tetap yang berurut x_1,...,x_n (untuk memudahkan, misalkan $0 \leq x_1 < ... < x_n \leq 1$), dan data ini memiliki model hubungan seperti pada persamaan (1) yang terdefinisi pada selang $x \in [0, 1]$. Tujuan dari analisis data adalah menduga fungsi regresi f pada tiap titik x.

Salah satu model regresi dengan pendekatan non parametrik yang dapat digunakan untuk menduga kurva regresi adalah regresi spline. Regresi spline adalah suatu pendekatan ke arah plot data dengan tetap memperhitungkan kemulusan kurva. Spline merupakan model polinomial yang tersegmen atau terbagi dimana sifat segmen inilah yang memberikan fleksibelitas yang lebih baik dibanding model polinomial biasa. Sifat ini memungkinkan model regresi spline menyesuaikan diri secara efektif terhadap karakteristik lokal dari data. Penggunaan spline difokuskan kepada adanya perilaku atau pola data, yang pada daerah tertentu mempunyai karakteristik yang berbeda dengan daerah lain.

Fungsi spline berorde ke-m dengan satu variabel penjelas adalah sembarang fungsi yang secara umum dapat disajikan dalam bentuk

$$f(x) = \beta_0 + \sum_{r=1}^{m-1} \beta_r X_r + \sum_{k=1}^{s} \beta_{(m-1).k} (X - K_k)_+^{m-1} \tag{2}$$

dengan fungsi *truncated* adalah

$$(X - K_k)_+^{m-1} = \begin{cases} (X - K_k)^{m-1} & ; X \geq K_k \\ 0 & ; X < K_k \end{cases}$$

dimana K_k adalah knot ke-k dari variabel X, k=1,2,...,s, dan s adalah banyaknya knot.

Dari bentuk matematis fungsi spline pada persamaan 2 menunjukkan bahwa spline merupakan model polinomial yang tersegmen (*piecewise polynomial*), tetapi spline masih bersifat kontinu pada knot-knotnya. Knot diartikan sebagai suatu titik fokus dalam fungsi spline sedemikian sehingga kurva yang dibentuk tersegmen pada titik tersebut. Oleh karena itu, spline adalah potongan polinomial mulus yang masih memungkinkan memiliki sifat tersegmen.

2.1 Fungsi Pemulus Spline

Berdasarkan fungsi regresi pada persamaan 1 dimana f adalah fungsi pemulus yang tidak spesifik dan $E[\epsilon] = 0$, Fahrmeir & Tuhtz (1994) menduga kurva pemulus $\widehat{f}(x_i)$ dapat diperoleh berdasarkan data amatan, yakni pasangan peubah penjelas dan peubah responnya. Penduga fungsi pemulus merupakan penduga fungsi yang mampu memetakan data dengan baik serta mempunyai ragam galat yang kecil. Oleh karena itu, dengan menggunakan data amatan sebanyak n, maka $f(x_i)$ diperoleh dengan meminimumkan fungsi *Penalized Least Square* (PLS), yaitu

$$\text{PLS} = \underbrace{\sum_{i=1}^{n} (y_i - f(x_i))^2}_{\text{(a)}} + \underbrace{\lambda \int [f"(x)]^2 dx}_{\text{(b)}} \tag{3}$$

dimana bagian (a) merupakan jumlah kuadrat sisaan atau fungsi jarak antara data dan dugaan, bagian (b) merupakan *roughness penalty*, yaitu ukuran kemulusan kurva dalam memetakan data, dan $0 < \lambda < 1$ adalah parameter pemulus, yaitu pengontrol keseimbangan antara kecocokan terhadap data (*Goodness-of-Fit*) dan kemulusan kurva (*penalty*). Apabila nilai λ besar mendekati 1 maka akan memberikan bobot *penalty* (kemulusan) yang besar dan mempunyai ragam yang kecil.

Dengan data amatan sebanyak n maka untuk menyelesaikan persamaan 3 digunakan algoritme numerik dari persamaan matrik berikut:

$$\text{PLS}(f) = (\mathbf{Y} - \mathbf{f})^T(\mathbf{Y} - \mathbf{f}) + \lambda \mathbf{f}^T\mathbf{K}\mathbf{f}$$

dimana $\mathbf{Y}^T = (y_1, ..., y_n)$, $\mathbf{f}^T = (f(x_1), ..., f(x_n)$, sedangkan $\mathbf{K}$ adalah matrik *penalty* yang mempunyai struktur spesifik, yaitu

$$\mathbf{K} = \mathbf{D}^T\mathbf{C}^{-1}\mathbf{D}$$

dimana $\mathbf{D} = (d_{ij})$ adalah matrik *upper triagonal* $[n-2, n]$ yang memiliki struktur sebagai berikut:

$$\begin{pmatrix} \lambda_1^{-1} & -(\lambda_1^{-1} + \lambda_2^{-1}) & \lambda_2^{-1} & 0 & \dots & 0 \\ 0 & \lambda_2^{-1} & -(\lambda_2^{-1} + \lambda_3^{-1}) & \lambda_3^{-1} & \dots & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & \dots & \dots & \lambda_{n-2}^{-1} & -(\lambda_{n-2}^{-1} + \lambda_{n-1}^{-1}) & \lambda_{n-1}^{-1} \end{pmatrix}$$

dan $\mathbf{C} = (c_{ij})$ adalah matrik *symmetric tridiagonal* $[n-2, n-2]$ yang memiliki struktur sebagai berikut:

$$\begin{pmatrix} 2(\lambda_1 + \lambda_2) & \lambda_2 & 0 & 0 & \dots & 0 \\ \lambda_2 & \dots & \dots & \dots & \dots & \dots \\ \vdots & \vdots & \vdots & \vdots & \ddots & \lambda_{n-2} \\ 0 & \dots & \dots & \dots & \lambda_{n-2} & 2(\lambda_{n-2} + \lambda_{n-1}) \end{pmatrix}$$

2.2 Pemilihan Parameter Pemulus Optimal

Eubank (1988) menyebutkan bahwa ukuran kinerja atas penduga kurva regresi dapat ditentukan dari rataan kuadrat sisaan (MSE), fungsi *loss* dan fungsi resiko, serta *Generalized Cross-Validation* (GCV). Rataan kuadrat sisaan merupakan ukuran kinerja yang paling sederhana, yaitu

$$\begin{aligned} \text{MSE}(\lambda) &= \frac{1}{n}(\mathbf{Y} - \mathbf{f})^T(\mathbf{Y} - \mathbf{f}), \text{ atau} \\ \text{MSE}(\lambda) &= \frac{1}{n}\sum_{i=1}^{n}(y_i - f(x_i))^2 \end{aligned} \tag{4}$$

Ukuran kinerja lainnya adalah fungsi kerugian dan fungsi resiko. Pada persamaan 3, pemilihan λ yang optimal sangat penting untuk mendapatkan model penduga kurva regresi yang baik. Oleh karena itu dipilih λ yang meminimumkan fungsi *loss* $L(\lambda)$ berikut:

$$L(\lambda) = \frac{1}{n}\sum_{i=1}^{n}(y_i - f(x_i))^2$$

dimana $L(\lambda)$ mewakili ukuran kedekatan dari y_i dan $f(x_i)$. Akan tetapi dalam regresi non parametrik, fungsi $f(x_i)$ tidak diketahui sehingga dari persamaan 3 ditentukan turunan pertama terhadap f dan diperoleh bentuk pemulus linier:

$$\begin{aligned} \widehat{\mathbf{f}} &= (\mathbf{I} + \lambda\mathbf{K})^{-1}\mathbf{Y} \\ &= \mathbf{SY} \end{aligned} \tag{5}$$

Fungsi resiko, $R(\lambda)$ adalah nilai harapan dari fungsi kerugian, yaitu

$$R(\lambda) = E[L(\lambda)] = E\left[\frac{1}{n}\sum_{i=1}^{n}(y_i - f(x_i))^2\right]$$

Untuk mengevaluasi f sebagai penduga bagi pengamatan baru y^* dimana $y_i^* = f(x_i) + \epsilon_i^*$ maka diperoleh $P(\lambda)$ sebagai fungsi resiko penduga, yaitu

$$\begin{aligned} P(\lambda) &= E\left[\frac{1}{n}\sum_{i=1}^{n}(y_i^* - f(x_i))^2\right] \\ &= E\left[\frac{1}{n}\sum_{i=1}^{n}((f(x_i) + \epsilon_i^*) - f(x_i))^2\right] \\ &= R(\lambda) + \sigma^2 \end{aligned}$$

Eubank (1988) menunjukkan bahwa ada hubungan antara $P(\lambda)$ dengan MSE(λ) yaitu

$$\widehat{P}(\lambda) = \text{MSE}(\lambda) + 2\widehat{\sigma}^2\frac{tr(\mathbf{H})}{n}$$

dimana $\mathbf{H}$ adalah $\mathbf{X}(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}$.

Generalized Cross-Validation (GCV) merupakan modifikasi dari *Cross-Validation* (CV) dimana

$$\text{CV}(\lambda) = \frac{1}{n}\sum_{i=1}^{n}\left[\frac{y_i - f(x_i)}{1 - h_{ii}}\right] \tag{6}$$

dan h_{ii} adalah elemen diagonal ke-i dari matrik $\mathbf{H}$. Dengan demikian, GCV adalah

$$\begin{aligned} \text{GCV}(\lambda) &= \frac{1}{n}\frac{\sum_{i=1}^{n}(y_i - f(x_i))^2}{[1 - n^{-1}tr(\mathbf{H})]^2} \\ &= \frac{\text{MSE}(\lambda)}{[n^{-1}tr(\mathbf{I} - \mathbf{H})]^2} \end{aligned} \tag{7}$$

Ketiga kriteria pengujian MSE(λ), $P(\lambda)$, dan GCV(λ) diharapkan memiliki nilai yang minimum sehingga model regresi spline dapat dikatakan memiliki λ yang optimal.

2.3 Derajat Bebas Smoothing Spline

Cantoni & Hastie (2000) menyatakan bahwa pemilihan faktor pemulus dalam regresi non parametrik dapat dilakukan melalui derajat bebas. Seperti telah diketahui bahwa *smoothing spline* dihitung untuk menyesuaikan vektor $\mathbf{y} = (y_1, ..., y_n)$ dengan model fit $\widehat{\mathbf{y}} = \widehat{\mathbf{f}} = \mathbf{SY}$ seperti yang tercantum pada persamaan 5. Parameter λ mengontrol kemulusan kurva regresi yang dapat juga didefinisikan melalui derajat bebas. Hastie & Tibshirani (1990) dalam Cantoni & Hastie (2000) mendefinisikan derajat bebas efektif pada *smoothing spline* sebagai

$$\text{df} = tr(\mathbf{S}) = \sum_{i=1}^{n}\frac{1}{1 + \lambda d_i} \tag{8}$$

dimana d_i adalah akar ciri dari matrik $\mathbf{K}$ (lihat persamaan 5). Persamaan 8 ini menunjukkan adanya hubungan terbalik antara λ dengan derajat bebas df.

3 Program R untuk Smoothing Spline

Di dalam program R, dengan sendirinya (*by default*) menggunakan *cross-validation* untuk memilih parameter pemulus. Format umum dari program R untuk menduga kurva regresi dengan pemulus spline adalah

```
smooth.spline(x, y = NULL, w = NULL, df, spar = NULL,
              cv = FALSE, all.knots = FALSE, nknots = NULL,
              keep.data = TRUE, df.offset = 0, penalty = 1,
              control.spar = list())
```

dimana	x	vektor dari variabel penjelas atau dapat juga berupa pasangan koordinat x dan y
	y	vektor variabel respon (tidak digunakan jika koordinat titik dinyatakan dalam x)
	w	vektor pembobot (*default*=1)
	df	derajat bebas seperti pada persamaan 8
	spar	parameter pemulus (0,1)
	cv	kriteria yang digunakan (*TRUE* berarti menggunakan CV seperti pada persamaan 6, *FALSE* berarti menggunakan GCV seperti pada persamaan 7)
	all.knot	jika *TRUE* maka semua titik x yang berbeda dianggap sebagai *knot*, dan jika *FALSE* (*default*) maka selang nilai $x[\ ]$ digunakan sebagai *knot* sesuai dengan banyaknya *knot* yang ditentukan pada *nknots*
	nknots	bilangan bulat yang menunjukkan banyaknya *knot* yang digunakan pada saat *all.knot* = *FALSE*
	keep.data	nilai logika untuk menentukan apakah data input diberikan pada hasil, jika *TRUE* (*default*) maka nilai fit dan residual tersedia dalam hasil
	df.offset	mengijinkan derajat bebas dinaikkan dalam kriteria GCV
	penalty	koefisien *penalty* untuk derajat bebas pada kriteria GCV
	control.spar	pilihan yang dapat diberikan pada saat menghitung *spar*, yaitu *low* adalah batas bawah (*default*=-1.5), *high* adalah batas atas (*default*=+1.5), *tol* adalah nilai toleransi mutlak yang digunakan (*default*=1E-4), *eps* adalah nilai toleransi relatif yang digunakan (*default*=0.0024), *trace* menentukan apakah setiap nilai hasil iterasi akan disimpan, dan *maxit* adalah bilangan bulat yang menunjukkan banyaknya iterasi maksimum (*default*=500).

Vektor $\mathbf{x}$ sedikitnya memiliki empat nilai yang berbeda (setelah mengalami pembulatan hingga 6 digit). Hubungan antara *spar* dan λ sebagai parameter pemulus adalah $\lambda = r * 256^{(3*spar-1)}$ dimana

$$r = \frac{tr(\mathbf{X}^T\mathbf{W}\mathbf{X})}{tr(\mathbf{\Sigma})}$$

$\mathbf{W}$ adalah matrik diagonal terboboti, dan $\mathbf{\Sigma} = (\sigma_{ij}) = \int B_i"(t)B_j''(t)dt$. Dari representasi B-spline $\mathbf{f} = \mathbf{Xc}$ dimana $\mathbf{c}$ adalah vektor koefisien spline, dan *penalized log likelihood* dapat dituliskan sebagai

$$L = (\mathbf{y} - \mathbf{f})^T\mathbf{W}(\mathbf{y} - \mathbf{f}) + \lambda\mathbf{c}^T\mathbf{\Sigma}\mathbf{c}$$

maka $\mathbf{c}$ merupakan jawaban dari persamaan

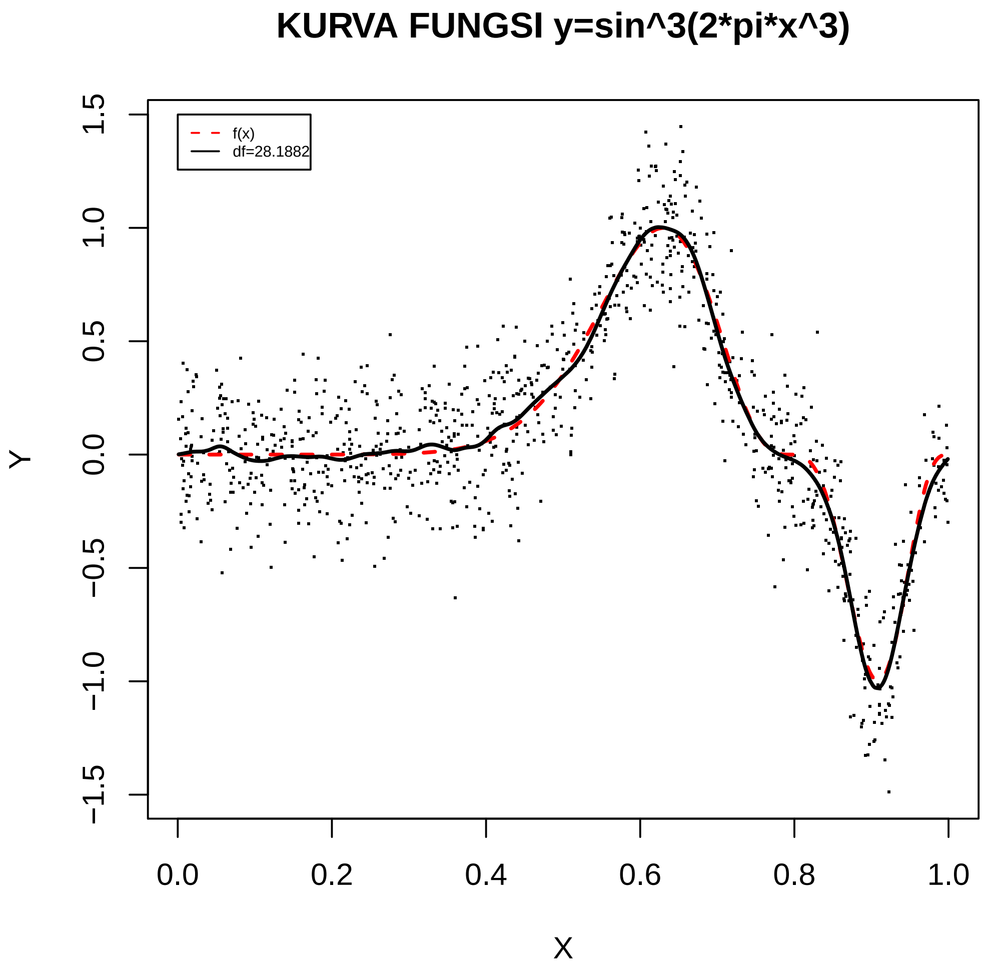
$$\left(\mathbf{X}^T\mathbf{W}\mathbf{X} + \lambda\mathbf{\Sigma}\right)\mathbf{c} = \mathbf{X}^T\mathbf{W}\mathbf{y}$$

Jika nilai *spar* dikosongkan dalam program R untuk *smoothing spline*, maka kriteria pemulusan menggunakan df. Jika keduanya *spar* dan *df* dikosongkan, maka program R menggunakan CV atau GCV untuk menentukan λ.

Untuk menguji parameter pemulus dalam pendugaan kurva regresi *smoothing spline*, dilakukan simulasi membangkitkan data seperti yang dilakukan oleh Faraway(2006) yaitu $y = f(x) = sin^3(2\pi x^3) + \epsilon$ dimana $\epsilon \sim$N(0,0.2) seperti di bawah ini. Data disimpan di dalam file bernama **sinus.dat**, sedangkan plot data dan fungsi sesungguhnya tercantum pada Gambar 1.

```
X  <- runif(1000,0,1)
t  <- sin(2*pi*X*X*X)
E  <- rnorm(1000,0,0.2)    ## error menyebar normal
F  <- t*t*t                ## fungsi sebenarnya f(x)
Y  <- F+E                  ## f(x)+error
dt <- data.frame(X,Y,F,E)
write.table(dt, file="sinus.dat", sep=" ",
   row.names = FALSE, col.names=TRUE)
```

Pendugaan kurva regresi dilakukan dengan menggunakan fungsi pemulus spline **smooth.spline** dan diperoleh hasil parameter pemulus seperti tercantum pada Tabel 1, yang juga menunjukkan hubungan antara nilai λ, *spar*, dan df. Sesuai dengan penjelasan program R untuk fungsi **smooth.spline**, nilai *spar* sejalan dengan nilai λ, tetapi memiliki hubungan terbalik dengan nilai *df*. Semakin besar nilai *df* maka semakin kecil nilai λ dan juga nilai *spar*. Tabel 1 juga menyajikan nilai parameter pemulus yang optimal yang diperoleh untuk menduga kurva regresi, yaitu *spar*=0.6062, λ=2.7359e-05, dan df=28.1882. Kurva dugaan regresi dengan menggunakan **smooth.spline** pada nilai parameter pemulus optimal disajikan pada Gambar 1. Kurva dugaan regresi yang dihasilkan sangat baik karena mendekati fungsi $f(x)$ yang sesungguhnya.



Gambar 1: Kurva fungsi $f(x)$ dan dugaannya menggunakan **smooth.spline**

Jika parameter pemulus diubah dimana nilai *df* diperbesar atau diperkecil dari nilai optimal (hal ini juga berarti nilai *spar* diperkecil atau diperbesar) maka kurva dugaan regresi spline semakin menjauh dari fungsi yang sebenarnya. Hal ini juga ditunjukkan oleh nilai GCV yang semakin besar dibanding nilai GCV pada saat parameter pemulus yang optimal sebesar 0.0411 (Tabel 1 dan Gambar 2).

Tabel 1: Nilai-nilai parameter pemulus yang dicobakan

Parameter	*df* = 2	*df* = 10	*Optimal*	*df* = 40	*df* = 100
spar	1.4999	0.8752	0.6062	0.5177	0.2613
λ	78.3652	0.0024	2.7359e-05	6.2673e-06	8.8179e-08
df	2.0297	9.9988	28.1882	40.0048	100.0103
Penalized	239.7506	49.355	38.7669	38.0589	34.9777
GCV	0.2407	0.0504	0.0411	0.0413	0.0432

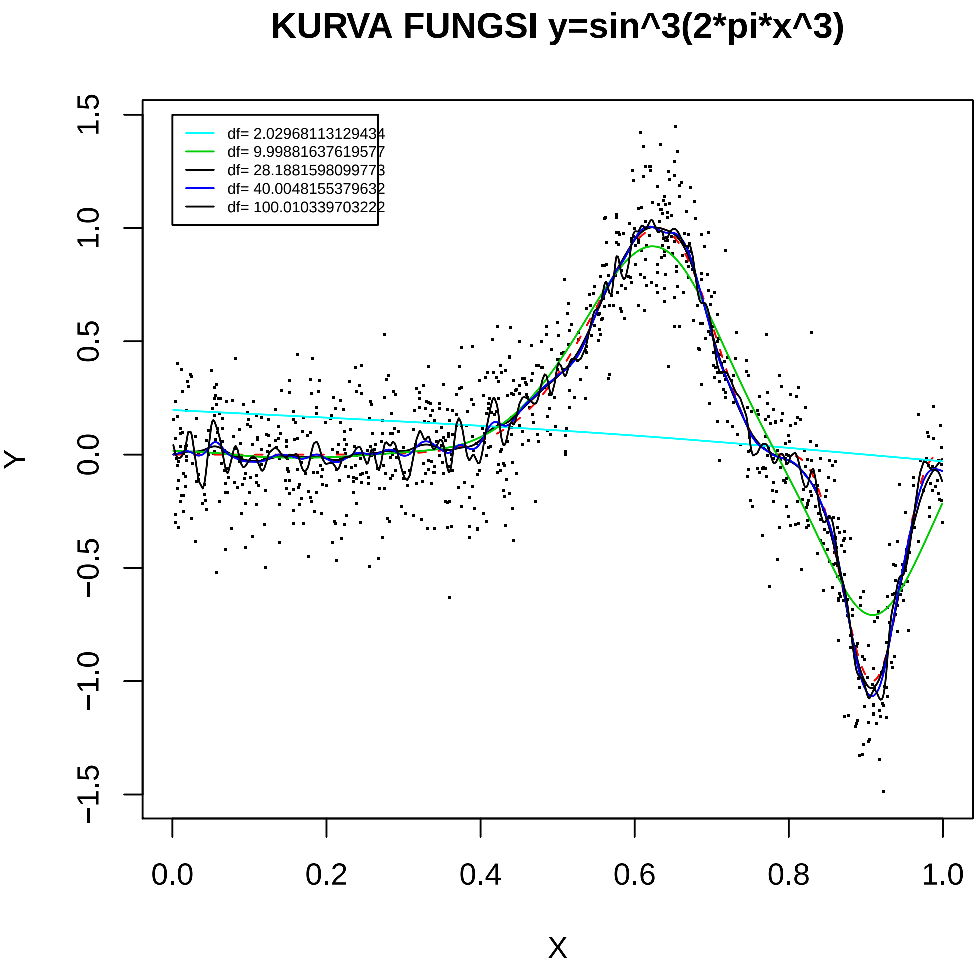
Program R untuk menduga kurva menggunakan *smoothing spline* pada nilai parameter pemulus optimal sebagai berikut:

```
d <- read.table(file="sinus.dat", header=T)
plot(Y~X, d, main="KURVA FUNGSI y=sin^3(2*pi*x^3)", pch=".")
j<-order(d$X)
lines(d$X[j],d$F[j], lty=2, lwd=2, col=2)
#
# parameter pemulus optimal
#
spl <- smooth.spline(d$X, d$Y)
df1 <- spl$df
lines(smooth.spline(d$X, d$Y, df=df1), lwd=2, col=1)
warna <- c(1,2)
clty <- c(2,1)
ket <- c("f(x)", "df=28.1882")
legend(0,1.5,legend = ket, col = warna,
       lty = clty, cex = .5, y.intersp = 1)
```

4 Kesimpulan

Pada makalah ini dibahas dan dicobakan pendugaan kurva regresi non parametrik dengan menggunakan metode *smoothing spline*. Nilai parameter pemulus memegang peranan penting dalam menentukan baik dan tidaknya kurva dugaan regresi yang dihasilkan. Besarnya nilai parameter pemulus sangat tergantung pada pola dan karakteristik data. Untuk data simulasi yang dicobakan, program R mampu mendapatkan nilai parameter pemulus yang optimal dan kurva dugaan yang dihasilkan sangat baik yaitu pada nilai-nilai *spar*=0.6062, λ=2.7359e-05, dan *df*=28.1882.

Terdapat hubungan yang erat antara parameter pemulus *spar*, λ, dan *df*. Sesuai dengan hasil percobaan dan tinjauan matematis dapat ditunjukkan bahwa semakin besar nilai *df* maka semakin kecil nilai *spar* dan λ, dan juga sebaliknya.



Gambar 2: Kurva fungsi $f(x)$ dan dugaannya pada beberapa nilai parameter pemulus

5 Daftar Pustaka

Cantoni, E & T.Hastie. 2000. ***Degrees of Freedom Tests for Smoothing Splines***. Statistics Department, Stanford University.

Eubank, R. 1988. ***Spline Smoothing and Nonparametric Regression***. Marcel Dekker. New York.

Faraway, J.J. 2006. ***Extending the Liniear Model with R***. Chapman & Hall, London.

Fahrmeir, L. & Tuhtz, G. 1994. ***Multivariate Statistical Modelling Based on Generalized Linier Models***. Springer-Verlag. New York

Hardle,W. 1990. ***Smoothing Techniques With Implementation in S***. Springer-Verlag. New York.

Koenker, R; Pin Ng; & S. Portnoy. 1993. ***Quantile Smoothing Splines***. University of Illinois.

Sasmitoadi, D. 2005. **Kajian Penggunaan Knot dan Orde pada Regresi Spline**. Jurusan Matematika, FMIPA Universitas Brawijaya.

Silverman, B.W. 1986. ***Density Estimation for Statistics and Data Analysis***. Chapman & Hall, London.

Tavio; Budiantara, I.N. & Kusuma, B. 2008. ***Spline Nonparametric Regression Analysis of Stress-Strain Curve of Confined Concrete***. Civil Engineering Dimension, Vol. 10, No. 1, March 2008, 14-27.

Venables, W.N & D.M. Smith. 2009. ***An Introduction to R: A Programming Environment for Data Analysis and Graphics, Version 2.10.1***. The R Development Core Team.