Statistika, Vol. 8 No. 1, 47 – 54
Mei 2008

Bootstrap Pada Regresi Linear dan *Spline Truncated*

HARISON DARMAWI[1] DAN BAMBANG WIDJANARKO OTOK[2]

[1] Tenaga Pengajar di Jurusan Matematika UNRI, Pekanbaru
e-mail: son_msi@yahoo.co.id
[2] Tenaga Pengajar di Jurusan Statistika, ITS, Surabaya
e-mail: bambang_wo@statistika.its.ac.id; otok_bw@yahoo.com

ABSTRAK

Pendekatan parametrik mengasumsikan bentuk model sudah ditentukan. Apabila tidak ada informasi apapun tentang bentuk kurva, maka pendekatan yang digunakan adalah pendekatan nonparametrik, salah satunya spline truncated. Karena pendekatan tidak tergantung pada asumsi bentuk kurva tertentu, sehingga memberikan fleksibilitas yang lebih besar. Tujuan penelitian ini adalah mengkaji bootstrap pada regresi linear dan regresi spline (truncated spline) dengan kajian simulasi. Hasil penelitian menunjukkan bahwa Fungsi optimal terjadi pada variansi yang kecil untuk sembarang pengamatan. Nilai MSE pada kurva truncated spline lebih kecil dibanding dengan regresi linear pada semua fungsi. Hal ini dapat diartikan bahwa kurva truncated spline lebih baik dibanding dengan regresi linear. Hal ini dapat dilihat dari simulasi estimator $g(t) = \sin(5\pi t)$ dan $g(t) = 5e^{-5t}$, truncated spline memberikan berbagai nilai titik knot, sehingga nilai MSE kecil dibanding regresi linear. Secara keseluruhan dengan kriteria MSE, Spline Truncated sesudah di bootstrap lebih baik dibanding dengan pendekatan regresi dan spline truncated.

Kata Kunci: Regresi Linear, Spline Truncated, Bootstrap, MSE

1. Pendahuluan

Analisis regresi memperlihatkan hubungan dan pengaruh variabel prediktor terhadap variabel respon. Misalnya y adalah variabel respon dan t adalah variabel prediktor, untuk n buah pengamatan, secara umum hubungan antara **y** dan **x** dapat ditulis sebagai berikut :

$$y_i = f(x_i) + \varepsilon_i \ , \ i = 1, 2, \ldots, n$$

dengan ε adalah sesatan random dan $f(x_i)$ merupakan kurva regresi.

Jika kurva regresi merupakan model parametrik maka disebut sebagai regresi parametrik dan apabila model yang diasumsikan ini benar, maka pendugaan parametrik sangat efisien, tetapi jika tidak, menyebabkan interpretasi data yang menyesatkan (Hardle, 1990). Pendekatan parametrik mengasumsikan bentuk model sudah ditentukan. Apabila tidak ada informasi apapun tentang bentuk $f(x_i)$, maka pendekatan yang digunakan adalah pendekatan nonparametrik. Karena pendekatan tidak tergantung pada asumsi bentuk kurva tertentu, sehingga memberikan fleksibelitas yang lebih besar. Dalam hal ini diasumsikan $f(x_i)$ termuat dalam ruang fungsi (Eubank, 1988).

Ada beberapa teknik estimasi dalam regresi nonparametrik antara lain pendekatan histogram, estimator *spline*, estimator kernel, estimator deret orthogonal, analisis wavelet dan lain-lain. Pendekatan estimator *spline* ada bermacam-macam antara lain *spline* original, *spline* type M, *spline* relaxed, *spline* terbobot dan lain-lain. Pendekatan *spline* mempunyai suatu basis fungsi. Basis fungsi yang biasa dipakai antara lain *spline truncated* dan B-*spline* (Botella and Shariff, 2003).

Wahba (1990) menunjukkan bahwa *spline* memiliki sifat-sifat statistik yang berguna untuk menganalisis hubungan dalam regresi. *Spline* adalah salah satu jenis *piecewise* polinomial, yaitu polinomial yang memiliki sifat tersegmen. Sifat tersegmen ini memberikan fleksibilitas lebih dari polinomial biasa, sehingga memungkinkan untuk menyesuaikan diri secara lebih efektif terhadap karakteristik lokal suatu fungsi atau data.

Dalam proses inferensi statistik, yang digunakan untuk melihat seberapa akurat ringkasan suatu data digunakan pendekatan *bootstrap*. Mooney (1997) melakukan perbandingan antara metode kuadrat terkecil dan metode *bootstrap* dengan hasil bahwa jika asumsi kenormalan terpenuhi maka kedua metode akan memberikan hasil yang relatif sama, tetapi jika asumsi kenormalan tidak terpenuhi kedua hasil memberikan hasil yang berbeda. Sedangkan Stone (1990) menunjukkan bahwa metode *bootstrap* merupakan perkembangan yang relatif baru secara teoritis, sehingga penerapannya dapat dimengerti. Permasalahan dalam metode *bootstrap* terletak pada sampel, salah satu untuk melihat keterbatasan dan kelebihannya memperlakukan sampel sebagai populasi dan menggunakan percobaan monte carlo untuk mengkonstruksikan estimator empiris distribusi sampling statistik.

Pada penelitian ini dibahas mengenai estimasi regresi linear dan *truncated spline* dan melakukan simulasi untuk membandingkan MSE regresi linear dan *spline truncated* sebelum dan sesudah di *bootstrap*.

2. Regresi Linear

Perubahan nilai suatu variabel tidak selalu terjadi dengan sendirinya, namun perubahan nilai variabel itu dapat pula disebabkan oleh berubahnya variabel lain yang berhubungan dengan variabel tersebut. Untuk mengetahui pola perubahan nilai suatu variabel yang disebabkan oleh variabel lain diperlukan alat analisis yang memungkinkan untuk membuat prediksi nilai variabel tersebut pada nilai tertentu variabel yang mempengaruhi. Prinsip dasar yang harus dipenuhi dalam membangun suatu persamaan regresi adalah bahwa antara variabel dependen dengan variabel independennya mempunyai sifat hubungan sebab akibat, baik yang didasarkan pada teori, hasil penelitian sebelumnya, ataupun yang didasarkan pada penjelasan logis tertentu.

Dalam persamaan regresi jika hanya mengandung satu variabel independent disebut Regresi Linear Sederhana dan jika dalam model regresi tersebut mengandung lebih dari satu variabel independent disebut Regresi Linear Berganda. Tujuan dan manfaat dalam analisis regresi antara lain : mendapatkan pola hubungan secara matematis antara variabel X dan Y, mengetahui besarnya perubahan variabel X terhadap Y dan memprediksi Y jika nilai X diketahui.

Regresi Linear Berganda merupakan perluasan dari regresi linear sederhana, yang bertujuan untuk mencari pola hubungan yang dapat digambarkan secara matematis antara satu variabel respon dengan beberapa variabel prediktor secara serentak.

Jika terdapat n pengamatan untuk variabel (Y) dan variabel bebas (X_i), $i = 1, 2, \ldots, n$ maka pola hubungan secara umum dapat ditulis sebagai berikut :

$$Y_i = \beta_0 + \beta_1 X_{1i} + \ldots + \beta_k X_{ki} + \varepsilon_i \qquad (1)$$

dengan :

Y_i = variabel respon

X_i = variabel prediktor/fixed

β_i = parameter

ε_i = unsur gangguan yang diasumsikan identik independen dan berdistribusi normal atau ε_i ~ IIDN$(0, \sigma^2)$

3. Spline Truncated

Spline merupakan potongan polinomial (piecewise polynomial), yaitu polynomial yang memiliki sifat tersegmen yang kontinu. Sifat ini yang memberikan fleksibilitas lebih daripada polinomial biasa, sehingga memungkinkan untuk menyesuaikan diri secara efektif terhadap karakteristik lokal dari fungsi atau data (Eubank,1998).

Spline adalah potongan polynomial order r, titik bersama dari potongan-potongan tersebut biasanya disebut dengan knots. *Spline* orde r dengan knots pada $\xi_1,\cdots,\xi_k$ didefinisikan suatu fungsi *s* dengan bentuk:

$$S(t)=\sum_{i=0}^{r-1}\theta_l t^i+\sum_{i=1}^{k}\delta_i (t-\xi_i)_+^{\,r-1} \tag{2}$$

Untuk suatu koefisien riil $\theta_0,\theta_1,\cdots,\theta_{r-1}$, $\xi_1,\cdots,\xi_k$ dan

$$(t-\xi_i)_+^{r-1}=\begin{cases}(t-\xi_i)^{r-1} & \text{jika } (t-\xi_i)\geq 0\\ 0 & \text{jika } (t-\xi_i)<0\end{cases} \tag{3}$$

Spline pada Persamaan (2) mempunyai sifat sebagai berikut:

(i) s merupakan potongan polynomial derajat r-1 pada setiap subinter $[\xi_i,\xi_{i+1}]$

(ii) s mempunyai turunan ke (r-2) yang kontinyu.

(iii) S mempunyai turunan ke (r-1) yang merupakan fungsi tangga dengan titik-titik lompatan pada $\xi_1,...,\xi_k$.

Apabila $S^r=(\xi_1,\cdots,\xi_k)$ menyatakan himpunan semua fungsi yang berbentuk (2), maka $S^r=(\xi_1,\cdots,\xi_k)$ adalah suatu ruang vektor. Apabila definisikan suatu *spline* alami berorde r = 2m dengan tititk-titik knots pada $(t_1,...,t_n)$ adalah *spline* yang memenuhi sifat (i)-(iii) juga memenuhi sifat (iv) berikut :

(iv) s adalah polinomial derajat m-1 diluar interval $[t_1,t_n]$ maka s memenuhi syarat batas alami (natural boundary condition), yaitu $s^{(j)}(a)=s^{(j)}(b)=0, j=m,...,2m-1$

Selanjutnya didefinisikan $NS^{2m}(t_1,...,t_n)$ adalah himpunan semua *spline* alami berorde-2m dengan titik knots pada $(t_1,...,t_n)$. adalah subruang dari ruang vektor $S^{2m}(t_1,...,t_n)$. Jika dalam persamaan (2) diambil nilai m= 2, maka diperoleh sebuah *spline* yang disebut dengan *spline* kubik. Jika diberikan bilangan real $t_1,...,t_n$ pada suatu interval [a,b] yang memenuhi $a<t_1<...<t_n<b$. Fungsi f terdefinisi dalam interval [a,b] dikatakan *Spline* kubik jika memenuhi syarat berikut:

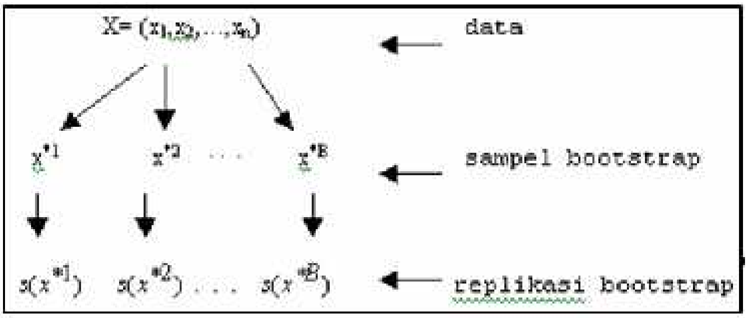
1. pada setiap interval $(a,t_1),(t_1,t_2),...,(t_n,b)$, f adalah polinomial kubik.
2. turunan pertama dan kedua dari f kontinyu pada setiap $t_i\in[a,b]$ dengan t_i titik *knot*.

Spline kubik pada interval [a,b] dikatakan *Spline* kubik alami jika turunan kedua dan ketiganya pada titik a dan b adalah nol (Green dan Silverman, 1994).

4. Bootstrap

Bootstrap suatu metode yang memungkinkan untuk mendapatkan model secara berulang-ulang dari hanya satu kumpulan data dalam ukuran sampel yang kecil. Sehingga didapat estimasi-estimasi parameter model untuk setiap pengulangan yang dilakukan (stabil) dengan standar error yang lebih rendah.

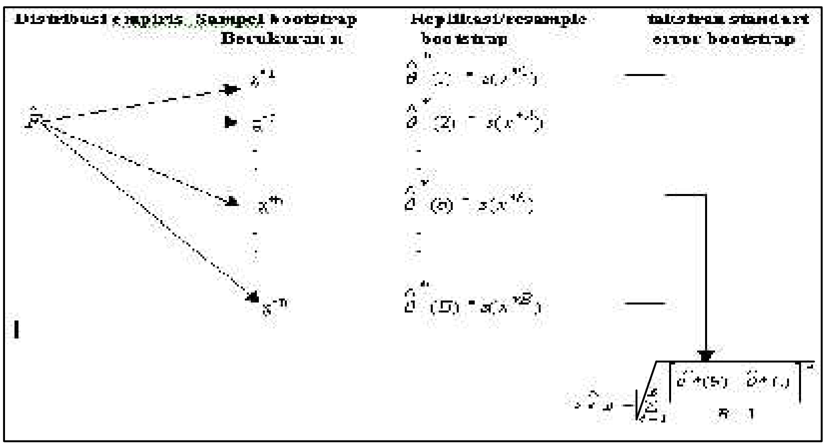
Metode *bootstrap* pertama kali dipelajari oleh B. Efron (1979). Metode *bootstrap* merupakan suatu metode penaksiran nonparametrik yang dapat menaksir parameter-parameter dari suatu distribusi, variansi dari sampel median, serta dapat menaksir tingkat kesalahan (*error*). Pada metode *bootstrap* dilakukan pengambilan sampel dengan pengembalian (*resamplimg with replacement*) dari sampel data.



Gambar 1. Skema Proses Bootstrap

Gambar 1., adalah skema proses *bootstrap* untuk menaksir standar error dari statistik s(x). Sebuah sampel asli (x) berukuran n yang terdiri dari $x_1, x_2, \ldots, x_n$. Kemudian sebanyak B sampel *bootstrap* berukuran n diambil (*resampel*) dengan pengembalian dari sampel asli (x), sehingga didapatkan himpunan data *bootstrap* ($x_1{}^*, x_2{}^*, \ldots, x_B{}^*$) yang terdiri dari anggota data asli, beberapa mungkin tidak akan muncul sama sekali, atau muncul hanya satu atau dua kali, tergantung randomisasi. Tanda $(*)$ menunjukkan x* bukan data aktual tetapi hasil dari randomisasi atau *resampel* dari x. tiap sampel *bootstrap* merupakan sampel acak yang saling independent. Replikasi *bootstrap* $s(x^{*1}), s(x^{*2}), \ldots, s(x^{*B})$ diperoleh dari menghitung nilai statistik s(x) pada masing-masing sampel *bootstrap*. Sehingga standar error s(x) adalah deviasi standar dari $s(x^{*1}), s(x^{*2}), \ldots, s(x^{*B})$ atau $\hat{\theta}_1^*, \hat{\theta}_2^*, \ldots, \hat{\theta}_B^*$.

Jumlah replikasi *bootstrap* (B) yang digunakan untuk mendapatkan taksiran standar error yang cukup baik biasanya pada selang 50 sampai 200 (Efron dan Tibshirani, 1993).



Gambar 2. Algoritma bootstrap untuk perhitungan standar error

Gambar 2 menunjukkan algoritma *bootstrap* untuk perhitungan standar error dari $\hat{\theta}$. Taksiran *bootstrap* untuk standar error dari $\hat{\theta}$, $se_F(\hat{\theta})$, adalah taksiran plug in yang menggunakan fungsi distribusi empiris $\hat{F}$ pada suatu distribusi yang tidak diketahui. Sehingga taksiran *bootstrap* $se_F(\hat{\theta})$ didefinisikan sebagai $se_{\hat{F}}(\hat{\theta})$.

5. Metodologi Penelitian

Data dalam penelitian ini digunakan data simulasi dengan bantuan program MINITAB dan SPLUS. Adapun langkah-langkah penelitian sebagai berikut:

1. Membangun model $y_i = g(t_i) + \varepsilon_i$, i=1,2,...,n dengan n=50, 100, 250. Dimana ε_i dibang-kitkan dari distribusi Normal dengan $\mu = 0$ dan σ=0,025, σ=0,5 dan σ = 1, t_i dibangkitkan dari distribusi Uniform(0,1), dengan fungsi $g(t_i) = 5e^{-5t}$ dan $g(t_i) = \sin(2\pi t)$.
2. Membandingkan nilai MSE model regresi dan *spline truncated* sebelum dan sesudah di *bootstrap.*

6. Analisis dan Pembahasan

6.1. Estimasi Regresi Linear dan Truncated Spline

Pandang persamaan (1), dan taksiran responnya adalah:

$$\hat{Y}_i = \hat{\beta}_0 + \hat{\beta}_1 X_{1i} + \ldots + \hat{\beta}_k X_{ki} \tag{4}$$

Masalah utama dalam analisa regresi adalah menaksir parameter atau koefisien regresi dan menyelidiki tingkat signifikansi dalam model secara serentak, kemudian menyelidiki secara individu. Metode kuadrat terkecil digunakan dengan tujuan untuk meminimumkan varians sehingga didapatkan penaksiran yang tak bias. Dalam persamaan regresi linear berganda, khususnya bila variabel prediktor lebih dari dua, akan lebih mudah apabila dengan menggunakan pendekatan matriks.

Sedangkan metode yang sering digunakan untuk menaksir parameter adalah OLS, yang prinsipnya meminimumkan jumlah kuadrat residual, atau secara matematis:

$$\sum_{i=1}^{n} \hat{\varepsilon}_i^2 = \varepsilon'\varepsilon \text{ minimum} \tag{5}$$

Taksiran $\hat{\varepsilon}_i$ dapat diperoleh dengan menyelesaikan turunan secara parsiil terhadap b_0, $b_1,\ldots,b_k$ dan menyamakan hasilnya dengan nol. Dari persamaan di atas didapatkan : ε = Y-βX dengan prinsip metode kuadrat terkecil maka :

$$\begin{aligned}\varepsilon^T\varepsilon &= (Y - \beta X)^T (Y - \beta X) \\ &= (Y^TY - Y^T\beta X - \beta^T X^T Y + X^T\beta^T\beta X) \\ &= (Y^TY - 2\beta^T X^T Y + \beta^T X^T X\beta)\end{aligned}$$

Dengan menurunkan ε'ε terhadap β secara parsiil berdasarkan aturan penurunan matrik akan diperoleh hasil sebagai berikut :

$$\frac{\delta\varepsilon^T\varepsilon}{\delta\hat{\beta}} = -2X^TY + 2X^TX\hat{\beta}$$

dengan menyamakan hasil diatas sama dengan nol maka diperoleh :

$$\begin{aligned}-2X^TY + 2X^TX\hat{\beta} &= 0 \\ X^TX\hat{\beta} &= X^TY \\ \hat{\beta} &= (X^TX)^{-1}(X^TY)\end{aligned}$$

Spline least squares merupakan generalisasi regresi polinomial (Eubank,1988), dimana estimasi kurva regresi *f* diperoleh melalui fungsi berikut:

$$s(t)=\sum_{j=1}^{m}\theta_j t^{j-1}+\sum_{j=1}^{k}\delta_j(t-\xi_j)_+^{m-1} \tag{6}$$

dalam persamaan (6) diatas, s merupakan *spline* orde-m dengan knots $\xi_1,\ldots,\xi_k$. himpunan dari semua fungsi ini, $S^m(\xi_1,\ldots,\xi_k)$) adalah suatu ruang vektor berdimensi m+k yang terdiri dari potongan polinomial orde-m yang memiliki m-2 turunan yang kontinu dan diskontinu pada turunan ke-(m-1) di titik ξ_j. Dengan memilih $\lambda=\{\xi_1,\ldots,\xi_k\}$, maka f dapat diestimasikan dengan mengestimasi koefisien-koefisien dari persamaan (6) salah satu metode untuk menyelesaikan hal tersebut adalah dengan menggunakan *least-squares*. Didefinisikan :

$$\begin{aligned}&x_1(t)=1,\\&x_2(t)=t,\\&\vdots\\&x_m(t)=t^{m-1},\\&x_{m+1}(t)=(t-\xi_1)_+^{m-1},\\&\vdots\\&x_{m+k}(t)=(t-\xi_k)_+^{m-1}\end{aligned} \tag{7}$$

dengan

$$\tilde{\boldsymbol{\beta}}=(\theta_1,\cdots,\theta_m,\delta_1,\cdots,\delta_k)^T \tag{8}$$

estimator *spline* least-square dari *f* adalah

$$f_\lambda=\sum_{j=1}^{m+k}\boldsymbol{\beta}_{\lambda j}x_j \tag{9}$$

dimana $\tilde{\boldsymbol{\beta}}_\lambda=(\beta_{\lambda 1},\ldots,\beta_{\lambda,m+k})^T$ adalah suatu minimizer dari

$$MSE(\beta;\lambda)=n^{-1}\sum_{i=1}^{n}(y_i-\sum_{j=1}^{m+k}\beta_j x_j(t_i))^2 \tag{10}$$

terhadap $\tilde{\beta}$ lebih jelasnya didefinisikan:

$$X(\lambda)=\{x_j(t_i)\}_{,i=1,\ldots,n;j=1,\ldots,m+k.} \tag{11}$$

maka $\tilde{\beta}_\lambda$ adalah suatu penyelesaian untuk persamaan normal:

$$X(\lambda)^T X(\lambda)\tilde{\beta}=X(\lambda)^T y, \tag{12}$$

dimana y= $(y_1,\ldots,y_n)^T$ jika X(λ) mempunyai rank m +k, maka:

$$\tilde{\boldsymbol{\beta}}_\lambda=[X(\lambda)^T X(\lambda)]^{-1}X(\lambda)^T\tilde{y} \tag{13}$$

dari (9) dan (13) terlihat bahwa dengan $\lambda=\{\xi_1,\ldots,\xi_k\}\}$, f_λ adalah estimator linear dari *f*.

6.2. Perbandingan Metode Pada Data Simulasi

Setelah membangun model $y_i=g(t_i)+\varepsilon_i$, maka langkah pertama didekati dengan kurva regresi dan *spline truncated*. Nilai MSE dengan berbagai variasi pada fungsi, pengamatan (n) secara lengkap tersaji pada Tabel 1 berikut.

Tabel 1. Nilai MSE Pada Regresi, *Spline Truncated* dan *Spline Bootstrap*

Fungsi	N	σ^2	MSE		
			Regresi	**Spline**	**Spline Bootstrap**
$5e^{-5t}$	50	0,025	0.098934	0.00058	**0.00054**
		0,5	0.174773	0.13982	**0.13007**
		1	0.592845	0.50940	**0.49840**
	100	0,025	0.097030	0.00072	**0.00066**
		0,5	0.227065	0.17331	**0.15684**
		1	0.678930	0.67544	**0.61675**
	250	0,025	0.079278	0.00073	**0.00066**
		0,5	0.298293	0.20838	**0.18718**
		1	0.792488	0.77183	**0.76443**
$\sin(2\pi t)$	50	0,025	0.093121	0.00127	**0.00124**
		0,5	0.198021	0.15802	**0.13578**
		1	0.545369	0.45707	**0.44408**
	100	0,025	0.040468	0.00125	**0.00120**
		0,5	0.213499	0.18962	**0.16563**
		1	0.670286	0.63835	**0.57318**
	250	0,025	0.011105	0.00016	**0.00158**
		0,5	0.263076	0.19760	**0.18147**
		1	0.888262	0.78974	**0.77430**

Berdasarkan Tabel.1, ternyata nilai MSE dengan banyaknya pengamatan yang semakin besar dan varians σ^2 konstan memberikan hasil yang semakin besar pada fungsi $g(t_i) = 5e^{-5t}$. Sedang pada fungsi $g(t_i) = \sin(2\pi t)$, nilai MSE dengan banyaknya pengamatan yang semakin besar dan varians σ^2 konstan memberikan hasil yang semakin kecil. Secara keseluruhan fungsi optimal terjadi pada varians σ^2 kecil dengan n sembarang. Sedangkan pada *truncated Spline*, nilai MSE dengan banyaknya pengamatan yang semakin besar dan varians σ^2 konstan memberikan hasil yang semakin kecil pada fungsi $g(t_i) = 5e^{-5t}$. Sedang pada fungsi $g(t_i) = \sin(2\pi t)$, nilai MSE dengan banyaknya pengamatan yang semakin besar pada varians σ^2 = 0,025 memberikan hasil yang semakin kecil, tetapi pada varians σ^2 = 0,5 dan 1 nilai MSE bervariasi. Secara keseluruhan dengan kriteria MSE, *Spline Truncated* sessudah di *bootstrap* lebih baik dibanding dengan pendekatan regresi dan *spline truncated.*

7. Kesimpulan

Berdasarkan hasil penelitian, maka estimator regresi linear adalah $\hat{\beta} = (X^T X)^{-1}(X^T Y)$ dan pada *spline truncated* adalah $\mathbf{X}(\lambda)^T \mathbf{X}(\lambda) = (\lambda)^T y$. Fungsi optimal terjadi pada variansi yang kecil untuk sembarang pengamatan. Nilai MSE pada kurva *truncated spline* lebih kecil dibanding dengan regresi linear pada semua fungsi. Hal ini dapat diartikan bahwa kurva *truncated spline* lebih baik dibanding dengan regresi linear. Hal ini dapat dilihat dari simulasi estimator $g(t) = \sin(5\pi t)$ dan $g(t) = 5e^{-5t}$, *spline truncated* memberikan berbagai nilai titik knot, sehingga nilai MSE kecil dibanding regresi linear. Secara keseluruhan dengan kriteria MSE, *spline truncated* sesudah di *bootstrap* lebih baik dibanding dengan pendekatan regresi dan *spline truncated.*

Daftar Pustaka

[1]. Botella, O., dan Shariff, K. 2003. B-*spline* methods in Fluid Dynamics. *International Journal of Computational Fluid Dynamics*, 17(2):133-149.

[2]. Efron, B. 1979. *Bootstrap* Method: Another look of the jackknife. *The Annals of Statistics*, 7:1-24.

[3]. Efron, B., dan Tibshirani, R. J. 1993. *An Introduction to the Bootstrap*. Chapman and Hall, Inc.

[4]. Eubank, R. L. 1988. *Spline Smooting and Nonparametric Regression*, Marcel Deker: New York.

[5]. Green, P. J., dan Silverman, B. W. 1994. *Nonparametric Regression and Generalized Linear Models: A Roughness Penalty Approach.* Chapman and Hall, London.

[6]. Hardle, W. 1990. Applied Nonparametric Regression, Cambrige University Press: New York.

[7]. Mooney, C. Z., dan Robert, D. D. 1993. *Bootstraping a nonparametric approach to Statistical inference*. Sage Publication, Inc, London.
[8]. Stone, C. J. 1990. Large Sample Inference for Log-*spline* Models. *The Annals of Statistics*, 18:717-741.