

PERBANDINGAN MODEL REGRESI NONPARAMETRIK DENGAN REGRESI SPLINE DAN KERNEL

Lilis Laome¹

¹Jurusan Matematika FMIPA Universitas Haluoleo Kendari 93232

email : lhi2s@yahoo.com

Abstrak

Tulisan ini membahas model regresi nonparametrik untuk data pertumbuhan anak balita. Ada dua metode estimasi regresi nonparametrik yang digunakan yaitu regresi spline dan kernel. Tujuan utama adalah membandingkan kedua metode untuk mengestimasi model regresi nonparametrik. Berdasarkan hasil numerik, diperoleh model regresi spline lebih baik dari pada model regresi kernel.

Kata kunci : Regresi nonparametrik, regresi spline, regresi kernel

Abstract

This paper study about using of nonparametric model for child growth data. It is discussed two nonparametric techniques called spline and kernel regression. The main goal is to compare the techniques used for prediction of the nonparametric regression models. According to the results of numerical studies, it is concluded that spline regression estimators are better than those of the kernel regression.

Keywords : Nonparametric regression, spline regression, kernel regression

I. Pendahuluan

Misalkan ada n pengamatan independen X_i, Y_i $_{i=1}^n$, yang diasumsikan mengikuti model :

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

dengan $\varepsilon_i \sim N(0, \sigma^2)$. Kurva regresi $f(x_i)$ pada persamaan (1) dapat diestimasi dengan pendekatan nonparametrik jika tidak ada informasi tentang bentuk $f(x_i)$.

Ada beberapa metode pendekatan regresi nonparametrik diantaranya spline, kernel, *k-nearest neighborhood* dan lain-lain. Diantara metode-metode pendekatan tersebut, regresi nonparametrik dengan pendekatan spline dan kernel merupakan metode yang sering digunakan. Kedua metode tersebut memiliki keunggulan masing-masing. Pendekatan kernel memiliki bentuk yang lebih fleksibel dan perhitungan matematisnya mudah disesuaikan [1]. Sedangkan pendekatan spline dapat menyesuaikan diri secara efektif terhadap data tersebut, sehingga didapatkan hasil yang mendekati kebenaran [2].

Olehnya itu dari kedua pendekatan spline dan kernel, sangat penting untuk mengetahui pendekatan regresi nonparametrik mana yang lebih baik, regresi spline atau regresi kernel untuk memodelkan fungsi pertumbuhan anak balita.

II. Tinjauan Pustaka

II.1 Estimator Spline

Secara umum, fungsi spline berorde p adalah sembarang fungsi yang dapat ditulis dalam

$$\text{bentuk : } f(x) = \sum_{i=1}^p \alpha_i x^i + \sum_{j=1}^h \delta_j (x - k_j)_+^p \quad (2)$$

$$\text{dengan } (x - k_j)_+^p = \begin{cases} (x - k_j)^p & , x \geq k_j \\ 0 & , x < k_j \end{cases}$$

α dan δ adalah konstanta real dan k_1, k_2, \dots, k_h adalah titik-titik knot. Jika $p = 2$, $p = 3$, dan $p = 4$, disubstitusikan kedalam persamaan (2) maka akan diperoleh fungsi spline secara berturut – turut dinamakan fungsi spline linear, spline kuadratik, dan spline kubik [3].

Misal diberikan $\mathbf{f} = (f(x_1), f(x_2), \dots, f(x_n))$ adalah vektor dengan nilai fungsi f pada titik-titik knot k_1, k_2, \dots, k_n . Estimasi spline $\hat{\mathbf{f}}_\lambda$ merupakan nilai estimasi pada $\mathbf{y} = (y_1, y_2, \dots, y_n)^T$ sebagai berikut :

$$\hat{\mathbf{f}}_\lambda = \begin{bmatrix} \hat{f}_\lambda(x_1) \\ \hat{f}_\lambda(x_2) \\ \vdots \\ \hat{f}_\lambda(x_n) \end{bmatrix} = (S_\lambda)_{n \times n} \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} \text{ atau } \hat{\mathbf{f}}_\lambda = S_\lambda \mathbf{y} \quad (3)$$

Dengan \hat{f}_λ adalah fungsi spline dengan titik-titik knot k_1, k_2, \dots, k_n untuk parameter penghalus $\lambda > 0$ dan S_λ adalah matrik smoother yang definit positif berukuran $n \times n$ [4].

II.2 Estimator Kernel

Penghalusan dengan pendekatan kernel yang selanjutnya dikenal sebagai penghalusan kernel (*kernel smoother*) sangat tergantung pada fungsi kernel dan bandwidth [5]. Taksiran kepadatan kernel sangat tergantung pada fungsi kernel yang digunakan, dinamakan fungsi kernel didefinisikan dengan :

$$K_h(x) = \frac{1}{h} K\left(\frac{x}{h}\right) \text{ untuk } -\infty < x < \infty \quad (4)$$

yang memenuhi : (i). $K(x) \geq 0$, (ii). $\int_{-\infty}^{\infty} K(x) dx = 1$, (iii). $\int_{-\infty}^{\infty} K^2(x) dx < \infty$, dan (iv).

$K(x)$ simetris di sekitar 0. Secara umum taksiran kernel didefinisikan sebagai berikut :

$$\hat{f}_\lambda(x) = \frac{1}{n} \sum_{i=1}^n K_h(x - X_i) = \frac{1}{nh} \sum_{i=1}^n K\left(\frac{x - X_i}{h}\right) \quad (5)$$

dimana h adalah derajat penghalus kernel yang disebut parameter bandwidth dan berperan untuk mengontrol penyebaran dari fungsi $\hat{f}_\lambda(x)$. Kriteria pemilihan fungsi kernel yang baik

berdasarkan pada resiko kernel minimum yang dapat diperoleh dari kernel optimal atau kernel-kernel dengan variansi minimum. Berikut macam-macam fungsi kernel [5] seperti pada table 1 di bawah:

Tabel 1. Macam-macam Fungsi Kernel

Kernel	K(u)
Epanechnikov	$\frac{3}{4} - u^2 + 1 I(u \leq 1)$
Quartic	$\frac{15}{16} 1 - u^2 + 2 I(u \leq 1)$
Tringular	$1 - u I(u \leq 1)$
Gaussian	$2\pi^{-\frac{1}{2}} \exp -\frac{u^2}{2} I(u < \infty)$
Uniform	$\frac{1}{2} I(u \leq 1)$

III. Hasil dan Pembahasan

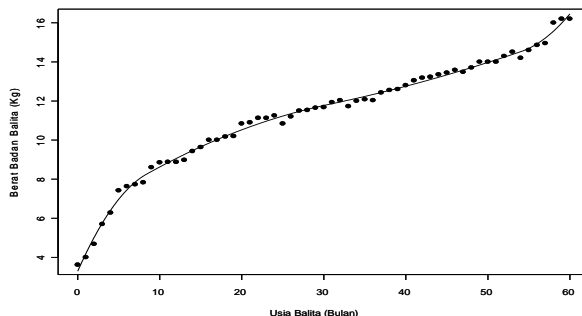
Data yang digunakan [1], dengan variabel penelitian adalah umur balita (X) dan berat badan balita (Y) sebanyak 284 anak laki-laki dan 254 anak perempuan. Langkah awal yang dilakukan sebelum melakukan analisis regresi adalah melakukan deskripsi data. Deskripsi data dilakukan dengan membuat plot data antara umur anak balita (X) dan berat badan anak balita (Y) baik laki-laki maupun perempuan. Plot data tersebut digunakan untuk menaksir fungsi data yang mendekati dan melihat bagaimana perubahan pola perilaku kurva. Selanjutnya melakukan pemodelan regresi nonparametrik dengan pendekatan spline dan kernel.

III.1 Model Balita Laki-laki

Pertama dilakukan pendekatan dengan spline. Untuk mendapatkan estimasi model yang optimal, terlebih dahulu dicari titik knot yang optimal dengan kriteria GCV minimum. Berdasarkan [3] diperoleh titik knot optimum adalah 8,4; 30,54; dan 54. Selanjutnya dengan titik knot tersebut, diestimasi fungsi pertumbuhan anak balita sebagai berikut :

$$\hat{Y} = 3,29 + 0,95X - 0,04X^2 + 0,039(X - 8,4)_+^2 + 0,004(X - 30,54)_+^2 + 0,029(X - 54)_+^2 \tag{6}$$

Spline Fit



Gambar 1. Estimasi pertumbuhan balita laki-laki dengan pendekatan spline

Ukuran kebaikan model R^2 diperoleh 99,55% dan nilai MSE adalah 0,042. Kemudian memenuhi asumsi identik, independen dan berdistribusi normal [3].

Kedua dilakukan pendekatan dengan kernel. Untuk mendapatkan estimasi model yang optimal, terlebih dahulu dicari nilai bandwidth yang optimal dengan kriteria GCV minimum. Berdasarkan pengolahan data dengan Matlab diperoleh nilai badwidth optimal dengan nilai GCV minimum sebagai berikut :

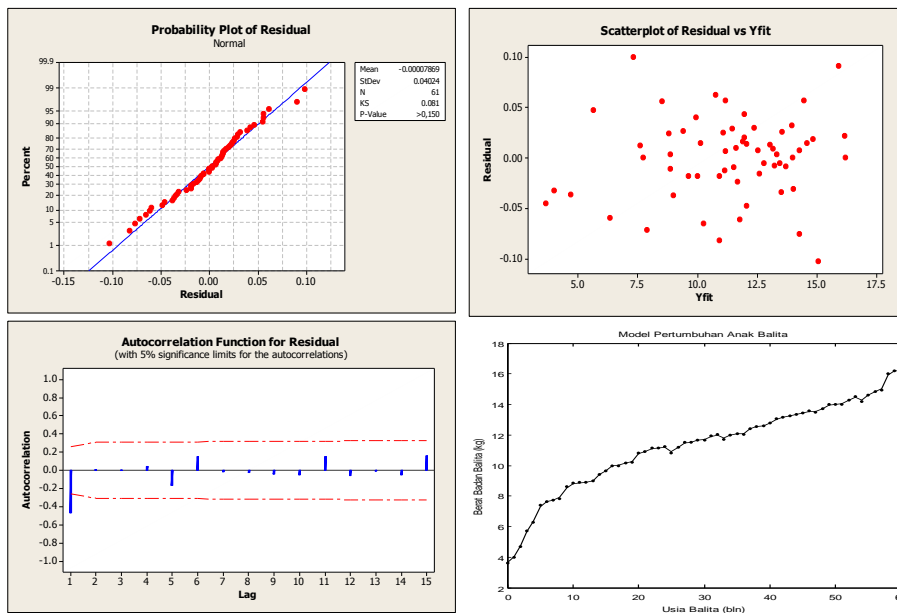
Tabel 2. Nilai bandwidth dan GCV

Bandwith (h)	GCV (h)
0,5000	0,0360*
0,6000	0,0361
0,7000	0,0365
0,8000	0,0371
0,9000	0,0383
1,0000	0,0399

* Nilai bandwidth optimal dengan GCV minimum

Berdasarkan nilai GCV minimum diperoleh model estimasi dengan pendekatan kernel :

$$\hat{Y} = \left[\frac{\sum_{i=1}^n \exp(-u^2) I(|u| \leq \infty) y_i}{\sum_{i=1}^n \exp(-u^2) I(|u| \leq \infty)} \right], u = \left(\frac{x - x_i}{0,5} \right) \quad (7)$$



Gambar 2. Uji residual dan estimasi dengan pendekatan kernel untuk balita laki-laki

Ukuran kebaikan model R^2 diperoleh 99,68% dan nilai MSE adalah 0,001. Kemudian diperiksa apakah residual memenuhi asumsi identik, independen dan berdistribusi normal. Berdasarkan Gambar 2 dapat ditunjukkan bahwa distribusi residual berdistribusi normal dengan p -value > 0,05. Plot residual terhadap estimasi Y (Y_{fit}) membentuk suatu tren tertentu sehingga asumsi variansi identik/konstan dianggap terpenuhi. Plot ACF dari residual terlihat ada yang keluar garis sehingga asumsi independen tidak terpenuhi.

III.2. Model Balita Perempuan

Dengan langkah yang sama seperti pada model laki-laki, dilakukan pemodelan dengan pendekatan spline. Selanjutnya diperoleh titik knot optimum adalah 9, 31, dan 53 dengan nilai GCV minimum adalah 0,11 , sehingga estimasi modelnya adalah :

$$\hat{Y} = 2,93 + 0,97X - 0,056X^2 - 0,057(X - 7)_+^2 + -0,0063(X - 31)_+^2 + 0,091(X - 53)_+^2 \tag{8}$$

Ukuran kebaikan model R^2 diperoleh 99,16%, MSE adalah 0,107 dan memenuhi asumsi identik, independen, dan berdistribusi normal [3].

Selanjutnya dilakukan pemodelan dengan pendekatan kernel, diperoleh bandwidth optimal disajikan pada table 3 dibawah ini :

Tabel 3. Nilai bandwidth dan GCV

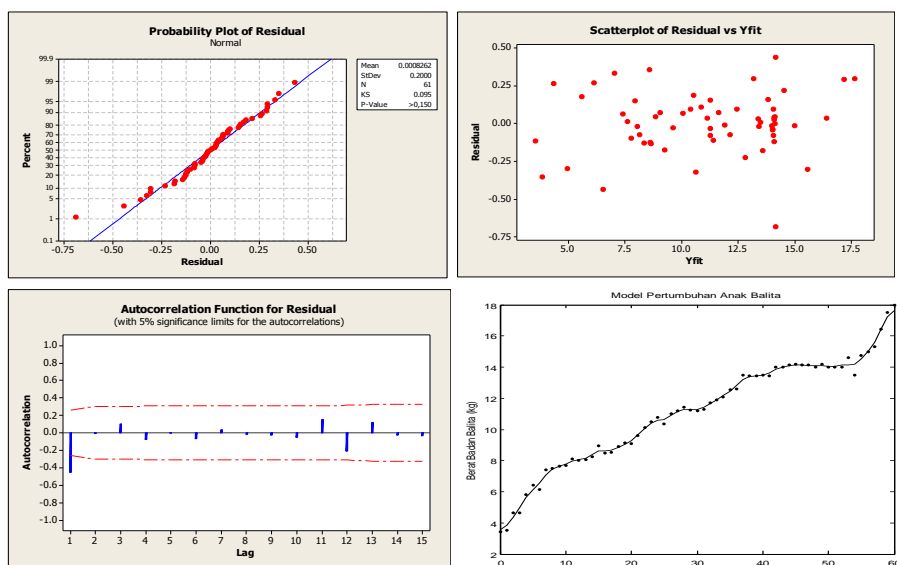
Bandwith (h)	GCV (h)
0,8000	0,1132
0,9000	0,1118
1,0000	0,1113*
1,1000	0,1116
1,2000	0,1129

* Nilai bandwidth optimal

Berdasarkan nilai bandwidth optimal diperoleh estimasi model sebagai berikut :

$$\hat{Y} = \left[\frac{\sum_{i=1}^n \exp(-u^2) I(|u| \leq \infty) y_i}{\sum_{i=1}^n \exp(-u^2) I(|u| \leq \infty)} \right], u = \left(\frac{x - x_i}{1} \right) \tag{9}$$

Ukuran kebaikan model R^2 diperoleh 99,26% dan MSE adalah 0,095. Kemudian diperiksa apakah residual memenuhi asumsi identik, independen, dan berdistribusi normal. Berdasarkan Gambar 3 diperoleh distribusi residual berdistribusi normal dengan p -value > 0,05. Plot residual terhadap estimasi Y (Y_{fit}) tidak membentuk suatu tren tertentu sehingga asumsi variansi identik/konstan dianggap terpenuhi. Plot ACF dari residual terlihat ada yang keluar garis sehingga asumsi independen tidak terpenuhi.



Gambar 3. Uji residual dan estimasi dengan pendekatan kernel pada balita perempuan

Berdasarkan uraian yang telah dipaparkan di atas, dari kedua pendekatan yang digunakan dapat diringkas pada Tabel 4 :

Tabel 4. Performan pendekatan regresi spline dan kernel untuk model balita laki-laki

Kriteria / Model Regresi	Spline	Kernel
R^2	99,55 %	99,68 %
MSE	0,042	0,001
Asumsi Residual		
- Identik	Memenuhi	Tidak memenuhi
- Independen	Memenuhi	Tidak memenuhi
- distribusi normal	Memenuhi	Memenuhi

Tabel 3. Performan pendekatan regresi spline dan kernel untuk model balita perempuan

Kriteria / Model Regresi	Spline	Kernel
R^2	99,17 %	98,43 %
MSE	0,107	0,039
Asumsi Residual		
- Identik	Memenuhi	Memenuhi
- Independen	Memenuhi	Tidak memenuhi
- distribusi normal	Memenuhi	Memenuhi

IV. Kesimpulan

Berdasarkan kriteria kebaikan model yaitu R^2 , MSE dan uji asumsi residual, pendekatan regresi spline lebih baik dari regresi kernel untuk pertumbuhan balita. Hal ini dapat dilihat dari uji asumsi residual yang menunjukkan bahwa pendekatan spline memenuhi semua asumsi residual.

V. Daftar Pustaka

- 1 Purnama, S.W. dan Wibowo, W. 2006. *Perbandingan Regresi Nonlinier dan Regresi Nonparametrik dalam Menaksir Fungsi Pertumbuhan Anak Balita*, Laporan Penelitian Dosen Muda. Surabaya : Institut Teknologi Sepuluh Nopember.
- 2 Eubank, R.L. 1998. *Spline Smoothing and Nonparametric Regression*. New York : Marcel Dekker.
- 3 Laome, L. 2009. *Regresi Nonparametrik Spline Linear dan Kuadratik dalam Menaksir Fungsi Pertumbuhan Anak Balita*. Paradigma, 13 (3), hal. 101-107.
- 4 Aidin, D. 2007. *A Comparison of the Nonparametric Regression Models using Smoothing Spline and Kernel Regression*. World Academy of Science, Engineering and Technology, 36, 253-257.
- 5 Hardle, W. 1990. *Smoothing Technique with Implementation in Statistics*. New York : Springer.