

Pemodelan Regresi Nonparametrik Spline *Truncated* Dan Aplikasinya pada Angka Kelahiran Kasar di Surabaya

Merly Fatriana Bintariningrum dan I Nyoman Budiantara
Jurusan Statistika, Fakultas Matematika dan Ilmu Pengetahuan Alam,
Institut Teknologi Sepuluh Nopember (ITS)
Jl. Arief Rahman Hakim, Surabaya 60111
E-mail: i_nyoman_b@statistika.its.ac.id

Abstrak—Surabaya sebagai ibukota Propinsi Jawa Timur menempati urutan kedua setelah Jakarta dalam jumlah penduduk. Pada tahun 2012 terjadi 40.343 jumlah kelahiran, dimana angka ini mencerminkan pertumbuhan penduduk yang sangat tinggi. Pertumbuhan penduduk yang begitu pesat dikhawatirkan akan menyebabkan ledakan pertumbuhan penduduk. Pada penelitian ini memodelkan angka kelahiran kasar (CBR) dengan 5 variabel yang diduga berpengaruh. Pemodelan dilakukan dengan metode regresi nonparametrik Spline truncated. Metode ini dipilih karena Spline merupakan metode yang fleksibel, model ini cenderung mencari sendiri estimasi data. Dalam pemodelan ini terdapat titik knot, yaitu titik yang menunjukkan perubahan data. Pemilihan titik knot optimum dilakukan dengan cara memilih nilai Generalized Cross Validation (GCV) yang paling minimum. Berdasarkan penelitian yang telah dilakukan, 5 variabel dinyatakan berpengaruh terhadap CBR yaitu persentase kepala keluarga yang berstatus pendidikan tidak sekolah/belum tamat SD, persentase kepala keluarga yang berstatus tidak bekerja, persentase kepala keluarga yang menikah pada umur 15-19 tahun, angka perkawinan kasar, dan angka migrasi masuk. Model regresi nonparametrik Spline truncated yang terbentuk memiliki koefisien determinasi sebesar 94,3%, serta nilai MSE yaitu 0,44.

Kata Kunci—CBR, Nonparametrik Spline Truncated, GCV

I. PENDAHULUAN

Pada saat ini, jumlah penduduk di Indonesia menempati urutan ketiga di dunia setelah Cina dan India. Jumlah penduduk yang begitu besar ini diakibatkan oleh pertumbuhan penduduk yang juga sangat pesat. Pertumbuhan penduduk adalah perubahan jumlah penduduk di suatu wilayah tertentu pada waktu tertentu dibandingkan waktu sebelumnya [1]. Jumlah penduduk, komposisi umur, dan laju pertumbuhan atau penurunan penduduk dipengaruhi oleh fertilitas (kelahiran), mortalitas (kematian), dan migrasi (perpindahan tempat), karena ketiga variabel tersebut merupakan komponen-komponen yang berpengaruh terhadap perubahan penduduk [2]. Pertambahan jumlah penduduk yang sangat pesat terjadi akibat jumlah kelahiran jauh lebih besar daripada kematian. Pada tahun 2011, Angka kelahiran kasar (CBR) di Indonesia justru jauh lebih tinggi dibandingkan negara Cina yang menempati jumlah penduduk terbanyak di dunia. CBR di Indonesia pada tahun 2011 adalah 19,63, sedangkan Cina hanya 11,93 [3].

Pada tahun 2012 terjadi 40.343 jumlah kelahiran di

Surabaya, dimana rata-rata kelahiran hidup kota Surabaya adalah 3000 kelahiran per bulan [4]. Jumlah penduduk yang sangat besar ini menimbulkan sebuah kekhawatiran meningkatnya berbagai permasalahan, seperti kriminalitas, tingkat pengangguran, dan penghidupan yang layak.

Pengendalian jumlah penduduk sangat diperlukan agar pertumbuhan jumlah penduduk dapat terkontrol dan menghindari terjadinya ledakan jumlah penduduk. Dalam hal ini yang dapat dilakukan untuk mengendalikan pertumbuhan jumlah penduduk adalah mengendalikan jumlah kelahiran. Dengan mengetahui pola hubungan faktor-faktor yang mempengaruhi angka kelahiran maka dapat menentukan langkah apa selanjutnya yang akan dilakukan.

Untuk mengetahui pola hubungan faktor-faktor yang mempengaruhi kelahiran dapat dilakukan dengan analisis regresi. Pada regresi nonparametrik, data dibiarkan mencari sendiri pola datanya sehingga subjektifitas dari peneliti dapat diminimalisir. Dengan demikian, pendekatan regresi nonparametrik memiliki fleksibilitas yang tinggi [5]. Salah satu kelebihan pendekatan Spline adalah model cenderung mencari sendiri estimasi data kemanapun pola data tersebut bergerak. Kelebihan ini terjadi karena dalam Spline terdapat titik-titik knot, yaitu titik-perpaduan bersama yang menunjukkan terjadinya perubahan pola perilaku data [6].

Data angka kelahiran kasar atau *Crude Birth Rate* dalam penelitian ini akan dimodelkan dengan menggunakan regresi Nonparametrik Spline *Truncated*. *Truncated* merupakan sebuah fungsi yang dapat diartikan sebagai fungsi potongan [7]. Metode tersebut dipilih karena berdasarkan plot pada studi awal yang telah dilakukan, diketahui bahwa plot variabel-variabel yang mempengaruhi angka kelahiran kasar di Surabaya polanya tidak diketahui bentuk polanya. Batasan masalah dalam penelitian ini adalah Data angka kelahiran kasar pada penelitian ini terbatas pada data kelahiran yang dilaporkan dan masuk ke Dinas Kependudukan dan Catatan Sipil Kota Surabaya pada tahun 2012.

II. TINJAUAN PUSTAKA

A. Regresi Nonparametrik Spline

Metode regresi nonparametrik merupakan metode regresi yang digunakan ketika kurva regresi antara variabel respon dan variabel prediktor tidak diketahui bentuknya atau polanya. Regresi nonparametrik Spline memiliki fleksibilitas yang tinggi dimana data diharapkan mencari

sendiri bentuk estimasi kurva regresinya tanpa dipengaruhi oleh subjektifitas peneliti. Secara umum model regresi nonparametrik secara dapat disajikan sebagai berikut

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

dengan $f(t_i)$ merupakan kurva regresi yang dihipotesiskan dengan fungsi spline berorde p dengan titik knot K_1, K_2, \dots, K_r yang dapat diberikan oleh persamaan.

$$f(t_i) = \sum_{j=0}^p \gamma_0 t_i^j + \sum_{j=1}^r \gamma_{p+j} (t_i - K_j)_+^p (2)$$

Apabila persamaan (1) disubstitusikan kedalam persamaan (2) maka akan diperoleh persamaan regresi nonparametrik spline sebagai berikut.

$$y_i = \sum_{j=0}^p \gamma_j t_i^j + \sum_{j=1}^r \gamma_{p+j} (t_i - K_j)_+^p + \varepsilon_i (3)$$

dimana $i = 1, 2, \dots, n$.

Fungsi $(t_i - K_j)_+^p$ merupakan fungsi *truncated* (potongan) yang diberikan oleh:

$$(t_i - K_j)_+^p = \begin{cases} (t_i - K_j)^p, & t_i \geq K_j \\ 0, & t_i < K_j \end{cases} (4)$$

dimana p adalah orde Spline dan K adalah knot yang terpilih.

B. Pemilihan Titik Knot Optimal

Untuk mendapatkan model regresi spline terbaik maka titik optimal dicari yang paling sesuai dengan data. Salah satu metode yang banyak dipakai dalam memilih titik knot optimal adalah *Generalized Cross Validation* (GCV). Untuk memperoleh titik knot optimal dapat dilihat dari nilai GCV yang paling minimum. Metode GCV secara umum didefinisikan sebagai berikut [8].

$$GCV(K_1, K_2, \dots, K_r) = \frac{MSE(K_1, K_2, \dots, K_r)}{(n^{-1} \text{trace}[I - A(K_1, K_2, \dots, K_r)])^2} (5)$$

dengan

$$MSE(K_1, K_2, K_3, \dots, K_r) = n^{-1} \sum_{i=1}^n (y_i - \hat{f}(x_i))^2 (6)$$

dimana K_1, \dots, K_r adalah titik knot pertama hingga knot ke- r .

METODOLOGI PENELITIAN

Sumber Data dan Variabel Penelitian

Data yang digunakan dalam penelitian ini adalah data sekunder yang diambil dari Dinas Kependudukan Catatan Sipil Kota Surabaya dan Buku Profil Perkembangan Kependudukan Kota Surabaya tahun 2012. Variabel respon yang digunakan pada penelitian ini adalah angka kelahiran kasar atau *Crude Birth Rate* (CBR) yang dihitung berdasarkan jumlah kelahiran yang terjadi di Surabaya selama tahun 2012. Variabel-variabel yang diduga berpengaruh didapatkan dari referensi baik penelitian sebelumnya maupun teori yang berhubungan dengan kelahiran (fertilitas). Variabel yang digunakan pada penelitian ini disajikan dalam Tabel 1 sebagai berikut.

Tabel 1.
Variabel Penelitian

Variabel	Keterangan
y	Angka kelahiran kasar (CBR)
x ₁	Persentase kepala keluarga yang berstatus pendidikan tidak sekolah/belum tamat SD
x ₂	Persentase kepala keluarga yang berstatus tidak bekerja
x ₃	Persentase kepala keluarga yang menikah pada umur 15-19 tahun
x ₄	Angka perkawinan kasar
x ₅	Angka migrasi masuk

Tabel 2
Statistika Deskriptif CBR dan 5 Faktor yang Diduga Berpengaruh

Variabel	Rata-Rata	Standar Deviasi	Minimum	Maksimum
Y	12,235	1,619	9,4	15,10
x ₁	2,232	1,100	1,100	5,630
x ₂	3,682	1,275	1,180	6,470
x ₃	0,1426	0,072	0,070	0,350
x ₄	3,903	2,158	0,6	10,7
x ₅	37,10	13,86	21,00	78,90

B. Langkah Analisis

Langkah-langkah analisis data yang dilakukan dalam penelitian ini ialah sebagai berikut :

1. Membuat *scatter plot* antara angka kelahiran kasar dengan masing-masing variabel yang diduga berpengaruh untuk mengetahui bentuk pola data. Apabila terdapat komponen nonparametrik, maka digunakan pendekatan regresi nonparametrik Spline.
2. Memodelkan data dengan menggunakan spline linear dengan satu, dua, tiga knot, serta kombinasi knot.
3. Memilih titik knot optimal berdasarkan nilai GCV yang paling minimum
4. Mendapatkan model regresi Spline dengan titik knot optimal
5. Menguji signifikansi parameter regresi Spline secara serentak
6. Melakukan uji parameter regresi Spline secara parsial
7. Menguji asumsi residual IIDN dari model regresi Spline
8. Menghitung nilai koefisien determinasi R² dan MSE
9. Menginterpretasikan model dan menarik kesimpulan.

III. HASIL DAN PEMBAHASAN

A. Deskriptif Karakteristik CBR di Surabaya

Surabaya sebagai ibukota propinsi Jawa Timur memiliki jumlah penduduk terbanyak, sekaligus memiliki jumlah kelahiran terbesar. Menurut Dinas Kependudukan dan Catatan Sipil Kota Surabaya jumlah penduduk Surabaya pada tahun 2012 adalah sebesar 3.124.890 jiwa, dan terjadi 40.343 jumlah kelahiran. Dalam penelitian ini diduga terdapat 5 variabel yang mempengaruhi CBR. Analisis ini mencakup 31 kecamatan yang ada di Surabaya. Analisis pertama yang digunakan adalah analisis deskriptif. CBR disajikan pada Tabel 2.

Rata-rata CBR adalah sebesar 12,235, hal ini dapat diartikan bahwa rata-rata terdapat 12 bayi lahir setiap 1000

penduduk. Standar deviasi atau keragaman dari data sebesar 1,619 yang dinilai data tidak terlalu beragam karena nilai tersebut terpaut jauh dari jangkauan data. Dimana minimum terdapat 9 kelahiran dan maksimum terdapat 15 kelahiran pada setiap 1000 penduduk.

B. Hubungan Antara CBR dengan Enam Variabel yang Diduga Berpengaruh

Hubungan antara CBR dengan enam variabel yang diduga berpengaruh dapat dilihat pada scatterplot pada Gambar 1.

Berdasarkan Tabel 3 tersebut diketahui bahwa nilai GCV paling kecil adalah 2.22, dengan titik knot optimum adalah.

$$K_1= 1,47 \quad K_2= 1,61 \quad K_3= 0,09, \\ K_4= 1.42 \quad K_5= 25,73$$

Nilai GCV ini akan dibandingkan dengan nilai GCV dengan 2 knot dan 3 knot, kemudian dipilih yang paling minimum.

C. Pemilihan Titik Knot Optimum dengan Dua Titik Knot

Setelah mendapatkan knot optimum dari satu titik knot maka selanjutnya adalah mencari titik knot optimum dengan dua titik knot. Percobaan dilakukan dengan cara yang sama, dan dipilih GCV yang paling minimum. Berikut adalah model regresi nonparametrik spline *truncated* untuk dua titik knot.

$$\hat{y} = \hat{y}_0 + \hat{y}_1x_1 + \hat{y}_2(x_1 - K_1)_+ + \hat{y}_3(x_1 - K_2)_+ + \hat{y}_4x_2 + \\ \hat{y}_5(x_2 - K_3)_+ + \hat{y}_6(x_2 - K_4)_+ + \dots + \hat{y}_{15}(x_5 - K_{10})_+$$

Berdasarkan Tabel 4 nilai GCV minimum adalah 2.10, dengan titik knot optimum adalah

$$(K_1=1.0; K_2=5.63), (K_3=1.18; K_4=6.47), \\ (K_5=0.07; K_6=0.35), (K_7=0.60; K_8=10.70), \\ (K_9= 21.00; K_{10}=78.90).$$

Nilai GCV tersebut lebih kecil dari hasil pemilihan titik knot optimum dengan satu titik knot.

D. Pemilihan Titik Knot Optimum dengan Tiga Titik Knot

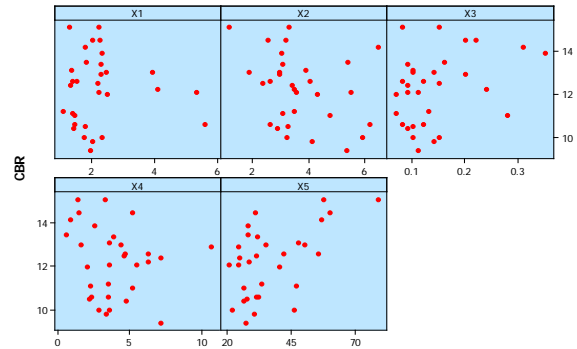
Setelah melakukan pemodelan dengan satu dan dua knot, pada kali ini akan mencoba mencari titik knot optimum dengan menggunakan tiga titik knot. Berikut adalah model regresi nonparametrik Spline *truncated* dengan tiga titik knot.

$$\hat{y} = \hat{y}_0 + \hat{y}_1x_1 + \hat{y}_2(x_1 - K_1)_+ + \hat{y}_3(x_1 - K_2)_+ + \\ \hat{y}_4(x_1 - K_3)_+ + \hat{y}_5x_2 + \dots + \hat{y}_{20}(x_5 - K_{15})_+$$

Berdasarkan Tabel 5, diketahui bahwa nilai GCV minimum adalah 1,38. Titik- titik knot optimum untuk model ini adalah.

$$(K_1=1.93; K_2=3.13; K_3=3.23), (K_4=2.15; K_5=3.56; K_6=3.66), \\ (K_7=0.12; K_8=0.20; K_9=0.21), (K_{10}=2.46; K_{11}=5.13; K_{12}=5.34), \\ (K_{13}=31.63; K_{14}=47.00; K_{15}=48.18).$$

Jika dibandingkan dengan percobaan sebelumnya dengan menggunakan satu titik knot dan dua titik knot, model ini memiliki nilai GCV yang lebih kecil. Sehingga dapat dikatakan bahwa model regresi nonparametrik Spline *truncated* ini lebih baik.



Gambar 1. Scatterplot Antara CBR dengan 5 Variabel yang Diduga Berpengaruh

Tabel 3.

Pemilihan Titik Knot Optimum dengan Satu Titik Knot						
No	x_1	x_2	x_3	x_4	x_5	GCV
1	1.19	1.29	0.08	0.81	22.18	2.50
2	1.28	1.40	0.08	1.01	23.36	2.40
3	1.38	1.50	0.09	1.22	24.54	2.34
4	1.47	1.61	0.09	1.42	25.73	2.22
5	1.56	1.72	0.10	1.63	26.91	2.23

Tabel 4.

Pemilihan Titik Knot Optimum dengan Dua Titik Knot						
No	x_1	x_2	x_3	x_4	x_5	GCV
1	1.10	1.18	0.07	0.60	21.00	3.11
	5.35	6.15	0.33	10.08	75.36	
2	1.10	1.18	0.07	0.60	21.00	2.10
	5.63	6.47	0.35	10.70	78.90	
3	1.19	1.29	0.08	0.81	22.18	2.93
	1.28	1.40	0.08	1.01	23.36	
4	1.19	1.29	0.08	0.81	22.18	2.96
	1.38	1.50	0.09	1.22	24.54	
5	1.19	1.29	0.08	0.81	22.18	3.04
	1.47	1.61	0.09	1.42	25.73	

Tabel 5.

Pemilihan Titik Knot Optimum dengan Tiga Titik Knot						
No	x_1	x_2	x_3	x_4	x_5	GCV
1	1.93	2.15	0.12	2.46	31.63	4.57
	3.04	3.45	0.19	4.93	45.81	
	5.45	6.25	0.34	10.29	76.54	
2	1.93	2.15	0.12	2.46	31.63	4.57
	3.04	3.45	0.19	4.93	45.81	
	5.54	6.36	0.34	10.49	77.72	
3	1.93	2.15	0.12	2.46	31.63	1.38
	3.13	3.56	0.20	5.13	47.00	
	3.23	3.66	0.21	5.34	48.18	
4	1.93	2.15	0.12	2.46	31.63	2.63
	3.13	3.56	0.20	5.13	47.00	
	3.32	3.77	0.21	5.55	49.36	
5	1.93	2.15	0.12	2.46	31.63	2.92
	3.13	3.56	0.20	5.13	47.00	
	3.41	3.88	0.21	5.75	50.54	

E. Pemilihan Titik Knot Optimum dengan Kombinasi Knot

Kombinasi knot adalah kombinasi antara satu titik knot, dua titik knot, dan tiga titik knot. Kombinasi ini digunakan untuk memilih titik knot optimum. Dalam pemilihan titik knot optimum pada model regresi nonparametrik Spline

truncated dengan kombinas knot ini dipilih nilai GCV yang paling minimum. Nilai GCV dengan kombinasi knot adalah sebagai berikut.

Berdasarkan Tabel 6 diketahui bahwa nilai GCV minimum dengan kombinasi knot adalah sebesar 1,38 dengan kombinasi 3,3,3,3,3. Kombinasi ini sama dengan pemilihan titik knot optimum dengan tiga knot. Sehingga nilai yang digunakan untuk analisis selanjutnya adalah nilai GCV dengan tiga titik knot.

F. Penaksiran Parameter Model Regresi Spline Truncated

Dari hasil pemilihan titik knot optimum yang dilakukan, maka model regresi dengan menggunakan tiga titik knot adalah yang terbaik. Hasil dari estimasi parameter dengan menggunakan tiga titik knot adalah sebagai berikut.

$$\hat{y} = 29.5 - 0.59 x_1 + 5.71(x_1 - 1.93)_+ - 134 (x_1 - 3.13)_+ + 133 (x_1 - 3.23)_+ - 7.57 x_2 + 9.59 (x_2 - 2.15)_+ - 16.9 (x_2 - 3.56)_+ + 14.4 (x_2 - 3.66)_+ - 48.9 x_3 + 140 (x_3 - 0.12)_+ - 940 (x_3 - 0.20)_+ + 856 (x_3 - 0.21)_+ - 1.33 x_4 + 1.68(x_4 - 2.46)_+ + 11.3(x_4 - 5.13)_+ - 12.3(x_4 - 5.34)_+ + 0.136 x_5 - 0.273 (x_5 - 31.63)_+ + 6.59 (x_5 - 47.00)_+ - 6.78 (x_5 - 48.18)_+$$

Model yang terbentuk tersebut memiliki koefisien determinasi sebesar 94,3%.

G. Pengujian Signifikansi Parameter Model Regresi Spline Truncated

Untuk melakukan pengujian signifikansi parameter dilakukan dengan uji serentak dan individu. Uji secara serentak menggunakan hipotesis sebagai berikut.

$$H_0 : \gamma_1 = \gamma_2 = \gamma_3 = \dots = \gamma_{20} = 0$$

$$H_1 : \text{minimal terdapat satu } \beta_k \neq 0 \text{ dimana } k=1,2,\dots,20.$$

Hasil ANOVA untuk model regresi nonparametrik Spline *truncated* secara serentak diberikan pada Tabel 7 sebagai berikut.

Berdasarkan hasil ANOVA pada Tabel 7 dapat diketahui bahwa p-value adalah sebesar 0.001, nilai ini kurang dari nilai α (0.05). Sehingga dalam hal ini dapat disimpulkan bahwa H_0 ditolak, maka minimal terdapat satu parameter yang signifikan terhadap variabel respon. Untuk dapat mengetahui parameter mana saja yang berpengaruh terhadap respon maka perlu dilakukan pengujian secara individu. Hipotesis untuk uji individu adalah sebagai berikut.

$$H_0 : \gamma_k = 0$$

$$H_1 : \gamma_k \neq 0 \text{ dimana } k=1,2,\dots,20$$

Tabel 8 adalah hasil pengujian signifikansi parameter secara individu. Berdasarkan Tabel 8 terlihat bahwa terdapat beberapa parameter yang tidak signifikan. Dari 20 parameter terdapat 5 parameter yang tidak signifikan. Meskipun terdapat 5 parameter yang tidak signifikan, namun secara keseluruhan keenam variabel yang diduga berpengaruh dapat disimpulkan bahwa semua berpengaruh terhadap variabel respon yaitu CBR.

H. Pengujian Asumsi Residual

Pada analisis regresi nonparametrik Spline *truncated*, residual yang terbentuk harus memenuhi asumsi IIDN. Berikut adalah hipotesis untuk asumsi identik.

$$H_0 : \sigma_1^2 = \sigma_2^2 = \dots = \sigma_n^2 = \sigma^2$$

$$H_1 : \text{minimal ada satu } \sigma_i^2 = \sigma^2, i=1,2,\dots,n$$

Tabel 6.
Pemilihan Titik Knot Optimum dengan Kombinasi Knot

No	Variabel	Banyak Titik Knot	Titik-Titik Knot	GCV
1	x_1	3	1.93; 3.13; 3.23	1.54
	x_2	1	1.61	
	x_3	1	0.09	
	x_4	3	2.46; 5.13; 5.34	
	x_5	2	21.0; 78.90	
2	x_1	3	1.93; 3.13; 3.23	1.38
	x_2	3	2.15; 3.56; 3.66	
	x_3	3	0.12; 0.20; 0.21	
	x_4	3	2.46; 5.13; 5.34	
	x_5	3	31.63; 47.00; 48.18	
3	x_1	3	1.93; 3.13; 3.23	1.56
	x_2	2	1.18; 0.07	
	x_3	2	6.47; 0.35	
	x_4	3	2.46; 5.13; 5.34	
	x_5	3	31.63; 47.00; 48.18	

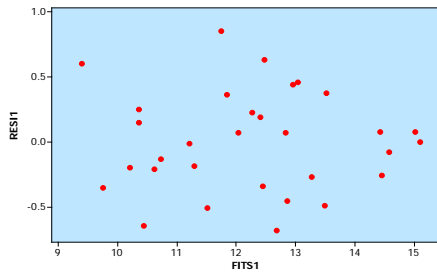
Tabel 7.
ANOVA Model Regresi Spline Secara Serentak

Sumber	Df	Sum of Square	Mean Square	F_{hitung}	P-value
Regresi	20	74.1992	3.7100	8.33	0.001
Error	10	4.4517	0.4452		
Total	30	78.6510			

Tabel 8.
Pengujian Parameter Model Regresi Secara Individu

Variabel	Parameter	Koef	p-val	Keputusan
x_1	γ_0	29.469	0.001	Signifikan
	γ_1	-0.589	0.573	Tidak Signifikan
	γ_2	5.705	0.020	Signifikan
	γ_3	-134.30	0.000	Signifikan
	γ_4	132.94	0.000	Signifikan
x_2	γ_5	-7.571	0.043	Signifikan
	γ_6	9.590	0.034	Signifikan
	γ_7	-16.898	0.055	Tidak Signifikan
	γ_8	14.361	0.081	Tidak Signifikan
	γ_9	-48.89	0.024	Signifikan
x_3	γ_{10}	139.59	0.006	Signifikan
	γ_{11}	-939.8	0.010	Signifikan
	γ_{12}	856.3	0.011	Signifikan
	γ_{13}	-1.3318	0.035	Signifikan
	γ_{14}	1.6808	0.046	Signifikan
x_4	γ_{15}	-11.327	0.064	Tidak Signifikan
	γ_{16}	-12.304	0.044	Signifikan
	γ_{17}	0.1361	0.075	Tidak Signifikan
	γ_{18}	-0.2726	0.023	Signifikan
	γ_{19}	6.594	0.001	Signifikan
x_5	γ_{20}	-6.780	0.001	Signifikan

Untuk mengetahui apakah residual identik atau tidak, dapat dilakukan dengan uji secara visual dan uji *Glejser*. Uji secara vsual dilakukan dengan cara membuat plot antara nilai dugaan respon \hat{y} dengan residual. Apabila tidak terdapat pola pada plot maka tidak terjadi heterokedastisitas. Berikut adalah hasil plot antara nilai dugaan respond \hat{y} dan residual.



Gambar 2 Scatterplot antara Fits dan Residual

Tabel 9.
ANOVA dari Uji *Glejser*

Sumber	Df	Sum of Square	Mean Square	F _{hitung}	P-value
Regresi	20	1.190612	0.059530	2.301	0.087
Error	10	0.258609	0.025860		
Total	30	1.449222			

Berdasarkan plot pada Gambar 2 terlihat pencarannya menyebar ke segala arah dan tidak membentuk adanya suatu pola, sehingga secara visual asumsi identik telah terpenuhi. Selain itu untuk mengetahui terjadi heterokedastisitas dapat dilakukan dengan uji *Glejser*. Uji *Glejser* dilakukan dengan cara meregresikan nilai mutlak dari residual dengan variabel prediktor yang signifikan terhadap model. Berikut adalah hasil dari uji *Glejser*.

Dari ANOVA pada Tabel 9 dapat diketahui bahwa p-value dari uji *Glejser* adalah sebesar 0.087, yaitu lebih besar dari nilai $\alpha(0.05)$. Sehingga dapat diputuskan bahwa H_0 gagal ditolak. Jadi dapat diartikan bahwa tidak terjadi heteroskedastisitas. Hal ini menunjukkan residual telah memenuhi asumsi identik.

Selain itu asumsi independen juga harus terpenuhi, hipotesisnya sebagai berikut.

$$H_0 : \rho = 0 \text{ (residual independen)}$$

$$H_1 : \rho \neq 0 \text{ (residual tidak independen)}$$

Untuk melakukan apakah residual independen atau tidak dapat dilakukan dengan cara melihat plot ACF dari residual. Apabila terdapat minimal satu autokorelasi pada lag yang keluar dari batas signifikansi maka H_0 ditolak, yaitu residual tidak independen. Berikut adalah hasil plot ACF dari residual.

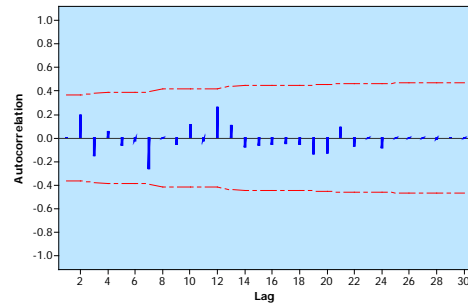
Berdasarkan plot ACF dari residual pada Gambar 3 terlihat bahwa autokorelasi pada semua lag berada di dalam batas atau bisa dikatakan tidak ada autokorelasi yang keluar dari batas. Sehingga dapat disimpulkan bahwa H_0 tidak ditolak, maka residual telah memenuhi asumsi independen. Selanjutnya asumsi yang harus dipenuhi adalah asumsi distribusi normal, hipotesisnya adalah sebagai berikut.

$$H_0 : F_0(x) = F(x) \text{ (residual berdistribusi normal)}$$

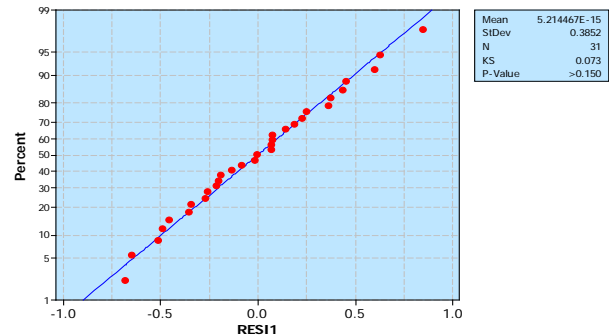
$$H_1 : F_0(x) \neq F(x) \text{ (residual tidak berdistribusi normal)}$$

Untuk melakukan pengujian distribusi normal dapat dilakukan dengan Uji Kolmogorov-Smirnov. Berikut adalah hasil pengujian dengan Kolmogorov-Smirnov.

Berdasarkan Gambar 4 terlihat bahwa p-value dari Uji Kolmogorov-Smirnov menunjukkan nilai >0.150 , nilai ini lebih dari nilai $\alpha(0.05)$. Maka dapat diputuskan H_0 tidak ditolak. Dapat disimpulkan bahwa residual telah berdistribusi normal.



Gambar 3 ACF dari Residual



Gambar 4. Hasil Uji Kolmogorov-Smirnov

1. Interpretasi Model

Langkah terakhir yang dilakukan adalah menginterpretasikan hasil dari model yang terbentuk untuk mengetahui seberapa besar perubahan CBR dari akibat perubahan variabel tertentu yang diduga. Misalnya untuk menginterpretasi variabel x_1 , apabila variabel x_2, x_3, x_4 , dan x_5 dianggap konstan, maka model tersebut dapat diinterpretasikan yaitu apabila persentase kepala keluarga berstatus pendidikan tidak sekolah/belum tamat SD berkisar antara 1.93% hingga 3.13% dan apabila persentase kepala keluarga berstatus pendidikan tidak sekolah/belum tamat SD pada keadaan ini naik 1%, maka CBR akan cenderung naik sebesar 0.0512. Wilayah yang termasuk dalam interval ini adalah Kecamatan Tegalsari, Sawahan, Genteng, Sukolilo, Pabean Cantian, Tandes, Benowo, Wiyung, Jambangan, Mulyorejo, Bulak, Pakal, dan Sambikerep. Selanjutnya apabila persentase kepala keluarga berstatus pendidikan tidak sekolah/belum tamat SD berkisar antara 3.13% hingga 3.23% dan apabila persentase kepala keluarga berstatus pendidikan tidak sekolah/belum tamat SD naik 1%, maka CBR akan cenderung turun sebesar 1.2888. Pada kondisi ini tidak ada wilayah yang masuk dalam interval ini. Apabila persentase kepala keluarga berstatus pendidikan tidak sekolah/belum tamat SD lebih dari 3.23% dan apabila persentase kepala keluarga berstatus pendidikan tidak sekolah/belum tamat SD naik 1%, maka CBR akan cenderung naik sebesar 0.0412. Wilayah yang masuk dalam interval ini adalah Kecamatan Tambak Sari, Krembangan, Kenjeran, dan Sukomanunggal.

Untuk menginterpretasi variabel x_2 , apabila variabel x_1, x_3, x_4 , dan x_5 dianggap konstan, maka model tersebut dapat diinterpretasikan yaitu apabila persentase kepala keluarga yang berstatus tidak bekerja kurang dari 2.15% maka pada kondisi ini apabila persentase kepala keluarga yang berstatus tidak bekerja naik 1%, maka CBR akan turun sebesar 0.0757. Wilayah yang termasuk dalam interval ini adalah

Kecamatan Kenjeran dan Benowo. Apabila persentase kepala keluarga yang berstatus tidak bekerja berkisar antara 2.15% hingga 3.56% dan apabila persentase kepala keluarga yang berstatus tidak bekerja pada keadaan ini naik 1%, maka CBR akan naik sebesar 0.0202. Wilayah yang masuk dalam interval ini adalah Kecamatan Karang Pilang, Wonocolo, Rungkut, Wonokromo, Sawahan, Gubeng, Sukolilo, Bubutan, Tandes, Lakarsantri, Wiyung, Dukuh Pakis, Gayungan, Mulyorejo, Sukomanunggal, Bulak, Pakal, dan Sambi Kerep. Selanjutnya interpretasi untuk variabel x_3 sampai x_5 dilakukann dengan cara yang sama seperti dijelaskan sebelumnya.

IV. KESIMPULAN DAN SARAN

A. Kesimpulan

Berdasarkan analisis dan pembahasan, maka dapat ditarik kesimpulan sebagai berikut.

1. Di Surabaya selama tahun 2012 terdapat 40.343 kelahiran. Rata-rata terdapat 12 bayi lahir setiap 1000 penduduk. Minimum terjadi sebanyak 9 kelahiran dan maksimum terjadi 15 kelahiran pada setiap 1000 penduduk. Keragaman data ditunjukkan dari nilai standar deviasi sebesar 1,619. Kecamatan yang memiliki angka kelahiran paling kecil adalah kecamatan Genteng yaitu 9,4. Sedangkan kecamatan dengan angka kelahiran tertinggi adalah Kecamatan Benowo dan Lakarsantri, yaitu sebesar 15,1.
2. Model regresi nonparametrik Spline *truncated* terbaik adalah dengan tiga titik knot. Model ini memiliki nilai GCV paling minimum yaitu 1,38 dengan R^2 sebesar 94,3% dan MSE yaitu 0,44. Berikut adalah model spline terbaik.

$$\hat{y} = 29.5 - 0.59 x_1 + 5.71(x_1 - 1.93)_+ - 134 (x_1 - 3.13)_+ + 133 (x_1 - 3.23)_+ - 7.57 x_2 + 9.59 (x_2 - 2.15)_+ + 16.9 (x_2 - 3.56)_+ + 14.4 (x_2 - 3.66)_+ - 48.9 x_3 + 140 (x_3 - 0.12)_+ - 940 (x_3 - 0.20)_+ + 856 (x_3 - 0.21)_+ + 1.33 x_4 + 1.68(x_4 - 2.46)_+ + 11.3(x_4 - 5.13)_+ + 12.3(x_4 - 5.34)_+ + 0.136 x_5 - 0.273 (x_5 - 31.63)_+ + 6.59 (x_5 - 47.00)_+ - 6.78 (x_5 - 48.18)_+$$
3. Berdasarkan uji parameter dapat diketahui bahwa semua variabel berpengaruh signifikan terhadap CBR. Variabel-variabel tersebut adalah persentase kepala keluarga yang berstatus pendidikan tidak sekolah/belum tamat SD, persentase kepala keluarga yang berstatus tidak bekerja, persentase kepala keluarga yang menikah pada umur 15-19 tahun, angka perkawinan kasar, dan angka migrasi masuk.

B. Saran

Saran yang dapat diberikan penulis adalah sebagai berikut.

1. Pada penelitian ini hanya terbatas pada tiga titik knot, selanjutnya dapat dikembangkan dengan lebih dari tiga knot.
2. Penelitian ini terbatas pada penggunaan regresi Spline *truncated* linier (orde satu). Untuk penelitian selanjutnya dapat dikembangkan dengan regresi orde dua dan tiga.
3. Data pada penelitian ini terbatas pada data *cross section*, untuk selanjutnya dapat dikembangkan dengan data longitudinal yaitu gabungan antara *cross section* dan *time series*.

DAFTAR PUSTAKA

- [1] Badan Pusat Statistika. (2010). *Laporan Hasil Sensus Penduduk 2010 Data Agregat Per Provinsi*. Jakarta: Badan Pusat Statistika.
- [2] Lucas, D., & Blake. (1982). *Pengantar Kependudukan*. Yogyakarta: UGM Press.
- [3] World Bank. (2011). *Public Data*. Retrieved January 14, 2011, from https://www.google.com/publicdata/explore?ds=d5bncppjof8f9_&met_y=sp_dyn_tfrt_in&idim=country:IDN&dl=id&hl=id&q=tingkat%20kelahiran%20di%20indonesia#!ctype=l&strail=false&bcs=d&nslm=h&met_y=sp_dyn_cbrt_in&scale_y=lin&ind_y=false&rdim=region&idim=country:
- [4] Dinas Kependudukan dan Catatan Sipil Kota Surabaya. (2012). *Buku Profil Perkembangan Kependudukan Kota Surabaya tahun 2012*. Surabaya: Dinas Kependudukan dan Catatan Sipil Kota Surabaya.
- [5] Eubank. (1988). *Spline Smoothing and Nonparametric Regression*. Marcel Dekker New York.
- [6] Budiantara, I. N. (2009). *Spline dalam Regresi Nonparametrik: Sebuah Pemodelan Statistika Masa Kini dan Masa Mendatang*. Surabaya: Jurusan Statistika, Fakultas Matematika dan Ilmu Pengetahuan Alam, Institut Teknologi Sepuluh Nopember.
- [7] Sugiantari, A. P., & Biantara, I. N. (2013). *Analisis Faktor-Faktor yang Mempengaruhi Angka Harapan Hidup di Jawa Timur Menggunakan Regresi Semiparametrik Spline*. Surabaya: Tugas Akhir, Jurusan Statistika, Fakultas Matematika dan Ilmu Pengetahuan Alam, Institut Teknologi Sepuluh Nopember.
- [8] Eubank, R. L., & Thomas, W. (1993). Detecting Heterocedasticity in Nonparametric Regression. *Journal of the American Statistical Association*, 387-392.