

# Pemodelan Indeks Pembangunan Gender dengan Pendekatan Regresi Nonparametrik Spline di Indonesia

Nurul Fajriyyah dan I Nyoman Budiantara

Jurusan Statistika, Fakultas MIPA, Institut Teknologi Sepuluh Nopember (ITS)

Jl. Arief Rahman Hakim, Surabaya 60111 Indonesia

*e-mail*: i\_nyoman\_b@statistika.its.ac.id

Indeks Pembangunan Gender (IPG) merupakan indeks pencapaian kemampuan dasar pembangunan manusia yang sama seperti Indeks Pembangunan Manusia (IPM) dengan memperhatikan ketimpangan gender. Indonesia memiliki IPG yang rendah jika dibandingkan dengan negara lain seperti Malaysia dan Australia. Dalam hal pembangunan manusia sering dibahas mengenai perbedaan gender, dimana berfokus pada bagaimana mencapai kesetaraan gender dengan meningkatkan kualitas sumber daya manusia tanpa membedakan laki-laki dan perempuan. Hal ini terkait dengan tujuan dari MDGs yaitu mendorong kesetaraan gender dan pemberdayaan perempuan. Untuk mengatasi permasalahan tersebut perlu diselidiki faktor-faktor yang diduga berpengaruh terhadap IPG di Indonesia menggunakan regresi nonparametrik spline. Pendekatan regresi nonparametrik spline dapat digunakan untuk memodelkan IPG di Indonesia karena pola data pada penelitian ini tidak membentuk suatu pola tertentu. Berdasarkan penelitian ini, model regresi nonparametrik spline terbaik adalah spline yang memiliki nilai GCV minimum yaitu kombinasi knot (1,2,1,3,3,3,2,3) dengan semua variabel signifikan yaitu Angka Partisipasi Sekolah (APS) SD/Sederajat penduduk perempuan, APS SMP/Sederajat penduduk perempuan, APS SMA/Sederajat penduduk perempuan, Angka Buta Huruf penduduk perempuan, Tingkat Partisipasi Angkatan Kerja penduduk perempuan, rasio jenis kelamin, rasio jenis kelamin saat lahir, dan persentase penduduk perempuan mempunyai keluhan kesehatan. Regresi spline linier menghasilkan  $R^2$  sebesar 99,81%.

**Kata Kunci**—GCV, IPG, Regresi Nonparametrik, Spline, Titik Knot

## I. PENDAHULUAN

SUATU negara dikatakan maju dapat dilihat dari tingginya pembangunan diseluruh aspek kehidupan bangsa dan negara. Dalam hal pembangunan manusia sering dibahas mengenai perbedaan gender, dimana berfokus pada bagaimana mencapai kesetaraan gender untuk meningkatkan kualitas sumber daya manusia tanpa membedakan laki-laki dan perempuan. Salah satu tujuan dari *Millenium Development Goals* (MDGs) adalah mendorong kesetaraan gender dan pemberdayaan perempuan [1]

Dalam *Women Studies* Ensiklopedia dijelaskan bahwa gender merupakan suatu konsep kultural, berupaya membuat perbedaan (*distinction*) dalam hal peran, perilaku, mentalitas, dan karakteristik emosional antara laki-laki dan perempuan yang berkembang dalam masyarakat. Secara umum gender dapat didefinisikan sebagai perbedaan peran, kedudukan dan sifat yang dilekatkan pada kaum laki-laki maupun perempuan melalui konstruksi secara sosial maupun kultural [2].

Perbedaan gender terlihat dari kecenderungan peran masing-masing, yaitu berperan dalam publik atau domestik. Peran publik diartikan dengan aktivitas yang dilakukan di luar rumah dan bertujuan mendapatkan penghasilan. Sedangkan peran domestik adalah aktivitas yang dilakukan di dalam rumah berkaitan dengan kerumahtanggaan dan tidak dimaksudkan untuk mendapat penghasilan [1].

Indeks Pembangunan Gender (IPG) merupakan indeks pencapaian kemampuan dasar pembangunan manusia yang sama seperti Indeks Pembangunan Manusia (IPM) dengan memperhatikan ketimpangan gender. IPG memiliki komponen pembentuk yang menentukan nilai dari IPG. Komponen pembentuk tersebut sama dengan yang digunakan dalam pengukuran IPM, yakni komponen dari dimensi kesehatan, pengetahuan, dan ekonomi [1]. Dari aspek kesehatan, kesenjangan gender dapat dilihat dari persentase penduduk yang mempunyai keluhan kesehatan. Dari aspek pengetahuan, kesenjangan gender dapat terlihat dari Angka Buta Huruf (ABH) serta Angka Partisipasi Sekolah (APS) di tiga jenjang pendidikan yaitu SD, SMP, dan SMA/ sederajat. Dari aspek ekonomi, kesenjangan gender dapat terlihat dari Tingkat Partisipasi Angkatan Kerja (TPAK). Sedangkan rasio jenis kelamin dan rasio jenis kelamin saat lahir merupakan indikator yang dapat menjelaskan mengenai orientasi gender [3].

Untuk mengetahui faktor apa saja yang mempengaruhi IPG di Indonesia, maka dilakukan pemodelan salah satunya dengan menggunakan regresi nonparametrik spline, karena data pada penelitian ini tidak membentuk suatu pola tertentu atau merupakan data nonparametrik. Model regresi nonparametrik spline merupakan model regresi yang mempunyai interpretasi statistik dan visual sangat khusus dan sangat baik. Pada penelitian ini akan diteliti IPG di Indonesia karena Indonesia termasuk negara yang memiliki angka IPG yang rendah jika dibandingkan dengan negara lain seperti Malaysia, Australia, dan Filipina. Variabel pada penelitian ini merujuk pada penelitian mengenai IPG yang dilakukan oleh [3] mengenai analisis faktor-faktor yang mempengaruhi IPG dengan menggunakan regresi probit serta [4] mengenai disparitas gender. Penelitian dengan menggunakan regresi nonparametrik spline sebelumnya pernah dilakukan oleh [5] mengenai faktor yang mempengaruhi kemiskinan di Jawa Tengah. Selain itu, [6] dalam penelitiannya mengenai angka kematian maternal di Jawa Timur.

II. TINJAUAN PUSTAKA

A. Indeks Pembangunan Gender

Dalam *Women Studies* Ensiklopedia dijelaskan bahwa gender merupakan suatu konsep kultural, berupaya membuat perbedaan (*distinction*) dalam hal peran, perilaku, mentalitas, dan karakteristik emosional antara laki-laki dan perempuan yang berkembang dalam masyarakat. Indeks Pembangunan Gender (IPG) merupakan indeks pencapaian kemampuan dasar pembangunan manusia yang sama seperti Indeks Pembangunan Manusia (IPM) dengan memperhatikan ketimpangan gender. IPG digunakan untuk mengukur pencapaian dalam dimensi yang sama dan menggunakan indikator yang sama dengan IPM, namun lebih diarahkan untuk mengungkapkan ketimpangan antara laki-laki dan perempuan. IPG dapat digunakan untuk mengetahui kesenjangan pembangunan manusia antara laki-laki dan perempuan [1].

IPG memiliki komponen pembentuk yang menentukan nilai dari IPG. Komponen pembentuk tersebut sama dengan yang digunakan dalam pengukuran IPM, yakni angka harapan hidup (mewakili dimensi kesehatan), angka melek huruf dan rata-rata lama sekolah (mewakili dimensi pengetahuan), serta sumbangan pendapatan (mewakili dimensi ekonomi) yang disajikan menurut jenis kelamin. Dengan kata lain, dinamika IPG dari waktu ke waktu sangat dipengaruhi oleh perubahan dari tiga komponen tersebut.

B. Regresi Nonparametrik Spline

Regresi nonparametrik adalah metode statistika yang digunakan untuk mengetahui hubungan antara variabel respon dan prediktor yang bentuk fungsinya tidak diketahui (kurva regresi tidak membentuk suatu pola), memiliki asumsi fungsi *smooth* (mulus), oleh karena itu regresi nonparametrik mempunyai fleksibilitas yang tinggi [7]. Model regresi nonparametrik secara umum dapat disajikan sebagai berikut:

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

dimana  $y_i$  merupakan variabel respon,  $x_i$  merupakan variabel prediktor,  $f(x_i)$  merupakan fungsi regresi yang tidak mengikuti pola tertentu serta  $\varepsilon_i \sim \text{IIDN}(0, \sigma^2)$ .

Spline adalah potongan polinomial yang mempunyai sifat tersegmen. Sifat tersegmen ini memberikan fleksibilitas lebih dari polinomial biasa, sehingga memiliki kelebihan untuk dapat menyesuaikan diri dengan lebih efektif terhadap karakteristik lokal suatu data. Misalnya terdapat data  $(x_{1i}, x_{2i}, \dots, x_{pi}, y_i)$ , hubungan antara  $(x_{1i}, x_{2i}, \dots, x_{pi})$  dan  $y_i$  (variabel respon) didekati dengan model regresi nonparametrik spline  $y_i = f(x_{1i}, x_{2i}, \dots, x_{pi}) + \varepsilon_i$ , dengan  $i = 1, 2, 3, \dots, n$ , dan  $f$  kurva regresi yang tidak diketahui bentuknya. Apabila kurva regresi  $f$  merupakan model aditif dan dihipotesis dengan fungsi spline maka diperoleh model regresi sebagai berikut:

$$y_i = \sum_{j=1}^p f(x_{ji}) + \varepsilon_i \quad ; i = 1, 2, 3, \dots, n \quad (2)$$

dimana,

$$f(x_{ji}) = \sum_{h=0}^q \beta_{hj} x_{ji}^h + \sum_{l=1}^m \beta_{(q+l)j} (x_{ji} - K_{lj})_+^q \quad (3)$$

$$\text{dengan } (x_{ji} - K_{lj})_+^q = \begin{cases} (x_{ji} - K_{lj})^q, & x_{ji} \geq K_{lj} \\ 0, & x_{ji} < K_{lj} \end{cases}$$

dan  $K_{1j}, K_{2j}, \dots, K_{mj}$  merupakan titik knot yang menunjukkan bentuk perubahan perilaku dari fungsi pada sub-sub interval tertentu. Nilai  $q$  merupakan derajat polinomial. Kurva polinomial derajat satu disebut kurva linier, kurva polinomial derajat dua disebut kurva kuadratik serta derajat tiga disebut kurva kubik. Persamaan (2) dapat diuraikan sebagai berikut:

$$y_i = \beta_{01} + \beta_{11}x_{1i} + \dots + \beta_{q1}x_{1i}^q + \alpha_{11}(x_{1i} - K_{11})_+^q + \dots + \alpha_{m1}(x_{1i} - K_{m1})_+^q + \beta_{02} + \beta_{12}x_{2i} + \dots + \beta_{q2}x_{2i}^q + \alpha_{12}(x_{2i} - K_{12})_+^q + \dots + \alpha_{m2}(x_{2i} - K_{m2})_+^q + \dots + \beta_{0p} + \beta_{1p}x_{pi} + \dots + \beta_{qp}x_{pi}^q + \alpha_{1p}(x_{pi} - K_{1p})_+^q + \dots + \alpha_{mp}(x_{pi} - K_{mp})_+^q$$

Estimasi untuk  $\beta$  sebagai berikut:

$$\hat{\beta} = (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{y} \quad (4)$$

C. Pemilihan Titik Knot Optimal

Model regresi spline terbaik merupakan model yang memiliki titik knot optimal. Titik knot merupakan titik perpaduan bersama dimana terdapat perubahan perilaku pada data. Salah satu metode yang digunakan untuk pemilihan titik knot optimal adalah *Generalized Cross Validation* (GCV). Metode GCV mempunyai sifat optimal asimtotik jika dibandingkan dengan metode lain, misalnya *Cross Validation* (CV) [8]. Model spline dengan nilai GCV terkecil dari titik knot optimal merupakan model spline yang terbaik [9]. Fungsi GCV sebagai berikut [7]:

$$GCV(k_1, k_2, \dots, k_j) = \frac{MSE(k_1, k_2, \dots, k_j)}{(n^{-1} \text{trace} [\mathbf{I} - \mathbf{A}(k_1, k_2, \dots, k_j)])^2} \quad (5)$$

dimana  $\mathbf{A}(k_1, k_2, \dots, k_j) = \mathbf{X}(\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T$  dan  $MSE(k_1, k_2, \dots, k_j)$  sebagai berikut.

$$MSE(k_1, k_2, \dots, k_j) = n^{-1} \sum_{i=1}^n (y_i - \hat{y}_i)^2 \quad (6)$$

dengan  $\hat{y} = \mathbf{A}(k_1, k_2, \dots, k_j) \mathbf{y}$ .

D. Pengujian Parameter Model Regresi

Pengujian parameter model regresi dapat dilakukan secara serentak dan secara individu.

1. Pengujian Parameter Model Secara Serentak

Pengujian parameter model secara serentak merupakan uji parameter kurva regresi secara simultan dengan menggunakan uji  $F$ . Hipotesis pada uji  $F$  sebagai berikut.

$$H_0 : \beta_1 = \beta_2 = \dots = \beta_{(q+m)} = 0$$

$$H_1 : \text{minimal ada } \beta_h \neq 0; \quad h = 1, 2, \dots, (q + m)$$

Statistik uji yang digunakan sebagaimana Persamaan 7.

$$F_{hitung} = \frac{MSR}{MSE} = \frac{SSR/df \text{ regresi}}{SSE/df \text{ error}} \quad (7)$$

dengan estimasi  $\mathbf{y}$  pada Persamaan 8.

$$\hat{y} = \mathbf{X}\hat{\beta} \quad (8)$$

Tolak  $H_0$  jika  $F_{hitung} \geq F_{\alpha:(q+m, n-(q+m)-1)}$  atau  $p\text{-value} < \alpha$  menunjukkan bahwa paling sedikit terdapat satu parameter yang tidak sama dengan nol.

2. Pengujian Parameter Model Secara Parsial

Uji parsial digunakan untuk mengetahui parameter yang berpengaruh signifikan secara individu terhadap model menggunakan uji  $t$ . Hipotesis pada uji  $t$  sebagai berikut:

$$H_0 : \beta_h = 0$$

$$H_1 : \beta_h \neq 0; h = 1, 2, \dots, (q + m)$$

Statistik uji  $t$  pada Persamaan 9.

$$t_{hitung} = \frac{\hat{\beta}_h}{SE(\hat{\beta}_h)} \quad (9)$$

Daerah kritis: tolak  $H_0$  jika  $|t_{hitung}| > t_{\left(\frac{\alpha}{2}, n-(q+m)-1\right)}$ , dimana

$n$  merupakan jumlah pengamatan dan  $(q + m)$  merupakan banyaknya parameter. Kebaikan model dapat dilihat dari nilai  $R^2$  pada Persamaan 10.

$$R^2 = \frac{SSR}{SST} \times 100\% \quad (10)$$

E. Pengujian Asumsi Residual Model Regresi

Pengujian asumsi residual (*goodness of fit*) model regresi dilakukan untuk mengetahui apakah residual yang dihasilkan sudah memenuhi ketiga asumsi yaitu identik, independen, dan berdistribusi normal.

1. Asumsi Residual Identik

Asumsi residual identik terpenuhi jika varians antar residual sama  $\sigma^2$  dan tidak terjadi heteroskedastisitas.

$$Var(y_i) = Var(\varepsilon_i) = \sigma^2; i = 1, 2, \dots, n \quad (11)$$

Asumsi identik terpenuhi jika sebaran plot tidak membentuk suatu pola tertentu (tersebar secara acak). Bila sebaran plot membentuk pola tertentu maka mengindikasikan terjadi heteroskedastisitas. Identifikasi heteroskedastisitas selain dilakukan dengan menggunakan metode grafis, dapat juga dilakukan dengan menggunakan uji Glejser. Hipotesis yang digunakan adalah sebagai berikut.

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

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

Statistik uji yang digunakan sebagaimana Persamaan 12.

$$F_{hitung} = \frac{\sum_{i=1}^n (|\hat{e}_i| - |\hat{e}|)^2 / (q + m) - 1}{\sum_{i=1}^n (|e_i| - |\hat{e}|)^2 / n - (q + m)} \quad (12)$$

Daerah penolakan yakni tolak  $H_0$  jika:

$$F_{hitung} > F_{\alpha:(q+m)-1, n-(q+m)} \text{ atau } p\text{-value} < \alpha. \text{ Dimana nilai } (q + m) \text{ adalah banyaknya parameter.}$$

2. Asumsi Residual Independen

Asumsi residual undependen adalah tidak terdapat korelasi pada residual yang ditunjukkan oleh nilai kovarian antara  $\varepsilon_i$  dan  $\varepsilon_j$  sama dengan nol. Persamaan untuk ACF adalah sebagai berikut [10].

$$\rho_k = \frac{Cov(e_t, e_{t+k})}{\sqrt{Var(e_t)}\sqrt{Var(e_{t+k})}} = \frac{\gamma_k}{\gamma_0} \quad (13)$$

dimana:

$\rho_k$  : korelasi antara  $e_t$  dan  $e_{t+k}$

$\gamma_k$  : kovarian antara  $e_t$  dan  $e_{t+k}$

$\gamma_0$  :  $Var(e_t) = Var(e_{t+k})$

Interval konfidensi untuk  $\rho_k$  sebagai berikut:

$$-t_{n-1, \alpha/2} SE(\hat{\rho}_k) < \rho_k < t_{n-1, \alpha/2} SE(\hat{\rho}_k) \quad (14)$$

Bila tidak terdapat  $\rho_k$  yang keluar dari batas signifikansi menunjukkan bahwa asumsi independen terpenuhi.

3. Asumsi Residual Normal

Asumsi residual Normal dapat dilakukan dengan menggunakan uji *Kolmogorov-Smirnov* dengan hipotesis sebagai berikut :

$H_0$ : Residual mengikuti distribusi Normal

$H_1$ : Residual tidak mengikuti distribusi Normal

Statistik uji yang digunakan yaitu uji *Kolmogorov-Smirnov*.

$$Z_{hitung} = Sup_x |F_n(x) - F_0(x)| \quad (15)$$

Daerah penolakan yaitu tolak  $H_0$  apabila  $Z_{hitung} > Z_\alpha$  atau  $p\text{-value} < \alpha$ .

III. METODOLOGI PENELITIAN

Data yang digunakan dalam penelitian ini adalah data sekunder. Data yang diambil merupakan data publikasi dari Badan Pusat Statistik (BPS) Republik Indonesia (RI) pada tahun 2012. Unit observasi yang digunakan dalam penelitian ini adalah 33 provinsi yang ada di Indonesia. Variabel yang digunakan adalah IPG ( $y$ ), Angka Partisipasi Sekolah (APS) SD/Sederajat penduduk perempuan ( $x_1$ ), Angka Partisipasi Sekolah (APS) SMP/Sederajat penduduk perempuan ( $x_2$ ), Angka Partisipasi Sekolah (APS) SMA/Sederajat penduduk perempuan ( $x_3$ ), Angka Buta Huruf (ABH) penduduk perempuan ( $x_4$ ), Tingkat Partisipasi Angkatan Kerja (TPAK) penduduk perempuan ( $x_5$ ), rasio jenis kelamin ( $x_6$ ), rasio jenis kelamin saat lahir ( $x_7$ ), dan persentase penduduk perempuan mempunyai keluhan kesehatan ( $x_8$ ). Langkah-langkah dalam analisis data meliputi.

1. Mendeskripsikan karakteristik IPG di Indonesia

2. Memodelkan IPG di Indonesia dengan pendekatan Spline.

Langkah yang digunakan yaitu

a. membuat *scatter plot* antara IPG di Indonesia dengan masing-masing variabel yang diduga berpengaruh.

b. memodelkan data dengan pendekatan spline satu, dua, tiga, dan kombinasi knot.

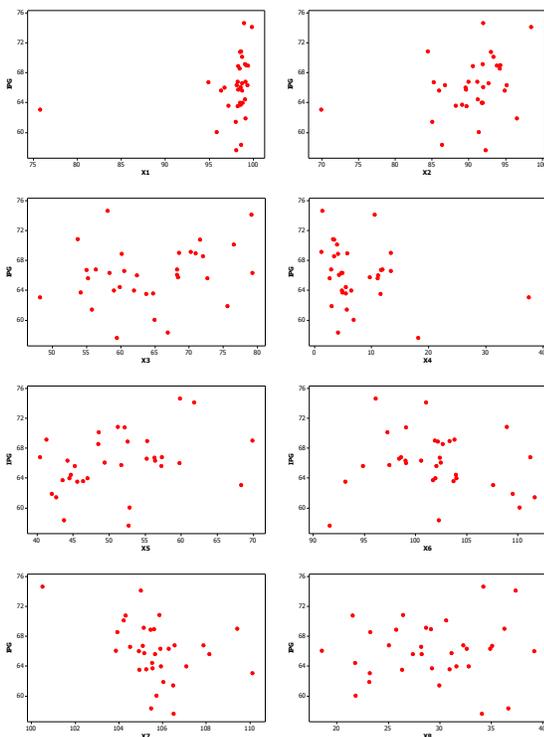
- c. memilih titik knot optimal berdasarkan nilai GCV yang paling minimum.
  - d. mendapatkan model regresi spline dengan titik knot optimal.
  - e. menguji signifikansi parameter regresi spline secara serentak.
  - f. melakukan uji parameter regresi spline secara parsial.
  - g. menguji asumsi residual.
  - h. menginterpretasikan model.
3. Menarik Kesimpulan

IV. ANALISIS DAN PEMBAHASAN

A. Karakteristik IPG di Indonesia Tahun 2012

Karakteristik IPG di Indonesia tahun 2012 meliputi rata-rata, varians, nilai minimum, dan nilai maksimum. Rata-rata IPG di Indonesia tahun 2012 sebesar 66,024 dengan varians sebesar 15,306. IPG minimum terdapat pada provinsi Nusa Tenggara Barat sebesar 57,58. Sedangkan IPG maksimum terdapat pada provinsi D.K.I Jakarta sebesar 74,66. Hal ini berarti bahwa angka IPG dari 33 provinsi di Indonesia pada tahun 2012 mencapai 57,580 hingga 74,660.

Berdasarkan scatterplot antara variabel  $y$  dengan  $x_1$  hingga  $x_8$  dapat diketahui bahwa pola hubungan antara variabel  $y$  dengan  $x_j$  hingga  $x_8$  tidak membentuk suatu pola tertentu sehingga hal ini mengindikasikan bahwa terdapat komponen nonparametrik. Oleh karena itu dapat dilanjutkan pada pemodelan menggunakan regresi nonparametrik spline. Pola hubungan antara variabel respon dan variabel prediktor ditampilkan pada Gambar 1.



Gambar 1. Pola Hubungan Antara Variabel Respon dan Prediktor

Pemodelan dengan menggunakan regresi nonparametrik spline satu, dua, tiga, dan kombinasi titik knot.

Dari hasil pemodelan didapatkan nilai GCV minimum dari masing-masing model regresi spline linier. Nilai GCV minimum dari pemodelan menggunakan knot satu, dua, tiga, dan kombinasi titik knot, dipilih nilai GCV paling minimum. Berikut merupakan nilai GCV untuk masing-masing titik knot yang disajikan pada Tabel 1.

Tabel 1.  
Nilai GCV Minimum dari Model Regresi Spline

Model Regresi Spline Linier	Nilai GCV
1 knot	9,7654
2 knot	3,7209
3 knot	0,4176
Kombinasi 3 knot	0,3013

Berdasarkan Tabel 1 diperoleh nilai GCV minimum dari pemodelan menggunakan knot satu, dua, tiga, dan kombinasi titik knot, dipilih nilai GCV paling minimum yaitu model dengan kombinasi titik knot dengan nilai GCV minimum sebesar 0,3013 dengan titik knot optimum pada masing-masing variabel prediktor sebagai berikut:

- $x_1 = 76,7624$
- $x_2 = 70,4647$  dan  $86,2514$
- $x_3 = 49,4965$
- $x_4 = 4,2049; 21,9943; \text{ dan } 26,4416$
- $x_5 = 42,8165; 57,2557; \text{ dan } 60,8655$
- $x_6 = 93,2567; 103,0771; \text{ dan } 105,5322$
- $x_7 = 100,6965 \text{ dan } 106,0029$
- $x_8 = 20,2333; 30,3329; \text{ dan } 32,8578$

B. Pengujian Parameter Model Regresi Spline

Pengujian parameter dapat diuji secara serentak dan secara individu. Pertama dilakukan pengujian parameter secara serentak lalu dilakukan pengujian parameter secara individu.

Pengujian parameter secara serentak merupakan uji parameter kurva regresi secara simultan menggunakan uji  $F$ . Hipotesis yang digunakan sebagai berikut.

$$H_0 : \beta_1 = \beta_2 = \dots = \beta_{26} = 0$$

$$H_1 : \text{minimal ada } \beta_h \neq 0; h = 1, 2, \dots, 26$$

Berikut merupakan tabel ANOVA dari model regresi spline yang ditampilkan pada Tabel 2.

Tabel 2.  
ANOVA Model Regresi Spline

Sumber Variasi	DF	SS	MS	F hitung
Regresi	26	488,8768	18,80295	
Error	6	0,9130154	0,1521692	123,5661
Total	32	489,7898	-	

Berdasarkan hasil tabel ANOVA pada Tabel 2 diperoleh nilai  $p$ -value sebesar 0,000002898. Karena  $p$ -value  $< \alpha$  dengan  $\alpha = 0,05$  maka dapat diambil keputusan tolak  $H_0$  yang menunjukkan bahwa minimal ada satu parameter yang berpengaruh signifikan terhadap model.

Selanjutnya dilakukan pengujian secara individu yang digunakan untuk mengetahui parameter yang berpengaruh signifikan secara individu terhadap model menggunakan uji  $t$ . Hipotesis yang digunakan sebagai berikut.

$$H_0 : \beta_h = 0$$

$$H_1 : \beta_h \neq 0; h = 1, 2, \dots, 26$$

Berikut merupakan hasil pengujian parameter secara individu dari model regresi spline yang ditampilkan pada Tabel 3.

Tabel 3.  
Hasil Pengujian Parameter Secara Individu

Variabel	Parameter	Estimator	<i>p</i> - value	Keterangan
<i>Constants</i>	$\beta_0$	-0,599	0,003	signifikan
	$\beta_1$	-22,649	0,004	signifikan
$x_1$	$\beta_2$	22,988	0,003	signifikan
	$\beta_3$	-23,898	0,002	signifikan
$x_2$	$\beta_4$	18,106	0,008	signifikan
	$\beta_5$	5,984	0,000	signifikan
$x_3$	$\beta_6$	-14,607	0,004	signifikan
	$\beta_7$	14,597	0,004	signifikan
$x_4$	$\beta_8$	1,157	0,006	signifikan
	$\beta_9$	-1,838	0,001	signifikan
	$\beta_{10}$	-5,467	0,000	signifikan
	$\beta_{11}$	-3,905	0,000	signifikan
$x_5$	$\beta_{12}$	-5,293	0,000	signifikan
	$\beta_{13}$	5,750	0,000	signifikan
	$\beta_{14}$	2,703	0,000	signifikan
	$\beta_{15}$	-3,418	0,000	signifikan
	$\beta_{16}$	1,474	0,010	signifikan
$x_6$	$\beta_{17}$	-1,784	0,005	signifikan
	$\beta_{18}$	-0,003	0,993	tidak sig*
	$\beta_{19}$	-0,832	0,113	tidak sig*
$x_7$	$\beta_{20}$	43,105	0,002	signifikan
	$\beta_{21}$	-44,529	0,002	signifikan
	$\beta_{22}$	1,407	0,006	signifikan
$x_8$	$\beta_{23}$	2,350	0,000	signifikan
	$\beta_{24}$	-2,429	0,000	signifikan
	$\beta_{25}$	0,376	0,078	tidak sig*
	$\beta_{26}$	-1,788	0,001	signifikan

\*) sig = signifikan

Berdasarkan hasil pengujian parameter secara individu pada Tabel 3 diperoleh hasil bahwa dari 26 parameter yang ada pada model regresi nonparametrik spline, terdapat 3 parameter yang tidak signifikan pada taraf signifikansi 0,05, karena *p*-value yang didapat lebih dari  $\alpha$  (0,05). Parameter  $\beta_0, \beta_1, \beta_2, \beta_3, \beta_4, \beta_5, \beta_6, \beta_7, \beta_8, \beta_9, \beta_{10}, \beta_{11}, \beta_{12}, \beta_{13}, \beta_{14}, \beta_{15}, \beta_{16}, \beta_{17}, \beta_{20}, \beta_{21}, \beta_{22}, \beta_{23}, \beta_{24},$  dan  $\beta_{26}$  berpengaruh signifikan terhadap model. Sedangkan parameter  $\beta_{18}, \beta_{19},$  dan  $\beta_{25}$  tidak berpengaruh signifikan terhadap model. Walaupun terdapat 3 parameter yang tidak signifikan, namun semua variabel berpengaruh signifikan terhadap IPG.

C. Pengujian Asumsi Residual

Pengujian asumsi residual dilakukan untuk mengetahui apakah residual yang dihasilkan telah memenuhi asumsi yakni identik, independen, dan berdistribusi normal. Asumsi identik terpenuhi bila varians antar residual sama dan tidak terjadi heteroskedastisitas. identifikasi heteroskedastisitas dapat dilakukan dengan menggunakan uji Glejser. Hipotesis yang digunakan sebagai berikut.

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

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

Berikut merupakan hasil pengujian asumsi residual identik

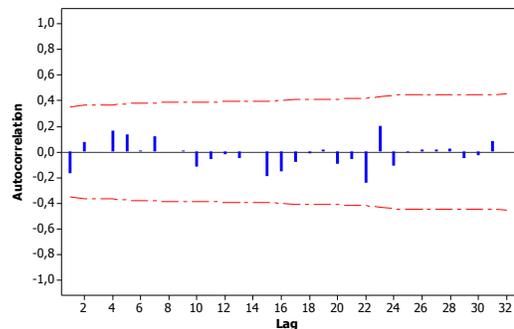
menggunakan uji Glejser yang ditampilkan pada Tabel 4.

Tabel 4.  
ANOVA Uji Glejser

Sumber Variasi	DF	SS	MS	F hitung
Regresi	26	0,3290614	0,01265621	
Error	6	0,1411147	0,02351911	0,5381245
Total	32	0,4701761	-	

Berdasarkan hasil tabel ANOVA pada Tabel 4 diperoleh nilai statistik uji *F* sebesar 0,538. Jika dibandingkan dengan nilai *F* tabel yaitu  $F(0,05;26,6)$  sebesar 2,47 maka *F* hitung lebih kecil dari nilai *F* tabel, oleh karena itu dapat diambil keputusan gagal tolak  $H_0$ . Hal ini diperkuat dengan *p*-value sebesar 0,873. Karena diperoleh *p*-value  $> \alpha$  dengan  $\alpha = 0,05$  maka dapat diambil keputusan gagal tolak  $H_0$  yang menunjukkan bahwa tidak terdapat heteroskedastisitas. Sehingga dapat dikatakan bahwa asumsi identik pada residual terpenuhi.

Asumsi residual independen merupakan asumsi kedua yang harus dipenuhi yaitu tidak terdapat korelasi pada residual yang ditunjukkan oleh nilai kovarian antara  $\epsilon_i$  dan  $\epsilon_j$  sama dengan nol. Uji asumsi independen dilakukan dengan pendekatan grafis menggunakan ACF plot sebagai berikut.



Gambar 2. Residual ACF Plot

Berdasarkan Gambar 2 dapat dilihat bahwa tidak terdapat  $\rho_k$  yang keluar dari batas signifikansi. Hal ini menunjukkan bahwa tidak terdapat kasus autokorelasi. Sehingga dapat disimpulkan bahwa asumsi residual independen terpenuhi.

Asumsi yang ketiga yaitu residual dari model regresi harus mengikuti distribusi Normal dengan mean nol dan varians  $\sigma^2$ . Uji asumsi distribusi Normal dapat dilakukan menggunakan uji *Kolmogorov-Smirnov* dengan uji hipotesis sebagai berikut.

$$H_0: \text{Residual mengikuti distribusi Normal}$$

$$H_1: \text{Residual tidak mengikuti distribusi Normal}$$

Berikut merupakan hasil pengujian asumsi residual Normal dengan menggunakan uji *Kolmogorov-Smirnov* yang ditampilkan pada Tabel 5.

Tabel 5.  
Hasil Uji *Kolmogorov-Smirnov*

<i>Kolmogorov-Smirnov</i>	<i>p</i> -value
0,146	0,073

Berdasarkan hasil uji *Kolmogorov-Smirnov* pada Tabel 5 diperoleh nilai  $Z_{\text{hitung}}$  sebesar 0,146. Jika dibandingkan dengan nilai  $Z_{\text{tabel}}$  sebesar 1,96 maka  $Z_{\text{hitung}}$  kurang dari  $Z_{\text{tabel}}$ , oleh

karena itu dapat diambil keputusan gagal tolak  $H_0$ . Hal ini diperkuat dengan  $p\text{-value} = 0,073$  maka  $p\text{-value} > \alpha (0,05)$ . Sehingga dapat diambil keputusan gagal tolak  $H_0$  yang menunjukkan bahwa residual berdistribusi normal.

*D. Interpretasi Hasil Model Regresi Spline Linier*

Dalam pemodelan menggunakan regresi spline linier menggunakan taraf signifikansi  $\alpha = 0,05$  didapat kesimpulan bahwa faktor-faktor yang mempengaruhi IPG di Indonesia adalah APS SD/Sederajat penduduk perempuan ( $x_1$ ), APS SMP/Sederajat penduduk perempuan ( $x_2$ ), APS SMA/Sederajat penduduk perempuan ( $x_3$ ), Angka Buta Huruf penduduk perempuan ( $x_4$ ), TPAK penduduk perempuan ( $x_5$ ), rasio jenis kelamin ( $x_6$ ), rasio jenis kelamin saat lahir ( $x_7$ ), dan persentase penduduk perempuan mempunyai keluhan kesehatan ( $x_8$ ). Model regresi nonparametrik spline untuk IPG di Indonesia seperti berikut.

$$\hat{y} = -0,599 - 22,649x_1 + 22,988(x_1 - 76,76)_+^1 - 23,898x_2 + 18,106(x_2 - 70,46)_+^1 + 5,984(x_2 - 86,25)_+^1 - 14,607x_3 + 14,597(x_3 - 49,5)_+^1 + 1,157x_4 - 1,838(x_4 - 4,2)_+^1 + -5,467(x_4 - 21,99)_+^1 - 3,905(x_4 - 26,44)_+^1 - 5,293x_5 + 5,75(x_5 - 42,82)_+^1 + 2,703(x_5 - 57,26)_+^1 + -3,418(x_5 - 60,86)_+^1 + 1,474x_6 - 1,784(x_6 - 93,26)_+^1 + -0,003(x_6 - 103,08)_+^1 - 0,833(x_6 - 105,53)_+^1 + 43,105x_7 - 44,529(x_7 - 100,7)_+^1 + 1,407(x_7 - 106,003)_+^1 + 2,35x_8 + -2,429(x_8 - 20,23)_+^1 + 0,376(x_8 - 30,33)_+^1 + -1,788(x_8 - 32,86)_+^1$$

Kemudian, interpretasi dari model spline terbaik adalah apabila variabel  $x_2, x_3, x_4, x_5, x_6, x_7,$  dan  $x_8$  diasumsikan konstan, maka pengaruh APS SD/Sederajat penduduk perempuan terhadap IPG sebagai berikut:

$$\hat{y} = -22,649x_1 + 22,988(x_1 - 76,76)_+^1 = \begin{cases} -22,649x_1, & x_1 < 76,76 \\ 0,339x_1 - 1764,56, & x_1 \geq 76,76 \end{cases}$$

Apabila APS SD/Sederajat penduduk perempuan berada dibawah 76,76 dan APS SD/Sederajat penduduk perempuan pada keadaan ini naik 1 satuan, maka IPG akan cenderung turun sebesar 22,649. Wilayah yang termasuk dalam interval ini adalah provinsi Papua. Sedangkan apabila APS SD/Sederajat penduduk perempuan lebih dari 76,76 dan APS SD/Sederajat penduduk perempuan pada keadaan ini naik 1 satuan, maka IPG akan cenderung naik sebesar 0,339. Wilayah yang termasuk dalam interval ini adalah semua provinsi di Indonesia kecuali provinsi Papua. Variabel lainnya dapat diinterpretasikan dengan cara yang sama dengan  $x_1$ .

V. KESIMPULAN DAN SARAN

Faktor-faktor yang mempengaruhi IPG di Indonesia pada tahun 2012 adalah APS SD/Sederajat penduduk perempuan ( $x_1$ ), APS SMP/Sederajat penduduk perempuan ( $x_2$ ), APS SMA/Sederajat penduduk perempuan ( $x_3$ ), Angka Buta Huruf penduduk perempuan ( $x_4$ ), TPAK penduduk perempuan ( $x_5$ ), rasio jenis kelamin ( $x_6$ ), rasio jenis kelamin saat lahir ( $x_7$ ), dan persentase penduduk perempuan mempunyai keluhan kesehatan ( $x_8$ ). Berdasarkan model spline linier yang optimal dengan kombinasi satu, dua, dan tiga knot didapatkan nilai  $R^2$  sebesar 99,81% dan nilai MSE sebesar 0,152 serta asumsi residual yang telah terpenuhi sehingga dapat dikatakan baik dalam pemodelan.

Saran yang dapat direkomendasikan pada penelitian selanjutnya dan untuk kebijakan pemerintah adalah perlunya dikembangkan regresi nonparametrik spasial karena adanya perubahan pola pada tiap provinsi sehingga mempengaruhi interpretasi, dan perlunya pengembangan program menjadi regresi spline kuadratik dan spline kubik dengan menggunakan kombinasi knot, serta diharapkan adanya peningkatan upaya dari pemerintah dalam meningkatkan kualitas penduduk tanpa membedakan gender untuk meningkatkan IPG di Indonesia, seperti peningkatan kualitas pendidikan pada provinsi Papua karena di provinsi Papua ketersediaan guru sangat minim dan tidak merata, banyak siswa kelas enam SD yang belum mampu membaca, minimnya fasilitas penunjang, serta masih rendahnya kualitas lulusan [11].

DAFTAR PUSTAKA

- [1] Kementerian Pemberdayaan Perempuan dan Perlindungan Anak. (2013). *Pembangunan Manusia Berbasis Gender 2013*. Jakarta: Kementerian Pemberdayaan Perempuan dan Perlindungan Anak.
- [2] Nurhaeni, I.D.A. (2009). *Kebijakan Publik Pro Gender*. Surakarta: UPT Penerbitan dan Percetakan UNS (UNS Press).
- [3] Fitarisca, A.V. (2014). *Analisis Faktor-faktor yang Mempengaruhi Indeks Pembangunan Gender (IPG) dengan menggunakan Regresi Probit*. Tugas Akhir S1 yang tidak dipublikasikan. Jurusan Statistika FMIPA Institut Teknologi Sepuluh Nopember Surabaya.
- [4] Hafizh, U.Q. (2013). *Pemodelan Disparitas Gender di Jawa Timur dengan Pendekatan Model Regresi Probit Ordinal*. Tugas Akhir S1 yang tidak dipublikasikan. Jurusan Statistika FMIPA Institut Teknologi Sepuluh Nopember Surabaya.
- [5] Merdekawati, I.P. (2013). *Pemodelan Regresi Spline Truncated Multivariabel Pada Faktor-Faktor yang Mempengaruhi Kemiskinan di Kabupaten/Kota Provinsi Jawa Tengah*. Surabaya: Institut Teknologi Sepuluh Nopember.
- [6] Arfan, N. (2014). *Pendekatan Spline untuk Estimasi Kurva Regresi Nonparametrik (Studi Kasus Pada Data Angka Kematian Maternal di Jawa Timur)*. Surabaya: Institut Teknologi Sepuluh Nopember.
- [7] Eubank, R.L. (1988). *Nonparametric Regression and Spline Smoothing*. New York: Marcel Dekker.
- [8] Wahba, G. (1990). *Spline Models For Observasion Data*. SIAM Pencilvania.
- [9] Budiantara, I.N. (2005). *Model Keluarga Spline Polinomial Truncated dalam Regresi Semiparametrik*. Surabaya: Berkala MIPA, Institut Teknologi Sepuluh Nopember.
- [10] Wei, W.W.S. (2006). *Time Series Univariate and Multivariate Methods*. Canada: Addison Wesley Publishing Company, Inc.
- [11] Joumilena, E. (2014). *Wajah Pendidikan Papua Masih Tampak Suram*. Papua: Tabloid Jubi.