

Pendekatan Regresi Semiparametrik Spline untuk Memodelkan Rata-Rata Umur Kawin Pertama (UKP) di Provinsi Jawa Timur

Dhira Audhia Pratiwi, I Nyoman Budiantara, dan Wahyu Wibowo

Jurusan Statistika, Fakultas Matematika dan Ilmu Pengetahuan Alam, Institut Teknologi Sepuluh Nopember (ITS)
Jl. Arief Rahman Hakim, Surabaya 60111 Indonesia

e-mail: dhiraaudhia@gmail.com, i_nyoman_b@statistika.its.ac.id, wahyu_w@statistika.its.ac.id

Abstrak—Penyebab utama penambahan penduduk yang besar setiap tahunnya adalah kelahiran. Kelahiran erat kaitannya dengan umur kawin pertama. Umur kawin pertama (UKP) adalah umur pada saat wanita menikah pertama kali. Jawa Timur memiliki rata-rata UKP wanita sebesar 22 tahun, dan merupakan terendah kedua di Pulau Jawa setelah Banten. Sebanyak lebih dari 50 persen wanita di Jawa Timur kawin di usia kurang dari 20 tahun pada tahun 2012. Terdapat 5 faktor yang diduga mempengaruhi UKP, yaitu persentase penduduk yang tinggal di pedesaan, rata-rata lama sekolah wanita, persentase penduduk usia 15 tahun ke atas yang bekerja, persentase penduduk miskin, dan gini ratio. Pemodelan UKP menggunakan regresi semiparametrik spline linier karena hubungan antara rata-rata UKP wanita Provinsi Jawa Timur dengan sebagian faktor yang diduga mempengaruhinya tidak membentuk pola tertentu dan sebagian lagi membentuk pola tertentu. Berdasarkan hasil analisis, terdapat 4 faktor yang berpengaruh signifikan terhadap rata-rata UKP, yaitu rata-rata lama sekolah wanita, persentase penduduk wanita yang tinggal di desa, persentase penduduk wanita usia 15 tahun ke atas yang bekerja, dan gini ratio. Model regresi semiparametrik spline menghasilkan koefisien determinasi sebesar 90,77%.

Kata Kunci—rata-rata umur kawin pertama, regresi semiparametrik, spline.

I. PENDAHULUAN

FENOMENA penambahan maupun penurunan penduduk yang besar setiap tahunnya dipengaruhi oleh tiga faktor, yaitu fertilitas (kelahiran), mortalitas (kematian), dan migrasi atau perpindahan penduduk [1]. Berdasarkan bentuk piramida penduduk Indonesia yang lebar pada bagian bawah, dapat dipastikan bahwa faktor utama yang mempengaruhi laju pertumbuhan penduduk adalah tingkat fertilitas (kelahiran). Kelahiran erat kaitannya dengan usia kawin pertama [2]. Semakin cepat seorang wanita kawin, maka masa reproduksinya akan semakin panjang dan kemungkinan banyaknya keturunan juga akan semakin tinggi [3]. Umur kawin pertama (UKP) adalah umur pada saat wanita menikah pertama kali atau pada saat mereka memulai hubungan dengan pasangan yang pertama [4].

Jawa Timur memiliki rata-rata UKP wanita sebesar 22 tahun, dan merupakan terendah kedua di Pulau Jawa setelah Provinsi Banten yang memiliki rata-rata UKP 21,5 tahun.

Berdasarkan data dari BKKBN pula, persentase wanita di Jawa Timur pada tahun 2012 yang berstatus kawin menurut usia pada perkawinan pertama, sebanyak lebih dari 50 persen wanita kawin di usia kurang dari 20 tahun. Hal ini tentunya mempengaruhi rendahnya rata-rata Umur Kawin Pertama wanita di Provinsi Jawa Timur.

Penelitian mengenai rata-rata UKP pernah dilakukan oleh Imawati (2012), Ananto (2014), Lehrer dan Chen (2013), dan Syilfi (2015). Secara keseluruhan, faktor yang mempengaruhi rata-rata UKP berkaitan dengan faktor ekonomi, sosial, dan budaya. Berdasarkan referensi pada penelitian-penelitian sebelumnya, terdapat beberapa variabel yang diduga mempengaruhi rata-rata UKP wanita di Provinsi Jawa Timur, variabel-variabel tersebut terkait aspek ekonomi, sosial, dan budaya.

Salah satu metode analisis yang dapat digunakan untuk mengetahui faktor-faktor apa saja yang mempengaruhi rata-rata UKP wanita adalah menggunakan analisis regresi. Terdapat tiga pendekatan yang dapat dilakukan untuk mengestimasi kurva regresi, yaitu regresi parametrik, regresi nonparametrik, dan regresi semiparametrik. Persentase penduduk wanita usia 15 tahun ke atas yang bekerja secara teori dapat mempengaruhi rata-rata umur kawin pertama di suatu provinsi, dimana semakin banyak wanita yang bekerja, maka rata-rata umur kawin pertama di suatu provinsi juga akan semakin tinggi, atau semakin tua wanita menikah. Namun berdasarkan data yang diperoleh, Kabupaten Pacitan yang memiliki persentase penduduk wanita usia 15 tahun ke atas yang bekerja tinggi, rata-rata UKP nya tergolong dalam *early marriage*, begitu juga dengan Kabupaten Sumenep. Fakta ini tentunya tidak sesuai dengan teori yang ada dan menunjukkan pola yang nonparametrik. Sementara pada data persentase penduduk miskin per kabupaten di Jawa Timur membentuk pola yang parametrik. Dimana semakin tinggi persentase penduduk miskinnya, maka rata-rata UKP nya semakin rendah. Hubungan antara rata-rata UKP wanita Provinsi Jawa Timur dengan sebagian faktor yang diduga mempengaruhinya tidak membentuk pola tertentu dan sebagian lagi membentuk pola tertentu sehingga digunakan pemodelan regresi semiparametrik dengan pendekatan spline.

II. TINJAUAN PUSTAKA

A. Regresi Semiparametrik Spline

Regresi semiparametrik merupakan salah satu metode statistika yang digunakan untuk mengetahui pola hubungan antara variabel respon dan prediktor dimana sebagian diketahui bentuk polanya, dan sebagian tidak diketahui bentuknya. Misal terdapat data berpasangan (x_i, y_i, t_i) dan hubungan antara x_i, y_i dan t_i diasumsikan mengikuti model regresi sebagai berikut.

$$y_i = \mathbf{x}_i \boldsymbol{\beta} + f(t_i) + \varepsilon_i, i = 1, 2, \dots, n \quad (10)$$

Regresi semiparametrik bisa juga disebut sebagai gabungan dari regresi parametrik dan regresi nonparametrik. Regresi parametrik adalah metode yang digunakan untuk mengetahui pola hubungan antara variabel respon dan prediktor dimana bentuk kurva regresi diketahui. Model regresi parametrik linear berganda dalam bentuk matriks adalah sebagai berikut.

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon} \quad (2)$$

dengan \mathbf{y} adalah vektor kolom berukuran $n \times 1$ yang elemennya berupa data respon, \mathbf{X} merupakan matrik berukuran $n \times (m+1)$ dengan elemen berupa data m prediktor, dan $\boldsymbol{\varepsilon}$ adalah vektor kolom berukuran $n \times 1$ dengan elemen berupa error random. Estimasi parameter pada regresi parametrik diperoleh dengan menggunakan metode estimasi Ordinary Least Square (OLS) sebagai berikut.

$$\hat{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}'\mathbf{y} \quad (3)$$

Sementara regresi nonparametrik merupakan metode statistika yang digunakan apabila pola data yang tidak diketahui bentuk kurva regresinya atau tidak terdapat informasi masa lalu yang lengkap tentang bentuk pola data [5]. Regresi nonparametrik spline merupakan regresi yang sangat fleksibel dalam memodelkan pola data. Model regresi nonparametrik spline secara umum adalah sebagai berikut.

$$y_i = f(t_i) + \varepsilon_i, i = 1, 2, 3, \dots, n \quad (4)$$

dimana y_i adalah variabel respon, t_i adalah variabel prediktor, $f(x_i)$ adalah fungsi regresi yang polanya tidak diketahui, serta $\varepsilon_i \sim \text{IIDN}(0, \sigma^2 \mathbf{I})$. Pada model regresi nonparametrik spline, kurva regresi dihipotesis dengan fungsi spline berorde p dengan titik knot k_1, k_2, \dots, k_r dalam bentuk.

$$f(t_i) = \sum_{j=0}^p \gamma_j t_i^j + \sum_{l=1}^r \gamma_{p+l} (t_i - k_l)_+^p \quad (5)$$

dengan γ adalah parameter model, p adalah orde spline dan k adalah banyak knot spline. Berdasarkan persamaan (4) dan (5) dapat diperoleh persamaan regresi seperti berikut.

$$y_i = \sum_{j=0}^p \gamma_j t_i^j + \sum_{l=1}^r \gamma_{p+l} (t_i - k_l)_+^p + \varepsilon_i, i = 1, 2, \dots, n \quad (6)$$

Fungsi *truncated* adalah sebagai berikut.

$$(t_i - k_l)_+^p = \begin{cases} (t_i - k_l)^p, & t_i \geq k_l \\ 0, & t_i < k_l \end{cases} \quad (7)$$

Estimator spline dalam regresi nonparametrik dapat menangani fungsi yang bersifat *smooth* [6]. Estimasi regresi nonparametrik spline *truncated* dapat diperoleh dengan

metode *Ordinary Least Square* (OLS). Hasil estimasi adalah sebagai berikut.

$$\hat{f}(\mathbf{k}) = (\mathbf{T}(\mathbf{k})'\mathbf{T}(\mathbf{k}))^{-1} \mathbf{T}(\mathbf{k})'\mathbf{y} \quad (8)$$

Dimana $\mathbf{T}(\mathbf{k})$ merupakan variabel prediktor komponen nonparametrik yang sudah mengandung titik knot masing-masing. Model regresi spline terbaik diperoleh dari titik knot yang optimal. Titik knot yang optimal diperoleh dari nilai GCV minimum. GCV merupakan modifikasi dari metode CV (Cross Validation) dan merupakan metode yang paling banyak dipakai dan disukai karena memiliki sifat optimal asimtotik. [7]. Titik knot optimal diperoleh dari nilai GCV minimum. Metode GCV dapat dituliskan sebagai berikut [8].

$$GCV(k) = \frac{MSE(k)}{[n^{-1} \text{trace}(\mathbf{I} - \mathbf{A}(\mathbf{k}))]^2} \quad (9)$$

dengan \mathbf{I} adalah matriks identitas dan n adalah jumlah pengamatan.

$$MSE(k) = n^{-1} \sum_{i=1}^n (y_i - \hat{y}_i)^2,$$

B. Ukuran Kebaikan Model

Setelah diperoleh model yang sesuai, perlu diukur seberapa besar kebaikan model tersebut. Nilai kebaikan model diberikan oleh nilai *R-square* yang ditunjukkan dengan persamaan sebagai berikut.

$$R^2 = \frac{SSR}{SST} = \frac{\mathbf{b}'\mathbf{X}(\mathbf{k})'\mathbf{y} - n\bar{y}^2}{(\mathbf{y}'\mathbf{y}) - n\bar{y}^2} \quad (11)$$

Dimana $X(k)$ merupakan gabungan variabel prediktor komponen parametrik dan komponen nonparametrik yang telah mengandung titik knot masing-masing.

C. Pengujian Parameter Model

1. Pengujian Parameter Model Secara Serentak

Uji serentak dilakukan untuk mengetahui signifikansi parameter model regresi secara bersama-sama. Berikut ini merupakan hipotesis untuk uji serentak.

$$H_0 : \beta_1 = \beta_2 = \dots = \beta_m = \gamma_1 = \gamma_2 = \dots = \gamma_{p+r} = 0$$

$$H_1 : \text{paling sedikit ada satu } \gamma_j \neq 0 \text{ atau } \beta_k \neq 0$$

$$j = 1, 2, \dots, p+r, k = 1, 2, \dots, m$$

dengan $p+r$ merupakan parameter regresi nonparametrik. Statistik uji yang digunakan adalah sebagai berikut.

$$F_{hit} = \frac{MSR}{MSE} \quad (12)$$

dimana $MSE = \frac{\mathbf{y}'\mathbf{y} - \mathbf{b}'\mathbf{X}(\mathbf{k})'\mathbf{y}}{n - (m + p + r + 1)}$ dan

$$MSR = \frac{\mathbf{b}'\mathbf{X}(\mathbf{k})'\mathbf{y} - n\bar{y}^2}{m + p + r}$$

Kesimpulan Tolak H_0 jika F_{hitung} lebih besar daripada $F_{\alpha; (m+(p+r), n-(m+(p+r)-1)}$ atau $p\text{-value} < \alpha$ yang menunjukkan bahwa paling sedikit terdapat satu parameter yang signifikan terhadap model.

2. Pengujian Parameter Model Secara Individu

Pengujian parameter secara individu dilakukan apabila pengujian parameter model secara serentak diperoleh kesimpulan bahwa minimal terdapat satu parameter yang signifikan. Berikut ini merupakan hipotesis dari pengujian parameter secara individu.

$$H_0 : \beta_k = 0$$

$$H_1 : \beta_k \neq 0, k = 1, 2, \dots, m$$

Sedangkan berikut merupakan hipotesis untuk parameter regresi nonparametrik.

$$H_0 : \gamma_j = 0$$

$$H_1 : \gamma_j \neq 0 ; j = 1, 2, \dots, p + r$$

Statistik uji disajikan sebagai berikut.

$$t_{hit} = \frac{\hat{\beta}_k}{se(\hat{\beta}_k)} \text{ (untuk parameter parametrik)} \quad (13)$$

$$t_{hit} = \frac{\hat{\gamma}_j}{se(\hat{\gamma}_j)} \text{ (untuk parameter nonparametrik)} \quad (14)$$

D. Pengujian Asumsi Residual

1. Asumsi Identik

Asumsi identik terpenuhi jika varians respon sama dengan varians error yaitu sama dengan σ^2 . Hal tersebut merupakan asumsi homoskedastisitas (varians sama) atau disebut dengan identik [9]. Kasus heteroskedastisitas dapat dideteksi menggunakan uji *glejser* dan dilakukan dengan cara meregresikan harga mutlak residual dengan variabel prediktor. Berikut ini merupakan hipotesis uji *glejser*.

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

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

Statistik uji yang digunakan adalah sebagai berikut.

$$F_{hit} = \frac{MSR_{glejser}}{MSE_{glejser}} \quad (15)$$

dimana $MSE_{glejser} = \frac{|\boldsymbol{\varepsilon}'|\boldsymbol{\varepsilon} - \mathbf{b}'\mathbf{X}(\mathbf{k})|\boldsymbol{\varepsilon}|}{n - (m + p + r) - 1}$ dan

$$MSR_{glejser} = \frac{\mathbf{b}'\mathbf{X}(\mathbf{k})|\boldsymbol{\varepsilon} - n|\boldsymbol{\varepsilon}|^2}{m + p + r}$$

H_0 ditolak apabila $F_{hit} > F_{\alpha, (m+(p+r), n-(m+(p+r))-1)}$ atau *p-value* kurang dari α . Apabila H_0 ditolak, maka dapat disimpulkan bahwa terjadi heteroskedastisitas yang artinya asumsi residual identik tidak terpenuhi.

2. Asumsi Independen

Asumsi klasik kedua yang harus dipenuhi adalah residual bersifat independen. Asumsi residual independen yaitu asumsi bahwa tidak ada korelasi antar residual atau adanya independensi pada residual yang ditunjukkan dengan nilai kovarian antara ε_i dan ε_j sama dengan nol.

Uji independen dilakukan menggunakan interval konfidensi untuk ρ_k dengan hipotesis sebagai berikut.

$$H_0 : cov(\varepsilon_i, \varepsilon_j) = 0$$

$$H_1 : \text{paling sedikit ada satu } cov(\varepsilon_i, \varepsilon_j) \neq 0, i \neq j$$

Rumus interval konfidensi adalah sebagai berikut.

$$-t_{n-1, \alpha/2} se(\hat{\rho}_k) < \rho_k < t_{n-1, \alpha/2} se(\hat{\rho}_k) \quad (16)$$

Apabila tidak terdapat autokorelasi yang keluar dari batas signifikansi, maka dapat disimpulkan bahwa asumsi residual independen terpenuhi.

3. Asumsi Distribusi Normal

Asumsi residual berdistribusi normal dilakukan untuk mengetahui apakah residual hasil regresi telah berdistribusi normal atau tidak. Cara lain dapat dilakukan untuk menguji asumsi residual berdistribusi normal adalah berdasarkan visual menggunakan *normal probability plot residual* yang cenderung mengikuti garis atau menggunakan uji Kolmogorov Smirnov. Apabila menggunakan uji Kolmogorov-Smirnov, berikut merupakan hipotesis yang digunakan.

$$H_0 : F_0(x) = S_N(x) \text{ (Residual berdistribusi Normal)}$$

$$H_1 : F_0(x) \neq S_N(x) \text{ (Residual tidak berdistribusi Normal)}$$

Statistik uji yang digunakan adalah

$$D = maks |F_0(x) - S_N(x)| \quad (15)$$

Daerah penolakan H_0 jika $|D| > q_{(1-\alpha)}$ dimana nilai $q_{(1-\alpha)}$ adalah berdasarkan tabel *Kolmogorov-Smirnov* atau dapat juga berdasarkan *p-value* $< \alpha$.

E. Umur Kawin Pertama

Umur kawin pertama adalah umur pada saat wanita menikah pertama kali atau pada saat wanita memulai kehidupan dengan pasangan yang pertama dinikahnya [2]. Semakin muda usia kawin pertama yang dilakukan seseorang, maka akan mempengaruhi tingkat fertilitas dan fekunditas (potensi fisik untuk melahirkan anak). Tingginya tingkat kelahiran akan menyebabkan meningkatnya pertumbuhan penduduk suatu [3]. BKKBN menetapkan usia kawin yang ideal untuk wanita adalah diatas 20 tahun dan 25 tahun untuk laki-laki. Umur Kawin Pertama (UKP) wanita dikelompokkan menjadi 4 kelompok, yaitu.

1. UKP wanita kurang dari 18 tahun, dikelompokkan ke dalam *Child Marriage*.
2. UKP wanita antara 18 tahun hingga kurang dari 20 tahun, disebut dengan *Early Marriage*.
3. UKP wanita antara 20 hingga kurang dari 22 tahun, disebut dengan *Marriage at Maturity*.
4. UKP wanita diatas 22 tahun, disebut dengan *Late Marriage*.

Pengelompokan UKP wanita tersebut merupakan pengelompokan secara individu atau perseorangan.

III. METODOLOGI PENELITIAN

A. Sumber Data

Data yang digunakan dalam penelitian ini adalah data sekunder yang diperoleh dari beberapa sumber yaitu "Hasil Survei Sosial Ekonomi Nasional Tahun 2012", "Data dan Informasi Kemiskinan Kabupaten/Kota Propinsi Jawa Timur Tahun 2012" yang dipublikasikan oleh Badan Pusat Statistik

Jawa Timur. Unit penelitian adalah 29 kabupaten yang ada di Jawa Timur.

B. Variabel Penelitian

Penelitian ini menggunakan respon y adalah rata-rata UKP wanita per kabupaten di Propinsi Jawa Timur. Variabel prediktor x adalah faktor-faktor yang diduga mempengaruhi rata-rata UKP wanita Propinsi Jawa Timur. Berikut merupakan variabel-variabel tersebut.

Tabel 1. Variabel Penelitian

Variabel	Keterangan	Skala
y	Rata-rata UKP wanita	Kontinu
x_1	Persentase penduduk wanita yang tinggal di pedesaan	Kontinu
x_2	Rata-rata lama sekolah wanita	Kontinu
x_3	Persentase penduduk wanita berusia 15 tahun ke atas yang bekerja	Kontinu
x_4	Persentase penduduk miskin	Kontinu
x_5	Gini Ratio	Kontinu

C. Langkah Penelitian

Mengacu pada tujuan penelitian, berikut merupakan langkah analisis dalam penelitian ini.

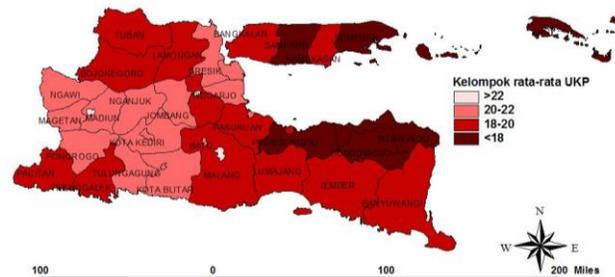
1. Tujuan pertama yaitu mengetahui karakteristik rata-rata Umur Kawin Pertama (UKP) wanita dan faktor-faktor yang diduga mempengaruhinya di Propinsi Jawa Timur dipenuhi dengan melakukan langkah-langkah sebagai berikut.
 - a. Mendeskripsikan data rata-rata Umur Kawin Pertama dan faktor yang diduga mempengaruhinya di Propinsi Jawa Timur.
 - b. Membuat *scatter plot* pada masing-masing variabel prediktor dengan variabel respon untuk mengetahui bentuk pola data.
 - c. Menentukan variabel yang mempunyai pola tertentu (parametrik) dan variabel yang tidak mempunyai pola tertentu (nonparametrik).
2. Tujuan kedua yaitu memodelkan rata-rata Umur Kawin Pertama (UKP) di Propinsi Jawa Timur menggunakan pendekatan regresi semiparametrik spline dipenuhi dengan melakukan langkah sebagai berikut.
 - a. Memodelkan variabel respon (UKP) menggunakan regresi semiparametrik spline linier dengan berbagai titik knot.
 - b. Memilih titik knot optimal berdasarkan nilai GCV yang paling minimum.
 - c. Mendapatkan model regresi semiparametrik spline terbaik dengan titik knot optimal.
 - d. Menguji kebaikan model dengan menghitung R^2 .
 - e. Melakukan uji signifikansi parameter secara serentak dan parsial.
 - f. Melakukan uji asumsi residual IIDN dari model regresi spline

Setelah melakukan analisis, maka membuat kesimpulan dari hasil analisis.

IV. ANALISIS DAN PEMBAHASAN

A. Karakteristik Data

Rata-rata Umur Kawin Pertama dapat dikategorikan menjadi 4 kategori, rata-rata UKP wanita kurang dari 18 tahun dikelompokkan ke dalam *Child Marriage*, rata-rata UKP wanita antara 18 tahun hingga kurang dari 20 tahun disebut dengan *Early Marriage*, rata-rata UKP wanita antara 20 hingga kurang dari 22 tahun disebut dengan *Marriage at Maturity*, dan rata-rata UKP wanita diatas 22 tahun disebut dengan *Late Marriage*.



Gambar 1. Peta tematik kategori rata-rata UKP Jatim

Berdasarkan Gambar 1, rata-rata UKP wanita di setiap kabupaten Provinsi Jawa Timur bervariasi, terdapat 4 kategori rata-rata UKP yang tersebar di seluruh provinsi.

Tabel 2. Jumlah Kabupaten pada Setiap Kategori UKP

Kategori	Jumlah Kab/Kota
<i>Child Marriage</i>	5
<i>Early Marriage</i>	13
<i>Marriage at Mature</i>	11
<i>Late Marriage</i>	-

Berdasarkan Tabel 2, keempat kategori rata-rata UKP tersebar di Jawa Timur. 5 kabupaten terkategori menjadi *child marriage*, 13 kabupaten terkategori ke dalam *early marriage*, 11 kabupaten/kota terkategori ke dalam *marriage at mature*, dan tidak terdapat kabupaten terkategori ke dalam *late marriage*.

Terdapat beberapa variabel yang diduga berpengaruh terhadap rata-rata umur kawin pertama di Provinsi Jawa Timur. Sebelum melakukan pemodelan, perlu diketahui karakteristik setiap variabel.

Tabel 3.

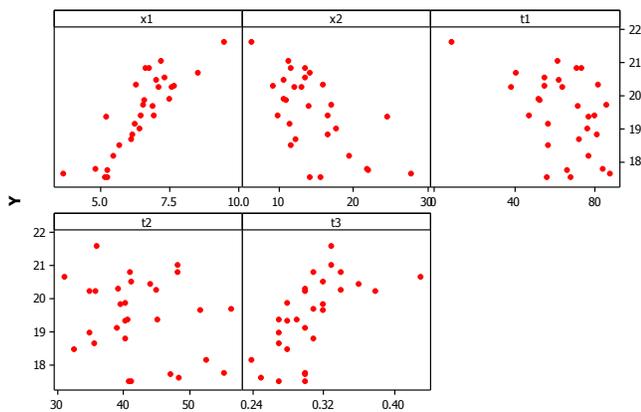
Karakteristik Variabel Penelitian					
Keterangan		Max	Min	Mean	Varian
x_1	Persentase penduduk wanita yang tinggal di pedesaan	87.34	8.74	63.86	292.79
x_2	Rata-rata lama sekolah wanita	9.49	3.66	6.492	1.376
x_3	Persentase penduduk wanita berusia 15 tahun ke atas yang bekerja	56.44	31.21	42.38	43.64
x_4	Persentase penduduk miskin	27.87	6.42	14.767	23.149
x_5	Gini Ratio	0.43	0.24	0.3069	0.00154

Berdasarkan Tabel 3, persentase penduduk yang tinggal di pedesaan tertinggi di Jawa Timur dari seluruh kabupaten

adalah 87,34%, yaitu di Kabupaten Sampang, dan terendah adalah 8,74% dimiliki oleh Kabupaten Sidoarjo. Kabupaten dengan rata-rata lama sekolah wanita terkecil adalah di Kabupaten Sampang, yaitu hanya selama 3,66 tahun dan tertinggi terjadi di Kabupaten Sidoarjo yaitu selama 9,49 tahun. Kabupaten dengan persentase penduduk wanita usia 15 tahun keatas yang bekerja tertinggi adalah di Kabupaten Pacitan yaitu sebesar 56,44% dan terendah Kabupaten Gresik yaitu sebesar 31,21%. Persentase penduduk miskin tertinggi terjadi di Kabupaten Sampang yang mencapai 27,87 persen dan terendah di Kabupaten Sidoarjo yaitu sebesar 6,42%. *Gini ratio* terkecil sebesar 0,24 terjadi di Kabupaten Pamekasan. Hal ini berarti ketimpangan kekayaan di Kabupaten Pamekasan termasuk dalam kategori ketimpangan rendah. Sementara *gini ratio* tertinggi sebesar 0,43 terjadi di Kabupaten Gresik.

B. Pola Data Antara Rata-rata UKP dengan Variabel yang Diduga Mempengaruhi

Berikut merupakan scatter plot yang menyatakan bentuk pola data antara rata-rata UKP dan variabel-variabel yang diduga mempengaruhinya.



Gambar 2. Scatterplot variabel prediktor terhadap respon

Berdasarkan Gambar 2, dapat diketahui bahwa sebagian dari lima pola data tidak membentuk suatu pola tertentu (persentase penduduk yang tinggal di desa, persentase penduduk wanita yang tinggal di desa, dan *gini ratio*), dan sebagian lagi membentuk pola tertentu (linier), yaitu persentase penduduk miskin dan rata-rata lama sekolah wanita. Dengan demikian, dalam pemodelan regresi akan digunakan pendekatan regresi semiparametrik.

C. Pemilihan Titik Knot Optimal

Nilai GCV minimum pada pemilihan titik knot optimal dengan satu titik knot, dua titik knot, tiga titik knot, dan kombinasi titik knot ditampilkan sebagai berikut.

Tabel 4.
Nilai GCV Minimum Setiap Pemilihan Titik Knot Optimal

Model	GCV
1 Knot	0,445144
2 Knot	0,406152
3 Knot	0,366851
Kombinasi Knot 3,3,2	0,360111

Berdasarkan Tabel 4, nilai GCV yang paling minimum dimiliki oleh model dengan kombinasi titik knot 3,3,2, yaitu sebesar 0,360111.

D. Penaksiran Parameter Model Regresi Semiparametrik Spline

Berdasarkan nilai GCV terkecil, didapatkan model terbaik yaitu model regresi nonparametrik spline dengan kombinasi knot 3,3,2. Berikut ini adalah hasil estimasi parameter model regresi semiparametrik spline terbaik.

$$\hat{y}_i = 7,689328929 + 0,832014645x_{i1} + 0,060511557x_{i2} - 0,001794196t_{i1} - 1,15824(t_{i1} - 64,88286)_+ - 1,926525(t_{i1} - 66,48694)_+ - 0,81288(t_{i1} - 69,6951)_+ + 0,052706t_{i2} - 0,78901(t_{i2} - 49,23143)_+ - 0,18206(t_{i2} - 49,74633)_+ + 1,031825(t_{i2} - 50,77612)_+ + 11,63324t_{i3} - 12,627886540(t_{i3} - 0,364082)_+ - 12,202034391(t_{i3} - 0,367959)_+$$

E. Pengujian Signifikansi Parameter Model Regresi Nonparametrik Spline

1. Uji Serentak

Pengujian secara serentak ini dilakukan untuk menguji estimasi parameter model secara bersamaan (simultan). Nilai α yang digunakan sebesar 0,05 dengan hipotesis sebagai berikut.

$$H_0 : \beta_1 = \beta_2 = \gamma_{11} = \gamma_{21} = \dots = \gamma_{33} = 0$$

$$H_1 : \text{paling sedikit ada satu } \gamma_j \neq 0 \text{ atau } \beta_k \neq 0$$

Berikut ini adalah hasil analisis ragam model regresi nonparametrik spline.

Tabel 5.
Analisis Ragam Uji Serentak

Sumber	df	SS	MS	Fhit	P-value
Regresi	13	35,27132	2,713179	11,34053	1,68E-05
Error	15	3,588693	0,2392462		
Total	28	38,86002			

Berdasarkan Tabel 5 didapat *p-value* 1,68E-05. Menggunakan taraf signifikansi (α) sebesar 0,05, keputusan yang dapat diambil dari uji serentak adalah tolak H_0 sehingga terdapat minimal satu variabel prediktor yang berpengaruh terhadap rata-rata UKP di Jawa Timur. Untuk mengetahui parameter mana yang signifikan dalam model, dilakukan pengujian secara individu.

2. Uji Individu

Pengujian secara individu dilakukan untuk mengetahui variabel yang berpengaruh signifikan terhadap jumlah pengangguran di Indonesia. Berikut merupakan hipotesis untuk parameter regresi parametrik.

$$H_0 : \beta_k = 0$$

$$H_1 : \beta_k \neq 0, k = 1, 2$$

Sedangkan berikut merupakan hipotesis untuk parameter regresi nonparametrik.

$$H_0 : \gamma_j = 0$$

$$H_1 : \gamma_j \neq 0 ; j = 1, 2, \dots, 11$$

Berikut ini adalah hasil pengujian signifikansi parameter model secara individu.

Berdasarkan Tabel 6, terdapat satu variabel yang tidak signifikan berengaruh terhadap rata-rata UKP Jawa Timur, yaitu x_2 (persentase penduduk miskin). Pemodelan kembali

dengan menghapus variabel yang tidak signifikan tidak perlu dilakukan karena sebenarnya variabel yang tidak berpengaruh signifikan terhadap rata-rata umur kawin pertama tersebut masih tetap dapat dimasukkan ke dalam model hanya tidak perlu diinterpretasi.

Tabel 6.
Hasil Uji Individu

Variabel	Parameter	Estimasi	P-Values	Kesimpulan
-	φ	7.689329	0.001442	Signifikan
x_1	β_1	0.832015	0.000442	Signifikan
x_2	β_2	0.060512	0.215171	Tidak Signifikan
t_1	γ_{11}	-0.00179	0.894285	Tidak Signifikan
	γ_{21}	-1.15824	0.010115	Signifikan
	γ_{31}	-1.926525	0.006228	Signifikan
	γ_{41}	-0.81288	0.007869	Signifikan
t_2	γ_{12}	0.052706	0.062187	Tidak Signifikan
	γ_{22}	-0.78901	0.067009	Tidak Signifikan
	γ_{32}	-0.18206	0.021699	Signifikan
	γ_{42}	1.031825	0.108969	Tidak Signifikan
t_3	γ_{13}	11.63324	0.025406	Signifikan
	γ_{23}	-12.6279	0.086487	Tidak Signifikan
	γ_{33}	-12.202	0.085652	Tidak Signifikan

F. Pengujian Asumsi Residual

1. Asumsi Identik

Uji identik pada residual dilakukan menggunakan uji Glejser dengan hipotesis sebagai berikut.

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

$$H_1 : \text{paling sedikit ada satu } \sigma_i^2 \neq \sigma^2$$

Berikut merupakan hasil uji Glejser pada residual.

Tabel 10.
Analisis Ragam Uji Glejser

Sumber	df	SS	MS	Fhit	P-value
Regresi	13	0.6085	0.046811	0.727414	0.714682
Error	15	0.9652	0.064353		
Total	28	1.5738			

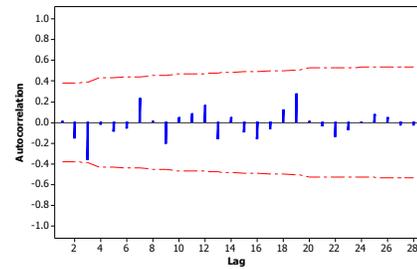
Berdasarkan Tabel 10 *p-value* adalah sebesar 1,50756 dan 0,20261. Menggunakan α (taraf signifikansi) sebesar 0,05, uji Glejser memberikan keputusan gagal tolak H_0 karena *p-value* > α . Residual yang dihasilkan model memenuhi asumsi identik.

2. Asumsi Independen

Asumsi residual selanjutnya yang harus terpenuhi adalah independen. Metode yang dapat digunakan untuk menguji asumsi independen adalah melalui plot *Autocorrelation Function* (ACF) dengan hipotesis sebagai berikut.

$$H_0 : \text{cov}(\varepsilon_i, \varepsilon_j) = 0$$

$$H_1 : \text{paling sedikit ada satu } \text{cov}(\varepsilon_i, \varepsilon_j) \neq 0, i \neq j$$



Gambar 3. ACF Residual

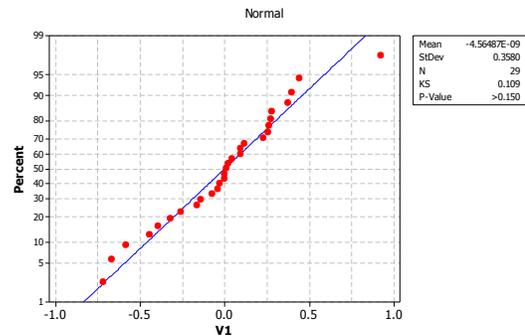
Berdasarkan Gambar 3, lag 1 hingga 29 tidak keluar dari batas signifikansi sehingga tidak terdapat kasus autokorelasi atau residual telah independen.

3. Asumsi Distribusi Normal

Pengujian asumsi distribusi normal menggunakan uji Kolmogorov-Smirnov dengan hipotesis sebagai berikut.

$$H_0 : \text{Residual berdistribusi normal}$$

$$H_1 : \text{Residual tidak berdistribusi normal}$$



Gambar 4. Hasil Uji Kolmogorov-Smirnov

Keputusan uji Kolmogorov-Smirnov dapat diambil melalui *p-value*. Berdasarkan Gambar 4, nilai $p > 0,150$. Dengan menggunakan taraf signifikansi (α) sebesar 0,05, uji Kolmogorov-Smirnov memberikan keputusan gagal tolak H_0 karena *p-value* > α sehingga residual model berdistribusi normal.

G. Nilai Koefisien Determinasi (R^2)

Model regresi terbaik memiliki koefisien determinasi sebesar 90,77%. Angka tersebut berarti bahwa model dapat menjelaskan variasi variabel respon berupa rata-rata UKP Jawa Timur sebesar 90,77%. Sedangkan sisa koefisien determinasi 9,23% dijelaskan oleh variabel prediktor lain yang tidak masuk dalam model

H. Interpretasi Model Regresi Nonparametrik Spline

Setelah dilakukan pengujian asumsi pada residual model regresi semiparametrik spline dan didapatkan bahwa semua asumsi terpenuhi, selanjutnya dilakukan interpretasi model. Interpretasi ini dibuat berdasarkan model untuk setiap variabel untuk memudahkan dalam interpretasi.

$$\hat{y}_i = 7,689328929 + 0,832014645x_{i1} + 0,060511557x_{i2} - 0,001794196t_{i1} - 1,15824(t_{i1} - 64,88286)_+ - 1,926525(t_{i1} - 66,48694)_+ - 0,81288(t_{i1} - 69,6951)_+ + 0,052706t_{i2} - 0,78901(t_{i2} - 49,23143)_+ - 0,18206(t_{i2} - 49,74633)_+ + 1,031825(t_{i2} - 50,77612)_+ + 11,63324t_{i3} - 12,627886540(t_{i3} - 0,364082)_+ - 12,202034391(t_{i3} - 0,367959)_+$$

1. Hubungan antara rata-rata lama sekolah wanita terhadap rata-rata Umur Kawin Pertama (UKP) di Provinsi Jawa Timur dengan asumsi variabel prediktor lain tetap adalah sebagai berikut.

$$\hat{y}_{x_1} = 7,689328929 + 0,832014645x_{t_1}$$

Apabila rata-rata lama sekolah wanita di Jawa Timur naik satu satuan, maka rata-rata Umur Kawin Pertama wanita juga akan naik sebesar 0,8320. Hal ini berarti bahwa semakin lama seorang wanita di Jawa Timur menjalani sekolah, maka semakin tua wanita tersebut menikah, atau wanita tersebut akan cenderung menunda usia pernikahannya.

2. Hubungan antara persentase penduduk wanita yang tinggal di pedesaan terhadap rata-rata Umur Kawin Pertama (UKP) di Provinsi Jawa Timur namun dengan asumsi variabel prediktor lain tetap adalah sebagai berikut.

$$y_{t_1} = -0.001794196t_{t_1} - 1.15824(t_{t_1} - 64.88286)_+ - 1.926525(t_{t_1} - 66.48694) - 0.81288(t_{t_1} - 69.6951)$$

$$= \begin{cases} -0.001794196t_1 & t_1 < 64.88286 \\ 75.134 - 1.16003t_1 & 64.88286 \leq t_1 < 66.48694 \\ 141.62 - 3.086t_1 & 66.48694 \leq t_1 < 69.6951 \\ 198.212 - 3.8985t_1 & t_1 \geq 69.6951 \end{cases}$$

Apabila persentase penduduk wanita yang tinggal di desa bernilai kurang dari 64,882% dan nilainya naik satu persen dengan asumsi prediktor lain tetap, maka rata-rata Umur Kawin Pertama (UKP) Jawa Timur akan turun sebesar 0,0017941. Terdapat 14 kabupaten yang termasuk ke dalam interval ini dan kebanyakan anggotanya merupakan kabupaten dengan kategori rata-rata umur kawin pertama *marriage at mature*. Apabila persentase penduduk wanita yang tinggal di desa bernilai antara 64,882% hingga 66,486% dan nilainya naik satu persen dengan asumsi prediktor lain tetap, maka rata-rata Umur Kawin Pertama (UKP) Jawa Timur akan turun sebesar 1,16003. Kabupaten yang termasuk ke dalam interval ini adalah Kabupaten Probolinggo. Kabupaten Probolinggo memiliki rata-rata umur kawin pertama terkategori dalam *child marriage*. Sementara apabila persentase penduduk wanita yang tinggal di desa bernilai antara 66,486% hingga 69,6951% dan nilainya naik satu persen dengan asumsi prediktor lain tetap, maka rata-rata Umur Kawin Pertama (UKP) Jawa Timur akan turun sebesar 3,086. Hanya terdapat satu kabupaten yang termasuk ke dalam wilayah interval ini yaitu Kabupaten Bondowoso yang memiliki rata-rata umur kawin pertama wanita terkategori ke dalam *child marriage*. Pada interval terakhir, apabila persentase penduduk wanita yang tinggal di desa bernilai lebih dari 69,6951% dan nilainya naik satu persen dengan asumsi prediktor lain tetap, maka rata-rata Umur Kawin Pertama (UKP) Jawa Timur akan turun sebesar 3,8985. Terdapat 13 wilayah atau kabupaten yang termasuk ke dalam interval ini dan didominasi oleh kabupaten yang memiliki kategori rata-rata umur kawin pertama *early marriage*.

Selanjutnya, interpretasi variabel komponen nonparametrik yaitu persentase penduduk wanita usia 15 tahun keatas yang bekerja dan *gini ratio* dilakukan dengan cara yang serupa seperti yang telah dijelaskan pada variabel persentase penduduk wanita yang tinggal di desa.

V. KESIMPULAN DAN SARAN

Terdapat 5 kabupaten yang terkategori menjadi *child marriage*, 13 kabupaten yang terkategori dalam *early marriage*, 11 kabupaten terkategori dalam *marriage at maturity*, dan tidak kabupaten yang terkategori dalam *late marriage*. Berdasarkan *scatterplot* antara rata-rata UKP Jawa Timur dengan setiap variabel yang diduga mempengaruhinya, terdapat 2 variabel komponen parametrik (rata-rata lama sekolah dan persentase penduduk miskin) dan 3 variabel komponen nonparametrik (persentase penduduk wanita yang tinggal di pedesaan, persentase penduduk wanita berusia 15 tahun keatas yang bekerja, dan *gini ratio*).

Model regresi semiparametrik spline terbaik untuk memodelkan rata-rata UKP Jawa Timur adalah dengan kombinasi titik knot 3,3,2. Dimana jumlah knot 3 untuk variabel persentase penduduk wanita yang tinggal di desa, 3 untuk persentase penduduk wanita usia 15 tahun ke atas yang bekerja, dan 2 untuk variabel *gini ratio*. Variabel yang tidak berpengaruh signifikan terhadap rata-rata Umur Kawin Pertama Provinsi Jawa Timur adalah persentase penduduk miskin. Koefisien determinasi yang dihasilkan oleh model terbaik adalah sebesar 90,77%.

Terdapat beberapa saran yang dapat diberikan oleh penulis yaitu bagi peneliti selanjutnya hendaknya menambahkan variabel lain yang diduga mempengaruhi rata-rata Umur Kawin Pertama di Jawa Timur, dan menambah kemungkinan titik knot. Sedangkan untuk pemerintah, rata-rata Umur Kawin Pertama di Jawa Timur masih rendah dibandingkan dengan daerah lain sehingga pemerintah dituntut untuk memaksimalkan program pencegahan pernikahan dini seperti PUP (Pendewasaan Usia Perkawinan) dan menambah sosialisasi mengenai program-program tersebut guna menambah kesadaran remaja untuk menunda usia perkawinannya. Apabila kebanyakan remaja sudah gagal menunda usia perkawinannya, maka akan lebih baik pemerintah memberikan sosialisasi hal-hal apa saja yang perlu disiapkan dalam sebuah pernikahan dan mewajibkan penggunaan alat kontrasepsi hingga usia istri mencapai 20 tahun untuk menghindari kelahiran anak di usia istri yang masih dibawah 20 tahun.

DAFTAR PUSTAKA

- [1] Mantra, I. B. (2000). Demografi Umum. Yogyakarta : Pustaka Belajar.
- [2] BAPPENAS. (2005). Laporan Perkembangan Tujuan Pembangunan Milenium Indonesia. Jakarta : BAPPENAS.
- [3] PKPP. (2012). Faktor-faktor yang Mempengaruhi Usia Kawin Pertama pada Perempuan di Indonesia. Jakarta : Tim Implementasi PKPP 2012 Kementerian Riset dan Teknologi Republik Indonesia.
- [4] BKKBN. (2013). Hubungan Antara Pengendalian Penduduk dengan Sosial Ekonomi di Propinsi Jawa Timur (Analisis Data Sensus Penduduk 2012, Susenas 2011-2012). Jawa Timur : BKKBN Jatim.
- [5] Eubank, R. 1988. *Spline Smoothing and Nonparametric Regression*. Marcel Dekker Inc. New York.

- [6] Budiantara, I N. (2009). *Spline Dalam Regresi Nonparametrik Dan Semiparametrik: Sebuah Pemodelan Statistika Masa Kini dan Masa Mendatang*, Pidato Pengukuhan Untuk Jabatan Guru Besar Dalam Bidang Ilmu Matematika Statistika dan Probabilitas, Pada Jurusan Statistika, Fakultas MIPA, Institut Teknologi Sepuluh Nopember, ITS Press, Surabaya.
- [7] Wahba.G. (1990). *Spline Models For Observational Data*. University Of Winsconsin at Madison.
- [8] Eubank, R.L. (1999), *Nonparametric Regression and Spline Smoothing Second Edition*, Marcel Deker, New York.
- [9] Gujarati, N. D. (2004). *Basic Econometrics fourth edition*. McGraw-Hill.