

Received: 1 Juni 2020
Revised: 25 Juni 2020
Accepted: 27 Juni 2020
Published: 30 Juni 2020

Pemodelan Fertilitas di Indonesia Tahun 2017 Menggunakan Pendekatan Regresi Nonparametrik Kernel dan Spline

Istiqomatul Fajriyah Yuliati ^{1, a)}, Pardomuan Robinson Sihombing ^{2, b)}

¹*Badan Kependudukan dan Keluarga Berencana Nasional, Jalan Permata No. 1, Jakarta Timur*

²*Badan Pusat Statistik, Jalan Dr. Sutomo No 6-8, Jakarta Pusat*

Email: ^{a)}isvayue@gmail.com, ^{b)}robinson@bps.go.id

Abstract

The aim of this study is to analyze the relationship pattern of Total Fertility Rate (TFR) with Contraceptive Prevalence Rate (CPR). The analysis that was often used for modeling is regression analysis. Regression analysis accordingly can be divided into two, parametric and nonparametric. Nonparametric regression methods that commonly used kernel and spline regression. In this study, the kernel regression used was kernel regression with the Nadaraya-Watson Estimator (NWE) method and the local polynomial estimator (LPE), while the spline regression used was smoothing spline and b-splines. The fitting curve showed that the best nonparametric regression model is the b-splines regression model with the degree is 2 and the number of knots are 5. This relates to the b-splines regression model having smooth curves and looked more suitable for the distribution of data than other regression curves. The selected b-spline regression model had a coefficient of determination $R^2 = 76.86\%$. Its means that the variation of the TFR variable explained by the b-spline regression model is 76.86%, while the remaining 23.14% is explained by other variables not included in the model.

Keywords: TFR, CPR, kernel regression, spline regression.

Abstrak

Tujuan dari penelitian ini adalah untuk menganalisis pola hubungan *Total Fertility Rate* (TFR) dengan *Contraceptive Prevalence Rate* (CPR). Analisis yang sering digunakan untuk pemodelan adalah analisis regresi. Analisis regresi menurut pendekatannya dapat dibedakan menjadi dua, parametrik dan nonparametrik. Metode regresi nonparametrik yang sering digunakan adalah regresi kernel dan spline. Pada penelitian ini untuk regresi kernel yang digunakan adalah regresi kernel dengan metode penaksir Nadaraya-Watson (NWE) dan penaksir polinomial lokal (LPE), sedangkan untuk regresi spline yang digunakan adalah *smoothing spline* dan b-splines. Hasil pengepasan kurva (*fitting curve*) menunjukkan bahwa model regresi nonparametrik terbaik adalah model regresi b-splines dengan *degree* 2 dan jumlah *knot* 5. Hal ini dikarenakan model regresi b-splines memiliki kurva yang halus dan terlihat lebih mengikuti sebaran data dibandingkan kurva model regresi lainnya. Model regresi b-splines terpilih memiliki nilai koefisien determinasi R^2 sebesar 76.86%, artinya besarnya variasi variabel TFR yang dijelaskan oleh model regresi b-splines sebesar 76.86%, sedangkan sisanya 23.14% dijelaskan oleh variabel lainnya yang tidak dimasukkan ke dalam model.

Kata-kata kunci: TFR, CPR, regresi kernel, regresi spline

PENDAHULUAN

Indonesia merupakan negara dengan jumlah penduduk ke empat terbesar di dunia setelah China, India, dan Amerika Serikat, yaitu sebesar 261.9 juta jiwa dengan laju pertumbuhan penduduk sebesar 1.34% (BPS, 2018). Masalah kependudukan yang dihadapi Indonesia saat ini tidak hanya terkait kuantitas penduduk karena jumlah penduduk yang besar dengan laju pertumbuhan yang relatif tinggi, tetapi juga mobilitas atau persebaran penduduk yang tidak merata dan kualitas penduduk yang masih rendah.

Jumlah penduduk beberapa tahun ke depan diprediksi akan terus bertambah dan diperkirakan akan terjadi ledakan penduduk tahun 2030, sehingga akan menjadi sebuah masalah besar jika tidak segera ditanggulangi oleh pemerintah dan tidak adanya kesadaran dari masyarakat. Pertumbuhan penduduk yang tinggi akan menghambat laju pembangunan dan menyebabkan tingkat kemiskinan menjadi tinggi. Tingginya pertumbuhan penduduk ini terjadi karena masih tingginya fertilitas di Indonesia, yang diukur melalui angka *Total Fertility Rate* (TFR) atau angka kelahiran total.

Mencapai penduduk tumbuh seimbang melalui upaya penurunan laju pertumbuhan penduduk (LPP) dan perwujudan keluarga berkualitas, telah ditetapkan secara nasional. Untuk mencapai kondisi penduduk tumbuh seimbang tersebut, diharapkan *Total Fertility Rate* (TFR) penduduk di Indonesia sebesar 2,1 per wanita pada tahun 2020. Namun, setelah melihat angka fertilitas total di Indonesia berdasarkan Survei Demografi dan Kesehatan Indonesia (SDKI) 2017 sebesar 2,4 per wanita, yang berarti seorang wanita di Indonesia rata-rata melahirkan 2.4 anak selama masa hidupnya, dikhawatirkan tidak sesuai dengan asumsi proyeksi penduduk yang dibangun ketika TFR pada tahun 2020 yaitu 2,1. Bukan pesimis untuk dapat mencapai angka TFR sesuai asumsi proyeksi tetapi sejarah menunjukkan secara konsisten posisi TFR yang stagnan di kisaran 2,6.



GAMBAR 1. Tren Angka Fertilitas Total, Indonesia 1991-2017

Angka TFR di Indonesia merupakan kontribusi dari setiap daerah atau provinsi. Kontribusi dan variasi yang menonjol terutama berasal dari wilayah timur Indonesia dengan TFR di atas angka nasional, yaitu rata-rata lebih dari 3 anak per wanita. TFR tertinggi terjadi di Nusa Tenggara Timur, Maluku, Papua, dan Papua Barat. Sedangkan, TFR terendah terjadi di Bali, Jawa Timur, DKI Jakarta, DI Yogyakarta dan Sulawesi Utara. Tingginya TFR ini disebabkan oleh salah satunya masih rendahnya partisipasi wanita kawin umur 15-49 tahun dalam penggunaan kontrasepsi modern, yang diukur melalui *Contraceptive Prevalence Rate* (CPR). Hal ini disebabkan oleh beberapa hal diantaranya kejadian kebutuhan keluarga berencana (KB) yang tidak terpenuhi (*unmet need KB*), kurang berjalannya program dan strategi kependudukan, keluarga berencana dan pembangunan keluarga (KKBPK) yang dilakukan pemerintah karena kurangnya Petugas KB di lini lapangan dan menurunnya dukungan pemerintah daerah terhadap program KKBPK sejak otonomi daerah, serta relatif tingginya angka putus pakai kontrasepsi.

Berdasarkan keterangan di atas, maka diperlukan suatu analisis hubungan antara angka fertilitas (TFR) di Indonesia dengan faktor yang mempengaruhinya, yaitu *Contraceptive Prevalence Rate* (CPR), menggunakan analisis regresi. Model regresi nonparametrik dipilih karena sifatnya yang sangat fleksibel dalam memodelkan pola data. Salah satu pendekatan dalam model regresi nonparametrik adalah regresi *kernel*, *smoothing spline* dan *b-splines*. Hasil analisis ini diharapkan dapat dijadikan bahan penyusunan kebijakan yang perlu dipertimbangkan oleh pemerintah, khususnya BKKBN, dalam membuat kebijakan yang terkait dengan upaya pengendalian penduduk.

METODE PENELITIAN

Sumber Data dan Variabel Penelitian

Data yang digunakan dalam penelitian ini adalah data sekunder yang bersumber dari hasil Survei Demografi dan Kesehatan Indonesia (SDKI) Tahun 2017. SDKI 2017 dilaksanakan bersama oleh Badan Pusat Statistik (BPS), BKKBN, dan Kementerian Kesehatan (Kemenkes). Dalam teknis pelaksanaannya, Pemerintah Indonesia dibantu oleh ICF melalui proyek Demographic and Health Surveys (DHS), yaitu program United States Agency for International Development (USAID) yang menyediakan dana dan bantuan teknis dalam pelaksanaan survei kependudukan dan kesehatan di banyak negara.

Objek pengamatan dalam penelitian ini merupakan data mengenai 34 provinsi yang ada di Indonesia, serta variabel yang digunakan terdiri dari variabel respon (y) yaitu angka kelahiran total atau *Total Fertility Rate* (TFR) dan variabel prediktor (x) yaitu *Contraceptive Prevalence Rate* (CPR). Pada umumnya pemakaian alat kontrasepsi, digunakan dengan alasan tidak ingin menambah anak sehingga akan mempengaruhi total kelahiran.

Regresi Nonparametrik

Regresi nonparametrik adalah salah satu metode statistika yang digunakan apabila pola hubungan antara variabel respon dan variabel prediktor tidak diketahui bentuk kurva regresinya. Dalam regresi nonparametrik, kurva regresi hanya diasumsikan mulus (*smooth*) dalam arti termuat pada suatu ruang fungsi tertentu sehingga mempunyai sifat fleksibilitas yang tinggi (Eubank, 1999).

Menurut Hardle (1994) model umum regresi nonparametrik dapat dituliskan sebagai berikut:

$$Y_i = m(X_i) + \varepsilon_i \quad ; i = 1, 2, \dots, n \quad (1)$$

dengan:

Y_i : variabel respon pengamatan ke- i

X_i : variabel prediktor pengamatan komponen nonparametrik ke- i

$m(X)$: fungsi regresi yang tidak diketahui

ε_i : residual ke- i , dimana $\varepsilon_1, \dots, \varepsilon_n$ i.i.d dan $E(\varepsilon_i) = 0$

Dalam setiap regresi nonparametrik, harapan bersyarat dari variabel Y relatif terhadap variabel X dapat ditulis $E(Y|X) = m(X)$ atau $E(Y|X = x) = \int y \frac{f(xy)}{f_x} dy$ (Hardle, 1994).

Penaksir Densitas Kernel

Fungsi penaksir kernel merupakan salah satu pendekatan untuk mengestimasi kurva regresi nonparametrik. Metode ini sering digunakan karena memiliki bentuk yang lebih fleksibel dan perhitungan matematisnya mudah dikerjakan. Penaksir kernel merupakan pengembangan dari estimator histogram yang diperkenalkan oleh Rosenblatt (1956) dan Parzen (1962) sehingga disebut penaksir densitas kernel Rosenblatt-Parzen (Hardle, 1994).

Secara umum fungsi penaksir kernel K dengan *bandwidth* h didefinisikan sebagai berikut:

$$K_h = \frac{1}{h} K\left(\frac{x}{h}\right), \text{ untuk } -\infty < x < \infty, h > 0 \quad (2)$$

Serta memenuhi:

- (i) $K(x) \geq 0$, untuk semua x
- (ii) $\int_{-\infty}^{\infty} K(x) dx = 1$
- (iii) $\int_{-\infty}^{\infty} x^2 K(x) dx = \sigma^2 > 0$
- (iv) $\int_{-\infty}^{\infty} x K(x) dx = 0$

Penaksir densitas kernel untuk fungsi densitas $f(x)$ didefinisikan sebagai berikut:

$$\hat{f}_h(x) = \frac{1}{n} \sum_{i=1}^n K_h(x - x_i) = \frac{1}{nh} \sum_{i=1}^n K\left(\frac{x-x_i}{h}\right) \tag{3}$$

Berdasarkan persamaan (3) di atas menunjukkan bahwa $\hat{f}_h(x)$ bergantung pada fungsi kernel K dan parameter h . Bentuk bobot kernel ditentukan oleh fungsi kernel K , sedangkan ukuran bobotnya ditentukan oleh parameter pemulus h yang disebut *bandwidth*.

Menurut Hardle (1994) ada beberapa jenis fungsi kernel, antara lain:

TABEL 1. Fungsi Kernel

Nama	k	μ	Kernel
Uniform	2	0	$\frac{1}{2} \mathbf{1}_{[-1,1]}(x)$
Triangle	2	0	$(1 - x) \mathbf{1}_{[-1,1]}(x)$
Normal	2	∞	$\frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2} x^2\right)$
Epanechnikov	2	1	$\frac{3}{4} (1 - x^2) \mathbf{1}_{[-1,1]}(x)$
Bisquare	2	2	$\frac{15}{16} (1 - 2x^2 + x^4) \mathbf{1}_{[-1,1]}(x)$
Triweight	2	3	$\frac{35}{32} (1 - 3x^2 + 3x^4 - x^6) \mathbf{1}_{[-1,1]}(x)$
Cauchy	2	0	$[\pi(1 + x^2)]^{-1}$
Picard	2	∞	$\frac{1}{2} \exp(- x)$

dengan μ disebut derajat kehalusan dari fungsi kernel dan $k = 0, \dots, \mu - 1$ merupakan order dari fungsi kernel.

Regresi Kernel

Regresi kernel merupakan salah satu teknik nonparametrik untuk memperkirakan ekspektasi bersyarat dari variabel acak dengan tujuan untuk menemukan hubungan non-linear antara variabel acak X dan Y , dan juga untuk memperoleh estimator menggunakan bobot yang sesuai.

Regresi Kernel Menggunakan Penaksir Nadaraya-Watson (NWE)

Salah satu teknik regresi nonparametrik untuk menaksir fungsi regresi $m(X)$ adalah dengan menggunakan penaksir Nadaraya-Watson (NWE). Secara terpisah di tahun yang sama 1964 Nadaraya dan Watson mempublikasikan metode penaksir $m(X)$, yang selanjutnya disebut metode Nadaraya-Watson (NWE) adalah sebagai berikut:

$$\hat{m}(X) = \frac{\sum_{i=1}^n K_h(X_i - x) Y_i}{\sum_{i=1}^n K_h(X_i - x)} \tag{4}$$

untuk X fixed, penaksir $\hat{\theta}$ yang meminimumkan

$$\sum_{i=1}^n (Y_i - \theta)^2 K_h(X_i - x) \tag{5}$$

memiliki bentuk $\sum_{i=1}^n a_i Y_i$. NWE merupakan minimizer persamaan (5) dimana:

$$a_i = \frac{K_h(X_i - x)}{\sum_{i=1}^n K_h(X_i - x)} \quad (6)$$

Regresi Kernel Menggunakan Penaksir Polinomial Lokal (LPE)

Teknik regresi nonparametrik lain untuk menaksir fungsi regresi $m(X)$ adalah dengan menggunakan penaksir Polinomial Lokal (LPE), yaitu adalah suatu metode regresi nonparametrik, dimana fungsi regresi $m(X)$ ditaksir menggunakan bentuk polinomial. Estimasi $m(X)$ menggunakan LPE akan melibatkan fungsi bobot yaitu fungsi kernel $K_h(X_i - x)$. Bentuk bobot polinomial lokal ditentukan oleh fungsi kernel $K(\cdot)$, sedangkan ukuran bobot ditentukan oleh parameter h yang disebut *bandwidth*.

Model regresi polinomial adalah sebagai berikut:

$$Y_i = \beta_0 + \beta_1(X_i - x) + \beta_2(X_i - x)^2 + \dots + \beta_p(X_i - x)^p + \varepsilon_i \quad (7)$$

Penaksir kuadrat terkecil dengan bentuk polinomial diperoleh sebagai berikut:

$$\sum_{i=1}^n (Y_i - \sum_{k=0}^p \beta_k (X_i - x)^k)^2 K_h(X_i - x) = \min_{\beta_0, \dots, \beta_p} ! \quad (8)$$

Maka penaksir koefisien regresi polinomial lokal adalah

$$\hat{\beta} = (X^T W X)^{-1} X^T W Y \quad (9)$$

dengan:

$$W = \text{diag}(W_{ii})$$

$$W_{ii} = K_h(X_i - x), Y = (Y_1, \dots, Y_n)^T$$

$$X = \begin{pmatrix} 1 & (X_1 - x) & (X_1 - x)^2 & \dots & (X_1 - x)^p \\ 1 & (X_2 - x) & (X_2 - x)^2 & \dots & (X_2 - x)^p \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & (X_n - x) & (X_n - x)^2 & \dots & (X_n - x)^p \end{pmatrix}$$

dengan $X^T W X$ invertible.

Maka, taksiran dari Y pada titik x adalah β_0 dan dapat diperoleh melalui persamaan sebagai berikut:

$$\hat{Y}(x) = \hat{\mu}(x; p, h) = \beta_0 = e_1^T (X^T W X)^{-1} X^T W Y \quad (10)$$

dimana $e_1^T = (1, 0, 0, \dots, 0)$ dengan $p + 1$ elemen.

Oleh karena itu, dalam menaksir $Y(x)$, terdapat tiga tahap yang harus dilakukan:

- a. Bangun matriks X untuk tiap $x \in (X_1, \dots, X_n)$
- b. Bangun matriks W untuk tiap $x \in (X_1, \dots, X_n)$
- c. Taksir $\hat{\mu}(x; p, h)$ dengan persamaan (10)

Pemilihan Bandwidth Optimal

Permasalahan utama pada estimasi kernel adalah pada pemilihan *bandwidth*, bukan pada pemilihan kernelnya. Menurut Hardle (1994) *bandwidth* adalah bentuk bobot kernel yang ditetapkan dengan K , dimana ukuran bobot ditentukan oleh h . *Bandwidth* h adalah parameter pemulus yang berfungsi untuk mengontrol kemulusan dari kurva yang diestimasi. *Bandwidth* yang terlalu kecil akan menghasilkan kurva yang *under-smoothing* yaitu sangat kasar dan sangat fluktuatif, dan sebaliknya *bandwidth* yang terlalu lebar akan menghasilkan kurva yang *over-smoothing* yaitu sangat mulus, tetapi tidak sesuai dengan pola data. *Bandwidth* dari kernel adalah parameter bebas yang menunjukkan pengaruh yang kuat pada perkiraan yang dihasilkan.

Pemilihan *bandwidth* yang optimum dilakukan dengan cara memperkecil tingkat kesalahan. Metode ini merupakan metode yang digunakan untuk menduga kesalahan prediksi. Semakin kecil tingkat kesalahannya semakin baik estimasinya. Untuk mengetahui ukuran tingkat kesalahan suatu estimator, salah satu cara dengan melihat MSE (*Mean Square Error*).

Regresi Spline

Metode regresi nonparametrik lain yang dapat digunakan untuk menduga kurva regresi adalah dengan menggunakan pendekatan regresi *spline*. Regresi *spline* adalah pendekatan ke arah plot data dengan tetap memperhitungkan kemulusan kurva. *Spline* merupakan suatu polinomial dimana segmen-segmen polinomial yang berbeda ditautkan pada *knot* $k_1, k_2, k_3, \dots, k_r$ dan kontinu sehingga bersifat fleksibel dibandingkan polinomial biasa. *Spline* merupakan model polinomial yang tersegmen (*piecewise polynomial* dengan sifat kontinu pada *knot-knotnya*. *Spline* mempunyai titik *knot* yang merupakan titik perpaduan bersama dimana terjadi perubahan perilaku kurva. *Spline* dapat menyesuaikan diri secara efektif terhadap data yang ada, sehingga didapatkan hasil yang mendekati kebenaran (Eubank, 1999).

Fungsi *spline* berorde m dengan satu variabel bebas dan basis *truncated* adalah sembarang fungsi yang secara umum dapat disajikan dalam bentuk:

$$E(y) = m(X) = \beta_0 + \sum_{r=1}^{m-1} \beta_r x^r + \sum_{j=1}^k \beta_{m-1+j} (x - k_j)_+^{m-1} + \varepsilon \quad (11)$$

dengan basis *truncated* :

$$(x - k_j)_+^{m-1} = \begin{cases} (x - k_j)^{m-1} & ; x \geq k_j \\ 0 & ; x \leq k_j \end{cases}$$

dimana k_j adalah *knot* ke- j dari variabel X , $j = 1, 2, \dots, s$, dan s adalah banyaknya *knot*.

Regresi Smoothing Spline

Penduga fungsi pemulus merupakan penduga fungsi yang mampu memetakan data dengan baik serta mempunyai ragam galat yang kecil. Oleh karena itu, dengan menggunakan data pengamatan sebanyak n , maka $m(X_i)$ diperoleh dengan meminimumkan fungsi *Penalized Least Square* (PLS), yaitu:

$$PLS = \underbrace{\sum_{i=1}^n (Y_i - m(X_i))^2}_a + \underbrace{\lambda \int [m''(x)]^2 dx}_b \quad (12)$$

dimana bagian (a) merupakan jumlah kuadrat sisaan atau fungsi jarak antara data dan dugaan, bagian (b) merupakan *roughness penalty*, yaitu ukuran kemulusan kurva dalam memetakan data, dan $0 < \lambda < 1$ adalah parameter pemulus, yaitu pengontrol keseimbangan antara kecocokan terhadap data (*Goodness of Fit*) dan kemulusan kurva (*penalty*). Apabila nilai λ besar mendekati 1, maka akan memberikan bobot *penalty* (kemulusan) yang besar dan mempunyai ragam kecil.

Regresi B-Splines

B-Splines adalah salah satu metode yang digunakan untuk menaksir kurva regresi nonparametrik. Model *B-Splines* dengan titik *knot* digunakan untuk menyelesaikan kelemahan model *spline* pada saat orde yang tinggi, titik *knot* banyak atau *knot* terlalu dekat yang akan membentuk matriks dalam perhitungan hampir singular, sehingga persamaan normal sulit diselesaikan. Metode yang digunakan dalam menaksir parameter *B-Splines* adalah metode Kuadrat terkecil (*Ordinary Least Square*) yang didefinisikan secara rekursif.

B-Splines merupakan fungsi *piecewise polynomial* dengan *support* lokal untuk derajat polinomial tertentu. *B-Splines* ke- j dengan derajat v berdasarkan *knot* sekuens t_0, \dots, t_u untuk $j = 1, \dots, v + u$ dinotasikan dengan formula rekursif berikut:

$$B_j(x; v) = \frac{x-t_j}{t_{j+v-1}-t_j} B_j(x; v-1) + \left(1 - \frac{x-t_{j+1}}{t_{j+v}-t_{j+1}} B_{j+1}(x; v-1) \right) \quad (13)$$

dengan:

$$B_j(x; 0) = \begin{cases} 1, & \text{jika } t_j \leq x \leq t_{j+1} \\ 0, & \text{lainnya} \end{cases}$$

B-Splines dinormalisasi berarti bahwa:

$$\forall x: \sum_{j=1}^{v+u} B_j(x; v) = 1 \quad (14)$$

Pada regresi mean $Y_i = f(x_i) + \varepsilon_i$, $i = 1, \dots, n$, kita selalu mengasumsikan berdistribusi independen dan identik (i.i.d) dengan rata-rata nol dan *finite* varians σ^2 . $f(\cdot)$ diasumsikan merupakan kombinasi yang halus yang dapat diaproksimasi dengan linier kombinasi dari *B-Splines*:

$$f(x) \approx \sum_{j=1}^m \alpha_j B_j(x; v) \quad (15)$$

dimana $\alpha = (\alpha_1, \dots, \alpha_j)$ merupakan vektor koefisien dari basis *B-splines* $B_j(\cdot; v)$ dengan derajat bebas v dan $u+1$ *equidistant knot* untuk $j = 1, \dots, v + u = m$.

Fungsi objektif regresi *B-Splines* adalah sebagai berikut:

$$\hat{\alpha} = \operatorname{argmin}_{\alpha} \left\{ \sum_{i=1}^n \left(Y_i - \sum_{j=1}^m \alpha_j B_j(x_i; v) \right)^2 \right\}$$

Maka $f(x)$ bisa ditaksir melalui persamaan berikut:

$$\hat{f}(x) \approx \sum_{j=1}^m \hat{\alpha}_j B_j(x; v)$$

Pemilihan Knot dalam Regresi Spline

Ukuran kinerja atas penduga kurva regresi dapat ditentukan dari rata-rata kuadrat sisaan (MSE), Eubank (1988) menyebutkan bahwa, fungsi *loss* dan fungsi resiko, serta *Generalized Cross-Validation* (GCV). Rataan kuadrat sisaan merupakan ukuran kinerja yang paling sederhana, yaitu:

$$MSE(\lambda) = \frac{1}{n} (Y - f)^T (Y - f) \quad (16)$$

atau

$$MSE(\lambda) = \frac{1}{n} \sum_{i=1}^n (y_i - f(x_i))^2 \quad (17)$$

Selain itu fungsi kerugian dan fungsi resiko dapat dijadikan ukuran kinerja model. Pada persamaan fungsi *Penalized Least Square* (PLS), pemilihan λ yang optimal sangat penting untuk mendapatkan model penduga kurva regresi yang baik. Oleh karena itu dipilih λ yang meminimumkan fungsi *loss* $L(\lambda)$ berikut:

$$L(\lambda) = \frac{1}{n} \sum_{i=1}^n (y_i - f(x_i))^2 \quad (18)$$

dimana $L(\lambda)$ mewakili ukuran kedekatan dari y_i dan $f(x_i)$. Namun, dalam regresi non parametrik, fungsi $f(x_i)$ tidak diketahui, sehingga harus ditentukan turunan pertama terhadap f . Fungsi resiko, $R(\lambda)$ adalah nilai harapan dari fungsi kerugian, yaitu:

$$R(\lambda) = E[L(\lambda)] = E \left[\frac{1}{n} \sum_{i=1}^n (y_i - f(x_i))^2 \right] \quad (19)$$

Maka diperoleh $P(\lambda)$ sebagai fungsi resiko penduga dari bagi pengamatan baru yaitu y^* , yaitu :

$$P(\lambda) = E \left[\frac{1}{n} \sum_{i=1}^n (y_i^* - f(x_i))^2 \right] \quad (20)$$

$$P(\lambda) = R(\lambda) + \sigma^2$$

Eubank (1999) menunjukkan bahwa ada hubungan antara $P(\lambda)$ dengan $MSE(\lambda)$, yaitu:

$$\hat{P}(\lambda) = MSE(\lambda) + 2\hat{\sigma}^2 \frac{tr(H)}{n} \quad (21)$$

Dimana H adalah $X(X^T X)^{-1} X$. *Generalized Cross-Validation* (GCV) merupakan modifikasi dari *Cross-Validation* (CV), yaitu:

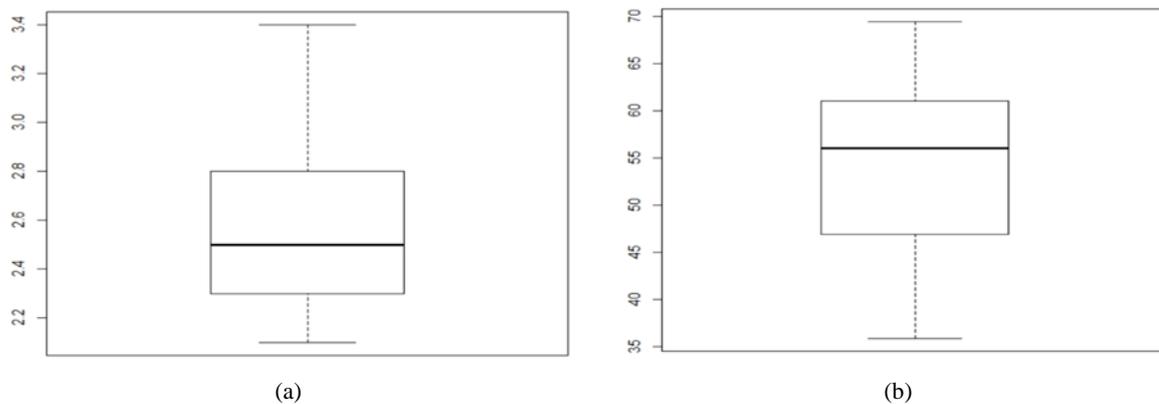
$$GCV(\lambda) = \frac{MSE(\lambda)}{[n^{-1}tr(I-H)]^2} \quad (22)$$

Model regresi spline dikatakan memiliki λ yang optimal jika ketiga kriteria pengujian $MSE(\lambda)$, $P(\lambda)$, dan $GCV(\lambda)$ memiliki nilai yang minimum. Cantoni & Hastie (2000) menyatakan bahwa pemilihan faktor pemulus dalam regresi nonparametrik dapat dilakukan melalui derajat bebas. Hastie & Tibshirani (1990) dalam Cantoni & Hastie (2000) mendefinisikan derajat bebas efektif pada smoothing spline sebagai berikut:

$$df = tr(S) = \sum_{i=1}^n \frac{1}{1+\lambda d_i} \quad (23)$$

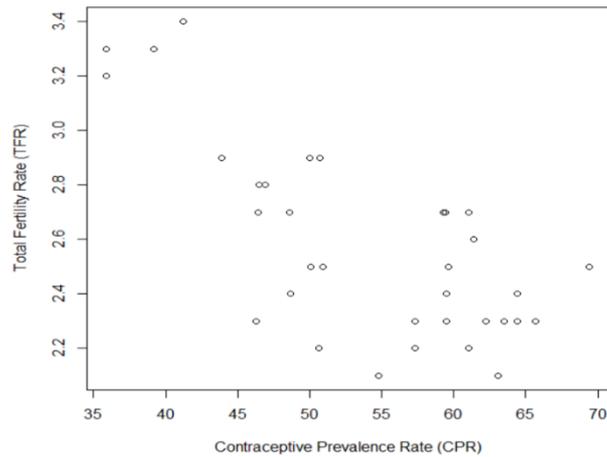
HASIL DAN PEMBAHASAN

Sebelum masuk dalam tahap analisis, eksplorasi data secara univariat dilakukan untuk melihat gambaran dari data yang digunakan dalam penelitian ini. Berdasarkan data yang ada menunjukkan bahwa rata-rata TFR tahun 2017 sebesar 2,6 anak per wanita dengan standar deviasi sebesar 0,4. Sedangkan, CPR tahun 2017 memiliki rata-rata sebesar 54% dengan standar deviasi sebesar 8,9. Angka TFR pada masing-masing provinsi di Indonesia tahun 2017 memiliki rentang yang cukup tinggi, yaitu antara 2.1 anak per wanita sampai dengan 3.4 anak per wanita. Angka CPR pada masing-masing provinsi di Indonesia tahun 2017 juga memiliki rentang yang cukup tinggi, yaitu antara 35.9% sampai dengan 69.4.



GAMBAR 2. a) Boxplot TFR, b) Boxplot CPR.

Pada boxplot di atas menunjukkan bahwa pada data TFR dan CPR di Indonesia tahun 2017 tidak memiliki *outlier*. Selanjutnya pada Gambar 3 di bawah ini menunjukkan bahwa pola hubungan antara TFR dan CPR tidak membentuk sebuah pola tertentu. Sehingga, analisis regresi nonparametrik lebih tepat digunakan untuk menggambarkan pola hubungan antara TFR dan CPR.



Sumber : Hasil Olahan *Software R-3.5.1*

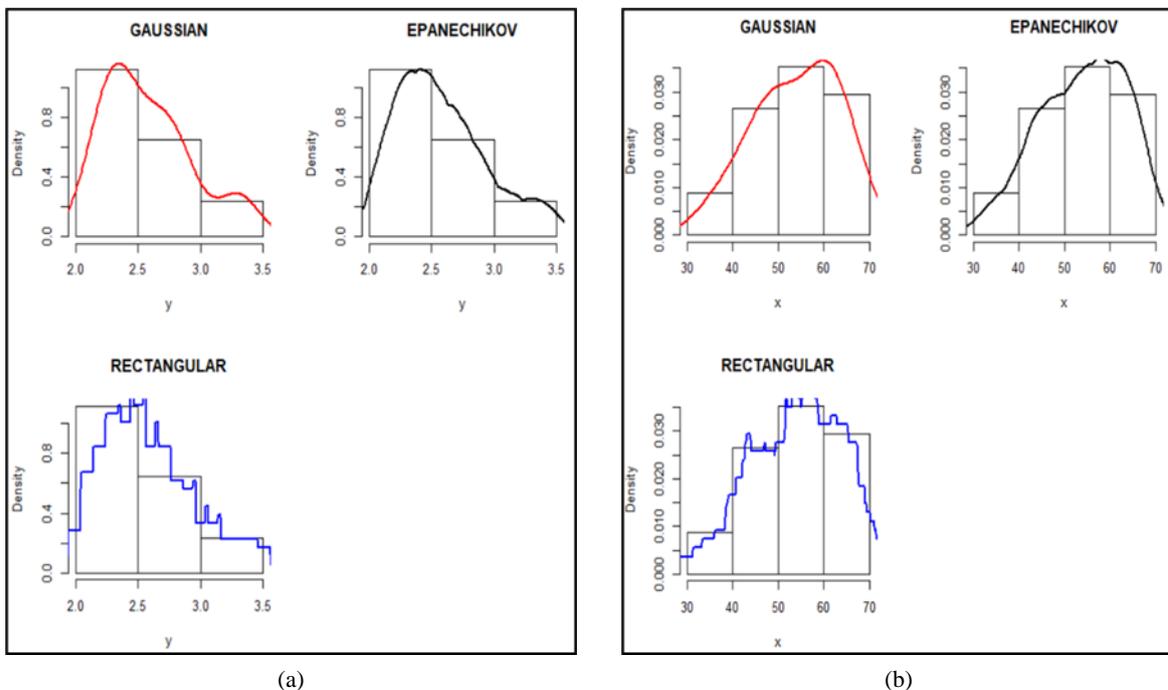
Gambar 3 Scatter Plot TFR dan CPR

Jika dicari keeratan hubungan antara TFR dan CPR dengan menggunakan korelasi pearson diperoleh hasil sebesar -0.7210 , artinya terdapat hubungan negatif yang kuat antara TFR dengan CPR. Jika TFR semakin rendah, maka CPR akan semakin tinggi, begitu pula sebaliknya.

Regresi Kernel

Estimasi Densitas Kernel

Estimasi densitas dilakukan dengan beberapa fungsi kernel (gaussian, rectangular dan epanechikov) untuk data TFR dan CPR. Estimasi densitas untuk setiap fungsi kernel menggunakan *bandwidth* optimal dan selanjutnya dilakukan perbandingan tingkat kemulusan kurva pada setiap fungsi kernel tersebut dengan hasil seperti pada Gambar 4.



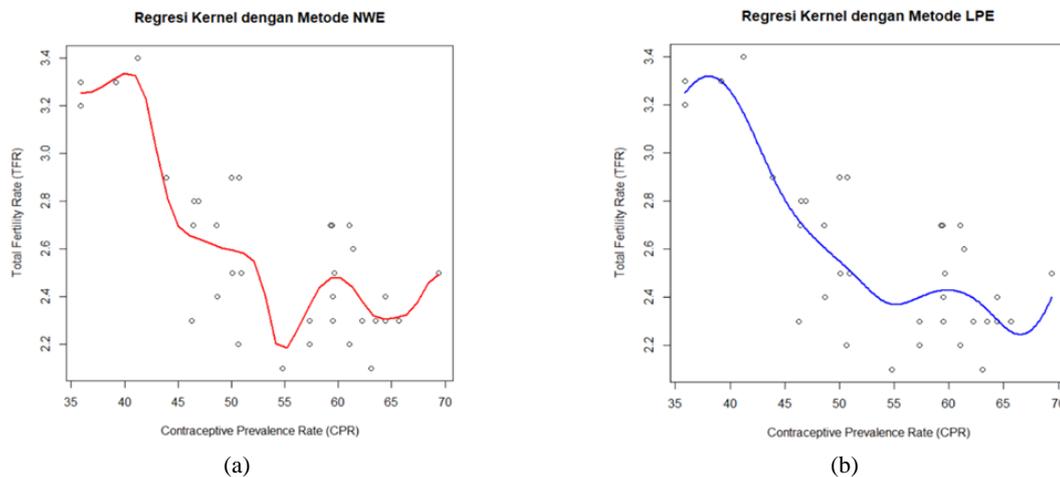
GAMBAR 4. a) Estimasi densitas Kernel untuk TFR, b) Estimasi densitas Kernel untuk CPR

Berdasarkan Gambar 4 (a) di atas menunjukkan untuk data TFR, fungsi kernel Gaussian dengan nilai h sebesar 0.1504074 merupakan fungsi yang memiliki kurva paling mulus dibandingkan dengan fungsi kernel lainnya. Sedangkan, fungsi kernel *Rectangular* dengan nilai h sebesar 0.2617175 terlihat sangat kasar. Begitu juga untuk data CPR pada Gambar 4.b di atas menunjukkan bahwa fungsi kernel Gaussian dengan nilai h sebesar 5.2 merupakan fungsi yang memiliki kurva paling mulus dibandingkan dengan fungsi kernel lainnya, sedangkan fungsi kernel *Rectangular* dengan nilai h sebesar 8 terlihat sangat kasar. Oleh karena itu, untuk selanjutnya regresi kernel pada penelitian ini akan menggunakan fungsi kernel Gaussian.

Regresi Kernel menggunakan Metode NWE dan LPE

Pemodelan regresi nonparametrik dengan pendekatan kernel didasarkan pada fungsi densitas kernel. Pendekatan kernel memiliki bentuk yang lebih fleksibel dan perhitungan matematisnya mudah disesuaikan. Ketepatan suatu pemulus kernel sebagai penaksir ditentukan oleh dua hal yaitu *bandwidth* (h) dan fungsi kernel yang digunakan sebagai bobot.

Regresi kernel menggunakan metode NWE dan LPE diperoleh nilai *bandwidth* optimal sebesar 3.932285, sehingga kurva regresinya adalah sebagai berikut:



GAMBAR 5. a) Kurva Regresi Kernel dengan Metode NEW, b) Kurva Regresi Kernel dengan Metode LPE

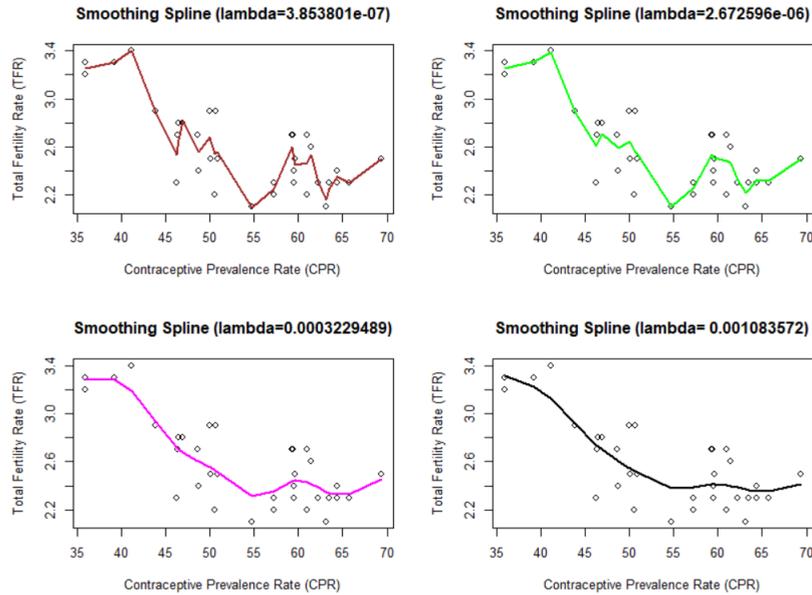
Berdasarkan Gambar 5(a) dan 5(b) di atas menunjukkan bahwa kurva regresi kernel dengan metode NWE lebih mengikuti sebaran data dibandingkan metode LPE. Namun, regresi kernel dengan metode LPE memiliki kurva yang *smooth* (halus) dibandingkan metode NWE. Sehingga dari kedua model tersebut model LPE lebih baik karena kurvanya lebih *smooth* (halus).

Regresi Spline

Regresi Smoothing Spline

Pemodelan regresi nonparametrik smoothing spline memiliki kurva yang sangat dipengaruhi oleh nilai dari parameter pemulus (λ). Apabila nilai parameter pemulus semakin besar, maka bentuk kurva akan mulus, namun hasil yang diperoleh belum tentu baik. Kurva regresi dikatakan baik apabila kurva regresi tersebut halus dan mendekati bentuk dari plot dengan bias yang kecil.

Beragam nilai parameter pemulus (λ) diujicobakan untuk regresi *smoothing spline*. Nilai parameter pemulus (λ) optimal adalah $\lambda = 0.0003229$, kerana memiliki nilai GCV (λ) paling kecil yaitu sebesar 0.0541. Pada kurva regresi *smoothing spline* di bawah ini menunjukkan bahwa semakin kecil nilai λ maka kurvanya semakin kasar. Sedangkan, jika dengan nilai λ semakin besar maka semakin mulus kurvanya.



Sumber : Hasil Olahan *Software R-3.5.1*

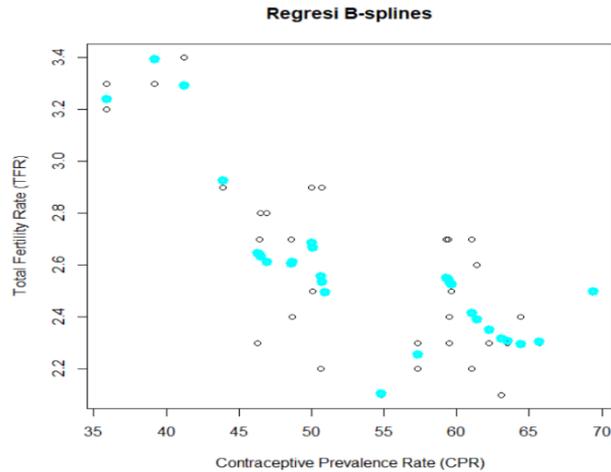
GAMBAR 6. Kurva Regresi *Smoothing Spline* berdasarkan nilai parameter yang diujicoba.

Berdasarkan kurva di atas maka regresi *smoothing spline* terbaik adalah model regresi *smoothing spline* dengan nilai parameter pemulus (λ) optimal sebesar 0.0003229 karena memiliki kurva yang *smooth* (halus) dan lebih dekat dengan sebaran data.

Regresi *B-splines*

Dalam membentuk model regresi *B-splines* terdapat tiga kriteria yang harus diperhatikan, yaitu menentukan orde untuk model, banyaknya *knot*, dan lokasi penempatan *knot*. *Knot* merupakan titik perpaduan bersama dimana terdapat perubahan perilaku pola pada interval yang berlainan. Untuk memperoleh model *B-spline* yang optimal (terbaik) maka perlu dipilih lokasi *knot* yang optimal pula. Kriteria yang digunakan dalam pemilihan *knot* yang optimal pada penelitian ini yaitu *Generalized Cross Validation* (GCV). Pasangan nilai GCV dan RSS minimum diperoleh model *degree 2* dengan jumlah *knot* sebanyak 5, dimana *knot* optimalnya adalah 43.6850; 49.8781; 50.1028; 58.9402; 59.4030. Oleh karena itu, model regresi *B-splines* terbaik adalah regresi *B-splines degree 2* dengan jumlah *knot* 5.

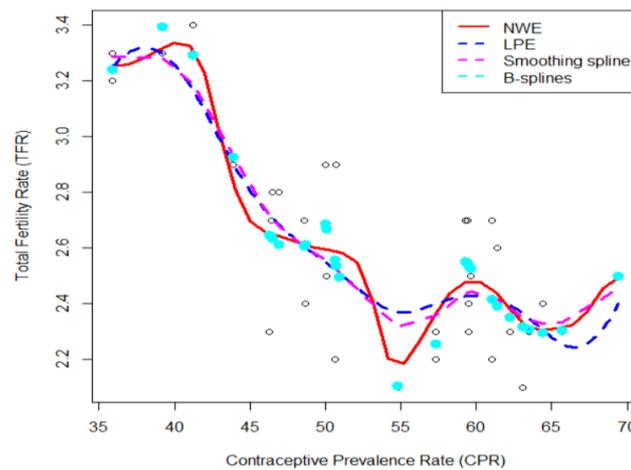
Adapun kurva dari regresi *B-splines* terbaik ialah kurva yang memiliki bentuk yang lebih *smooth* (halus) dengan nilai *r-square* yang lebih tinggi dan didapat hasil kurva terbaik adalah sebagai berikut:



GAMBAR 7. Kurva Regresi *B-splines* dengan *degree* 2 dan jumlah *knot* 5

FITTING CURVE REGRESI NONPARAMETRIK

Pendekatan regresi nonparametrik kernel dengan metode NWE dan LPE, *smoothing spline*, dan *B-splines* yang telah diperoleh pada analisis sebelumnya, yang menggambarkan pengaruh dari CPR terhadap CPR, maka perbandingan kurvanya adalah sebagai berikut:



GAMBAR 8. *Fitting Curve* Regresi Kernel dengan Metode NWE, LPE, *Smoothing Spline* dan *B-splines*

Pada kurva di atas menunjukkan bahwa model regresi nonparametrik terbaik yang menggambarkan pola hubungan TFR dan CPR adalah model regresi *B-splines* dengan *degree* 2 dan jumlah *knot* 5. Hal ini dikarenakan kurva tersebut *smooth* (halus) dan terlihat lebih mengikuti sebaran data dibandingkan kurva model regresi lainnya. Uji koefisien regresi secara simultan (uji F) menunjukkan bahwa model regresi *B-splines degree* 2 dengan jumlah *knot* 5 merupakan model yang signifikan, karena memiliki nilai *p-value* lebih kecil dari tingkat kesalahan ($\alpha = 0.05$). Selanjutnya, uji koefisien regresi secara parsial (uji t) menunjukkan bahwa semua koefisien regresi berarti atau signifikan kecuali α_1 , karena memiliki nilai *p-value* lebih kecil dari tingkat kesalahan ($\alpha = 0.05$). Nilai koefisien determinasi R^2 sebesar 76.86%, artinya besarnya variasi variabel TFR yang dijelaskan oleh model regresi *B-splines* sebesar 76.86%, sedangkan sisanya 23.14% dijelaskan oleh variabel lainnya yang tidak dimasukkan ke dalam model. Sehingga dapat disimpulkan bahwa terdapat hubungan antara *Total Fertility Rate* (TFR) dengan *Contraceptive Prevalence Rate* (CPR).

Adapun persamaan regresinya adalah sebagai berikut:

$$\hat{y} = 3.2397 + 0.418B_1(x; 2) - 0.827B_2(x; 2) - 0.5442B_3(x; 2) - 1.6158B_4(x; 2) - 0.6726B_5(x; 2) - 1.1698B_6(x; 2) - 0.74B_7(x; 2)$$

KESIMPULAN

Berdasarkan hasil dan pembahasan di atas dapat disimpulkan bahwa:

1. Terdapat hubungan negatif antara TFR dan CPR, dimana nilai korelasi pearson sebesar - 0.7210. Jika TFR semakin rendah, maka CPR akan semakin tinggi, begitu pula sebaliknya.
2. Pengepasan kurva untuk membandingkan model regresi nonparametrik kernel dengan metode NWE dan LPE, *smoothing spline* dan *B-spline* menunjukkan model terbaik yang menggambarkan pola hubungan TFR dan CPR adalah model regresi *B-splines* dengan *degree* 2 dan jumlah *knot* 5. Hal ini dikarenakan kurva tersebut *smooth* (halus) dan terlihat lebih mengikuti sebaran data dibandingkan kurva model regresi lainnya.
3. Model regresi *B-splines* dengan *degree* 2 dan jumlah *knot* 5 memiliki nilai koefisien determinasi R^2 sebesar 76.86%, artinya besarnya variasi variabel TFR yang dijelaskan oleh model regresi *B-splines* sebesar 76.86%, sedangkan sisanya 23.14% dijelaskan oleh variabel lainnya yang tidak dimasukkan ke dalam model.

UCAPAN TERIMA KASIH

Terima kasih kepada Bapak Yudhie Andriyana, M.Sc., Ph.D. atas masukan dan bimbingannya untuk penelitian ini.

REFERENSI

- Andriyana, Yudhie. 2015. Materi Ajar Statistika Nonparametrik. Bandung : Universitas Padjadjaran.
- BKKBN, Kemenkes, BPS dan USAID. 2018. Survei Demografi dan Kesehatan Indonesia 2017. Jakarta : Badan Kependudukan dan Keluarga Berencana Nasional (BKKBN).
- BPS. 2018. Statistik Indonesia : *Statistical Yearbook of Indonesia*. Jakarta : Badan Pusat Statistik (BPS).
- Eubank, R. 1999. *Nonparametric Regression and Spline Smoothing*. New York : Marcel Dekker
- Hardle, W. 1990. *Applied Nonparametric Regression*. Cambridge University Press