Jurnal Ekonomi dan Statistik Indonesia
2022, Vol. 2, No. 2, 172 – 181
http://dx.doi.org/10.11594/jesi.02.02.05
E ISSN : 2777-0028

Research Article

Penerapan Analisis Regresi Nonparametrik dengan Pendekatan Regresi Kernel dan Spline

Ade Famalika[1]*, Pardomuan Robinson Sihombing[2]

[1]Universitas Bina Insan, Indonesia
[2]Badan Pusat Statistik, Indonesia

Article history:
Submission August 2022
Revised August 2022
Accepted August 2022

**Corresponding author:*
E-mail:
adefamalika@univbinainsan.ac.id

ABSTRACT

Penelitian ini bertujuan untuk menerapkan regresi nonparametrik menggunakan regresi kernel dan spline. Regresi kernel menggunakan metode penaksir Nadaraya-Watson (NWE) dan penaksir Polinomial Lokal (LPE), sedangkan untuk regresi spline adalah *smoothing spline* dan *b-splines*. Metode ini diterapkan dalam menganalisis pola hubungan Pertumbuhan Produksi Industri (PPI) dan Tingkat Pajak Perusahaan (TPP). 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. TPP berpengaruh signifikan negatif terhadap PPI artinya kenaikan TPP akan menurunkan PPI. Oleh sebab itu perlu kebijakan yang komprehensif dalam menerapkan nilai TPP agar tetap dapat meningkatkan produktivitas industri.

Keywords: *kernel, LPE, nonparametrik, spline, TPP*

Pendahuluan

Analisis regresi merupakan salah satu analisis dalam statistika yang digunakan untuk menyelidiki pola hubungan fungsional antara variabel respon dan variabel prediktor (mencari bentuk estimasi kurva regresi). Terdapat dua pendekatan dalam analisis regresi untuk mengestimasi kurva regresi, yaitu pendekatan regresi parametrik dan regresi nonparametric (Yani, 2017).

Pertimbangan untuk memilih metode mana yang akan digunakan tergantung dari informasi tentang bentuk fungsi f. Pendekatan metode parametrik digunakan apabila ada informasi sebelumnya tentang bentuk fungsi f, dengan kata lain bentuk hubungan regresinya diketahui dan diasumsikan mengikuti pola data regresi (Suparti et al., 2018).

Sedangkan pendekatan metode nonparametrik digunakan jika bentuk fungsi f tidak diketahui, sehingga memberikan fleksibel yang lebih besar. Bentuk pola hubungan dapat diidentifikasi berdasarkan pada informasi masa lalu atau scatter plot data (Hardle, 1994). Dalam beberapa kasus, tidak semua pola data dapat didekati dengan pendekatan parametrik,

How to cite:
Famalika, A. & Sihombing, P. R. (2022). Penerapan Analisis Regresi Nonparametrik dengan Pendekatan Regresi Kernel dan Spline. *Jurnal Ekonomi dan Statistik Indonesia*. *2* (2), 172 – 181. doi: 10.11594/jesi.02.02.05

dimana dapat diakibatkan karena tidak dimilikinya informasi mengenai bentuk atau pola hubungan antara variabel penjelas dengan variabel respon, sehingga pendekatan nonparametrik menjadi lebih relevan digunakan. Penggunaan regresi nonparametrik dengan pendekatan kernel atau spline telah digunakan oleh beberapa penelitian terdahulu diantaranya Sanusi et al (2019) pada studi kasus berat badan lahir rendah di Rumah Sakit Ibu dan Anak Siti Fatimah Makassar dan. Selain itu Wulandary et al (2020) pada pemodelan rata-rata lama sekolah dan pengeluaran perkapita di Indonesia.

Pada penelitian ini analisis regresi nonparametrik diterapkan untuk melihat pola hubungan antara Pertumbuhan Produksi Industri (PPI) dan Tingkat Pajak Perusahaan (TPP) yang terjadi di beberapa Negara di dunia. Hal ini karena dalam proses pembangunan, sektor industri dijadikan sebagai prioritas pembangunan yang diharapkan mempunyai peranan sebagai leading sektor atau sektor pemimpin bagi pembangunan sektor-sektor lainnya (Arsyad, 2010). Leading sektor maksudnya adalah dengan pembangunan industri maka memacu dan mengangkat pembangunan sektor-sektor lainnya seperti sektor pertanian dan sektor jasa. Menurut Lewis dalam Todaro (2006), pengaruh pertumbuhan ekonomi terhadap penyerapan tenaga kerja dimulai dari investasi di sektor industri, dan akumulasi modal secara keseluruhan di sektor modern akan menimbulkan perluasan output pada sektor modern tersebut.

Pertumbuhan produksi industri dapat dipengaruhi oleh beberapa faktor, salah satunya adalah tingkat pajak perusahaan atau rata-rata jumlah tarif pajak perusahaan yang ada dalam suatu negara. Pajak adalah suatu kewajiban seluruh warga negara karena pajak memiliki peranan penting bagi kemajuan perekonomian suatu negara. Untuk membiayai kebutuhan negara, pemerintah sangat mengandalkan potensi penerimaan pajak sebagai sumber pembiayaan terbesar. Salah satu sasaran terbesar pemerintah dalam memungut pajak adalah perusahaan atau Badan Usaha, karena menggeluti dunia bisnis dan berbagai kegiatan usaha. Dalam menganalisis hubungan kedua tersebut salah satunya adalah dengan menggunakan analisis regresi.

Metodologi

Data

Data yang digunakan adalah data sekunder tentang Pertumbuhan Produksi Industri (Y) dan Tingkat Pajak Perusahaan (X) masing-masing dalam satuan persen pada beberapa negara di dunia pada Tahun 2021 yang diakses di website https://id.tradingeconomics.com/stocks.

Data pertumbuhan produksi industri yang dimaksud dalam penelitian ini adalah data yang digunakan untuk mengetahui ataupun membandingkan perkembangan produksi industri yang terjadi pada masing-masing negara di seluruh dunia.

Langkah-langkah Analisis

1. Membuat scatter plot antara Pertumbuhan Produksi Industri (PPI) dan Tingkat Pajak Perusahaan (TPP).
2. Melakukan pemodelan data dengan menggunakan pendekatan nonparametrik regresi kernel, dengan langkah-langkah sebagai berikut:
   a. Menentukan penaksir densitas kernel;
   b. Menentukan bandwidth yang optimal;
   c. Melakukan estimasi regresi kernel menggunakan metode NWE;
   d. Melakukan estimasi regresi kernel menggunakan metode LPE;
3. Melakukan pemodelan data dengan menggunakan pendekatan nonparametrik regresi *spline* dengan langkah-langkah sebagai berikut:
   a. Menentukan nilai parameter pemulus (λ) yang optimal;
   b. Melakukan estimasi regresi *smoothing spline* berdasarkan jumlah *knot*s dan nilai parameter pemulus (λ) yang optimal;
   c. Menentukan jumlah *knot* yang optimal;
   d. Menentukann titik *knot* sebanyak jumlah *knot* yang optimal;
   e. Melakukan estimasi regresi *B-splines* dengan mengkombinasikan beberapa pilihan *knot*.

4. Membandingkan hasil *fitting curve* yang dihasilkan dari pendekatan metode kernel dan *spline.*
5. Memilih model regresi nonparametrik terbaik:
   a. Melakukan pengujian hipotesis koefisien regresi nonparametrik;
   b. Menghitung nilai R^2

Regresi Nonparametrik

Statistik nonparametrik disebut juga statistik bebas sebaran/ distribusi. Statistik nonparametrik tidak mensyaratkan bentuk sebaran parameter populasi. Statistik nonparametrik dapat digunakan pada data yang memiliki sebaran normal atau tidak. Statistik nonparametrik biasanya digunakan untuk melakukan analisis pada data nominal atau ordinal. Dalam regresi nonparamterik kurva regresi hanya diasumsikan mulus (smooth) dalam arti termuat dalam suatu ruang fungsi tertentu sehingga mempunyai sifat fleksibilitas yang tinggi (Eubank, 1999). Model regresi nonparametrik secara umum adalah sebagai berikut:

$$Y_i = f\ (t_i) + \varepsilon_i \quad ; i = 1, \dots n \quad (1)$$

dengan Y_i adalah variable respon, t_i merupakan variabel prediktor, $f\ (t_i)$ adalah kurva regresi yang tidak diketahui bentuknya, dimana ε_i adalah error random berdistribusi normal, dengan mean nol dan varians σ^2. Estimasi fungsi regresi nonparametrik dilakukan berdasarkan data pengamatan dengan menggunakan teknik smoothing. Terdapat beberapa teknik smoothing dalam model regresi nonparametrik antara lain penaksir kernel dan penaksir spline (Eubank, 1999).

Estimasi 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. Estimator kernel merupakan pengembangan dari estimator histogram yang diperkenalkan oleh Rosenblatt (1956) dan Parzen (1962) sehingga disebut estimator densitas kernel Rosenblatt-Parzen (Hardle, 1994). Secara umum fungsi penaksir kernel K dengan bandwidth h didefinisikan sebagai :

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

dengan memenuhi:

- $K(x) \geq 0, \forall x$
- $\int_{-\infty}^{\infty} K(x)dx = 1$
- $\int_{-\infty}^{\infty} x^2\ K(x)dx = \sigma^2 > 0$
- $\int_{-\infty}^{\infty} xK(x)dx = 0$

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

$$\hat{f}_h(x) = \frac{1}{n}\sum_{i=1}^{n} K_h(x - x_i)$$

$$= \frac{1}{nh}\sum_{i=1}^{n} K\left(\frac{x - x_i}{h}\right) \quad (3)$$

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

1. Kernel Uniform: $K(x) = \frac{1}{2} I\ (|x| \leq 1)$
2. Kernel Triangle: $K(x) = (1 - |x|)I\ (|x| \leq 1)$
3. Kernel Epanechnikov: $K(x) = \frac{3}{4}(1 - x^2)I\ (|x| \leq 1)$
4. Kernel Kuartik: $K(x) = \frac{15}{16}(1 - x^2)^2 I\ (|x| \leq 1)$
5. Kernel Triweight: $K(x) = \frac{35}{32}(1 - x^2)^3 I\ (|x| \leq 1)$
6. Kernel Cosinus : $K(x) = \frac{\pi}{4}\cos\left(\frac{\pi}{2}x\right) I\ (|x| \leq 1)$
7. Kernel Gaussian: $K(x) = \frac{1}{\sqrt{2\pi}}\left(\frac{1}{2}(-x^2)\right) \quad ; \quad -\infty < x < \infty$

dengan x disebut sebagai derajat kehalusan dari fungsi kernel dan $k = 0, \dots, x - 1$ merupakan order dari fungsi kernel, serta I disebut sebagai indikator.

Penaksir Nadaraya-Watson (NWE)

Salah satu teknik regresi non parametrik untuk menaksir fungsi regresi $(m(.))$ adalah dengan menggunakan penaksir Nadaraya-Watson. Secara terpisah di tahun yang sama 1964

Nadaraya dan Watson mempublikasikan metode penaksir m(.), yang selanjutnya disebut metode Nadaraya – Watson Estimator (NWE) sebagai berikut:

$$\widehat{m}(x) = \frac{\sum_{i=1}^{n} K\left(\frac{x - X_i}{h}\right) Y_i}{\sum_{k=1}^{n} \left(\frac{x - X_k}{h}\right)} \quad (4)$$

dengan:

$\widehat{m}(x)$ = fungsi taksiran regresi
x = variabel prediktor yang nilainya tidak teramati namun akan digunakan untuk menaksir Y
X_i = variabel prediktor pada data ke-i
Y_i = variabel respon pada data ke-i
$K(.)$ = fungsi kernel
n = ukuran sampel/ banyak pengamatan
h = lebar *bandwidth*

Cara lain untuk memperoleh penduga Nadaraya-Watson $\widehat{m}(x)$, yaitu dengan meminumkan kuadrat kesalahan (*square error*) yang diberi bobot, yaitu sebagai berikut:

$$E_{constant}(x) = \sum_{i=1}^{n} (Y_i - a_0)^2 \ w\left(\frac{x - X_i}{h}\right) \quad (5)$$

dengan $w\left(\frac{x-X_i}{h}\right)$ adalah pembobotnya. Dengan menurunkan $E_{constant}(x)$ terhadap a_0, maka akan diperoleh:

$$\widehat{m}(x) = a_0 = \frac{\sum_{i=1}^{n} w\left(\frac{x - X_i}{h}\right) Y_i}{\sum_{i=1}^{n} \left(\frac{x - X_i}{h}\right)} \quad (6)$$

Pada persamaan tersebut ditetapkan bahwa $w\left(\frac{x-X_i}{h}\right)$ sebagai $K\left(\frac{x-X_i}{h}\right)$. Dari cara ini, penduga Nadaraya-Watson menjadi sebuah metode untuk mendugai $m(x)$ dengan menggunakan $w\left(\frac{x-X_i}{h}\right)$ sebagai fungsi pembobot. Metode ini disebut sebagai metode pengepasan konstanta secara lokal (*method of fitting constants locally*) karena $m(x)$ terdiri dari konstantakonstanta lokal. Oleh karena itu teknik ini disebut pendugaan konstanta lokal.

Penaksir Polinomial Lokal (LPE)

Aproksimasi $m(x)$ secara lokal dengan sebuah konstanta $m(x) \approx m_0$ melalui kuadrat terkecil lokal

$$\sum_{i=1}^{n} (Y_i - \beta)^2 \ K_h(X_i - x) = \min_{\beta}! \quad (7)$$

$$\frac{\partial}{\partial \beta} \sum_{i=1}^{n} (Y_i - \beta)^2 \ K_h(X_i - x) \stackrel{!}{=} 0 \quad (8)$$

Sehingga diperoleh

$$\beta = m_0 = m(X) = \frac{\frac{1}{n}\sum_{i=1}^{n} K_h(X_j - x) Y_i}{\frac{1}{n}\sum_{i=1}^{n} K_h(X_i - x)} \quad (9)$$

disebut juga Penaksir Kuadrat Terkecil Lokal (PKTL).

Misalkan model regresi polinomial sebagai berikut:

$$Y_i = \beta_0 + \beta_1(X_i - x) + \beta_2(X_i - x)^2 + \ldots + \beta_p(X_i - x)^p + \varepsilon_i \quad (10)$$

- Gunakan Kernel sebagai pengganti bobot
- dikatakan lokal karena setiap bobot diberikan pada setiap titik x –
- penaksir kuadrat terkecil dengan bentuk polinomial diperoleh sebagai berikut

$$\sum_{i=1}^{n} \left(Y_i - \sum_{k=0}^{p} \beta_k (X_i - x)^k \right)^2 K_h(X_i - x) = \min_{\beta_0,\ldots,\beta_p}! \quad (11)$$

Maka penaksir koefisien regresi polinomial lokal adalah

$$\hat{\beta} = (X^T W X)^{-1} X^T W Y \quad (12)$$

dimana $W = diag(W_{ii})$ dengan $W_{ii} = K_h(X_i - x)$, $Y = (Y_1, \ldots, Y_n)^T$ dan

$$X = p\begin{pmatrix} 1 & (X_1 - x) & (X_1 - x)^2 & \ldots & (X_1 - x)^p \\ 1 & (X_2 - x) & (X_2 - x)^2 & \ldots & (X_2 - x)^p \\ : & : & : & : & : \\ . & . & . & . & . \\ 1 & (X_n - x) & (X_n - x)^2 & \ldots & (X_n - x)^p \end{pmatrix}$$

dengan syarat $X^T W Y$ *invertible*.

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

$$\hat{Y}(x) = \hat{\mu}(x; p, h) = \beta_0 = e_1^T \ (X^T W X)^{-1} X^T W Y \quad (13)$$

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

Sehingga untuk menaksir $Y(x)$, terdapat tiga tahap yang harus dilakukan yaitu;

1. Bangun matriks X untuk $x \in (X_1, \ldots, X_n)$
2. Bangun matriks W untuk $x \in (X_1, \ldots, X_n)$
3. Taksir $\hat{\mu}(x; p, h)$ dengan menggunakan Persamaan (13).

Pemilihan Bandwidth Optimal

Memilih bandwidth yang sesuai (parameter smoothing) adalah bagian penting dari regresi nonparametrik. Untuk mendapatkan bandwidth yang tepat maka harus ditemukan keseimbangan antara varians dan bias. Permasalahan utama pada kernel smoothing bukan terletak pada pemilihan kernel tetapi pada pemilihan bandwidth. Pemilihan bandwidth optimum lebih ditekankan pada penyeimbangan antara bias dan varians.

Menurut Hardle (1994), 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.

Salah satu metode pemilihan bandwidth optimal adalah menggunakan metode Cross Validation. Metode ini merupakan metode yang digunakan untuk menduga kesalahan prediksi. Berikut fungsi dari Cross Validation:

$$CV(h) = n^{-1} \sum_{j=1}^{n} \left[Y_j - \widehat{m}_{h.j}(X_j)\right]^2 w(X_j) \quad (14)$$

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

Salah satu model regresi dengan pendekatan nonparametrik yang dapat digunakan untuk menduga kurva regresi adalah regresi *spline*. Regresi *spline* adalah suatu pendekatan ke arah plot data dengan tetap memperhitungkan kemulusan kurva. *Spline* merupakan suatu polinomial dimana segmen-segmen polinomial yang berbeda digabungkan bersama pada *knots* $k_1, k_2, \ldots, k_r$ dan kontinu sehingga bersifat fleksibel dibandingkan polinomial biasa. Jadi dapat dikatakan bahwa *spline* merupakan model polinomial yang tersegmen (*piecewise polynomial*), tetapi *spline* masih bersifat kontinu pada *knot-knot*nya. *Spline* mempunyai titik *knot* yang maksudnya adalah titik perpaduan bersama dimana terjadi perubahan perilaku kurva. *Spline* dapat menyesuaikan diri secara efektif terhadap data tersebut, sehingga didapatkan hasil yang mendekati kebenaran (Eubank, 1999).

Bentuk umum regresi spline orde ke m adalah sebagai berikut:

$$f(x) = \beta_0 + \sum_{j=1}^{m} \beta_j x^j + \sum_{k=1}^{N} \beta_{j+k} (x - K_k)_+^{m-1} + \varepsilon \quad (15)$$

jika menggunakan data amatan sebanyak n, maka bentuk matriks dari Persamaan (15) dapat ditulis:

$$f(x) = X_1\delta_1 + (X - K)\delta_2 + \varepsilon \quad (16)$$

dengan

$$X_1 = \begin{bmatrix} 1 & x_1 & x_1^2 & \cdots & x_1^m \\ 1 & x_2 & x_2^2 & \cdots & x_2^m \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & x_n & x_n^2 & \cdots & x_n^m \end{bmatrix}; \quad \delta_1 = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_m \end{bmatrix};$$

$$\delta_1 = \begin{bmatrix} \beta_{m+1} \\ \beta_{m+2} \\ \vdots \\ \beta_{m+N} \end{bmatrix}; \quad \varepsilon = \begin{bmatrix} \varepsilon_0 \\ \varepsilon_1 \\ \vdots \\ \varepsilon_n \end{bmatrix}$$

$$X - K = \begin{bmatrix} (x_1 - k_1)^m & (x_1 - k_2)^m & \cdots & (x_1 - k_N)^m \\ (x_2 - k_1)^m & (x_2 - k_2)^m & \cdots & (x_2 - k_N)^m \\ \vdots & \vdots & \ddots & \vdots \\ (x_n - k_1)^m & (x_n - k_2)^m & \ldots & (x_n - k_N)^m \end{bmatrix}$$

Regresi B-Spline

B-splines merupakan fungsi *piecewise polynomial* dengan suport lokal untuk derajat polinomial tertentu (Boor, 2001). *B-splines* ke-*j* dengan derajat V berdasarkan *knot* sekuens $t_0, \ldots, t_u$ untuk $j = 1, \ldots, v + u$ dinotasikan dengan formula rekursif berikut:

$$B_j(x;v) = \frac{x - t_j}{t_{j+v-1} - t_j} B_j(x;v-1)$$
$$+ \left(1 - \frac{x - t_{j+1}}{t_{j+v-1} - t_{j+1}} B_{j+1}(x;v-1)\right) \quad (17)$$

dengan

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

B-splines dinormalisasi, berarti bahwa

$$\forall\, x : \sum_{j=1}^{v+u} B_j\,(x;v) = 1 \quad (18)$$

Fungsi objektif regresi *B-splines*:

$$\hat{\alpha} = \arg min_a \left\{ \sum_{i=1}^{n} \left(Y_i - \sum_{j=1}^{m} \alpha_j B_j(x;v) \right)^2 \right\} \quad (19)$$

Ada 3 kriteria yang harus diperhatikan dalam membentuk model regresi *spline* dengan basis fungsi *B-spline* yaitu menentukan order untuk model, banyaknya *knot*, dan lokasi penempatan *knot* (Rahasia dkk, 2020). Order untuk model dapat ditentukan berdasarkan pola umum yang terjadi pada data, sedangkan banyaknya *knot* dan lokasi *knot* ditentukan berdasarkan perubahan pola di daerah tertentu pada kurva. Untuk memperoleh model *B-spline* yang optimal (terbaik) maka perlu dipilih lokasi *knot* yang optimal pula (Gun et al., 2018).

Smoothing Spline

Cantoni & Hastie (2000) menyatakan bahwa pemilihan faktor pemulus dalam regresi nonparametrik dapat dilakukan melalui derajat bebas. Seperti telah diketahui bahwa *smoothing spline* dihitung untuk menyesuaikan vector $y = (y_1, \ldots, y_n)$ dengan model fit $y = \hat{y} = \hat{f} = SY$. Parameter λ mengontrol kemulusan kurva regresi yang dapat juga didefinisikan melalui derajat bebas. Hastie & Tibshirani (1990) dalam Cantoni & Hastie (2000) mendefinisikan derajat bebas efektif pada smoothing spline sebagai berikut:

$$df = tr(S) = \sum_{i=1}^{n} \frac{1}{1 + \lambda d_i} \quad (20)$$

Dimana d_i adalah akar ciri dari matrik K. Persamaan (20) menunjukkan adanya hubungan terbalik antara λ dengan derajat bebas df.

Hasil dan Pembahasan

Statistik Deskriptif

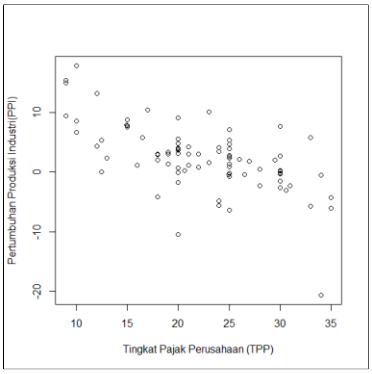
Statistik deskriptif digunakan untuk melihat ukuran-ukuran pemusatan data sehingga dapat melihat perkembangan nilai variabel penelitian selama periode penelitian.

Tabel 1. Statistik Deskriptif Variabel

Statistik	PPI	TPP
Min	-20.60	9.00
Max	17.80	35.00
Mean	2.606	21.93
Median	2.70	22.00
Standar Deviasi	5.56	6.77

Hubungan antara PPI dan TPP

Pada Gambar 1 menunjukkan bahwa pola hubungan pertumbuhan produksi industri dan tingkat pajak perusahaan tidak linier dan tidak membentuk sebuah pola tertentu. Sehingga, analisis regresi nonparametrik lebih tepat digunakan untuk menggambarkan pola hubungan pertumbuhan produksi industri dan tingkat pajak perusahaan.

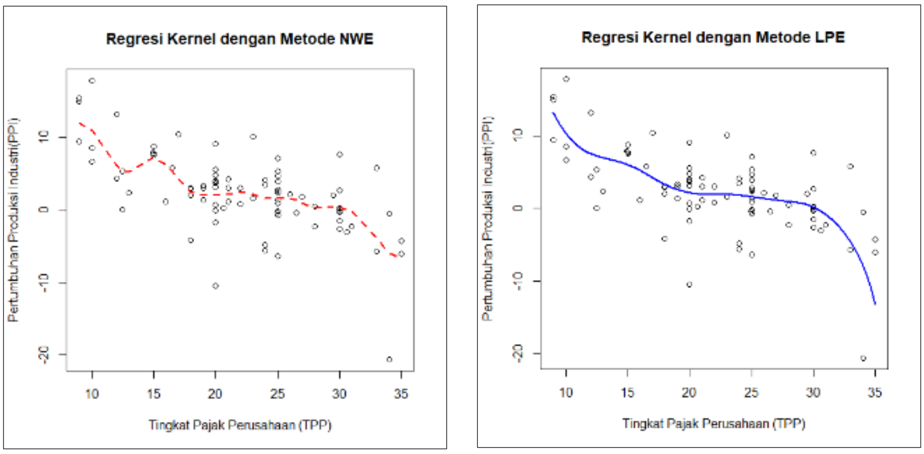


Gambar 1. Pola Hubungan PPI dan TPP

Jika dicari keeratan hubungan antara PPI dan TPP dengan menggunakan korelasi pearson maka didapat hasil sebesar -0.631 yang artinya bahwa terdapat hubungan negatif yang kuat antara PPI dan TPP. Jika PPI semakin rendah, maka TPP akan semakin tinggi, begitu pula sebaliknya.

Regresi Kernel dengan Metode NWE dan LPE

Estimasi kernel yaitu regresi kernel dengan penaksir Nadaraya-Watson (NWE) dan penaksir Lokal Polinomial (LPE). 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 digunaan sebagai bobot. Regresi kernel menggunakan metode NWE dan LPE diperoleh nilai bandwidth optimal sebesar 2.7756, sehingga kurva regresinya adalah seperti pada Gambar 2.



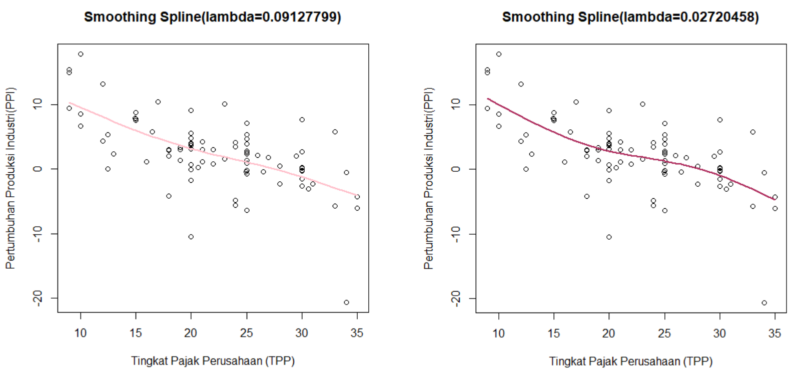
Gambar 2. Kurva Regresi Kernel dengan Metode NWE dan LPE

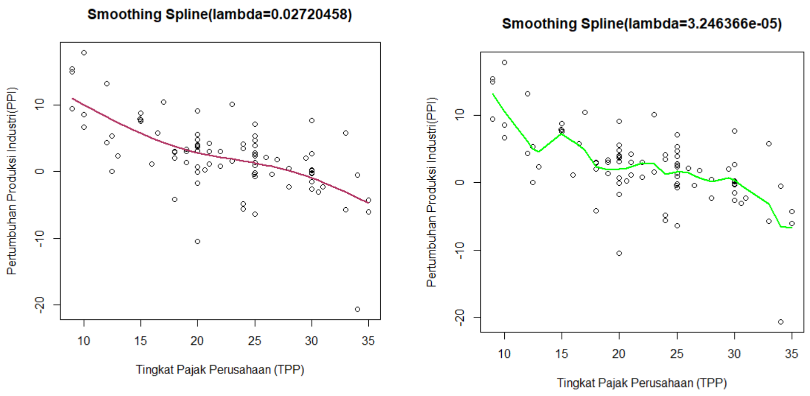
Berdasarkan Gambar 2 menunjukkan bahwa kurva regresi kernel dengan menggunakan *bandwidth* optimal diperoleh bahwa metode NWE lebih mengikuti sebaran data dibandingkan LPE. Namun, regresi kernel metode LPE memiliki kurva yang smooth (halus) dibandingkan metode NWE.

Regresi Smoothing Spline

Pendekatan regresi non parametrik selain kernel adalah pendekatan regresi non parametrik spline. Pemodelan regresi nonparametrik smoothing spline memiliki kurva yang sangat dipengaruhi oleh nilai dari parameter pemulus (λ) .

Beragam nilai parameter pemulus (λ) diujicobakan untuk regresi *smoothing spline*, Pada Gambar 3 di bawah ini, *smoothing spline* terbaik adalah model regresi *smoothing spline* dengan nilai parameter pemulus (λ) optimal adalah lambda (λ) = 0.02720458 karena memiliki nilai CGV (λ) yang paling kecil yaitu sebesar 18.68391, memiliki kurva yang smooth (halus) dan lebih dekat dengan sebaran data.





Gambar 3. Kurva Regresi Smoothing Spline Berdasarkan Nilai Parameter Pemulus

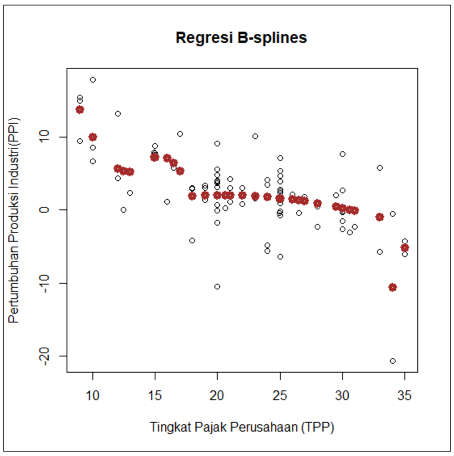
Regresi Smoothing B-Spline

Model *B-Spline* dengan titik *knot* digunakan untuk menyelesaikan kelemahan model spline pada saat orde yang tinggi, titik *knot* yang banyak atau *knot*s yang terlalu dekat yang akan membentuk matriks dalam perhitungan yang hampir singular sehingga persamaan normal sulit diselesaikan. Ada tiga kriteria yang harus diperhatikan dalam membentuk model regresi *B-Spline* yaitu menentukan orde untuk model, banyaknya *knot*, dan lokasi penempatan *knot*. 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). Nilai GCV yang minimum dari model regresi *B-Spline* dengan beberapa ordo model dan titik *knot* disajikan pada Tabel 2.

Tabel 2. Nilai GCV dan RSS dari Beberapa Titik Knot

Degree	Jumlah Knot	Knot Optimal	GCV	RSS
1	1	25.6324	18.2609	1445.9711
	2	32.6040 33.9573	17.8214	1377.3713
	3	12.8053 32.9997 33.9991	17.8902	1349.1705
	4	14.91555 15.00078 32.70776 33.47590	17.9821	1322.8222
	5	14.42783 17.99997 18.00062 32.98952 33.99950	17.6746	1267.9016
2	1	23.93341	18.6895	1444.4639
	2	32.99995 33.78484	18.0134	1358.4572
	3	13.29562 32.99784 33.99319	18.3409	1349.2144
	4	15.41626 16.50019 32.99501 33.99833	18.3566	1316.8191
	5	17.61025 18.00041 19.00011 32.99541 33.98879	18.3186	1281.0380

Tabel 2 menunjukkan bahwa pasangan nilai GCV dan RSS minimum diperoleh model degree 2 dengan jumlah *knot* sebanyak 5. 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 adalah pada Gambar 4.

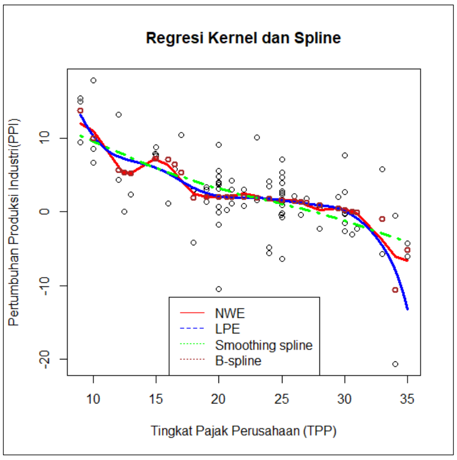


Gambar 4. Kurva Regresi B-splines dengan degree 2 dan jumlah knot 5

Fitting Curve Regresi Nonparametrik

Perbandingan kurva regresi nonparametrik antara pendekatan regresi kernel, NWE, LPE, *smoothing spline* dan *b-spline* yang telah diperoleh adalah pada Gambar 5.

Pada kurva di atas menunjukkan bahwa model regresi nonparametrik terbaik yang menggambarkan pola hubungan PPI dan TPP adalah model regresi B-splines dengan degree 2 dan jumlah knot 5. Hal ini dikarenakan kurva tersebut terlihat lebih mengikuti sebaran data dibanbandingkan kurva model regresi lainnya. Selanjutnya, hasil estimasi parameter model regresi B-splines disajikan pada Tabel 3.



Gambar 5. Kurva regresi kernel, NWE, LPE, smoothing spline dan b-spline

Tabel 3. Hasil Estimasi Parameter Regresi B-Spline

Parameter	Koefisien	SE	*t-value*	*p-value*	Simultan *p-value*	*R-square*
α_0	13.736	1.98	6.938	9.84E-10		
α_1	-11.846	4.472	-2.649	0.009746		
α_2	-4.164	2.853	-1.459	0.148413		
α_3	-11.805	2.362	-4.997	3.42E-06	1.384e-10	0.5229
α_4	-11.186	2.999	-3.73	0.000359		
α_5	-14.927	2.898	-5.15	1.87E-06		
α_6	-33.539	6.32	-5.307	9.94E-07		
α_7	-18.936	3.456	-5.479	4.94E-07		

Nilai *p-value* yang diperoleh pada Tabel 3 lebih kecil dari tingkat kesalahan (α = 0.05) dengan koefisien bertanda negative artinya kenaikan TPP akan menurunkan PPI. Hal ini senada dengan penelitian Utami dan Widanaputra (2017) yang menyatakan kenaikan pajak akan menurunkan struktur modal dan akan berimplikasi kepada produktivitas perusahaan. Rahayuningsih dan Saprudin (2019) mendapatkan hasil yang sama dimana terdapat hubungan negatif antara pajak dan laba perusahan. Nilai koefisien determinasi R^2 sebesar 52.29%, artinya besarnya variasi variabel PPI yang dijelaskan oleh variabel TPP dengan model regresi *B-Spline* sebesar 52.29%, sedangkan sisanya dijelaskan oleh variabel lainnya yang tidak dimasukkan ke dalam model. Sehingga persamaan regresinya dapat ditulis sebagai berikut:

$$\hat{y} = 13.736 - 11.846b_1 - 4.164b_2 - 11.805b_3 - 1.186b_4 - 14.927b_5 - 33.539b_6 - 18.936b_7$$

Kesimpulan

Perbandigan model regresi nonparametrik kernel dengan metode NWE dan LPE, smoothing spline dan B-splines dengan *fitting curve* menunjukkan bahwa model regresi *b-spline* dengan derajat 2 dan *knot* 5 merupakan model yang terbaik. Pertumbuhan Produksi Industri (PPI) dan Tingkat Pajak Perusahaan (TPP) memiliki hubungan yang negatif, dengan nilai koefisien determinasi sebesar 52.29 persen.

Saran untuk penelitian selanjutnya dapat menambahkan variabel yang potensial mempenagruhi PPI. Selain itu dapat membandingkan dengan berbagai metode regresi nonparametrik lainnya.

Daftar Pustaka

Arsyad, Lincoln. 2010. *Ekonomi Pembangunan, Edisi 5*. UPP STIM YKPN Yogyakarta. Yogyakarta.

Cantoni, E & T.Hastie. 2000. *Degrees of Freedom Tests for Smoothing Splines*. Statistics Department, Stanford University.

Carl de Boor. 2001. *A Practical Guide to Splines*. New York : Springer.

Eubank, R. 1999. *Nonparametric Regression and Spline Smoothing*. New York : Marcel Dekker.

Gan Y, Sun Z, Chen Z, Zhang X, Liu Y. 2018. Enhancement of the material point method using B-spline basis functions. *International Journal for Numerical Methods in Engineering.* 113(3): 411–431.

Hardle, W. 1994. *Applied Nonparametric Regression*. New York : Cambridge University Press.

Rahasia, Z. dkk. 2020. Pemodelan Data Time Series dengan Pendekatan Regresi Nonparametrik B-Spline. *Jurnal Matematika dan Pendidikan Matematika*. 11(1): 2579-7646.

Rahayuningsih, Tiyas; Saprudin, Saprudin. Pengaruh Beban Pajak Tangguhan Dan Laba Terhadap Beban Pajak Tahun Berjalan Pada Perusahaan Pertambangan. 2019. Journal of Information System, Applied, Management, Accounting and Research.3(4): 89-95

Sanusi, W., Syam, R., Adawiyah, R. 2019. Model Regresi Nonparametrik dengan Pendekatan Spline (Studi Kasus: Berat Badan Lahir Rendah di Rumah Sakit Ibu dan Anak Siti Fatimah Makassar). *Journal of Mathematics, Computations, and Statistics*. 2(1): 70-81.

Suparti, S., Santoso, R., Prahutama, A., Devi, A. R. 2018. *Regresi Nonparametrik*. Ponorogo: Wade Group.

Todaro, Michael and Smith,C Stephen. 2006. *Pembangunan Ekonomi Edisi Kesembilan Jilid 2.* PT Gelora Aksara Pratama. Jakarta.

Utami, LNS., Widanaputra, AAGP. 2017. Pengaruh Tarif Pajak, Profitabilitas, Likuiditas dan Ukuran Perusahaan Terhadap Struktur Modal Perusahaan Manufaktur Di BEI. E-Jurnal Akuntansi .20(1): 352-379

Wulandary, S., Purnama, D I. 2020. Perbandingan Regresi Nonparametrik Kernel NWE Dan B-Splines Pada Pemodelan Rata-Rata Lama Sekolah Dan Pengeluaran Perkapita Di Indonesia. *Journal Of Probability And Statistics*. 1(2): 89-97.

Yani, N. W. M. N. 2017. Aplikasi Model Regresi Semiparametrik Spline Truncated. *Jurnal Matematika*. 6(1): 65-73.