REGRESI NONPARAMETRIK *SPLINE TRUNCATED* TERHADAP TINGKAT PENGANGGURAN TERBUKA DI PULAU SUMATERA TAHUN 2014-2019

KARYA ILMIAH

OLEH

MARDIYAH MUHGNI
NIM 1603136583

PROGRAM STUDI S1 STATISTIKA
JURUSAN MATEMATIKA
FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM
UNIVERSITAS RIAU
2021

REGRESI NONPARAMETRIK *SPLINE TRUNCATED* TERHADAP TINGKAT PENGANGGURAN TERBUKA DI PULAU SUMATERA TAHUN 2014-2019

Mardiyah Muhgni, Harison

Program Studi S1 Statistika
Jurusan Matematika
Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Riau
Kampus Bina Widya, Pekanbaru 28293

mardiyah.muhgni6580@student.unri.ac.id

ABSTRACT

The problem of unemployment is an important concern for local governments. The average of an open unemployment rate in Sumatra Island in 2014-2019 was 5.4% which was still above the national target of 4.8%. The factors that are thought to have an effect are population density, economic growth, labor force participation rate, and regional minimum wage. In this study, panel data from 2014-2019 were used in 10 provinces on the island of Sumatra. The scatter diagram of each variable does not show a certain pattern so that spline truncated nonparametric regression was used. This regression is also carried out by looking at the optimum knot point of the generalized cross validation value. The study shows that the variables that have a significant effect are population density, labor force participation rate, and regional minimum wage. The variable that does not have significant effect is economic growth.

Keywords: *Open unemployment rate, multivariable linear regression, nonparametric, spline truncated, generalized cross validation*

ABSTRAK

Permasalahan pengangguran menjadi perhatian penting bagi pemerintah daerah. Rata-rata tingkat pengangguran terbuka di Pulau Sumatera pada tahun 2014-2019 5,4% masih berada di atas target nasional sebesar 4,8%. Faktor-faktor yang diduga berpengaruh yaitu kepadatan penduduk, pertumbuhan ekonomi, tingkat partisipasi angkatan kerja, dan upah minimum regional. Dalam penelitian ini digunakan data panel dari tahun 2014-2019 di 10 provinsi di Pulau Sumatera. Diagram pencar dari setiap variabel tidak menunjukan pola tertentu sehingga digunakan regresi nonparametrik *spline truncated*. Regresi ini juga dilakukan dengan melihat titik knot optimum dari nilai *generalized cross validation*. Penelitian ini menunjukkan bahwa variabel yang berpengaruh secara signifikan adalah kepadatan penduduk, tingkat partisipasi angkatan kerja, dan upah minimum regional. Variabel yang tidak berpengaruh signifikan adalah pertumbuhan ekonomi.

Kata kunci: Tingkat pengangguran terbuka, regresi linear multivariabel, nonparametrik, *spline truncated, generalized cross validation*

1. PENDAHULUAN

Banyaknya jumlah angkatan kerja yang tidak bekerja serta aktif mencari pekerjaan disebut tingkat pengangguran. Penduduk yang sudah masuk dalam angkatan kerja namun tidak mempunyai pekerjaan, lagi mempersiapkan usaha, lagi mencari pekerjaan, dan telah mempunyai pekerjaan namun belum mulai bekerja disebut pengangguran terbuka. Tingkat pengangguran terbuka merupakan persentase jumlah pengangguran terhadap jumlah angkatan kerja (BPS, 2019).

Pengangguran merupakan permasalahan yang terjadi dalam suatu perekonomian di suatu daerah, sehingga permasalahan ini harus segera ditangani oleh pemerintah. Tingkat pengangguran yang tinggi akan menimbulkan berbagai permasalahan sosial, dengan adanya pengangguran maka pendapatan masyarakat akan berkurang sehingga menimbulkan kemiskinan. Ketiadaan pendapatan menyebabkan para penganggur akan mengurangi pengeluaran konsumsinya sehingga dapat mengganggu taraf kesehatan keluarga (Sukirno, 2016). Ada beberapa faktor yang diduga dapat mempengaruhi Tingkat Pengangguran Terbuka (TPT) seperti Kepadatan Penduduk (KP), Pertumbuhan Ekonomi (PE), Tingkat Partisipasi Angkatan Kerja (TPAK), dan Upah Minimum Regional (UMR).

Ditinjau dari *trend* TPT di Pulau Sumatera, terdapat pola hubungan yang berubah pada tiap sub interval tertentu, sehingga penelitian ini ingin mengetahui pengaruh PE, KP, TPAK, dan UMR terhadap TPT dengan menggunakan regresi nonparametrik *spline*. Regresi *spline* merupakan pendekatan nonparametrik yang dapat memodelkan hubungan antara variabel dependen dan independen, dimana dengan pendekatan ini akan membentuk fungsi polinomial berorde m yang memiliki sifat terpotong-potong (Yozza & Afrimayani, 2019).

Penelitian sebelumnya mengenai regresi nonparametrik *spline* pernah dilakukan oleh Nurcahayani et al (2019), Utama et al (2015), Muhgni et al (2020), Hidayat et al (2015) dan banyak peneliti lainnya.

2. REGRESI LINEAR MULTIVARIABEL, REGRESI NONPARAMETRIK, DAN REGRESI NONPARAMETRIK *SPLINE TRUNCATED*

Regresi linear multivariabel merupakan regresi yang melibatkan lebih dari satu variabel bebas (Montgomery et al. 2012). Misalkan diberikan n observasi $(x_{11}, x_{12}, \dots, x_{k1}, y_1)$, $(x_{21}, x_{22}, \dots, x_{k2}, y_2), \dots$, $(x_{1n}, x_{2n}, \dots, x_{kn}, y_n)$maka model regresi linear multivariabel dapat ditulis pada persamaan (1) berikut:

$$y_i = \beta_0 + \sum_{j=1}^{k} \beta_j x_{ji} + \varepsilon_i, \quad i = 1,2, \dots, n \tag{1}$$

dengan y_i menyatakan variabel dependen pengamatan ke-i, β_0 intersep, β_j koefisien regresi pada variabel x_j, x_{ji} variabel independen ke-j untuk pengamatan ke-i, ε_i residual dengan asumsi $\varepsilon_i \sim NID(0, \sigma^2)$.

Regresi nonparametrik merupakan metode statistik yang digunakan untuk menentukan hubungan antara variabel dependen dan variabel independen di mana bentuk fungsi tidak diketahui dengan kurva regresi yang hanya diasumsikan halus

dan terkandung dalam ruang fungsi tertentu (Budiantara, 2005). Misalkan x adalah variabel independen dan y adalah variabel dependens untuk n pengamatan, maka model regresi nonparametrik secara umum dirumuskan pada persamaan (2) berikut:

$$y_i = f(x_i) + \varepsilon_i \,, \tag{2}$$

y_i menyatakan variabel dependen, $f(x_i)$ fungsi regresi nonparametrik dan ε_i kesalahan yang terdistribusi normal, independen, dengan *mean* 0 dan *varians* σ^2 dari $i = 1,2, \dots, n$ (Eubank, 1999).

Regresi *spline* merupakan salah satu jenis polinomial yang memuat titik-titik knot sehingga memungkinkan untuk beradaptasi lebih efektif dengan karakteristik lokal dari suatu fungsi. *Spline* memiliki kemampuan untuk memperkirakan perilaku data yang cenderung berbeda pada interval yang berbeda serta memiliki fleksibilitas yang tinggi. Kemampuan untuk melihat data ini dapat dilakukan dengan fungsi yang telah melekat di estimator dan bagian yang disebut titik knot pada regresi nonparametrik. Titik knot dapat menunjukkan perubahan fungsi pola perilaku pada interval yang berbeda (Syisliawati et al. 2016). Fungsi *spline* derajat m secara umum dapat disajikan pada persamaan (3) berikut:

$$f(x_i) = \sum_{k=0}^{m} \beta_{ik} x_i^k + \sum_{u=1}^{r} \beta_{i(m+u)} \, (x_i - K_u)_+^m, \tag{3}$$

dengan β_i menyatakan koefisien, x_i variabel independen ke-i, m orde spline, k_u titik knot, dan

$$(x_i - K_u)_+^m = \begin{cases} (x_i - K_u)^m, x \geq K_u, \\ 0 \qquad\quad , x < K_u. \end{cases}$$

Asumsi kesalahan ε_i berdistribusi normal independen dengan rata-rata 0 dan variansi σ^2, dengan y_i dalam model regresi juga berdistribusi normal dengan rata-rata $\bar{y}$ dan variansi σ^2 (Syisliawati et al. 2016). Dengan menggunakan metode kuadrat terkecil maka bentuk ε_i pada persamaan (2) menjadi:

$$\sum_{j=1}^{n} {\varepsilon_i}^2 = \sum_{i=1}^{n} (y_j - f(x_i))^2.$$

Nilai R^2 dan $\bar{R}^2$ merupakan ukuran yang sederhana dan sering digunakan untuk mengukur kesesuaian uji pada sebuah garis regresi (Gujarati, 2004). Nilai R^2 memberikan gambaran tentang kesesuaian variabel independen dalam memprediksi variabel dependen. Adapun perhitungan nilai R^2 sebagai berikut:

$$R^2 = \frac{\sum (y_i - \hat{y}_i)^2}{\sum (y_i - \bar{y})^2}, \tag{4}$$

dengan y_i menyatakan observasi ke-i, $\hat{y}_i$ adalah prediksi respon ke-i, $\bar{y}$ rata-rata data pengamatan, dimana sifat R^2 merupakan besaran yang non-negatif dan batasnya $0 \leq R^2 \leq 1$. Untuk melengkapi kelemahan pada R^2 dapat menggunakan $\bar{R}^2$ yakni dengan mempertimbangkan jumlah sampel data dan jumlah variabel yang digunakan. Adapun perhitungan nilai $\bar{R}^2$ sebagai berikut:

$$\bar{R}^2 = 1 - (1 - R^2) \times \frac{n-1}{n-k}, \tag{5}$$

dengan n menyatakan banyaknya data observasi, dan k banyaknya variabel independen.

Salah satu langkah penting dalam pemodelan regresi terpotong nonparametrik *spline* adalah memilih titik knot optimal. Titik knot yang besar berarti penduga fungsi yang diperoleh akan lebih halus tetapi kemampuan pemetaan datanya tidak cukup baik. Titik knot yang terlalu kecil berarti estimator fungsi yang diperoleh akan lebih kasar (*under smoothing*) sehingga diperlukan titik yang optimal. Pemilihan parameter *smoothing* dilakukan untuk menentukan model estimator kurva regresi terbaik. Salah satu metode yang digunakan untuk memilih titik knot optimal adalah metode *Generalized Cross Validation* (*GCV*) (Wu & Zhang, 2006). Rumus metode *GCV* dapat dituliskan sebagai berikut:

$$GCV_i(k) = \frac{n^{-1}\sum_{j=1}^{n}[y_i - \hat{y}_i]^2}{\{1 - \frac{tr(\boldsymbol{A_k})}{\boldsymbol{n}}\}^2}, \tag{6}$$

dengan $GCV_i(K)$ menyatakan nilai *GCV* pada subjek ke-i pada knot ke-k, n_i banyaknya pengamatan pada subjek ke-i, y_{ij} data aktual subjek ke-i, $\hat{y}_i$ hasil dari estimasi subjek ke-i, $tr(\boldsymbol{A}_k)$ menyatakan $tr[\mathbf{X}(\mathbf{X}'\mathbf{X})^{-\mathbf{1}}\mathbf{X}']$. Pemilihan titik knot yang optimal dapat dilakukan dengan melihat nilai $GCV_i(k)$ yang paling minimum.

Pengujian signifikansi parameter model regresi *spline* dapat dilakukan dengan menggunakan uji serentak. Pengujian ini dilakukan secara serentak dengan parameter yang ada dalam model (Draper & Smith, 1998). Hipotesis untuk uji serentak sebagai berikut:

$H_0: \beta_1 = \beta_2 = \cdots = \beta_p = 0,$

$H_1: \beta_k \neq 0, \;\; k = 1,2, \ldots, p.$

Statistik uji yang digunakan yaitu:

$$F_{hitung} = \frac{MS_{Reg}}{s^2}. \tag{7}$$

Keputusan: Tolak H_0 jika F_{hitung} lebih besar dari F_{tabel} $(F_{\alpha;(k-1,n-k)})$.

Parameter yang signifikan secara individu terhadap model maka dapat dilakukan dengan menggunakan uji parsial (Draper & Smith, 1998). Hipotesis untuk uji parsial adalah sebagai berikut:

$H_0: \beta_k = 0,$

$H_1: \beta_k \neq 0, \;\; k = 1,2, \ldots, p.$

Statistik uji yang digunakan:

$$t_{hitung} = \frac{\widehat{\beta}_k}{st.error(\widehat{\beta}_k)}. \tag{8}$$

Keputusan: Tolak H_0 jika $|t_{hitung}|$ lebih besar dari t_{tabel} $(t_{\alpha/2;,n-k})$ atau $p - value < \alpha$.

Pengujian normalitas dapat menggunakan uji Shapiro-Wilk. Uji ini dikembangkan oleh Samuel Shapiro dan Martin Wilk pada tahun 1965. Uji ini tergantung pada korelasi antara data yang diberikan dan kecocokan angka normalnya dengan statistik uji yang dirumuskan pada persamaan (9) berikut (Thode, 2002):

$$W = \frac{\sum_{i=1}^{\frac{n}{2}} a_{n-i+1}\left(x_{(n-i+1)} - x_{(i)}\right)}{(n-1)\left(\frac{\sum_{i=1}^{n}(x_i - \bar{x})}{n-1}\right)}, \tag{9}$$

dengan $x_{(i)}$ merupakan nilai sampel terbesar ke$-i$.

Homoskedastisitas menyatakan bahwa nilai variansi faktor kesalahan adalah sama dan dapat juga dijelaskan bahwa populasi Y berkaitan dengan berbagai nilai X yang memiliki nilai variansi yang sama. Uji Glejser merupakan salah satu metode untuk menguji penyebaran residual. Pengujian dilakukan dengan cara meregresikan antara nilai absolut error terhadap semua variabel independen. Asumsi homoskedastisitas terpenuhi jika variabel independen tidak signifikan mempengaruhi nilai *absolut error*, dengan asumsi seperti pada persamaan (10) berikut (Gujarati, 2004):

$$var(\varepsilon_i) = {\sigma_i}^2. \tag{10}$$

Multikolinearitas merupakan adanya hubungan linear di antara variabel-variabel independen dalam model regresi (Gujarati, 2004). Kecepatan dari meningkatnya varians dan kovarians dapat dilihat dengan *Variance Inflating Factor* (VIF), yang didefinisikan pada persamaan (11) berikut:

$$VIF = \frac{1}{(1 - r_i^2)}. \tag{11}$$

Nilai VIF ini menunjukkan variansi dari sebuah estimator ditingkatkan oleh adanya multikolinearitas, r_i^2 merupakan koefisien determinasi dari regresi variabel bebas ke-i. Suatu aturan baku menyatakan bahwa jika *VIF* suatu variabel lebih dari 10, maka variabel tersebut dikatakan sangat kolinear. Dalam model regresi dihindari terjadinya multikolinearitas dalam variabel-variabel independen (Gujarati, 2004).

Uji Durbin-Watson merupakan salah satu cara yang dapat digunakan untuk mendeteksi autokorelasi yang dikembangkan oleh ahli statistik Durbin dan Watson, yang didefinisikan sebagai berikut (Gujarati, 2004):

$$\mathrm{dw} = \frac{\sum_{t=2}^{t=n}(\hat{u}_t - \hat{u}_{t-1})^2}{\sum_{t=1}^{t=n}\hat{u}_t^2}, \tag{12}$$

jika H_0 adalah dua sisi yang tidak ada autokorelasi baik positif ataupun negatif, maka kriteria penolakan H_0 untuk nilai dw pada persamaan (13) sebagai berikut:

$0 < \mathrm{dw} < \mathrm{dl}$ (H_0 ditolak),
$\mathrm{du} < \mathrm{dw} < 4 - \mathrm{du}$ (H_0 tidak ditolak),
$\mathrm{dl} < \mathrm{dw} < \mathrm{du}$ atau $4 - \mathrm{dl} \leq \mathrm{dw} \leq 4 - \mathrm{du}$ (pengujian tidak meyakinkan).

3. METODOLOGI PENELITIAN

Artikel ini dilakukan dengan menggunakan data TPT, KP, PE, TPAK, dan UMR di Pulau Sumatera pada tahun 2014-2019 yang diperoleh dari BPS Provinsi Riau. Tahapan-tahapan dalam artikel ini adalah sebagai berikut:

1. Pengumpulan data yang akan digunakan untuk penelitian ini.
2. Membuat deskripsi TPT di pulau Sumatera dan faktor yang mempengaruhinya.
3. Membuat Visualisasi data dengan membuat diagram pencar variabel dependen terhadap variabel independen.
4. Mengolah data dengan menggunakan metode regresi linear multivariabel dan metode regresi nonparametrik *spline*.
5. Memilih titik knot optimal berdasarkan nilai *GCV* yang paling minimum.
6. Menguji asumsi model.
7. Membandingkan nilai *R-Square* dan *Adjusted R-Square* antara regresi linear multivariable dengan regresi nonparametrik *spline*.
8. Menarik kesimpulan dan saran.

4. TINGKAT PENGANGGURAN TERBUKA BERDASARKAN REGRESI NONPARAMETRIK *SPLINE TRUNCATED*

Pemilihan titik knot optimum yang dilakukan pada variabel-variabel yang diduga berpengaruh terhadap TPT dengan menggunakan *GCV*. Titik knot optimum menggunakan satu titik knot, dua titik knot, tiga titik knot, dan kombinasi titik knot, dan yang akan dipilih titik knot dengan nilai *GCV* minimum. Dari persamaan (6) didapatkan hasil dengan menggunakan *software* R yang disajikan pada Tabel 1.

Tabel 1. Pemilihan titik knot optimum

Jumlah Titik Knot	*GCV*	Titik Knot				Jumlah Parameter
		x_1	x_2	x_3	x_4	
1	0,772	169,041	3,398	67,978	2,177	9
2	0,772	67,000	-0,730	63,060	1,350	13
		169,041	3,398	67,978	2,177	
3	0,722	75,163	-0,400	63,453	1,416	17
		213,939	5,214	70,142	2,540	
		230,265	5,874	70,929	2,672	
kombinasi (3,3,3,1)	0,648	75,163	-0,4	63,453	2,177	15
		213,939	5,214	70,142		
		230,265	5,874	70,929		

Berdasarkan Tabel 1 di atas nilai *GCV* paling minimum adalah 0,648 yang berada pada kombinasi titik knot dengan jumlah titik knot sebanyak tiga, tiga, tiga, dan satu dengan jumlah parameternya adalah 15 parameter termasuk $\hat{\beta}_0$.

Selanjutnya adalah melakukan pengujian parameter untuk mengetahui variabel yang berpengaruh secara signifikan terhadap TPT di Pulau Sumatera pada Tahun 2014-2019. Pengujian serentak dilakukan secara serentak untuk menguji estimasi parameter model dengan $\alpha = 0{,}05$. Berikut hasil dari ANOVA untuk model regresi nonparametrik *spline* berdasarkan pada persamaan (7) dengan menggunakan *software* R yang disajikan dalam Tabel 2. berikut:

Tabel 2. ANOVA model regresi *spline*

Sumber Variasi	df	*Sum of Square*	*Mean Square*	F_{hitung}	$p - value$
Regresi	14	110,248	7,875	15,498	9×10^{-13}
Error	45	22,865	0,508		
Total	59	133,113			

Berdasarkan Tabel 2 dapat dilihat bahwa bahwa $F_{hitung} > F_{tabel}$ maka tolak H_0. Ini berarti bahwa terdapat minimal satu parameter yang berpengaruh signifikan terhadap TPT di Pulau Sumatera pada tahun 2014-2019.

Pengujian parameter individu dapat dilakukan dengan menggunakan uji parsial pada persamaan (8) hasilnya diperoleh dengan menggunakan *software* R seperti pada Tabel 3.

Tabel 3. Pengujian parameter individu

Variabel	Parameter	Estimator	T_{hitung}	$p - value$	Kriteria
Konstan	$\hat{\beta}_0$	-0,040	-0,784	0,437	Tidak Signifikan
x_1	$\hat{\beta}_{11}$	0,213	3,211	0,002	Signifikan
	$\hat{\beta}_{12}$	-0,200	-2,952	0,005	Signifikan
	$\hat{\beta}_{13}$	-0,115	-3,524	0,001	Signifikan
	$\hat{\beta}_{14}$	0,147	3,102	0,003	Signifikan
x_2	$\hat{\beta}_{21}$	-2,599	-0,994	0,326	Tidak Signifikan
	$\hat{\beta}_{22}$	2,154	0,804	0,425	Tidak Signifikan
	$\hat{\beta}_{23}$	1,309	1,873	0,068	Tidak Signifikan
	$\hat{\beta}_{24}$	0,187	0,164	0,870	Tidak Signifikan
x_3	$\hat{\beta}_{31}$	-0,144	-1,774	0,083	Tidak Signifikan
	$\hat{\beta}_{32}$	-0,317	-3,372	0,002	Signifikan
	$\hat{\beta}_{33}$	2,557	2,551	0,014	Signifikan
	$\hat{\beta}_{34}$	-3,250	-2,085	0,043	Signifikan
x_4	$\hat{\beta}_{41}$	0,571	1,164	0,251	Tidak Signifikan
	$\hat{\beta}_{42}$	-2,589	-2,808	0,007	Signifikan

Tabel 3 menunjukkan bahwa ada beberapa variabel yang tidak signifikan dikarenakan $p - value < \alpha$. Variabel yang sama sekali tidak signifikan terhadap TPT di Pulau Sumatera pada tahun 2014-2019 adalah variabel PE.

Selanjutnya dilakukan pengujian asumsi residual normalitas dengan menggunakan uji $\text{Shapiro} - \text{Wilk}$ seperti pada persamaan (9). Pengujian ini bertujuan untuk mengetahui apakah model regresi telah memenuhi asumsi normal. Regresi yang baik harus memiliki residual berdistribusi normal. Berikut disajikan hasil pengujian normalitas dalam Tabel 4.

Tabel 4. Pengujian asumsi normalitas

Model Regresi	W	$p - value$
Regresi Linear Berganda	0,972	0,188
Regresi *Spline* Berganda	0,982	0,517

Hipotesis:
H_0 menyatakan data mengikuti distribusi normal,
H_1 menyatakan data tidak mengikuti distribusi normal.
daerah keputusan: Tolak H_0 jika $p - value < \alpha_{(0,05)}$.
Berdasarkan Tabel 4 dapat disimpulkan bahwa dari kedua model didapatlah nilai $p - value > \alpha$ ini berarti H_0 diterima yang menyatakan bahwa kedua model mengikuti distribusi normal.

Pengujian asumsi homoskedastisitas dilakukan dengan menggunakan uji Glejser seperti pada persamaan (10). Pengujian ini dengan meregresikan absolut residual terhadap variabel independen, hasilnya disajikan pada Tabel 5.

Tabel 5. Pengujian asumsi homoskedastisitas

Sumber Variasi	df	*Sum of Square*	*Mean Square*	F_{hitung}	$p - value$
Regresi	14	2,381	0,170	1,204	0,306
Error	45	6,356	0,121		
Total	59	8,737			

Hipotesis :
H_0 menyatakan data memiliki nilai varian yang sama,
H_1 menyatakan data tidak memiliki nilai varian yang sama,
daerah keputusan: Terima H_0 jika $F_{hitung} < F_{tabel}$.
Berdasarkan Tabel 5 dapat disimpulkan bahwa nilai F_{hitung} sebesar 1,204, sedangkan nilai $F_{tabel} = F_{0,05(14,45)}$ sebesar 1,92, karena $F_{hitung} < F_{tabel}$ berarti bahwa data sudah memiliki nilai varian yang sama maka asumsi ini sudah terpenuhi.

Model regresi dapat dikatakan bebas dari multikolinearitas jika memiliki nilai $VIF < 10$. Berikut disajikan hasil dari VIF dari setiap variabel pada Tabel 6.

Tabel 6. Pengujian non-multikolinearitas

Variabel	Regresi Linear Berganda	Regresi *Spline* Berganda
	VIF	VIF
KP	1,104	1,291
PE	1,474	1,323
TPAK	1,333	1,172
UMR	1,036	1,221

Berdasarkan hasil dari Tabel 6 dapat disimpulkan bahwa nilai VIF dari semua variabel kecil dari 10. Ini berarti bahwa model regresi bebas dari multikolinearitas.

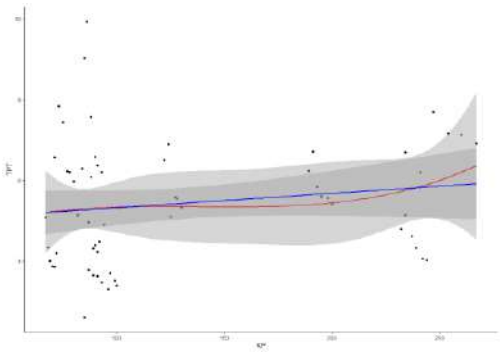
Pengujian non-autokorelasi dengan menggunakan uji $\text{Durbin} - \text{Watson}$ seperti pada persamaan (12). Tabel $\text{Durbin} - \text{Watson}$ dengan $n = 60$, dan $k = 4$ maka diperoleh nilai $\text{dl} = 1{,}444$ dan $\text{du} = 1{,}727$, hasil dari nilai $\text{Durbin} - \text{Watson}$ disajikan pada Tabel 7.

Tabel 7. Pengujian non-autokorelasi dengan statistik $\text{Durbin} - \text{Watson}$

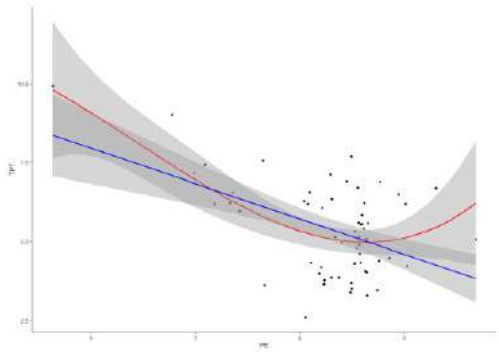
Model Regresi	dw	Keterangan	kesimpulan
Regresi Linear Berganda	1,719	$\text{du} < \text{dw} < 4 - \text{du}$	Bebas autokorelasi
Regresi *Spline* Berganda	2,266	$\text{du} < \text{dw} < 4 - \text{du}$	Bebas autokorelasi

Berdasarkan Tabel 7 dapat disimpulkan bahwa model regresi bebas dari autokorelasi karena nilai $\text{du} < \text{dw} < 4 - \text{du}$.

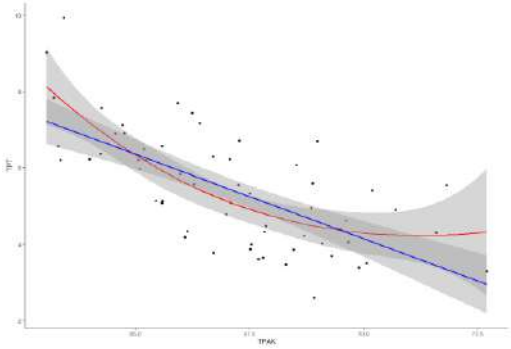
Berikut akan disajikan *plot* hasil dari regresi linear berganda dan regresi nonparametrik *spline* pada Gambar 1.



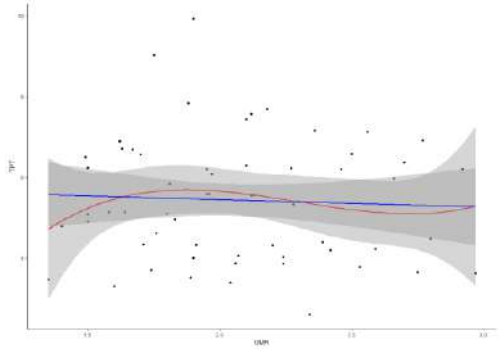
(a). *Plot* data variabel KP terhadap TPT tahun 2014-2019



(b). *Plot* data variabel PE terhadap TPT tahun 2014-2019



(c). *Plot* data variabel TPAK terhadap TPT tahun 2014-2019



(d). *Plot* data variabel UMR terhadap TPT tahun 2014-2019

Gambar 1. Perbandingan regresi linear berganda dan regresi nonparametrik *spline*

Keterangan:

Merupakan garis regresi linier berganda,

Merupakan garis regresi nonparametrik *spline*.

Berdasarkan Gambar 1 di atas dapat dilihat bahwa pada gambar (a) dan (b) persebaran data tidak terlalu menyebar, sedangkan pada gambar (c) dan (d) terlihat jelas bahwa data menyebar. Dari ke empat gambar tersebut dapat dilihat pula bahwa dengan menggunakan regresi linear berganda hanya akan menghasilkan garis linier dan banyak terdapat *outlier*, sedangkan dengan menggunakan regresi nonparametrik *spline* garis regresi lebih mendekati data aktual. Berdasarkan persamaan (4) dan (5) disajikan hasil dari nilai R^2 dan $\bar{R}^2$ pada Tabel 8.

Tabel 8. Nilai R^2 dan $\bar{R}^2$

Model Regresi	R^2	$\bar{R}^2$
Regresi Linear Berganda	0,621	0,594
Regresi Nonparametrik *Spline*	0,828	0,816

Berdasarkan Tabel 8 regresi nonparametrik *spline* memiliki nilai R^2 dan $\bar{R}^2$ yang lebih tinggi dibandingkan dengan regresi linear berganda. Hal ini berarti bahwa perhitungan TPT lebih sesuai dengan menggunakan regresi nonparametrik *spline*.

Model regresi nonparametrik *spline* terbaik berdasarkan pada persamaan (3) menggunakan kombinasi knot (3,3,3,1) adalah sebagai berikut:

$$\begin{aligned}\hat{y} = {} & -0{,}040 + 0{,}213x_1 - 0{,}200(x_1 - 75{,}163)_+^1 - 0{,}115(x_1 - 213{,}939)_+^1 + \\ & 0{,}147(x_1 - 230{,}265)_+^1 - 2{,}599x_2 + 2{,}154(x_2 + 0{,}400)_+^1 + 1{,}309(x_2 - \\ & 5{,}214)_+^1 + 0{,}187(x_2 - 5{,}874)_+^1 - 0{,}144x_3 - 0{,}317(x_3 - 63{,}453)_+^1 + \\ & 2{,}557(x_3 - 70{,}142)_+^1 - 3{,}250(x_3 - 70{,}929)_+^1 + 0{,}571x_4 - 2{,}589(x_4 - \\ & 2{,}177)_+^1.\end{aligned}$$

Berdasarkan model di atas dapat dijelaskan bahwa untuk nilai x_1 akan berpengaruh positif ketika nilai berada di bawah $213{,}939$ dan di atas $230{,}265$, dan berpengaruh negatif ketika berada di atas $213{,}939$ dan di bawah $230{,}265$. Nilai x_2 akan berpengaruh positif ketika nilai berada di atas $5{,}214$, dan berpengaruh negatif ketika berada di bawah $5{,}214$. Nilai x_3 akan berpengaruh positif ketika nilai berada di atas $70{,}142$ dan di bawah $70{,}929$, dan berpengaruh negatif ketika nilai berada di bawah $63{,}453$ dan di atas $70{,}929$. Nilai x_4 akan berpengaruh positif ketika nilai berada di bawah $2{,}177$, dan berpengaruh negatif ketika berada di atas $2{,}177$.

5. KESIMPULAN

Berdasarkan hasil penelitian yang telah dilakukan didapat kesimpulan bahwa dari empat variabel independen yang telah diuji terdapat tiga variabel yang signifikan yaitu variabel kepadatan penduduk, tingkat partisipasi angkatan kerja, dan upah minimum regional. Variabel yang tidak signifikan terhadap tingkat pengangguran terbuka di Pulau Sumatera yaitu variabel pertumbuhan ekonomi. Pemodelan regresi nonparametrik *spline truncated* terbaik yaitu didapatkan dengan menggunakan kombinasi titik knot (3,3,3,1).

DAFTAR PUSTAKA

BPS. (2019). *Keadaan angkatan kerja di Indonesia*. Jakarta: BPS-Statistics Indonesia.

Budiantara. (2005). Model keluarga spline polinomial truncated dalam regresi semiparametrik. *Berkala Ilmiah Matematika dan Ilmu Pengetahuan Alam*, ***15*** (3), 55–61.

Draper, N. R., & Smith, H. (1998). *Applied regression analysis* (3rd ed.). New York: John Wiley & Sons.

Eubank, R. L. (1999). *Nonparametric regression and spline smoothing* (2nd ed.). New York: Marcel Dekker.

Gujarati. (2004). *Basic econometrics* (4nd ed.). New York: McGraw-Hill/Irvin.

Hidayat, R., Yuliani, & Sam, M. (2015). Model regresi nonparametrik dengan pendekatan spline truncated. *Prosiding Seminar Nasional*, ***3*** *(1)*, 203–210.

Montgomery, D. C., Peck, E. A., & Vining, G. G. (2012). *Introduction to linear regression analysis* (5th ed.). New Jersey: John Wiley & Sons.

Muhgni, M., Fadly, F., Adnan, A., & Harison. (2020). Pemodelan tingkat pengangguran terbuka di Pulau Sumatera dengan menggunakan regresi nonparametrik spline truncated. *Jurnal Sains Matematika dan Statistika*, ***6*** (1), 133–144.

Nurcahayani, H., Budiantara, I. N., & Zain, I. (2019). Nonparametric truncated spline regression on modelling mean years schooling of regencies in Java. *AIP Conference Proceedings*, ***2194*** (1), 1–8.

Sukirno, S. (2016). *Makroekonomi teori pengantar* (3nd ed.). Jakarta: PT Raja Grafindo Persada.

Syisliawati, Wibowo, W., & Budiantara, I. N. (2016). Regression spline truncated curve in nonparametric regression. *International Conference on Research, Implementation and Education of Mathematics and Science*, ***1*** (1), 115–122.

Thode, H. C. (2002). *Testing for normality*. Boca Raton FL: CRC press.

Utama, S. S., Suparti, & Rahmawati, R. (2015). Pemodelan tingkat pengangguran terbuka di Jawa Tengah menggunakan regresi spline. *Jurnal Gaussian*, ***4*** (1).

Wu, H., & Zhang, J. T. (2006). *Nonparametric regression methods for longitudinal data analysis: mixed-effects modeling approaches*. Hoboken, New Jersey: John Wiley & Sons.

Yozza, H., & Afrimayani. (2019). *Korelasi model hubungan dua variabel dengan analisis regresi spline*. ***8*** (1), 67–74.