

MODEL HUBUNGAN JUMLAH PENGANGGURAN DAN INDEKS KEDALAMAN KEMISKINAN DI PULAU SUMATERA TAHUN 2019 MENGGUNAKAN REGRESI NONPARAMETRIK *SPLINES*

Aida Meimela¹

1. Departemen Statistik Fakultas Matematika Universitas Padjadjaran
Email: aida.mey@bps.go.id

Abstract

Poverty does not only focus on decreasing the number of poor people. There is an important thing that must also be considered, namely the Poverty Gap Index (P1). From year to year, the poverty gap index for all regencies / cities in Sumatra tends to stagnate. While the island of Sumatra is the second island with the largest population in Indonesia. This should be a serious concern for the government. One of the factors that influence the poverty gap index is unemployment. The more people who are unemployed, can increase the poverty gap index. Therefore we need to model the relationship between the number of unemployment and poverty gap index. The approach used is nonparametric regression modeling where the residual value is not normally distributed. The model is smoothing splines regression and quantile splines regression (median, $\tau = 0, 5$). Meanwhile, to see the best model performance by looking at the RMSE values of both models. From the results of the study it was found that the quantile regression smoothing splines model was better because the RMSE value was lower than the regression smoothing splines.

Keywords: poverty gap, unemployment, quantile regression

JEL Classification: I32, J64, C21

1. PENDAHULUAN

Kemiskinan tidak hanya fokus untuk mengurangi jumlah penduduk miskin tetapi ada faktor lain yang juga perlu diperhatikan. Salah satunya adalah tingkat Indeks Kedalaman Kemiskinan (P1). Semakin besar indeks kedalaman kemiskinan maka semakin jauh pula rata-rata pengeluaran penduduk dari garis kemiskinan.

Salah satu pulau di Indonesia dengan jumlah penduduk terbanyak kedua adalah pulau Sumatera. Hampir sepertiga jumlah provinsi di Indonesia berada di Pulau Sumatera. Berdasarkan proyeksi penduduk tahun 2020, jumlah penduduk di Pulau Sumatera adalah 21,89 persen dari jumlah penduduk Indonesia. Sementara itu, pada tahun 2019 jumlah penduduk miskin di Pulau Sumatera sebesar 10,54 persen dari total jumlah penduduk miskin di Indonesia. Angka ini bisa dikatakan cukup besar.

Oleh karenanya pemerintah perlu melakukan upaya untuk mengatasi kemiskinan di Pulau Sumatera. Tidak hanya terpaku pada kemiskinan di Pulau Jawa dan daerah terpencil. Hal ini karena indeks kedalaman kemiskinan kabupaten/kota yang ada di Pulau Sumatera cenderung stagnan dari tahun ke tahun.

2. TINJAUAN PUSTAKA DAN HIPOTESIS

Beberapa penelitian dilakukan untuk melihat faktor-faktor apa saja yang mempengaruhi indeks kedalaman kemiskinan. Ahmaddien (2019) melakukan penelitian terkait faktor determinan keparahan dan kedalaman kemiskinan Jawa Barat dengan regresi data panel. Hasil penelitian menemukan pertumbuhan ekonomi, Tingkat Partisipasi Angkatan Kerja (TPAK) dan Indeks Pembangunan Manusia (IPM) berpengaruh signifikan terha-

dap indeks kedalaman kemiskinan. Penelitian yang lain juga dilakukan oleh Tambun & Herawaty (2018) terkait pemodelan faktor-faktor yang mempengaruhi indeks kedalaman kemiskinan dan indeks keparahan kemiskinan kabupaten/kota di Sumatera Utara menggunakan regresi data panel. Hasil penelitian menunjukkan bahwa hanya variabel pengeluaran rumah tangga, angka melek huruf, dan rata-rata lama sekolah yang berpengaruh signifikan terhadap indeks kedalaman kemiskinan.

Oleh karena itu dalam penelitian ini akan dilakukan pemodelan hubungan indeks kedalaman kemiskinan dengan jumlah pengangguran terbuka. Di mana semakin banyak jumlah pengangguran, maka akan semakin besar pula kesenjangan rata-rata pengeluaran penduduk terhadap garis kemiskinan. Pada umumnya untuk melihat hubungan variabel dependen dan independen menggunakan analisis regresi. Akan tetapi karena data yang digunakan adalah kabupaten/kota maka data akan cenderung bervariasi bahkan ada *outlier*. Kondisi ini mengakibatkan *residual* tidak berdistribusi normal. Sehingga harus dilakukan pemodelan dengan regresi nonparametrik yakni regresi *smoothing splines* dan regresi kuantil *smoothing splines* (Craig & Ng, 2001). Dari kedua model tersebut selanjutnya akan dilihat model mana yang baik untuk memodelkan hubungan antara jumlah pengangguran terbuka dan indeks kedalaman kemiskinan di kabupaten/kota di Pulau Sumatera.

3. METODE PENELITIAN

Sumber Data dan Variabel Penelitian

Data yang digunakan dalam penelitian bersumber dari Badan Pusat Statistik. Variabel yang digunakan adalah indeks kedalaman kemiskinan (P1) dan jumlah penduduk yang berusia 15 tahun ke atas kegiatan utama

nya seminggu yang lalu adalah pengangguran. Pengangguran yang dimaksud di sini adalah jumlah pengangguran terbuka.

Objek penelitian adalah 154 kabupaten/kota di Pulau Sumatera tahun 2019. Pada penelitian ini variabel respon adalah indeks kedalaman kemiskinan dan variabel prediktor adalah jumlah pengangguran terbuka. Data diolah menggunakan *software R 3.6.3 package quantreg*.

Indeks kedalaman kemiskinan (P1) adalah ukuran rata-rata kesenjangan pengeluaran masing-masing penduduk miskin terhadap garis kemiskinan. Sementara definisi jumlah pengangguran terbuka sesuai dengan konsep ILO yang tertuang dalam *an ILO Manual on Concept and Methods*. Pengangguran terbuka adalah penduduk yang berumur 15 tahun ke atas yang tidak punya pekerjaan dan sedang mencari pekerjaan, sedang mempersiapkan usaha, tidak mencari pekerjaan karena merasa tidak mungkin mendapatkan pekerjaan atau sudah punya pekerjaan tetapi belum mulai bekerja (Badan Pusat Statistik, 2020).

Regresi Nonparametrik

Regresi nonparametrik adalah salah satu alat statistik yang digunakan untuk mengetahui hubungan antara dua variabel acak atau lebih. Misalkan Y adalah variabel respon dan X adalah variabel prediktor. Misalkan suatu data $\{(X_i, Y_i)\}, i = 1, \dots, n$. Maka secara umum hubungan antara X dan Y mengikuti model regresi berikut:

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

Di mana $m(X_i)$ adalah fungsi regresi yang belum diketahui dan akan ditaksir. ε_i adalah variabel acak yang menggambarkan variasi Y disekitar $m(x)$ (Hardle, 1990). Jika fungsi $m(X_i)$ tidak diketahui apakah mengikuti bentuk kurva tertentu, maka dilakukan pendekatan regresi nonparametrik. Hal

ini karena regresi nonparametrik akan mencari sendiri bentuk estimasi dari $m(x)$ dari data yang ada (Eubank, 1999). Cara ini lebih flexibel dibanding regresi parametrik karena tidak memerlukan informasi apapun tentang sebaran data.

Regresi Smoothing Splines

Regresi *smoothing splines* adalah pemulusan data dengan menggunakan fungsi *splines* sehingga diperoleh jumlah kuadrat residual yang bernilai kecil. Salah satu cara agar fungsi cukup halus adalah memberikan pemulus (*penalty*). Dalam Hastie & Tibshirani (1990), estimasi fungsi *splines* $f(x_i)$ adalah meminimisasi:

$$\sum_{i=1}^n [y_i - f(x_i)]^2 + \lambda \int_0^1 [f''(x)]^2 dx$$

Di mana $\sum_{i=1}^n [y_i - f(x_i)]^2$ adalah jumlah kuadrat dari residual. $\int_0^1 [f''(x)]^2 dx$ adalah ukuran kemulusan yang disebut *roughness penalty* dan λ adalah parameter pemulus. Nilai λ berada diantara 0 sampai ∞ . Jika λ mendekati nol, maka kurva cenderung terlihat kasar akan tetapi ketepatan model tinggi.

Di mana kondisi jarak antara data dan nilai dugaan sangat dekat. Sebaliknya jika nilai λ menuju tak hingga kelengkungan kurva sangat kecil dan mulus, akan tetapi ketepatan model rendah. Jarak antara data dan nilai dugaan sangat jauh sehingga nilainya akan sama dengan *roughness penalty*.

Regresi Kuantil Smoothing Splines

Pendekatan regresi dari kuantil *smoothing splines* adalah model regresi yang tidak hanya pada ukuran pemusatan variabel respon, tetapi juga pada berbagai kuantil dengan menggunakan teknik *smoothing splines* (Koenker, 2011; Koenker, Ng, & Portnoy, 1994). Regresi kuantil *smoothing splines* diperoleh dengan meminimumkan (Mulyani, 2017):

$$\sum_{i=1}^n \rho_{\tau}(y_i - f(x_i)) + \lambda \int_0^1 |f''(x)| dx$$

Di mana nilai persamaan dari $\rho_{\tau} = [\tau - 1 (u < 0)]u$ adalah *check function* yang dipelopori oleh Koenker & Gilbert Bassett (1978). $\tau \in [0,1]$ dan $\lambda \in \mathbb{R}_+$, dengan nilai $0 = x_0 < x_1 < \dots < x_n < x_{n+1} = 1$. Di mana x_i adalah titik knot dan f'' adalah fungsi kontinu dalam selang $[0,1]$ dan berupa *splines* linier natural dengan bentuk fungsi :

$$f(x) = \begin{cases} a_i + b_i(x - x_i) & x \in [x_i, x_{i+1}] \\ a_0 + b_0(x - x_i) & x \in [x_0, x_1] \end{cases}$$

dimana $i = 1, \dots, n$
 nilai koefisien $a_i = y_i$;
 $b_i = a_{i+1} - a_i / x_{i+1} - x_i$.

Jika nilai $h_i = x_{i+1} - x_i$ di mana $i = 1, \dots, n - 1$ maka koefisien b_i dapat juga ditulis:

$$b_i = a_{i+1} - a_i / h_i$$

Dengan memperhatikan *constraint* maka $a_0 = a_i$; $b_0 = b_i$ dan $b_{n-1} = b_n$ maka fungsi persamaan $f(x_{i+1}) = a_i + b_i h_i = a_{i+1}$. Fungsi objektif regresi kuantil *smoothing splines* dapat ditulis sebagai berikut: (Mulyani, 2017):

$$\min \sum_{i=1}^n \rho_{\tau}(y_i - f(x_i)) + \lambda \sum_{i=1}^{n-1} |f'(x_{i+1}) - f'(x_i)|$$

Pemilihan Parameter Penghalus

Dalam membentuk regresi *smoothing splines* yang optimal tergantung pada parameter pemulus atau lambda (λ). Dalam penelitian ini untuk menyelesaikan parameter pemulus digunakan *Generalized Cross Validation (GCV)*. GCV diperkenalkan oleh Craven & Wahba (1979). Kriteria pemilihan pa-

parameter dengan GCV dirumuskan sebagai berikut (Mulyani, 2017):

$$GCV(\lambda) = \frac{\sum_{i=1}^n (y_i - \hat{f}(x_i))^2}{n \left[1 - \frac{1}{n} \text{tr}[A(\lambda)]\right]^2}$$

Tahapan Penelitian

Langkah analisis yang dilakukan dalam penelitian ini adalah sebagai berikut:

- 1) Melakukan spesifikasi model lewat *scatter plot* antara indeks kedalaman kemiskinan dan jumlah pengangguran terbuka. Selanjutnya untuk menentukan teknik regresi yang akan digunakan uji asumsi normalitas. Jika tidak memenuhi asumsi, maka dilakukan pendekatan nonparametrik yakni dengan regresi *smoothing splines*.
- 2) Melakukan pemodelan regresi *smoothing splines* di mana nilai parameter pemulus dipilih dengan metode GCV.
- 3) Pemodelan regresi kuantil *smoothing splines* dengan mendeteksi ada tidaknya data *outlier*. Salah satu cara untuk mendeteksi dengan menggunakan *boxplot* dari *residu-*

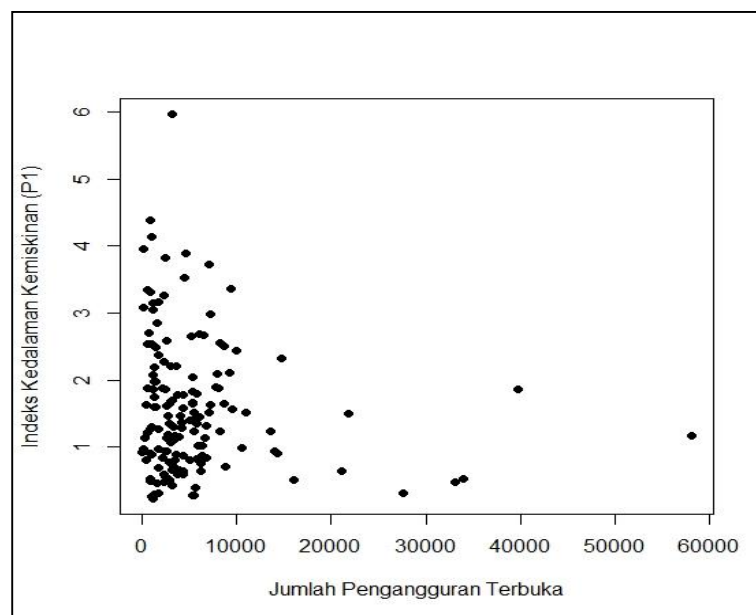
al model regresi *smoothing splines*.

- 4) Menentukan nilai parameter pemulus (λ) dari nilai SIC yang paling minimum. Di mana model regresi kuantil *smoothing splines* (median) nilai $\tau = 0,5$
- 5) Membuat kurva regresi kuantil *smoothing splines*.
- 6) Menghitung nilai RMSE kemudian membandingkan nilai RMSE antara model regresi *smoothing splines* dan regresi kuantil *smoothing splines* pada $\tau = 0,5$. Model yang memiliki nilai RMSE lebih kecil merupakan model yang lebih efisien.

4. ANALISIS DATA DAN PEMBAHASAN

Spesifikasi Model

Tahapan ini adalah langkah awal untuk melihat spesifikasi model dari plot sebaran data. Jika pola sebaran data tidak dapat ditentukan fungsi *slope* dan *intersept* nya, maka dapat digunakan pendekatan nonparametrik. Bentuk pola sebaran data kedua variabel dapat dilihat pada Gambar 1.



Gambar 1. Pola Hubungan Jumlah Pengangguran Terbuka dan Indeks Kedalaman Kemiskinan (P1) Kabupaten/Kota di Pulau Sumatera Tahun 2019

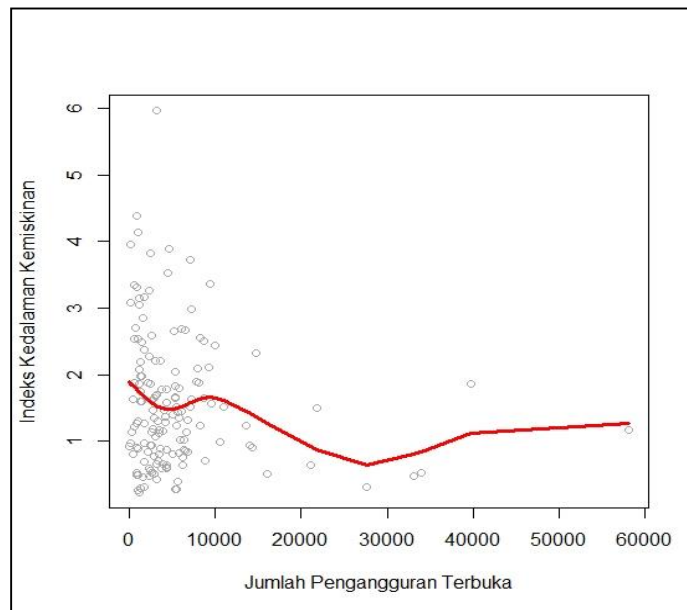
Sumber: Hasil Pengolahan

Dari Gambar 1 terlihat jika sebaran data mengikuti pola yang tidak diketahui fungsinya. Di mana terlihat data cenderung mengumpul di sisi kiri. Namun untuk memastikan, dilakukan pengujian asumsi parametrik yakni *residual* berdistribusi normal dengan uji Saphiro Wilk.

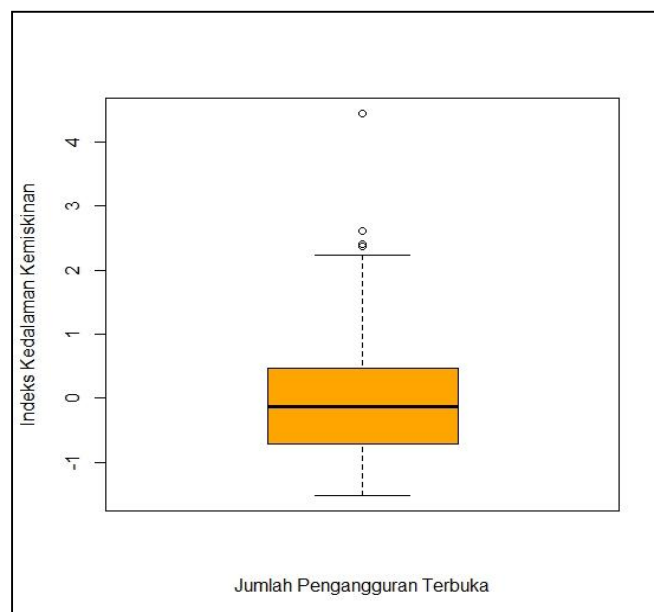
Hipotesis uji ini adalah H_0 : residual berdistribusi normal. Sedangkan H_1 : residual tidak berdistribusi normal.

Tolak H_0 jika $p\text{-value} < \alpha$ (0,05), maka dapat ditarik kesimpulan *residual* tidak berdistribusi normal. Hasil pengujian didapat bahwa nilai $p\text{-value}$ sebesar $1,963 \exp^{-7}$.

Dengan demikian dapat simpulkan bahwa *residual* tidak berdistribusi normal. Oleh karenanya untuk melihat hubungan jumlah pengangguran terbuka dan indeks kedalaman kemiskinan (P1) menggunakan pendekatan regresi nonparametrik.



Gambar 2. Kurva Regresi *Smoothing Splines*
Sumber: Hasil Pengolahan



Gambar 3. *Boxplot Residual* Regresi *Smoothing Splines*
Sumber: Hasil Pengolahan

Regresi Smoothing Splines

Nilai parameter dari λ yang optimum diperoleh dengan teknik *Generalized Cross Validation* (GCV). Hasil penghitungan diperoleh:

Tabel 1. Parameter Regresi *Smoothing Splines*

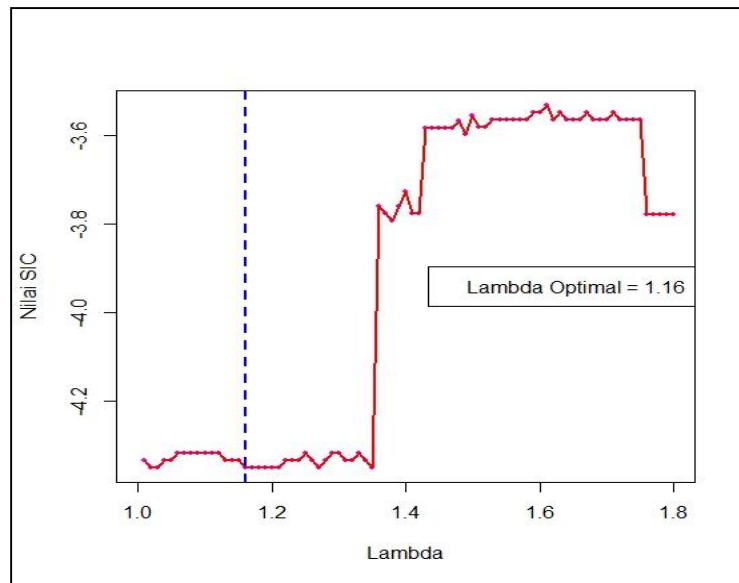
Parameter	Nilai
Spar	1,186336
Lambda	0,0007533958
GCV	1,011754

Parameter optimum pada tabel 1 digunakan pada pemodelan regresi

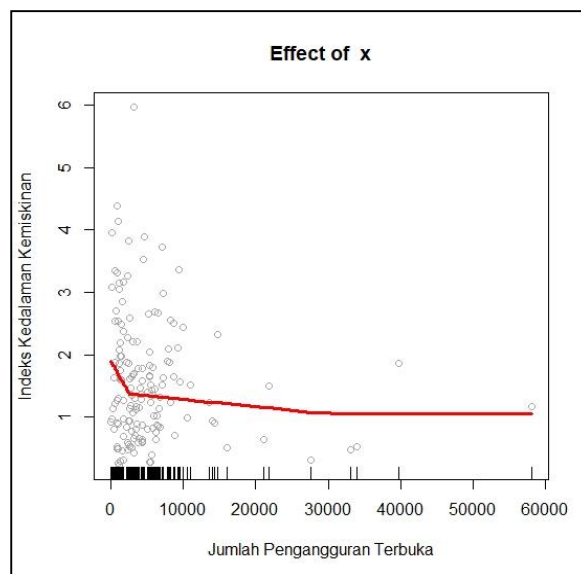
smoothing splines. Pemodelan yang terbentuk seperti Gambar 2.

Regresi Kuantil *Smoothing Splines* (Median)

Jika dilihat pada Gambar 2, kurva regresi *smoothing splines* kurang dapat menjelaskan secara detail sebaran data apalagi jika ada data *outlier*. Jika hal ini terjadi maka sebaiknya menggunakan pendekatan regresi kuantil *smoothing splines* yang mampu mengakomodir adanya *outlier* dan melihat sebaran data pada seluruh garis kuantil yang diinginkan.



Gambar 4. Plot SIC dan Lambda Regresi *Smoothing Splines* (Median, $\tau = 0,5$)
 Sumber: Hasil Pengolahan



Gambar 5 . Kurva Regresi Kuantil *Smoothing Splines* (Median, $\tau = 0,5$)
 Sumber: Hasil Pengolahan

Salah satu cara untuk mengidentifikasinya adalah melihat *box plot residual* dari model regresi *smoothing splines*. Hasil pengolahan diperoleh bahwa adanya data *outlier* di sisi kanan distribusi.

Penentuan lambda optimum pada regresi kuantil *smoothing splines* (median) menggunakan kriteria *Schwarz Information Criteria* (SIC) dengan nilai $\tau = 0,5$. Hasil menunjukkan nilai lambda optimum sebesar 1,16 (Gambar 4).

Lambda optimal sebesar 1,16 digunakan pada regresi kuantil *smoothing splines* digunakan untuk pemodelan. Sehingga pada plot data terlihat seperti dibawah ini.

Seperti yang telah dijelaskan sebelumnya regresi kuantil *smoothing splines* dapat digunakan pada berbagai nilai kuantil. Dalam penelitian ini indeks kedalaman kemiskinan dibagi menjadi lima kategori yakni sangat rendah, rendah, sedang, tinggi dan sangat tinggi. Dengan demikian kuantil dapat dibagi menjadi 4 nilai yakni 0,2; 0,4; 0,6 dan 0,8. Hasil pe-

modelan pada keempat nilai kuantil tersebut dapat dilihat pada gambar 6 berikut ini.

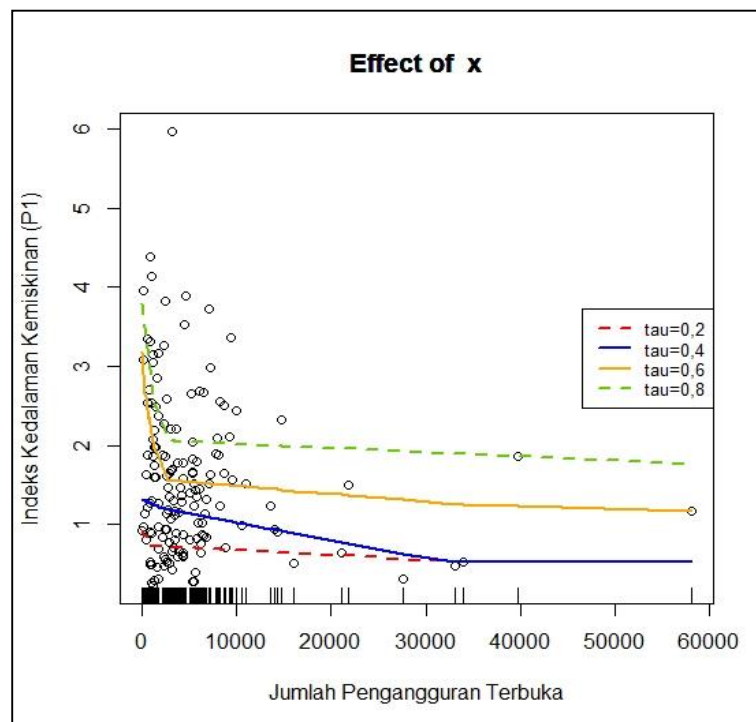
Perbandingan Model

Adanya data *outlier* maka pendekatan regresi kuantil *smoothing splines* lebih efisien dibanding regresi *smoothing splines*. Performa kedua model ini dapat dibandingkan dengan melihat nilai *Root Means Square Error* (RMSE). Semakin kecil nilai RMSE maka model semakin baik.

Tabel 2. Perbandingan Model

Model Regresi	RMSE
Regresi <i>Smoothing Splines</i>	0,967
Regresi Kuantil <i>Smoothing Splines</i> (Median)	0,703

Dari tabel 2 diketahui nilai RMSE model regresi *smoothing splines* sebesar 0,967. Angka ini lebih besar dibandingkan dengan RMSE regresi kuantil *smoothing splines*. Hal ini dikarenakan adanya *outlier* di mana regresi *smoothing splines* sangat sensitif terhadap hal tersebut yang mengakibatkan nilai *residual* besar.



Gambar 6 . Kurva Regresi Kuantil *Smoothing Splines* ($\tau = 0,2; 0,4; 0,6$ dan $0,8$)
 Sumber: Hasil Pengolahan

5. KESIMPULAN, IMPLIKASI, SARAN, DAN BATASAN

Dari hasil penelitian, maka dapat ditarik kesimpulan sebagai berikut:

- 1) Regresi kuantil *smoothing splines* memiliki kelebihan karena dapat memberikan nilai berbagai kuantil sesuai dengan keinginan.
- 2) Regresi kuantil *smoothing splines* lebih robust terhadap data outlier.
- 3) Nilai RMSE model regresi kuantil *smoothing splines* (median, $\tau = 0,5$) lebih kecil dibanding RMSE model regresi *smoothing splines*.

DAFTAR PUSTAKA

- Ahmaddien, I. (2019). Faktor Determinan Keparahan dan Kedalaman Kemiskinan Jawa Barat dengan Regresi Data Panel. *Forum Ekonomi*, 21(1), 87–96. <https://doi.org/10.29264/jfor.v21i1.5225>
- Badan Pusat Statistik. (2020). *Keadaan Angkatan Kerja di Indonesia Agustus 2019*. Jakarta: Badan Pusat Statistik.
- Craig, S. G., & Ng, P. T. (2001). Using Quantile Smoothing Splines to Identify Employment Subcenters in a Multicentric Urban Area. *Journal of Urban Economics*, 49(1), 100–120. <https://doi.org/10.1006/juec.2000.2186>
- Craven, P., & Wahba, G. (1979). Smoothing Noisy Data with Spline Functions: Estimating the Correct Degree of Smoothing by The Method of Generalized Cross-Validation. *Numerische Mathematik*, 31, 377–403. <https://doi.org/10.1007/BF01404567>
- Eubank, R. (1999). *Nonparametric Regression and Spline Smoothing*. New York: Marcel Dekker.
- Hardle, W. (1990). *Applied Nonparametric Regression*. Cambridge University Press.
- Hastie, T. ., & Tibshirani, R. J. (1990). *Generalized Additive Models*. London: Chapman and Hall.
- Koenker, R. (2011). Additive Models for Quantile Regression: Model Selection and Confidence Bands. *Brazilian Journal of Probability and Statistics*, 25(3), 239–262. <https://doi.org/10.1214/10-BJPS131>
- Koenker, R., & Gilbert Bassett, J. (1978). Regression Quantile. *Econometrica*, 46(1), 33–50. <https://doi.org/10.2307/1913643>
- Koenker, R., Ng, P. T., & Portnoy, S. (1994). Quantile Smoothing Splines. *Biometrika*, 81(4), 673–680. <https://doi.org/10.1093/biomet/81.4.673>
- Mulyani, S. (2017). *Pemodelan Hubungan Indeks Pembangunan Manusia dan Persentase Penduduk Miskin Menggunakan Regresi Kuantil Smoothing Splines*. Tesis: Universitas Padjadjaran.
- Tambun, J. M. S., & Herawaty, R. (2018). Pemodelan Faktor-faktor yang Mempengaruhi Indeks Kedalaman Kemiskinan dan Indeks Keparahan Kemiskinan Kabupaten/Kota di Sumatera Utara Menggunakan Regresi Data Panel. *Jurnal Administrasi Publik*, 6(1), 100–110. <https://doi.org/https://doi.org/10.31289/publika.v6i1.1574>