



ARTIKEL PENELITIAN

Analisis Efektivitas Model *Geographically Weighted Quantile Regression* (GWQR) Dalam Penanganan Outlier: Data Simulasi Teridentifikasi Heterogenitas Spasial

Febrya Christin Handayani Buan^{1,*}, Zofar Agluis Banunaek¹, Widya Reza²

¹Program Studi Agroteknologi, Universitas Timor, Kefamenanu-NTT, Indonesia

²Program Studi Matematika, Institut Teknologi Batam, Batam-Kepulauan Riau, Indonesia

*Penulis korespondensi: putrybuan@unimor.ac.id

Diterima: 10 Oktober 2023; Direvisi: 08 November 2023; Disetujui: 11 November 2023; Dipublikasi: 09 Februari 2024.

Abstrak: Regresi kuantil klasik bersifat global dengan hasil estimasi parameter yang general, kondisi heterogenitas spasial tidak dapat ditangkap oleh model ini. Penggunaan model lokal dengan atribut spasial dapat mengakomodir karakteristik antar lokasi pengamatan. Model lokal regresi kuantil disebut model *Geographically Weighted Quantile Regression* (GWQR). Pengujian efektivitas lebih lanjut dari model ini diperlukan dengan pemanfaatan data simulasi. Penelitian ini menggunakan data simulasi bangkitan dengan ukuran sampel ($n = 50, 100, 150$), berdistribusi uniform dengan interval $(0, 1)$ yang terkontaminasi *outlier* 5%, 10%, 15%, dengan variabel prediktor $(x = 4)(\beta_1, \beta_2, \beta_3, \beta_4)$, dan ukuran kuantil sebesar 0.05, 0.25, 0.50, 0.75, 0.95. Selanjutnya dibandingkan dengan model lokal regresi klasik yaitu model *Geographically Weighted Regression* (GWR). Model GWR dapat mengakomodir heterogenitas spasial tetapi permasalahan *outlier* diabaikan. Pengukuran efektivitas model berdasarkan nilai *Root Mean Square Error* (RMSE). Dari hasil penelitian diperoleh model GWQR merupakan model terbaik dibandingkan model GWR dalam mengatasi permasalahan *outlier* hingga besaran kontaminasi *outlier* 15% yang teridentifikasi heterogenitas spasial. Kondisi ini ditunjukkan dengan nilai RMSE model GWQR yang semakin mendekati nol seiring bertambahnya ukuran sampel, hasil ini berbanding terbalik dari model GWR. Selain itu, pada penelitian ini diperoleh besaran nilai kuantil 0.95 merupakan penghasil nilai estimasi parameter terbaik dibandingkan nilai kuantil lainnya, karena kauntal 0,95 konsisten menghasilkan nilai RMSE terendah disetiap ukuran sampel yang berbeda.

Kata Kunci: *Geographically Weighted Quantile Regression*, *Geographically Weighted Regression*, Heterogenitas Spasial, *Root Mean Square Error*, Simulasi.

Abstract: Classical quantile regression is global generalized parameter estimation results, spatial heterogeneity conditions cannot be captured by this model. The use of local models with spatial attributes can accommodate the characteristics between observation locations. The local quantile regression model is called the *Geographically Weighted Quantile Regression* (GWQR) model. Further testing of the effectiveness of this model is required by utilizing simulation data. This study uses simulated data generation with sample sizes ($n = 50, 100, 150$), uniformly distributed with intervals $(0, 1)$ contaminated with 5%, 10%, 15% outliers, with predictor variables $(x =$

4) $(\beta_1, \beta_2, \beta_3, \beta_4)$, and quantile sizes of 0.05, 0.25, 0.50, 0.75 and 0.95. Furthermore, it is compared with the local classical regression model, namely the Geographically Weighted Regression (GWR) model. The GWR model can accommodate spatial heterogeneity but the outlier problem is ignored. Measurement of model effectiveness is based on the Root Mean Square Error (RMSE) value. From the results of the study, it was obtained that the GWQR model is the best model compared to the GWR model in overcoming outlier problems up to the amount of 15% outlier contamination identified spatially in simulated data. This condition is shown by the RMSE value of the GWQR model which is getting closer to 0 as the sample size increases and the percentage of outliers, this result is inversely proportional to the GWR model. In addition, this study found that the 0.95 quantile value is the best producer of parameter estimation values compared to other quantile values, because the 0.95 quantile consistently produces the lowest RMSE value in each different sample size.

Keywords: Geographically Weighted Quantile Regression, Geographically Weighted Regression, Root Mean Square Error, Simulation, Spatial Heterogeneity.

1. Pendahuluan

Margin of error suatu kondisi yang sering dialami dalam tahapan pengumpulan data yang mengakibatkan data pengamatan menyimpang terlalu jauh dari pola data, sehingga sebaran data tidak akan simetris, dan teridentifikasi terjadi *outlier*/pencilan [1, 2]. Jika kondisi ini dibiarkan maka akan berdampak pada tidak terpenuhinya asumsi kenormalan galat dalam pengujian asumsi kalsik model regresi [3]. Solusi penyelesaian masalah *outlier* dapat dilakukan dengan memisahkan data menjadi kuantil-kuantil tertentu yang memiliki nilai dugaan parameter yang berbeda untuk setiap kuantil, dan tidak terpengaruh terhadap *outlier*. Solusi tersebut dikenal sebagai regresi kuantil, dengan prosedur meminimumkan jumlah nilai mutlak sisaan (*absolute residual*) yang disebut *Least Absolute Deviation* (LAD) [4, 5].

Analisis regresi bersifat global menghasilkan nilai parameter secara *general* untuk setiap lokasi pengamatan [6]. Ketika kondisi ragam galat tidak homongan maka pemodelan dengan model regresi global tidak dapat menangkap secara baik sebaran data yang teridentifikasi heterogenitas spasial oleh karena itu, untuk analisis lebih lanjut diperlukan pemodelan dengan pemanfaatan model regresi lokal yang dapat mengakomodir heterogenitas spasial [7]. Model regresi lokal mengimplementasikan data spasial sebagai data atribut yang memiliki unsur geografis lokasi pengamatan, sehingga estimasi parameter yang dihasilkan akan berbeda-beda di setiap unit lokasi pengamatan [8, 9]. Model lokal dari regresi kuantil dan pengembangannya disebut model *Geographically Weighted Quantile Regression* (GWQR) [10].

Penelitian yang menerapkan model GWQR untuk estimasi parameter telah dilakukan, untuk menjawab permasalahan ilmu pengetahuan terapan dalam memanfaatkan model GWQR dalam penanganan berbagai nilai *outlier* yang kemudian di petakan [11–14]. Dengan demikian, permasalahan *outlier* terselesaikan oleh model GWQR dengan berbagai nilai kuantil pada data yang teridentifikasi heterogenitas spasial. Berdasarkan penelitian terdahulu maka diperlukan penelitian untuk mengkaji lebih dalam terkait tingkat keefektifan model GWQR dalam penanganan permasalahan *outlier* pada beberapa sampel terhadap beberapa nilai *outlier* berdasarkan data simulasi yang terkonfirmasi heterogenitas spasial yang akan dibandingkan dengan model lokal regresi klasik yaitu *Geographically Weighted Regression* (GWR) yang dapat mengakomodasi heterogenitas spasial tetapi tidak untuk *outlier*.

Variabel prediktor yang akan diamati pada penelitian ini berjumlah ($x = 4$), sedangkan nilai kuantil sebesar 0.05, 0.25, 0.50, 0.75, dan 0.95 yang bertujuan untuk mendapatkan gambaran pola sebaran data pada kuantil bawah, tengah, dan atas serta untuk mendapatkan nilai-nilai ekstrim. Sedangkan besaran presentase *outlier* yang diamati adalah 5%, 10%, dan 15%, yang dilakukan

pengujian untuk sampel ($n = 50, 100, \text{ dan } 150$). Pengujian efektivitas diukur berdasarkan nilai *Root Mean Square Error* (RMSE).

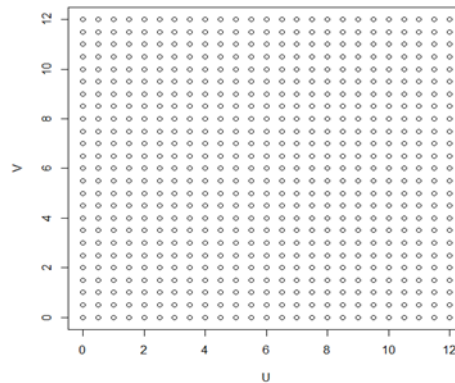
2. Metode Penelitian

2.1. Sumber Data

Data simulasi diperoleh dengan cara membangkitkan data sebanyak sampel $n = 50, 100, 150$ untuk kasus yang terdapat outlier berdasarkan model.

$$y_i = \beta_0(u_i, v_i) + \beta_1(u_i, v_i)x_i + \varepsilon_i \quad i = 1, 2, \dots, 100 \quad (2.1)$$

Outlier diberikan pada variabel x_i dengan persentase *outlier* sebesar 5%, 10%, dan 15%. Data lokasi digambarkan sebagai sebuah wilayah persegi yang memiliki koordinat kartesius dengan panjang sisi 12 unit. Sistem koordinat dibentuk melalui titik asal pada sudut kiri bawah, dengan lokasi pertama merupakan koordinat (0,0) sehingga terdapat 100 titik lokasi pengamatan dengan jarak 0,5 antara dua titik horizontal dan vertikal.



Gambar 2.1: Koordinat spasial untuk lokasi (u_i, v_i)

2.2. Tahapan Analisis

Studi simulasi pada penelitian ini dilakukan menggunakan data bangkitan dengan bantuan software RStudio. Algoritma simulasi dijelaskan sebagai berikut:

1. Membangkitkan data variabel prediktor $x_i (i = 1, 2, \dots, 4)$ sebanyak sampel yang $n = 50, 100, 150$ berdistribusi *uniform* dengan interval (0,1).
2. *Outlier* diberikan pada variabel prediktor x^* yang berdistribusi *uniform* dengan interval (4,5) secara acak dengan persentase *outlier* $n\delta$ ($\delta = 5\%, 10\%, \text{ dan } 15\%$).
3. Melakukan identifikasi *outlier* dengan boxplot.
4. Membangkitkan galat $\varepsilon_i \sim N(0, 1); i = 1, 2, \dots, n$.
5. Menghitung nilai $\beta_0(u_i, v_i)$ dan $\beta_1(u_i, v_i)$ yang teridentifikasi heterogenitas spasial dengan persamaan $\beta_0(u_i, v_i) = 1 + (\frac{1}{6})(u + v)$ dan $\beta_1(u_i, v_i) = u/3$.
6. Menghitung nilai variabel respon y_i dengan menambahkan nilai $x_i, \varepsilon_i, \beta_0(u_i, v_i)$ dan $\beta_1(u_i, v_i)$.
7. Melakukan pengujian heterogenitas spasial dengan *Breusch-Pagan Tests*.
8. Melakukan pengujian efektivitas model GWQR berdasarkan uji *bartlett* dan nilai standar error.
9. Menghitung nilai RMSE pada model GWQR yang terbentuk dan melakukan uji kesamaan galat.

3. Hasil dan Pembahasan

3.1. Pembangkitan Data Simulasi

Banyaknya variabel prediktor (x) yang terdapat *outlier* sebesar $n\delta$. Pada penelitian ini dimisalkan $n = 100$ dan $\delta = 5\%$, 10% , dan 15% dibangkitkan terlebih dahulu untuk $\delta = 5\%$ sehingga $n\delta = 100 \times 5\% = 5$ yang artinya akan mempertimbangkan adanya *outlier* sejumlah 5 dari 100 pengamatan. Titik amatan *outlier* dapat dihitung dengan langkah berikut:

$$j = n(1 - \delta) + 1 = 100(1 - 5\%) + 1 = 96 \text{ (outlier berada di amatan ke 96)}$$

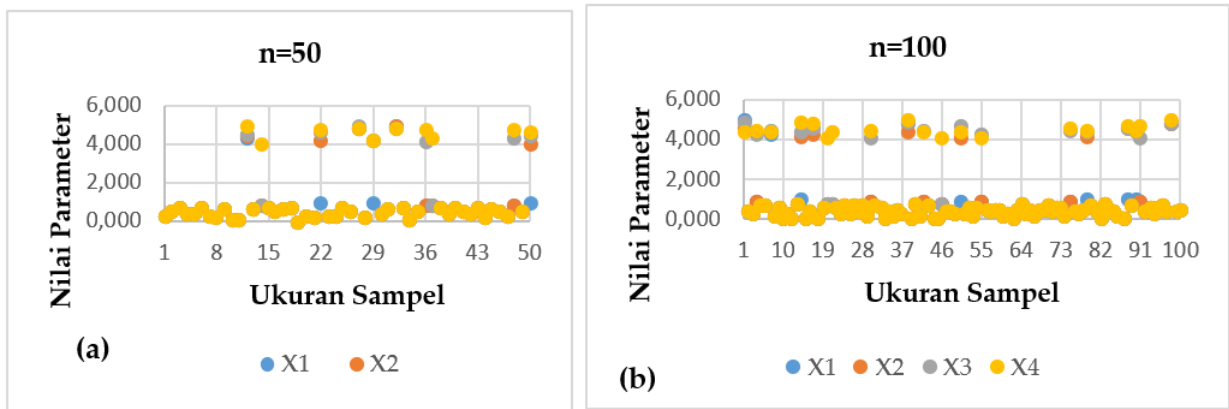
$$j = n(1 - \delta) + 2 = 100(1 - 5\%) + 2 = 97 \text{ (outlier berada di amatan ke 97):}$$

$$j = n(1 - \delta) + 10 = 100(1 - 5\%) + 5 = 100 \text{ (outlier berada di amatan ke 100)}$$

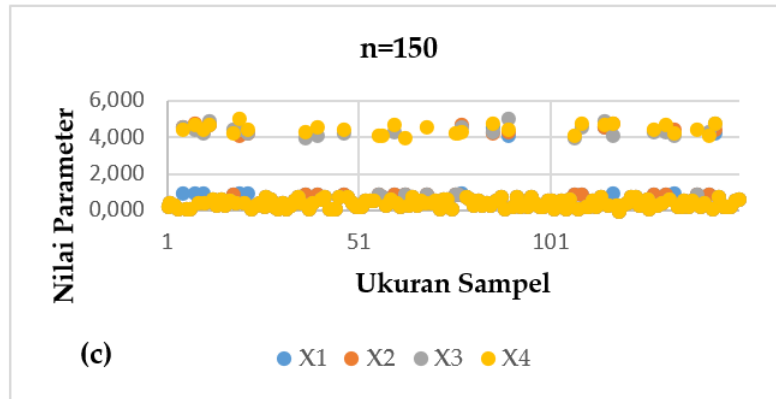
Simulasi dilakukan dengan cara membangkitkan data sampel yang didasari pada model GWR dan GWQR sebagai berikut:

$$y_i = \beta_0(u_i, v_i) + \beta_1(u_i, v_i)x_1 + \beta_2(u_i, v_i)x_2 + \beta_3(u_i, v_i)x_3 + \beta_4(u_i, v_i)x_4 + \varepsilon_i \quad (3.2)$$

dimana ε_i merupakan galat bangkitan yang berdistribusi normal $\varepsilon_i \sim N(0, 1)$, dan variabel prediktor x_i dibangkitkan berdasarkan distribusi *uniform* dengan interval (0,1). Hasil bangkitan data simulasi pada penelitian ini dilakukan untuk setiap ukuran sampel yang tercermin pada Gambar 3.2, 3.3. Berdasarkan data bangkitan terlihat jelas bahwa setiap ukuran sampel telah teridentifikasi adanya *outlier*, untuk memastikan maka dilakukan pendeteksian dengan *boxplot*.



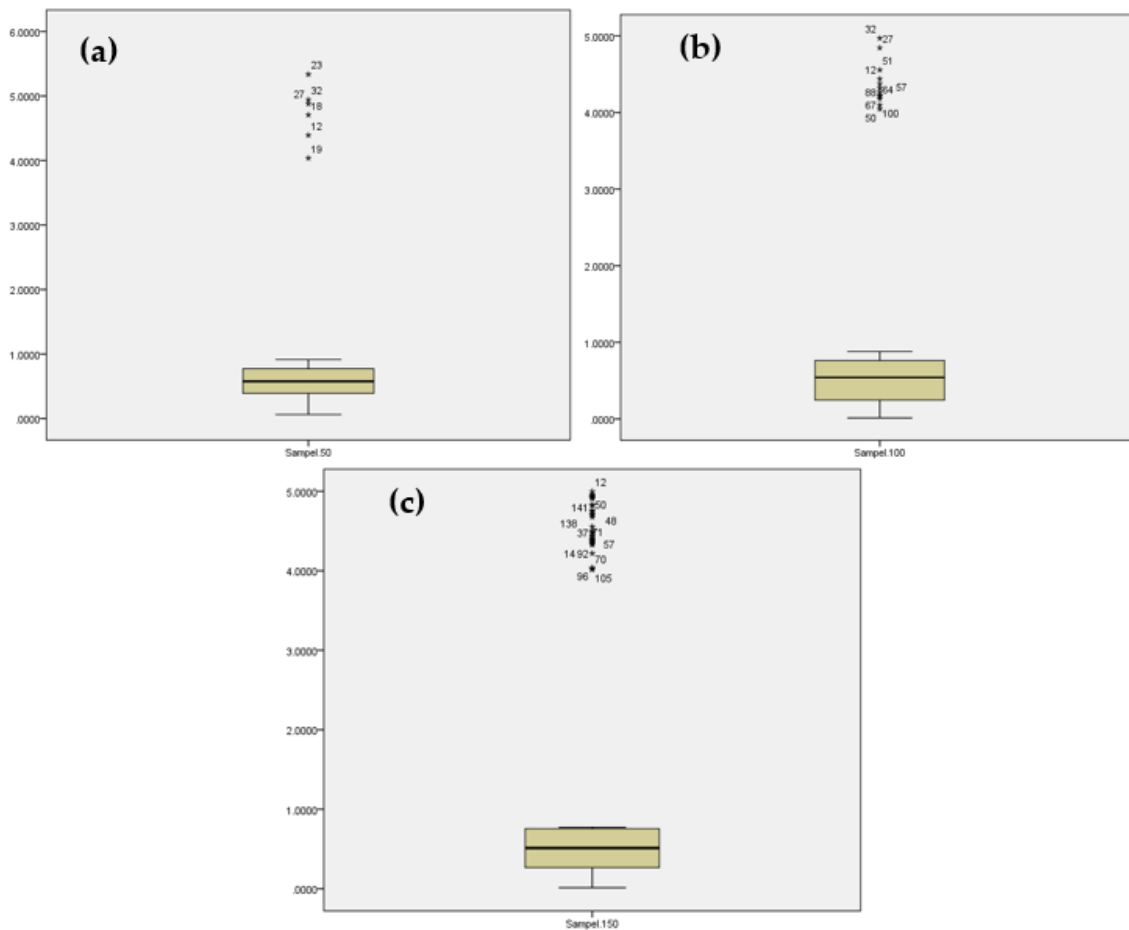
Gambar 3.2: Sebaran data bangkitan ukuran sampel $n = 50$ (a), ukuran sampel $n = 100$ (b)



Gambar 3.3: Sebaran data bangkitan ukuran sampel $n = 150$ (c)

3.2. Identifikasi Outlier

Tahapan pendeteksian outlier pada data bangkitan dilakukan dengan menggunakan boxplot yang disajikan pada Gambar 3.4.



Gambar 3.4: Boxplot dengan ukuran sampel $n = 50$ (a), ukuran sampel $n = 100$ (b), ukuran sampel $n = 150$ (c)

Berdasarkan hasil output *boxplot* dari data bangkitan untuk setiap ukuran sampel amatan terkonfirmasi adanya *outlier* yang menyebabkan sebaran data menjadi tidak simetris yang tercermin dari panjang garis *whiskers* bagian atas dan bawah yang tidak simetris, dan nilai median pada *boxplot* setiap sampel amatan tidak berada tepat di tengah *boxplot* dengan sebaran data cenderung menjalar ke kanan (*positive skewness*).

3.3. Pengujian Heterogenitas Spasial

Data bangkitan yang terindikasi adanya heterogenitas spasial antar lokasi pengamatan dapat diketahui dengan pengujian *Breusch-Pagan Tests*, berdasarkan hipotesis sebagai berikut:

$H_0 : \sigma_{u_1, v_1}^2 = \sigma_{u_2, v_2}^2 = \dots = \sigma_{u_n, v_n}^2 = \sigma^2$ (tidak terdapat heterogenitas spasial)

$H_1 : \text{minimal ada satu } \sigma_{u_i, v_i}^2 \neq \sigma^2$ (terdapat heterogenitas spasial)

dimana $i = 1, 2, n$.

Statistik uji:

$$BP = \left(\frac{1}{6}\right) h^T Z(Z^T Z)^{-1} Z^T h \sim \chi_{(x+1)}^2 \quad (3.3)$$

dimana h : vektor dengan elemen $h_i = ((e_i^2)/\sigma^2 - 1)$, Z : matriks χ berukuran $n \times (p+1)$ untuk setiap pengamatan dengan x merupakan banyaknya variabel prediktor.

Kriteria pengambilan keputusan adalah terima H_0 jika nilai statistik uji $BP \leq \chi_{(p+1)}^2$ atau jika $p - \text{value} \geq \alpha$, dimana $\chi_{(p+1)}^2$ merupakan titik kritis uji χ_{p-k-1}^2 dengan taraf kesalahan α dan derajat bebas $(p + 1)$ sehingga dapat disimpulkan adanya heterogenitas spasial.

Hasil pengujian *BP* diperoleh $p - \text{value}$ sebesar $0,001 < \alpha(0,05)$ sehingga disimpulkan bahwa, pada data bangkitan benar adanya keberagaman antar lokasi pengamatan. Heterogenitas spasial mengakibatkan pembentukan model global regresi kuantil tidak dapat digunakan karena setiap lokasi akan memberikan nilai yang berbeda-beda, sehingga diperlukan model kuantil lokal yang mengakomodasi heterogenitas spasial.

3.4. Kajian Efektivitas Metode GWQR

Efektivitas metode GWQR dalam penanganan *outlier* pada data bangkitan diuji dengan beberapa pengujian, yang pertama uji *Bartlett*. Uji *Bartlett* dilakukan pada model GWR dan GWQR untuk mengetahui apakah terdapat keragaman galat dari setiap kuantil. Hasil uji *Bartlett* disajikan pada Tabel 1.

Tabel 1: Nilai Ragam Galat

Hasil Uji Barlett				
Sample size	Method	5% Outliers VE	10% Outliers VE	15% Outliers VE
n=50	GWR	1.829	1.854	1.943
	GWQR 0.05	1.024	1.021	1.012
	GWQR 0.25	1.076	1.053	1.077
	GWQR 0.50	1.020	1.023	1.028
	GWQR 0.75	1.065	1.063	1.089
	GWQR 0.95	1.015	1.015	1.015
Barlett = 3645.632				
n=100	GWR	2.012	2.034	2.056
	GWQR 0.05	1.223	1.023	1.034
	GWQR 0.25	1.032	1.076	1.077
	GWQR 0.50	1.020	1.023	1.028
	GWQR 0.75	1.065	1.062	1.089
	GWQR 0.95	1.015	1.075	1.016
Barlett = 4645.632				
n=150	GWR	2.249	2.651	2.756
	GWQR 0.05	1.183	1.024	1.271
	GWQR 0.25	1.023	1.026	1.320
	GWQR 0.50	1.024	1.023	1.132
	GWQR 0.75	1.020	1.013	1.312
	GWQR 0.95	1.010	1.015	1.161
Barlett = 7324.947				

Uji *Bartlett* dilakukan terhadap setiap ukuran sampel dengan hipotesis sebagai berikut:

$H_0 : \sigma_{e_1}^2 = \sigma_{e_2}^2 = \dots = \sigma_{e_k}^2$ (ragam galat sama)

$H_1 : \sigma_{(e_i)}^2 \neq \sigma_{e_j}^2$ (paling tidak terdapat satu pasang (i, j) ragam galat yang berbeda)

Hasil uji untuk model GWR dan GWQR masing-masing kuantil untuk setiap ukuran sampel berdasarkan presentase *outlier* dengan $\alpha = 0.05$ diperoleh titik kritis $\chi_{(0.05, 18-1)}^2$ sebesar 27,587112 sehingga H_0 ditolak karena $b > \chi_{(\alpha, k-1)}^2$ yang disimpulkan bahwa ragam galat dari model GWR dan GWQR tersebut berbeda. Dengan kata lain model GWR dan GWQR untuk setiap ukuran sampel amatan pada kondisi adanya heterogenitas spasial memberikan hasil pendugaan yang berbeda. Tetapi pada model GWR menghasilkan nilai ragam galat yang tinggi dibandingkan model GWQR Selanjutnya, dilakukan pengujian efektivitas berdasarkan nilai standar error setiap parameter $\beta_1, \beta_2, \beta_3, \beta_4$ untuk ukuran sampel $n = 50, 100, 150$, hasil pengujian disajikan pada Tabel 2.

Tabel 2: Nilai Standard Error ($\hat{\beta}_i$)

Sample size	Method	5% Outliers				10% Outliers				15% Outliers			
		SE b1	SE b2	SE b3	SE b4	SE b1	SE b2	SE b3	SE b4	SE b1	SE b2	SE b3	SE b4
n=50	GWR	2.08	2.84	2.09	2.54	2.41	2.05	2.81	2.84	2.52	2.53	2.51	2.65
	GWQR 0.05	1.84	1.58	1.65	1.72	1.60	1.62	1.58	1.59	1.50	1.75	1.43	1.54
	GWQR 0.25	1.83	1.34	1.52	1.65	1.57	1.53	1.67	1.55	1.65	1.53	1.52	1.64
	GWQR 0.50	1.80	1.76	1.64	1.43	1.82	1.63	1.89	1.43	1.75	1.32	1.23	1.53
	GWQR 0.75	1.73	1.84	1.64	1.53	1.58	1.62	1.76	1.34	1.60	1.50	1.20	1.63
	GWQR 0.95	1.56	1.65	1.50	1.45	1.43	1.30	1.45	1.29	1.29	1.21	1.01	1.13
n=100	GWR	2.31	2.40	2.65	2.42	2.32	2.45	2.48	2.54	2.45	2.54	2.32	2.64
	GWQR 0.05	1.54	1.86	1.73	1.64	1.59	1.83	1.64	1.74	1.38	1.46	1.44	1.54
	GWQR 0.25	1.42	1.20	1.34	1.47	1.21	1.24	1.32	1.23	1.30	1.43	1.45	1.33
	GWQR 0.50	1.40	1.32	1.52	1.54	1.36	1.22	1.38	1.54	1.45	1.33	1.21	1.25
	GWQR 0.75	1.34	1.43	1.39	1.38	1.27	1.30	1.25	1.42	1.44	1.53	1.23	1.43
	GWQR 0.95	1.26	1.30	1.32	1.24	1.20	1.20	1.21	1.20	1.28	1.20	1.18	1.17
n=150	GWR	2.56	2.64	2.76	2.54	2.74	2.46	2.54	2.12	2.54	2.54	3.23	2.14
	GWQR 0.05	1.12	1.23	1.05	1.04	1.14	1.13	1.09	1.05	1.13	1.23	1.10	1.00
	GWQR 0.25	1.02	1.00	1.04	1.17	1.01	1.23	1.20	1.03	1.10	1.00	0.78	0.33
	GWQR 0.50	1.01	1.02	1.05	1.14	1.06	1.12	1.08	1.14	1.05	1.03	0.54	0.85
	GWQR 0.75	1.83	1.82	1.89	1.88	1.87	1.80	1.84	1.82	1.85	1.89	0.86	0.73
	GWQR 0.95	0.66	0.64	0.60	0.64	0.65	0.66	0.61	0.67	0.61	0.68	0.65	0.61

Dari hasil pengujian diperoleh model GWR menghasilkan nilai *standard error* tertinggi untuk setiap ukuran sampel, hal ini membuktikan bahwa model GWQR efektif dalam penyelesaian masalah *outlier*. Penelitian serupa yang dilakukan oleh Herawati (2020), menekankan bahwa model global regresi kuantil merupakan model yang efektif menyelesaikan permasalahan *outlier* dibandingkan model *OLS* yang dilihat dari nilai *standard error* berdasarkan data simulasi tidak teridentifikasi spasial [15]. Kondisi yang sama juga terjadi pada pengujian dengan model lokal regresi kuantil yang mengakomodasi adanya heterogenitas spasial pada penelitian ini.

Kuantil 0.95 memberikan nilai *standard error* terendah dibandingkan nilai kuantil 0.5, 0.25, 0.50, 0.75, untuk setiap ukuran sampel. Akan tetapi, secara spesifik ukuran sampel $n = 150$ menghasilkan nilai *standard error* terkecil dibandingkan ukuran sampel lainnya, kondisi ini menjelaskan permasalahan *outlier* terselesaikan seiring dengan bertambahnya jumlah sampel. Berdasarkan temuan tersebut dapat disimpulkan bahwa model GWQR terbukti merupakan yang efektif dibandingkan model GWR, dan kuantil 0.95 merupakan ukuran kuantil yang paling akurat dibandingkan ukuran kuantil lainnya.

3.5. Pengujian Root Mean Square Error (RMSE)

Setelah dilakukan pengujian kesamaan ragam, maka pengujian selanjutnya dilakukan dengan menguji nilai RMSE yang merupakan nilai rata-rata dari jumlah kuadrat kesalahan untuk setiap kuantil terhadap beberapa presentase *outlier*, yang dapat dihitung dengan persamaan berikut [16]:

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2} \tag{3.4}$$

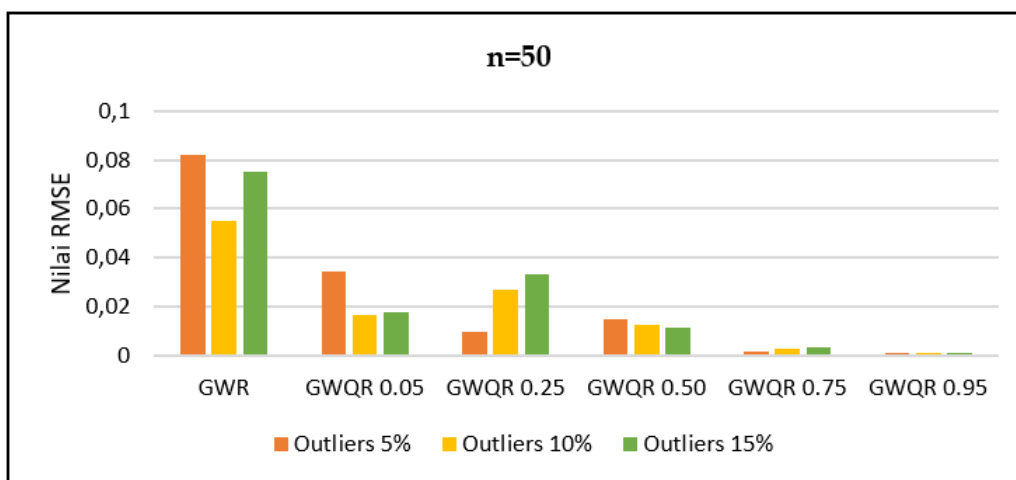
Nilai RMSE untuk model GWR dan GWQR disajikan pada Tabel 3. dari hasil tersebut dapat dilihat bahwa nilai RMSE dari kedua model dengan setiap ukuran sampel terhadap berbagai nilai kuantil yang terkontaminasi presentase *outlier* yang berbeda menghasilkan nilai yang variatif. Hasil analisis menunjukkan nilai RMSE model GWR tertinggi, hal ini membuktikan bahwa penyelesaian masalah *outlier* pada data heterogenitas spasial perlu dilakukan dengan model lokal regresi kuantil (GWQR) karena dengan adanya model regresi kuantil maka data akan dibagi dalam berbagai bagian

kuantil2 tertentu sehingga permasalahan *outlier* teratasi [17]. Model GWR merupakan model lokal dari regresi klasik yang dapat mengakomodir heterogenitas spasial pada data, akan tetapi tidak dapat mengakomodir permasalahan *outlier* sehingga sifat kebaikan model tidak terpenuhi yang berdampak pada tingginya nilai RMSE yang dihasilkan, hal ini disebabkan karena *outlier* pada data amatan yang diabaikan. [18]. Nilai kuantil 0.95 secara konsisten menghasilkan nilai RMSE terendah untuk setiap ukuran sampel terhadap presentase kontaminasi *outlier*.

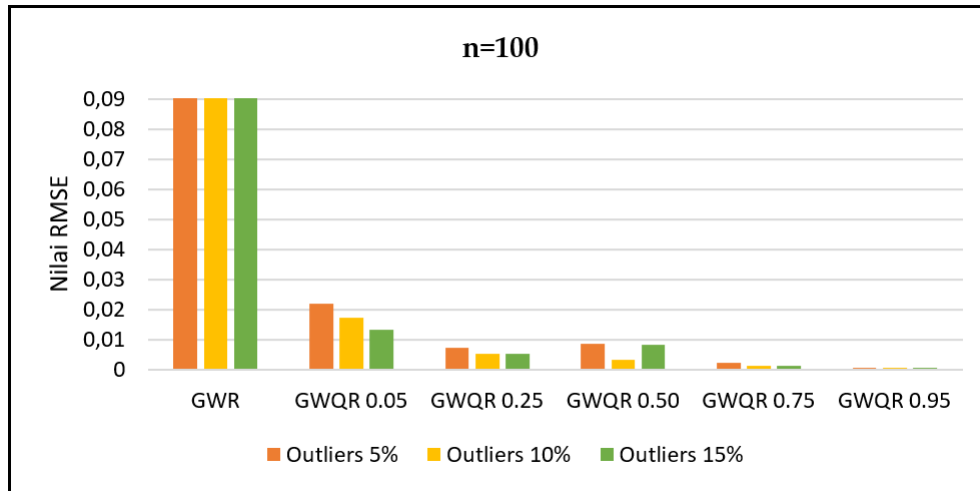
Tabel 3: Nilai RMSE

RMSE	Sample size	Method	5% Outliers	10% Outliers	15% Outliers
n=50		GWR	1.0023	1.0349	1.0753
		GWQR 0.25	0.0097	0.0670	0.0534
		GWQR 0.50	0.0651	0.0127	0.0112
		GWQR 0.75	0.0014	0.0027	0.0033
		GWQR 0.95	0.0009	0.0006	0.0008
	n=100		GWR	1.3422	1.5364
		GWQR 0.05	0.0219	0.0174	0.0134
		GWQR 0.25	0.0072	0.0053	0.0053
		GWQR 0.50	0.0086	0.0034	0.0082
		GWQR 0.75	0.0023	0.0014	0.0013
		GWQR 0.95	0.0007	0.0005	0.0003
n=150		GWR	1.6472	1.8353	1.9342
		GWQR 0.05	0.0125	0.0141	0.0106
		GWQR 0.25	0.0014	0.0018	0.0016
		GWQR 0.50	0.0021	0.0018	0.0017
		GWQR 0.75	0.0009	0.0013	0.0015
		GWQR 0.95	0.0001	0.0004	0.0002

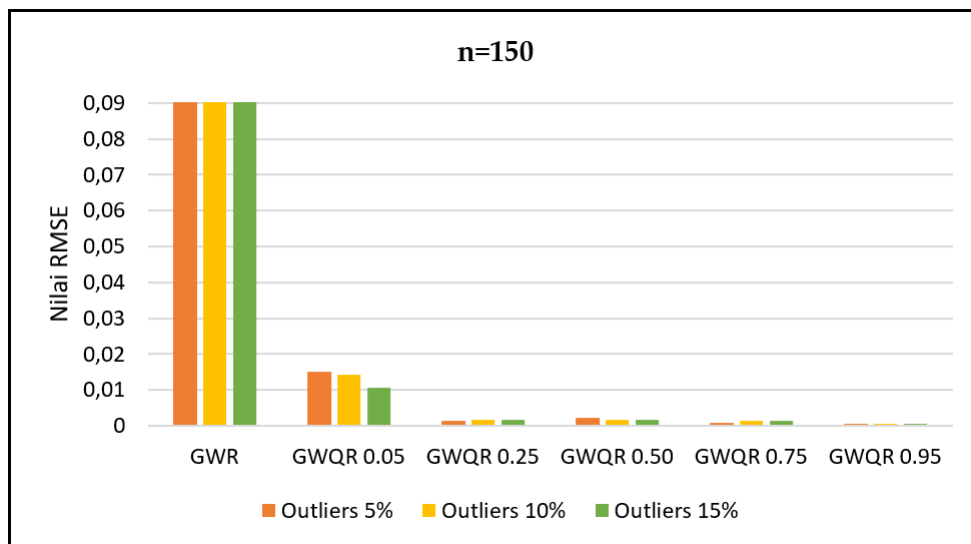
Visualisasi lebih lanjut untuk hasil pengujian kebaikan model GWR dan GWQR yang lebih komperensif, maka disajikan dalam bentuk grafik yang disajikan pada Gambar 3.5, 3.6, dan 3.7.

Gambar 3.5: Nilai RMSE Model GWR dan GWQR, Ukuran Sampel $n = 50$

Berdasarkan gambar 3.5, 3.6 dan 3.7, terlihat jelas bahwa nilai RMSE semakin berukuran kecil seiring dengan bertambahnya ukuran sampel. Selain itu, nilai RMSE dalam pengujian model GWQR



Gambar 3.6: Nilai RMSE Model GWR dan GWQR, Ukuran Sampel $n = 100$



Gambar 3.7: Nilai RMSE Model GWR dan GWQR, Ukuran Sampel $n = 150$

data bangkitan diperoleh nilai yang lebih kecil dari satu yang mendekati nol, sehingga model GWQR terbukti merupakan model terbaik dalam mengatasi permasalahan *outlier* terkhususnya pada data yang teridentifikasi spasial hingga kontaminasi *outlier* 15%. Dilihat dari nilai kuantil maka kuantil 0.95 merupakan kuantil yang memberikan nilai parameter terbaik dibandingkan dengan kuantil 0.05, 0.25, 0.50, dan 0.75.

Di sisi lain, model GWR memiliki nilai RMSE yang tinggi dibandingkan model GWQR. Hal ini menunjukkan bahwa dalam pemodelan dengan data heterogenitas spasial yang terkontaminasi *outlier* diperlukan analisis dengan model GWQR yang dapat mengakomodir permasalahan tersebut, sehingga model yang dihasilkan merupakan model terbaik yang dapat merepresentasikan kondisi aktual. Penelitian terdahulu telah dilakukan oleh Kurniawati (2019), dalam memodelkan IPM di Pulau Jawa pada tahun 2017 dan studi simulasi dengan data bangkitan untuk menganalisis efektivitas model GWQR yang dibandingkan dengan model regresi kuantil yang merupakan model global yang tidak dapat mengakomodir adanya heterogenitas spasial. Peneliti berfokus untuk menguji satu ukuran sampel saja ($n = 625$) dengan ukuran kontaminasi outlier adalah 5%, 10%, 15% dan 20%. Dari hasil penelitian diperoleh bahwa model GWQR merupakan model terbaik dalam

dalam penanganan *outlier* dibandingkan model regresi kuantil [19]. Perbedaan signifikan penelitian ini dengan penelitian terdahulu terletak pada model pembanding yang digunakan, ukuran sampel, dan presentase *outlier*. Penelitian ini berfokus pada model lokal regresi yang dapat menangkap heterogenitas spasial tetapi mengabaikan *outlier*. Hasil penelitian diperoleh model GWQR konsisten mengatasi permasalahan *outlier* yang dibuktikan dengan nilai RMSE yang semakin kecil seiring bertambahnya ukuran sampel hingga ukuran kontaminasi *outlier* sebesar 15%.

4. Kesimpulan

Model GWQR terbukti menyelesaikan permasalahan *outlier* pada data yang teridentifikasi heterogenitas spasial yang diuji pada ukuran sampel berbeda dengan nilai kuantil 0.05, 0.25, 0.50, 0.75, 0.95 hingga besaran kontaminasi *outlier* 15% jika dibandingkan dengan model *Geographically Weighted Regression* (GWR) yang tidak dapat mengakomodir permasalahan *outlier*. Nilai RMSE yang diperoleh kurang dari satu dan mendekati nol, sehingga terbukti model GWQR merupakan model yang tepat. Permasalahan *outlier* pada data penelitian yang teridentifikasi heterogenitas spasial dapat dimodelkan dengan model GWQR karena sifat kebaikan penduga akan terpenuhi sehingga model yang diperoleh merepresentasikan kondisi aktual.

Referensi

- [1] P. R. Sihombing, S. Suryadiningrat, D. A. Sunarjo, and Y. P. A. C. Yuda, "Identifikasi data outlier (pencilan) dan kenormalan data pada data univariat serta alternatif penyelesaiannya," *J. Ekon. Dan Stat. Indones.*, vol. 2, no. 3, pp. 307–316, 2023. [View Online](#).
- [2] S. Indra, D. Vionanda, and R. Sriningsih, "Pendeteksian data pencilan dan pengamatan berpengaruh pada beberapa kasus data menggunakan metode diagnostik," *J. Math. UNP*, vol. 1, no. 2, pp. 67–74, 2013. [View Online](#).
- [3] H. Aguinis, R. K. Gottfredson, and H. Joo, "Best-practice recommendations for defining, identifying, and handling outlier," *Organ. Res. Methods*, vol. 16, no. 2, pp. 270–301, 2013. [View Online](#).
- [4] R. Koenker and G. Bassett, "No title regression quantiles," *J. Chem. Inf. Model.*, vol. 53, no. 9, pp. 1689–1699, 2013. [View Online](#).
- [5] M. Y. Matdoan and A. M. Balami, "Estimasi parameter regresi kuantil dengan fungsi spline truncated pada kasus demam berdarah dengue di kota surabaya," *J. MSA (Mat. dan Stat. serta Apl.)*, vol. 7, no. 1, p. 44, 2019. [View Online](#).
- [6] N. Lutfiani and S. Mariani, "Pemodelan geographically weighted regression (gwr) dengan fungsi pembobot kernel gaussian dan bi-square," *UNNES J. Math.*, vol. 5, no. 1, pp. 82–91, 2019. [View Online](#).
- [7] W. S. Lestari, G. Pawitan, and M. Jaya, "Analisis data spasial menggunakan metode geographically weighted regression: Studi kasus data pdrb per kapita di provinsi jawa timur," *relax Semin. Nas. Stat. IV "Peranan Stat. di Bid. Eksplor. Energi Indones. Dep. Stat. FMIPA Univ. Padjajaran. Jatinangor, 14 Sept. 2014*, pp. 1–11, 2014. [View Online](#).
- [8] F. C. H. Buan, R. Fitriani, and Nurjannah, "Estimating gross regional domestic product (grdp) district / city in east nusa tenggara with spatial dynamic panel data," *Wacana, Journal Sosial dan Humaniora*, vol. 24, no. 4, pp. 190–194, 2021. [View Online](#).
- [9] A. Maulana, R. Meilawati, and V. Widiastuti, "Pemodelan indeks pembangunan manusia (ipm) metode baru menurut provinsi tahun 2015 menggunakan geographically weighted regression (gwr)," *Indones. J. Appl. Stat.*, vol. 2, no. 1, p. 21, 2019. [View Online](#).
- [10] V. Y. Chen, W. Avenue, and S. A. Matthews, "regression: An enhancement based on the bootstrap approach," *Journal of Physics: Conference Series*, vol. 52, no. 4, 2021. [View Online](#).

- [11] N. Kurniawati, H. Pramoedyo, and S. Astutik, "Geographically weighted quantile regression modelling on human development," *International Journal of Humanities, Religion and Social Science*, vol. 52, no. 4, pp. 642–661, 2021. [View Online](#).
- [12] V. Y. J. Chen, W. S. Deng, T. C. Yang, and S. A. Matthews, "Geographically weighted quantile regression (gwqr): An application to u.s. mortality data," *Geogr. Anal.*, vol. 44, no. 2, pp. 134–150, 2012. [View Online](#).
- [13] W. Wang, S. Xu, and T. Yan, "Structure identification and model selection in geographically weighted quantile regression models," *Spat. Stat.*, vol. 26, pp. 21–37, 2018. [View Online](#).
- [14] V. Subramaniam and S. Wasiuzzaman, "Geographical diversification, firm size and profitability in malaysia: A quantile regression approach," *Heliyon*, vol. 52, no. 4, pp. 642–661, 2021. [View Online](#).
- [15] N. Herawati, "The effectiveness of quantile regression in dealing with potential outlier," *BAREKENG J. Ilmu Mat. dan Terap.*, vol. 14, no. 2, pp. 301–308, 2020. [View Online](#).
- [16] T. Chai and R. R. Draxler, "Root mean square error (rmse) or mean absolute error (mae)? - arguments against avoiding rmse in the literature," *Geosci. Model Dev.*, vol. 7, no. 3, pp. 1247–1250, 2014. [View Online](#).
- [17] A. Anisa, A. Islamiyati, S. Sahriman, J. Massalesse, and B. Aprilia, "Model regresi kuantil spline orde dua dalam menganalisis perubahan trombosit pasien demam berdarah," *Jambura J. Math.*, vol. 5, no. 1, pp. 38–45, 2023. [View Online](#).
- [18] J. A. Yacim and D. G. B. Boshoff, "A comparison of bandwidth and kernel function selection in geographically weighted regression for house valuation," *Int. J. Technol.*, vol. 10, no. 1, pp. 58–68, 2019. [View Online](#).
- [19] N. Kurniawati, *Pemodelan Geographically Weighted Quantile Regression (Gwqr) Pada Data Indeks Pembangunan Manusia (Ipm) Di Pulau Jawa Tahun 2017 Dan Data Simulasi (Studi Kasus Dan Simulasi)*. Universitas Brawijaya, 2019. [View Online](#).

Format Sitasi IEEE:

F. C. H. Buan, Z. A. Banunaek, W. Reza, "Analisis Efektivitas Model *Geographically Weighted Quantile Regression* (GWQR) Dalam Penanganan Outlier: Data Simulasi Teridentifikasi Heterogenitas Spasial", *Jurnal Diferensial*, vol. 6(1), pp. 40-51, 2024.

This work is licensed under a [Creative Commons "Attribution-ShareAlike 4.0 International"](#) license.

