PENERAPAN METODE REGRESI KUANTIL PADA KASUS PELANGGARAN ASUMSI KENORMALAN SISAAN

Ferra Yanuar, Hazmira Yozza dan Izzati Rahmi

Jurusan Matematika Universitas Andalas
Email: ferrayanuar@yahoo.co.id, hyozza@gmail.com, izzatirahmihg@gmail.com,

ABSTRACT

Quantile regression is a regression method with the approach of separating or dividing the data into any particular quantiles. The method is by minimizing the absolute weighted residual asymmetric and estimate the quantile functions conditional on distribution of data. Quantile regression parameter estimation does not require parametric assumptions. This study aims to apply the quantile method for data that violates the assumption of residual normality. Small size of data were generated from various distribution which residual designed with chi squared distribution. This study resulted manyproposed models that divided over several quantiles selected. The values of regression coefficient estimation were close to the initial value. This study found that the proposed model was good enough and could be accepted.

Keywords: *Quantile Regression, Residual Normality, Regression Coefficient*

PENDAHULUAN

Analisis regresi adalah metode statistika yang bertujuan untuk memodel kan hubungan antara variabel tak bebas (Y) dengan satu atau lebihvariabel bebas (X) dalam suatu sistem. Hubungan antara variabel-variabel tersebut biasanya di nyatakan dalam suatu model regresi yang secara umum dinyatakan sebagai Y = f(x) + e, dengan e menyatakan komponen sisaan (*error*). Model tersebut menghubungkan variabel bebas dan variabel tak bebas melalui suatu parameter yang dinamakan sebagai parameter regresi dinotasikan dengan β.

Untuk menduga nilai parameter regresi ini biasanya digunakan metode kuadrat terkecil (MKT), dimana prinsip kerjanya adalah meminimumkan jumlah kuadrat sisaan nilai observasi terhadap rata-ratanya. Metode MKT ini dapat diterapkan jika beberapa asumsi terpenuhi, seperti asumsi kenormalan, non-multikolineritas, kehomogenan ragam sisaan dan non-auto korelasi. Semua asumsi harus terpenuhi supaya didapatkan penduga parameter yang bersifat BLUE (*Best Linier Unbiased Estimator*). Tetapi metode ini sensitif terhadap penyimpangan asumsi tersebut, misalnya data tidak memenuhi asumsi kenormalan, varians data tak homogen (heteroskedastisitas), terdapat masalah multikolinearitas, autkorelasi dan sebagai nya. Pada data yang memiliki masalah tersebut tidak bisa diterapkan metode MKT untuk menduga parameter modelnya.

Metode regresi median kemudian muncul untuk mengatasi kelemahan MKT, metode ini mengganti pendekatan rataan pada MKT menjadi median. Hal ini dilakukan dengan mempertimbangkan apabila data berbentuk lonceng atau tidak simetris. Tetapi pada kenyataannya, pendekatan regresi median juga dianggap kurang informatif karena regresi ini hanya melihat pada dua kelompok data. Padahal ada kemungkinan data bisa terbagi menjadi lebih dari dua kelompok, sehingga berkembanglah metode regresi kuantil (Wu & Liu, 2009; Yanuar, 2013).

Metode regresi kuantil ini merupa kan salah satu metode regresi dengan pendekatan memisahkan atau membagi data menjadi kuantil-kuantil tertentu, dengan meminimumkan sisaan mutlak berbobot yang tidak simetris dan menduga fungsi kuantil bersyarat pada suatu sebaran data. Artikel ini bertujuan menerangkan tentang penerapan metode regresi kuantil dalam menduga parameter model untuk data yang tidak memenuhi asumsi kenormalan sisaan dengan menggunakan data simulasi yang berukuran kecil.

METODE PENELITIAN

Data yang digunakan dalam kajian ini adalah data bangkitan dengan sisaannya tidak berdistribusi normal, sesuai dengan tujuan utama dari kajian ini yaitu menerapkan penggunakan regresi kuantil pada sekelompok data yang tidak memenuhi asumsi kenormalan sisaan.

Data kajian ini terdiri dari tiga variabel independen (X_1, X_2 dan X_3) dan satu variabel dependen (Y), dimana variabel independen (X_1, X_2 dan X_3) bersifat stokastik (tidak deterministik). Variabel independen, X_1 dan X_2 menyebar menurut sebaran t dengan derajat kebebasan 1, atau X_1 dan $X_2 \sim t_{(1)}$. Variabel independen, X_3 menyebar menurut sebaran normal baku atau $X_3 \sim N(0,1)$, sedangkan variabel dependen (Y) ditetapkan nilai Y = X_1 + X_2 + e, dengan e ~ $\chi^2_{(1)}$. Selanjutnya sebanyak 20 buah data dibangkitkan untuk setiap variabel. Untuk membangkitkan dan analisa data, kajian ini menggunakan bantuan software R.3.2.4.

Adapun metode analisis yang digunakan adalah regresi kuantil karena metode secara teorinya mampu mengatasi pelanggaran asumsi kenormalan (Davino *et al.*, 2014). Regresi kuantil merupakan teknik statistika yang digunakan untuk menduga hubungan antara peubah tak bebas dengan peubah penjelas pada fungsi kuantil bersyarat tertentu. Seperti pada metode kuadrat terkecil, yang meminimum kan jumlah kuadrat galat dan menduga model dengan menggunakan fungsi rata-rata bersyarat, regresi kuantil meminimum kan galat mutlak berbobot yang tidak simetris dan menduga fungsi kuantil ber syarat pada suatu sebaran data.

Suatu peubah acak Y dengan fungsi sebaran peluang $F\ (Y)\ = P(Y \leq y)$, di mana untuk setiap 0 < τ < 1, terdapat fungsi invers, $\mathrm{F}^{-1} = inv\{y: |F_y(y|X) \geq \tau\}$ yang merupakan kuantil ke-τ dari Y. Jika rata-rata contoh merupakan solusi dari masalah: $min_{\mu \in \mathbb{R}} \sum_{i=1}^{n} (y_\mathrm{i} - \mu)^2$ maka untuk $\mu(x_\mathrm{i}) = x_\mathrm{i}^{'}\beta$ yang merupakan rata-rata bersyarat dari y dengan x diketahui, nilai β dapat diduga dengan menyelesaikan:

$$min_{\beta \in \mathbb{R}} \sum_{i=1}^{n} (y_\mathrm{i} - x_\mathrm{i}^{'}\beta)^2$$

Selanjutnya model berkembang menjadi median contoh yang dinyatakan:

$$min_{\beta \in \mathbb{R}} \sum_{i=1}^{n} |y_\mathrm{i} - x_\mathrm{i}^{'}\beta|$$

Berdasarkan konsep median estimasi bagi β dari regresi kuantil ke-τ diperoleh dengan meminimumkan jumlah nilai mutlak dari galat dengan pembobot τ untuk galat positif dan pembobot (1 -τ) untuk galat negatif yaitu (Yanuar, 2013; Bentzien& Friederichs, 2014) :

$$min_{\beta \in \mathbb{R}} \sum_{i \epsilon i\,|y_i \geq x_\mathrm{i}^{'}\beta} \tau |y_\mathrm{i} - x_\mathrm{i}^{'}\beta| + min_{\beta \in \mathbb{R}} \sum_{i \epsilon i\,|y_i < x_\mathrm{i}^{'}\beta} (1 - \tau)|y_\mathrm{i} - x_\mathrm{i}^{'}\beta|$$

Atau dapat ditulis lagi menjadi :

$$min_{\beta \in \mathbb{R}} \sum_{i=1}^{n} \rho_\tau (y_\mathrm{i} - \mathrm{Q}_\tau(Y|X))$$

Dengan :
τ menyatakan indeks kuantil ∈(0, 1)
ρ_τ merupakan *loss function* yang asimetrik
$Q_\tau(Y|X) = x_i'\beta$, yaitu fungsi kuantil ke-τ dari *Y* dengan syarat *X*

Adapun indikator kebaikan model dilihat dari nilai R^2 (koefisien determinasi) yang dihasilkan untuk setiap kuantil (Furno, 2014).

HASIL DAN PEMBAHASAN

Hasil Penelitian

Pada bagian ini akan diuraikan hasil analisis data dengan menggunakan metode kuantil sebagaimana yang diuraikan pada bagian Data dan Metode di atas. Tabel 1 di bawah ini menampilkan hasil pendugaan parameter model, yaitu β_0, β_1, β_2 dan β_3 dengan menggunakan data hasil bangkitan untuk kuantil ke-**τ** = 0,05, 0,25, 0,50, 0,75 dan 0,95.

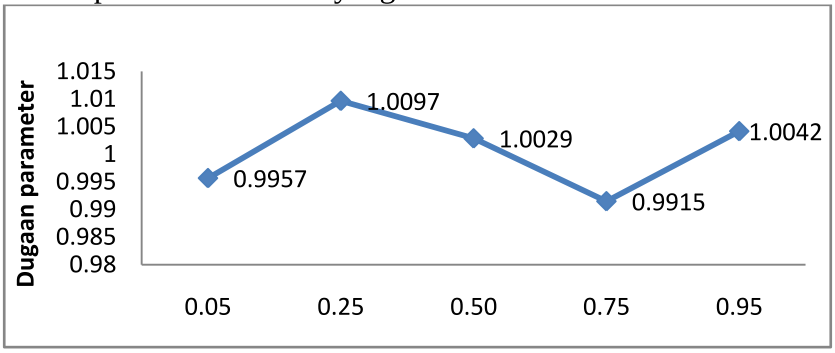
Tabel 1. Nilai Parameter Dugaan dengan Regresi Kuantil

Kuantil ke-τ	Parameter	Nilai Dugaan	Nilai *p*	R^2
0.05	β_0	0,0069	0,542	0,9439
	β_1	0,9957	0,000	
	β_2	0,9957	0,000	
	β_3	0,9639	0,000	
0.25	β_0	0,0465	0,305	0,923
	β_1	1,0097	0,000	
	β_2	0,9503	0,000	
	β_3	0,9741	0,000	
0.50	β_0	0,1304	0,017	0,8966
	β_1	1,0029	0,000	
	β_2	0,9612	0,000	
	β_3	1,0227	0,000	
0.75	β_0	0,1520	0,039	0,8639
	β_1	0,9915	0,000	
	β_2	0,8859	0,000	
	β_3	1,0045	0,000	
0.90	β_0	0,6079	0,000	0,821
	β_1	1,0042	0,000	
	β_2	0,8031	0,000	
	β_3	0,8940	0,015	

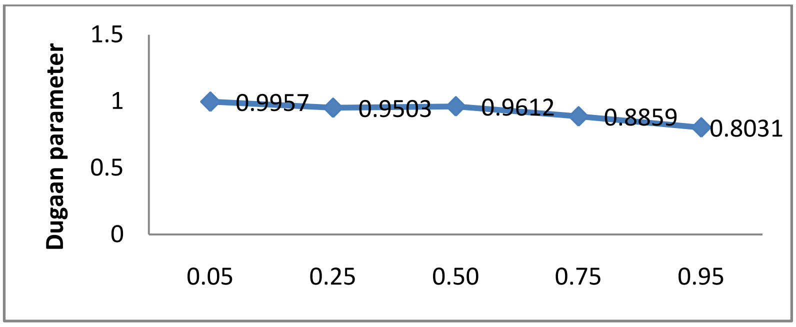
Berdasarkan tabel 1 diketahui bahwa semua nilai koefisien regresi untuk semua kuantil terpilih adalah signifikan pada taraf keyakinan 95% (karena nilai *p* kurang dari 0,05), nilai $\beta_0$tidak diperhitung kan. Hal ini berarti bahwa pada model dugaan yang diperoleh, ketiga variabel independen (X_1 , X_2 dan X_3)signifikan mempengaruhi variabel dependen (*Y*), pada semua nilai kuantil terpilih.

Tabel 1 juga menginformasi kan adalah bahwa nilai koefisien model (β_1,β_2

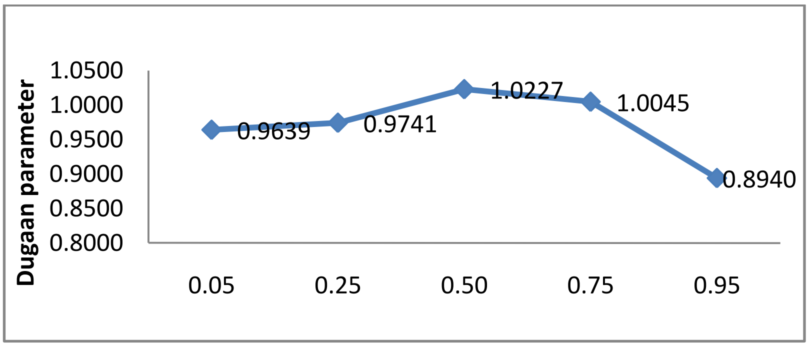
dan β_3) pada kuantil ke 0,05 masing-masing adalah 0,9957, 0,9957 dan 0,9639. Ketiga nilai tersebut telah cukup dekat dengan nilai awal yang ditetapkan yaitu 1 untuk ketiga nilai parameter. Hasil yang sama juga diperoleh pada kuantil ke 0,25; 0,50; 0,75 dan 0,95. Tren perubahan nilai koefisien regresi (β_1, β_2 dan β_3) untuk semua kuantil terpilih diilustrasikan pada Gambar 1.



Gambar a. Koefisien β_1



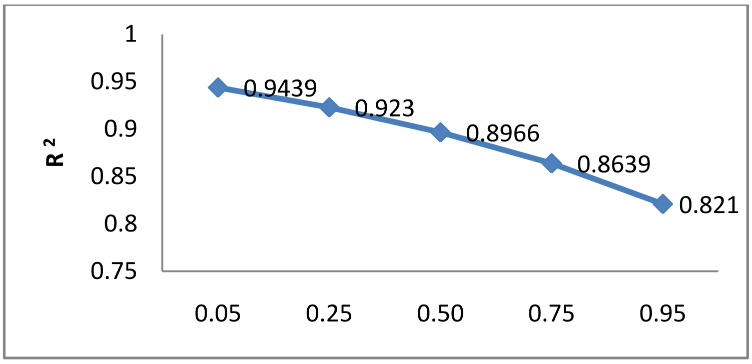
Gambar b. Koefisien β_2



Gambar c. Koefisien β_3

Gambar 1. Nilai Dugaan Koefisien Regresi untuk Setiap Kuantil:

Tabel 1 juga menginformasikan nilai R^2 untuk setiap model dugaan pada semua nilai kuantil terpilih, yaitu terdapat pada kolom paling kanan. Analisis ini meng hasilkan nilai R^2 cukup besar yaitu di atas 80%, dengan nilai R^2 terbesar adalah pada model dugaan pada kuantil ke-0,05. Sebagaimana diketahui bahwa nilai R^2ini juga menyatakan besarnya kebaikan suatu model dugaan. Dengan cara ini semua model dugaan yang dihasilkan dari metode kuantil ini sudah memenuhi kriteria kebaikan model. Gambar 2 meng ilustrasikan perubahan nilai R^2untuk setiap nilai kuantil yang diamati. Dapat terlihat bahwa ada kecendrungan penurunan nilai R^2 dengan bertambahnya nilai kuantil, tetapi penurunannya kecil.



Gambar 2. Nilai Dugaan R^2 untuk Setiap Kuantil.

KESIMPULAN

Artikel ini bertujuan menunjukkan penerapan metode regresi kuantil pada data yang melanggar asumsi kenormalan sisaan. Regresi kuantil merupakan metode penduga parameter regresi dengan pendekatan memisahkan atau membagi data menjadi kuantil-kuantil tertentu. Metode pendekatanyang dilakukan adalah dengan meminimumkan sisaan mutlak berbobot yang tidaksimetris dan menduga fungsi kuantil bersyarat pada sebaran data tersebut.

Data simulasi yang digunakan pada kajian ini dibangkitkan dari berbagai distribusi dengan sisaannya berdistribusi chi kuadrat. Kajian ini menghasilkan beberapa model dugaan yang dibagi-bagi atas beberapa kuantil terpilih. Koefisien regresi dugaan pada berbagai kuantil terpilih menghasilkan nilai dugaan hampir mendekati nilai awal yang ditetapkan. Nilai R^2 untuk setiap model dugaan juga cukup baik yaitu di atas 80%, artinya keragaman respon dapat diterangkan oleh model dugaan sebesar lebih dari 80%. Hal ini mengindikasikan bahwa model dugaan yang dihasilkan untuk setiap kuantil adalah cukup baik, dengan demikian model dugaan tersebut dapat diterima.

DAFTAR PUSTAKA

Bentzien, S & Friederichs,P. 2014. **Decomposition and graphical portrayal of the quantile score.** *Quarterly Journal of the Royal Meteorological Society*, 140 (683) : 1924–1934.

Davino, C., Furno, M. & Vistocco, D. 2014. **Quantile Regression Theory and Applications**. John Wiley & Sons, Ltd.

Furno M. 2011.**Goodness of Fit and Misspecification in Quantile Regression.** Journal Of Educational And Behavioral Statistics, 36 (1): 105-131

Wu, Y. & Liu, Y. 2009. **Variable Selection in Quantile Regression**. Statistica Sinica, 19: 801-817.

Yanuar, F. 2013. **Quantile Regression Approach to Determine the Indicator of Health Status.** Scientific Research Journal, 1 : 17 – 23.