

REGRESI KUANTIL MEDIAN UNTUK MENGATASI HETEROSKEDASTISITAS PADA ANALISIS REGRESI

IDA AYU PRASETYA UTHAMI¹, I KOMANG GDE SUKARSA²,
I PUTU EKA NILA KENCANA³

^{1, 2, 3} Jurusan Matematika FMIPA Universitas Udayana, Bukit Jimbaran-Bali
e-mail: ¹tamieemoetss@yahoo.co.id, ²sukarsakomang@yahoo.com
³i.putu.enk@gmail.com

Abstract

In regression analysis, the method used to estimate the parameters is Ordinary Least Squares (OLS). The principle of OLS is to minimize the sum of squares error. If any of the assumptions were not met, the results of the OLS estimates are no longer best, linear, and unbiased estimator (BLUE). One of the assumptions that must be met is the assumption about homoscedasticity, a condition in which the variance of the error is constant (same). Violation of the assumptions about homoscedasticity is referred to heteroscedasticity. When there exists heteroscedasticity, other regression techniques are needed, such as median quantile regression which is done by defining the median as a solution to minimize sum of absolute error. This study intended to estimate the regression parameters of the data were known to have heteroscedasticity. The secondary data were taken from the book Basic Econometrics (Gujarati, 2004) and analyzing method were performed by EViews 6. Parameter estimation of the median quantile regression were done by estimating the regression parameters at each quantile τ th, then an estimator was chosen on the median quantile as regression coefficients estimator. The result showed heteroscedasticity problem has been solved with median quantile regression although error still does not follow normal distribution properties with a value of R^2 about 71 percent. Therefore it can be concluded that median quantile regression can overcome heteroscedasticity but the data still abnormalities.

Keywords: Heteroscedasticity, Median Quantile Regression, EViews

1. Pendahuluan

Dalam analisis regresi, metode yang digunakan untuk melakukan estimasi parameter adalah Metode Kuadrat Terkecil (MKT). Prinsip dari metode ini adalah meminimumkan jumlah kuadrat *error* (residu). Namun, terdapat beberapa asumsi di dalam analisis regresi yang harus dipenuhi dalam melakukan estimasi dengan metode kuadrat terkecil. Asumsi tersebut antara lain, data harus mengikuti sebaran normal, homoskedastisitas, tidak ada multikolinearitas dan tidak ada autokorelasi. Apabila semua asumsi dapat terpenuhi maka hasil estimasi dengan

¹ Alumni Mahasiswa Jurusan Matematika FMIPA Universitas Udayana

^{2,3} Staf Pengajar Jurusan Matematika FMIPA Universitas Udayana

metode kuadrat terkecil dapat memenuhi sifat *Best, Linear, Unbiased, Estimator* (BLUE). Tetapi jika terdapat satu atau lebih asumsi yang tidak terpenuhi, maka hasil estimasi yang diperoleh tidak dapat memenuhi sifat BLUE.

Salah satu asumsi penting yang harus dipenuhi adalah asumsi homoskedastisitas (*homoscedasticity*). Homoskedastisitas berarti bahwa varian dari *error* bersifat konstan [4]. Asumsi ini menyatakan peubah respon memiliki varian yang sama sepanjang nilai peubah bebas, pelanggaran terhadap asumsi homoskedastisitas disebut dengan heteroskedastisitas (*heteroscedasticity*). Heteroskedastisitas umumnya terjadi pada data *cross section*, yaitu data yang diambil pada satu waktu, yang mewakili berbagai ukuran (kecil, sedang, dan besar).

Pada saat melakukan estimasi dengan metode kuadrat terkecil kemudian terjadi heteroskedastisitas, maka hasil estimasi yang diperoleh tidak lagi memenuhi sifat BLUE. Sehingga diperlukan metode alternatif lain dalam estimasi parameter yang dapat mengatasi adanya data dengan varian *error* yang tidak konstan atau mengandung heteroskedastisitas. Salah satu metode estimasi parameter yang bisa mengatasi heteroskedastisitas adalah metode regresi kuantil median. Regresi ini bertujuan untuk memperluas ide-ide dalam estimasi model fungsi kuantil bersyarat, dimana distribusi kuantil bersyarat dari peubah respon dinyatakan sebagai fungsi dari kovariat yang diamati [1]. Analisis ini sangat berguna untuk sebaran bersyarat yang tak simetrik, padat di ekor sebarannya, atau sebarannya terpotong.

Berdasarkan latar belakang tersebut, maka penulis ingin mengetahui bagaimanakah kemampuan regresi kuantil median mengatasi heteroskedastisitas dalam melakukan estimasi parameter pada analisis regresi.

Regresi Kuantil Median

Regresi kuantil median pertama kali diperkenalkan oleh Roger Koenker dan Gilbert Basset tahun 1978 [2]. Kuantil dapat dioperasikan dengan penyusunan atau pengurutan sampel pengamatan sehingga lebih mudah menentukan letaknya dan dapat mendefinisikan kuantil melalui alternatif yang sederhana sebagai masalah optimisasi. Sama seperti metode kuadrat terkecil yang dapat mendefinisikan rata-rata sampel sebagai solusi untuk masalah me-minimumkan jumlah kuadrat *error*, regresi kuantil median dapat mendefinisikan median sebagai solusi untuk masalah meminimumkan jumlah absolut *error*. Simetri dari nilai fungsi absolut linier mengimplikasikan bahwa minimum jumlah absolut *error* harus sama dengan jumlah *error* yang positif dan negatif, sehingga menjamin bahwa terdapat jumlah pengamatan yang sama di atas dan di bawah median. Karena simetri nilai absolut menghasilkan median, diharapkan dengan meminimumkan jumlah absolut *error* terboboti yang tidak simetri, secara sederhana memberikan perbedaan bobot untuk *error* positif dan negatif akan menghasilkan kuantil [3].

$$\min_{\xi \in \mathbb{R}} \sum \rho_{\tau}(y_i - \xi) \quad (1)$$

dimana fungsi $\rho_{\tau}(\cdot)$ merupakan fungsi absolut yang menghasilkan sampel kuantil ke- τ sebagai solusinya.

Untuk mendefinisikan regresi kuantil median dalam sebuah analogi, metode kuadrat terkecil menawarkan sebuah model dengan sampel acak $\{y_1, y_2, y_3, \dots, y_n\}$ diperoleh persamaan berikut:

$$\min_{\mu \in \mathbb{R}} \sum_{i=1}^n (y_i - \mu)^2 \quad (2)$$

maka diperoleh rata-rata sampel yang merupakan pengestimasi dari rata-rata populasi tidak bersyarat EY. Jika skalar $\mu = X' \beta$ maka persamaan (2) menjadi:

$$\min_{\beta \in \mathbb{R}^p} \sum_{i=1}^n (y_i - X_i' \beta)^2 \quad (3)$$

kemudian berkembang menjadi median contoh yang dinyatakan pada persamaan (4) sebagai berikut:

$$\min_{\beta \in \mathbb{R}^p} \sum_{i=1}^n |y_i - X_i' \beta| \quad (4)$$

secara umum dispesifikasikan dalam fungsi kuantil bersyarat ke- τ dengan mempertimbangkan penduga bagi $\beta(\tau)$ yaitu $(\hat{\beta}(\tau))$ sehingga diperoleh ide bahwa masalah tersebut dapat dinyatakan dengan,

$$\min_{\beta \in \mathbb{R}^p} \sum_{i=1}^n \rho_{\tau}(y_i - Q_{\tau}(Y|X)) \quad (5)$$

dengan:

τ = indeks kuantil $\in (0,1)$

ρ_{τ} = loss function yang asimetrik

$Q_{\tau}(Y|X) = X' \beta(\tau)$ = fungsi kuantil ke- τ dari Y dengan syarat X

Sehingga untuk meminimalkan kuantil tak bersyarat dengan estimasi regresi kuantil median atau yang ditunjukkan pada persamaan (6):

$$\hat{\beta}(\tau) = \arg \min_{(\tau)} \{ \sum_i \rho_{\tau}(Y_i - X_i' \beta(\tau)) \} \quad (6)$$

2. Metode Penelitian

Data yang digunakan dalam penelitian ini adalah data yang diperoleh dari buku “*Basic Econometrics*” (Gujarati, 2004). Data yang terdapat di dalam buku ini adalah data mengenai *Passenger Car Milage* dengan jumlah pengamatan sebanyak 81 buah pengamatan. Peubah respon (Y) yang digunakan adalah *Average Miles per Gallon (MPG)*, kemudian peubah bebasnya adalah *Top Speed (SP)* untuk (X_1) , *Engine Horsepower (HP)* untuk (X_2) , *Cubic Feet of Cupe Space (VOL)* untuk (X_3) , *Vehicle Weight (WT)* untuk (X_4) . Pemeriksaan heteroskedastisitas dilakukan dengan menggunakan bantuan program SPSS 17.0 dan Minitab 14, sedangkan untuk melakukan estimasi dengan Regresi Kuantil Median menggunakan program EViews 6.

Penelitian diawali dengan melakukan estimasi dengan analisis metode kuadrat terkecil (MKT). Langkah selanjutnya adalah melakukan pengujian terhadap asumsi kenormalan, dengan Uji *Anderson-Darling*. Setelah dilakukan pengujian terhadap asumsi kenormalan kemudian mendeteksi adanya heteroskedastisitas dengan melihat nilai *P-value* dari hasil pengujian dengan Uji

Levene. Apabila ditemukan adanya heteroskedastisitas pada data, maka estimasi parameter dilakukan dengan *Regresi Kuantil Median*. Langkah pertama dalam estimasi dengan regresi kuantil median adalah melakukan estimasi pada setiap kuantil ke- τ untuk masing-masing koefisien regresi. Sehingga diperoleh nilai dugaan untuk setiap koefisien regresi sebanyak k -kuantil, dimana k adalah banyaknya kuantil. Langkah selanjutnya adalah memilih nilai dugaan pada median kuantil sebagai hasil estimasi untuk setiap koefisien regresinya. Setelah hasil estimasi dengan regresi kuantil median, selanjutnya dilakukan pengujian kembali terhadap asumsi kenormalan dan asumsi homoskedastisitas. Pengujian terhadap asumsi kenormalan dilakukan dengan menggunakan uji *Anderson Darling*, sedangkan untuk melihat adanya heteroskedastisitas pada *error* hasil estimasi dengan regresi kuantil median dilakukan dengan uji Park.

3. Hasil dan Pembahasan

Melakukan estimasi dengan Metode Kuadrat Terkecil (MKT)

Hasil estimasi dari metode kuadrat terkecil diperoleh model $MPG_i = 181,87 - 1,18SP_i + 0,36HP_i - 0,01VOL_i - 1,81WT_i$ dengan nilai R^2 sebesar 0,88. Nilai R^2 tersebut mengindikasikan bahwa peubah-peubah bebas yang digunakan dapat menjelaskan peubah respon (Y) sebesar 88%. Namun untuk mengetahui bahwa hasil estimasi tersebut mengandung heteroskedastisitas atau tidak maka selanjutnya akan dilakukan pengujian asumsi klasik terhadap asumsi kenormalan dan asumsi homoskedastisitas. Jika salah satu asumsi saja tidak terpenuhi maka hasil estimasi dengan MKT tidak dapat lagi dipercaya.

Pengujian Asumsi Klasik terhadap *Error* Hasil Estimasi dengan MKT

Hasil yang diperoleh dari pengujian terhadap asumsi kenormalan adalah *error* tidak mengikuti sebaran normal dengan nilai *P-value* sebesar 0,006 yang lebih kecil dari α sebesar 0,05. Kemudian untuk hasil dari pengujian terhadap asumsi homoskedastisitas adalah data mengandung heteroskedastisitas atau varian dari *error* tidak konstan dengan nilai *P-value* sebesar 0,000 yang lebih kecil dari α . Selain itu plot antara *error* yang terstandarisasi dengan nilai dugaannya juga mengindikasikan adanya heteroskedastisitas, dimana plot yang dihasilkan menunjukkan pola sebaran yang meningkat yaitu secara terus-menerus bergerak menjauhi nol. Berdasarkan hasil pengujian asumsi klasik tersebut kedua asumsi tidak terpenuhi, maka hasil estimasi dengan MKT tidak lagi dapat dipercaya, sehingga selanjutnya akan dilakukan estimasi dengan regresi kuantil median.

Estimasi dengan Regresi Kuantil Median

Sama seperti MKT yang mendefinisikan rata-rata sampel sebagai solusi untuk meminimumkan jumlah kuadrat *error*, regresi kuantil median mendefinisikan median sebagai solusi untuk meminimumkan jumlah absolut

error. Langkah pertama dalam estimasi dengan regresi kuantil median adalah melakukan estimasi pada setiap kuantil ke- τ untuk masing-masing koefisien regresi. Sehingga diperoleh nilai dugaan untuk setiap koefisien regresi sebanyak k -kuantil, dimana k adalah banyaknya kuantil. Selanjutnya memilih nilai dugaan pada median kuantil sebagai nilai dari koefisien regresi dugaannya. Hasil estimasi dengan regresi kuantil median ditunjukkan pada tabel 1, 2, dan 3.

Tabel 1 Hasil Estimasi Koefisien β_0 pada setiap Kuantilnya

Peubah	Kuantil	Koefisien	Standar Error	t-Statistik	P-value
C	0,05	58,8147	23,1336	2,5424	0,0130**
	0,10	75,0954	24,517	3,063	0,0030**
	0,15	76,8604	27,8753	2,7573	0,0073**
	0,20	66,5982	26,9787	2,4685	0,0158**
	0,25	66,6291	26,1761	2,5454	0,0129**
	0,30	81,3593	25,5973	3,1784	0,0021**
	0,35	83,3175	26,2327	3,1761	0,0022**
	0,40	88,6259	24,5033	3,6169	0,0005**
	0,45	83,3664	25,1979	3,3085	0,0014**
	0,50	105,3749	25,6544	4,1075	0,0001**
	0,55	110,6154	28,3546	3,9011	0,0002**
	0,60	116,8092	28,8408	4,0501	0,0001**
	0,65	157,5947	77,8709	2,0238	0,0465**
	0,70	207,3845	107,7401	1,9249	0,0580 ^{ns}
	0,75	233,6692	51,4708	4,5398	0,0000**
	0,80	269,4404	44,8173	6,012	0,0000**
	0,85	267,9866	49,861	5,3747	0,0000**
0,90	287,5208	43,5911	6,5959	0,0000**	
0,95	340,5571	48,1111	7,0786	0,0000**	

Pada tabel 1 hasil pengujian koefisien regresi dugaan untuk β_0 menunjukkan hampir semua koefisien-koefisien regresi dugaan signifikan, karena nilai P-value yang lebih kecil dari α . Terdapat satu koefisien regresi dugaan yang tidak signifikan, yaitu koefisien β_0 pada kuantil 0,700 dengan nilai P-value 0,0580. Kemudian nilai koefisien dugaan yang dipilih adalah nilai koefisien dugaan pada median kuantil yaitu 105,3749 dengan nilai P-value 0,0001 yang signifikan pada α sebesar 0,05.

Tabel 2 Hasil Estimasi Koefisien β_1 pada setiap Kuantilnya

Peubah	Kuantil	Koefisien	Standar Error	t-Statistik	P-value
X ₁	0,05	-0,0022	0,2321	-0,0093	0,9926 ^{ns}
	0,10	-0,161	0,2526	-0,6375	0,5257 ^{ns}
	0,15	-0,1786	0,2879	-0,6204	0,5369 ^{ns}
	0,20	-0,0692	0,2746	-0,2521	0,8016 ^{ns}
	0,25	-0,07	0,2658	-0,2632	0,7931 ^{ns}
	0,30	-0,212	0,2611	-0,8121	0,4193 ^{ns}
	0,35	-0,2303	0,266	-0,8659	0,3893 ^{ns}
	0,40	-0,274	0,2484	-1,1029	0,2736 ^{ns}
	0,45	-0,2137	0,2548	-0,8387	0,4042 ^{ns}
	0,50	-0,4361	0,2603	-1,6752	0,0980 ^{ns}
	0,55	-0,4829	0,2855	-1,6913	0,0949 ^{ns}
	0,60	-0,5457	0,2899	-1,8822	0,0636 ^{ns}
	0,65	-0,9482	0,7703	-1,231	0,2221 ^{ns}
	0,70	-1,4207	1,0628	-1,3367	0,1853 ^{ns}
	0,75	-1,6796	0,5131	-3,2733	0,0016**
	0,80	-2,0293	0,4514	-4,4956	0,0000**
	0,85	-2,0124	0,5122	-3,9286	0,0002**
0,90	-2,2194	0,4469	-4,9657	0,0000**	
0,95	-2,7589	0,4968	-5,5532	0,0000**	

Hasil estimasi terhadap koefisien β_1 pada tabel 2 ditunjukkan hanya terdapat lima koefisien dugaan yang signifikan, yaitu pada kuantil 0,750 sampai dengan kuantil 0,950. Koefisien regresi dugaan yang tidak signifikan terdapat 14 koefisien, ditunjukkan pada kuantil 0,050 sampai dengan 0,700. Median kuantil yang ditunjukkan pada kuantil 0,500 menunjukkan nilai koefisien regresi dugaan untuk β_1 sebesar (-0,4361), nilai tersebut tidak signifikan dengan nilai *P-value* 0,0980.

Tabel 3 Hasil Estimasi Koefisien β_2 pada setiap Kuantilnya

Peubah	Kuantil	Koefisien	Standar Error	t-Statistik	P-value
X ₂	0,05	-0,0075	0,0733	-0,1021	0,9190 ^{ns}
	0,10	0,036	0,0762	0,4719	0,6383 ^{ns}
	0,15	0,0424	0,0865	0,4902	0,6254 ^{ns}
	0,20	0,0088	0,0847	0,1037	0,9177 ^{ns}
	0,25	0,0146	0,0835	0,1743	0,8621 ^{ns}
	0,30	0,0612	0,0816	0,75	0,4556 ^{ns}
	0,35	0,0728	0,0851	0,8546	0,3955 ^{ns}
	0,40	0,0884	0,0798	1,1083	0,2712 ^{ns}
	0,45	0,0708	0,0817	0,866	0,3892 ^{ns}
	0,50	0,1319	0,0815	1,6171	0,1100 ^{ns}
	0,55	0,1377	0,0887	1,5526	0,1247 ^{ns}
	0,60	0,1561	0,0897	1,7405	0,0858 ^{ns}
	0,65	0,3021	0,2642	1,1436	0,2564 ^{ns}
	0,70	0,4537	0,3642	1,2457	0,2167 ^{ns}
	0,75	0,5344	0,1845	2,8967	0,0049**
	0,80	0,6487	0,1642	3,9511	0,0002**
	0,85	0,6336	0,1828	3,4661	0,0009**
0,90	0,7094	0,1588	4,4657	0,0000**	
0,95	0,8799	0,1785	4,9292	0,0000**	

Jumlah koefisien yang signifikan pada koefisien β_2 sama dengan jumlah koefisien yang signifikan pada koefisien β_1 , dimana hanya terdapat lima koefisien dugaan yang signifikan. Koefisien tersebut signifikan pada kuantil 0,75 sampai dengan 0,95 dengan nilai *P-value* yang lebih kecil dari α . Kemudian pada median kuantil diperoleh nilai koefisien regresi dugaan untuk β_2 sebesar 0,1319 dengan *P-value* sebesar 0,1100 tidak signifikan pada α sebesar 0,05.

Hasil estimasi untuk koefisien β_3 diperoleh nilai dugaan pada median kuantilnya sebesar 0,0108 dengan nilai *P-value* sebesar 0,5564, nilai tersebut tidak signifikan pada α sebesar 0,05. Kemudian hasil estimasi untuk koefisien β_4 diperoleh sebesar (-1,2959) signifikan dengan nilai *P-value* 0,000. Model yang diperoleh dari hasil estimasi dengan regresi kuantil median adalah:

$$MPG_i = 105,37 - 0,44SP_i + 0,13HP_i + 0,01VOL_i - 1,30WT_i$$

dengan nilai R_2 sebesar 0,71, nilai tersebut mengindikasikan bahwa peubah bebas yang digunakan dapat menjelaskan peubah respon sebesar 71%.

Pengujian Asumsi Klasik terhadap Error Hasil Estimasi dengan Regresi Kuantil Median

Sama seperti MKT, setelah melakukan estimasi dengan regresi kuantil median selanjutnya akan dilakukan pengujian terhadap asumsi kenormalan dan asumsi homoskedastisitas. Pengujian asumsi kenormalan dilakukan dengan uji *Anderson Darling*, dimana hasil yang diperoleh menunjukkan bahwa *error* hasil estimasi dengan regresi kuantil median tidak mengikuti sebaran normal. Kemudian untuk melihat apakah *error* yang dihasilkan memiliki varian yang sama atau tidak dapat diketahui dengan menggunakan uji Park yaitu dengan meregresikan logaritma natural dari *error* kuadrat dengan logaritma natural dari masing-masing peubah bebasnya. Hasil uji Park disajikan dalam tabel 4:

Tabel 4. Hasil uji Park

Regresi	P-value	$\alpha = 0,05$	Kesimpulan
$\text{Ln}(\varepsilon^2)$ dengan $\text{Ln}(X_1)$	0,088	$\text{P-value} > \alpha$	Homoskedastisitas
$\text{Ln}(\varepsilon^2)$ dengan $\text{Ln}(X_2)$	0,200	$\text{P-value} > \alpha$	Homoskedastisitas
$\text{Ln}(\varepsilon^2)$ dengan $\text{Ln}(X_3)$	0,808	$\text{P-value} > \alpha$	Homoskedastisitas
$\text{Ln}(\varepsilon^2)$ dengan $\text{Ln}(X_4)$	0,901	$\text{P-value} > \alpha$	Homoskedastisitas

4. Kesimpulan dan Saran

Berdasarkan hasil penelitian mengenai kemampuan regresi kuantil median dalam mengatasi heteroskedastisitas pada analisis regresi diperoleh kesimpulan bahwa hasil estimasi dengan MKT tidak memenuhi asumsi klasik, sehingga dilakukan estimasi dengan regresi kuantil median. Estimasi dengan regresi kuantil median dilakukan dengan mengestimasi koefisien regresi pada setiap kuantilnya, kemudian dipilih nilai dugaan pada median kuantil sebagai hasil estimasi. Hasil estimasi dengan regresi kuantil median diperoleh nilai R^2 sebesar 71%. Setelah dilakukan pengujian terhadap asumsi kenormalan diketahui *error* tidak mengikuti sebaran normal, tetapi asumsi homoskedastisitas dapat terpenuhi. Berdasarkan hasil tersebut, maka regresi kuantil median terbukti dapat mengatasi heteroskedastisitas.

Adapun saran pada penelitian selanjutnya adalah membangkitkan data dengan varian *error* yang tidak konstan untuk dapat melihat lebih jelas kemampuan analisis regresi kuantil median dalam mengatasi heteroskedastisitas. Selain itu penelitian selanjutnya dapat dilakukan dengan membandingkan metode regresi kuantil median dengan metode lain yang bisa mengatasi heteroskedastisitas, seperti *Generalized Least Square*.

Daftar Pustaka

- [1] John, O. O. and Nduka, E.C. 2009. "Quantile Regression Analysis as a Robust Alternative to Ordinary Least Squares", *The Journal of Scientia Africana*, Vol. 8, No. 2, pp. 61-65 published by: University of Port Harcourt.
- [2] Koenker, R. and Basset, G. 1978. "Regression Quantile", *Econometrica*. January, 46:1, pp. 33-50
- [3] Koenker, R. and Hallock, K.F. 2001. "Quantile Regression", *The Journal of Economic Perspectives*, Vol. 15, No. 4, pp. 143-156 Published by: American Economic Association.
- [4] Sarwoko. 2005. *Dasar-Dasar Ekonometrika*. Yogyakarta: Penerbit Andi Yogyakarta