Jurnal Matematika UNAND
Vol. **VIII** No. **1** Hal. 291 – 298
Edisi Mei 2019
ISSN : 2303–291X
©Jurusan Matematika FMIPA UNAND

PERBANDINGAN METODE REGRESI KUANTIL DAN METODE BAYES DALAM MENGESTIMASI PARAMETER MODEL REGRESI LINIER SEDERHANA DENGAN GALAT HETEROSKEDASTISITAS

RIZKY EFFENDI, MAIYASTRI, RITA DIANA
Program Studi S1 Matematika,
Fakultas Matematika dan Ilmu Pengetahuan Alam, Universitas Andalas,
Kampus UNAND Limau Manis Padang, Indonesia.
email : rizkyeffendi.math@gmail.com

Diterima 9 Maret 2019 Direvisi 7 April 2019 Dipublikasikan 7 Mei 2019

Abstrak. Salah satu metode estimasi parameter yang paling sering digunakan adalah Metode Kuadrat Terkecil (MKT). Ada beberapa asumsi-asumsi yang harus dipenuhi agar estimasi parameter dikatakan baik. Salah satu asumsi yang harus dipenuhi dalam estimasi MKT, yaitu galatnya mempunyai varian konstan (homoskedastisitas). Pelanggaran terhadap asumsi ini mengakibatkan varian menjadi tidak konstan (heteroskedastisitas) sehingga varian dari estimasi MKT yang diperoleh menjadi tidak efisien. Dengan demikian, diperlukan metode alternatif untuk mengatasi masalah heteroskedastisitas. Metode alternatif yang dapat digunakan dalam mengestimasi parameter dengan kasus heteroskedastisitas adalah Metode Regresi Kuantil dan Metode Bayes. Penelitian ini bertujuan untuk memperoleh nilai absolut bias dan MSE yang kecil. Hasil perbandingan estimasi parameter menunjukkan bahwa Metode Regresi Kuantil memiliki nilai absolut bias dan MSE yang kecil daripada Metode Bayes.

Kata Kunci: Heteroskedastisitas, Metode Regresi Kuantil, Metode Bayes

1. Pendahuluan

Dalam analisis regresi diperlukan suatu metode untuk mengestimasi parameter regresi, salah satu metode estimasi yang paling sering digunakan adalah Metode Kuadrat Terkecil (MKT). Ada beberapa asumsi yang harus dipenuhi agar estimasi parameter dikatakan baik. Salah satu asumsi yang harus dipenuhi dalam estimasi MKT, yaitu galatnya mempunyai varian yang konstan (homoskedastisitas). Pelanggaran terhadap asumsi homoskedastisitas disebut heteroskedastisitas, yang artinya varian galat bersifat tidak konstan. Konsekuensi terjadinya heteroskedastisitas dapat mengakibatkan varian dari estimasi MKT yang diperoleh menjadi tidak efisien. Dengan demikian, diperlukan metode alternatif untuk mengatasi masalah heteroskedastisitas.

Metode regresi kuantil median kemudian muncul untuk mengatasi masalah heteroskedastisitas pada MKT. Metode regresi kuantil median mendefinisikan median sebagai solusi untuk masalah meminimumkan jumlah sisa mutlak [2]. Pada kenyataannya, pendekatan regresi kuantil 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. Selain Metode Regresi Kuantil, ada juga metode alternatif yang dapat mengatasi heteroskedastisitas yaitu Metode Bayes. Dalam Metode Bayes, seluruh parameter yang tidak diketahui dipandang sebagai variabel acak yang memiliki distribusi awal yang disebut distribusi *prior* [1]. Setelah pengamatan dilakukan, informasi dalam distribusi *prior* dikombinasikan dengan informasi dari data sampel melalui teori Bayes dan hasilnya dinyatakan sebagai distribusi *posterior* yang akan menentukan estimasi parameter.

2. Kajian Pustaka

2.1. *Heteroskedastisitas*

Salah satu asumsi regresi linier yang harus dipenuhi adalah homogenitas varian dari galat (homoskedastisitas). Homoskedastisitas berarti bahwa varian dari galat model regresi bersifat konstan (tetap). Jika terjadi pelanggaran terhadap asumsi homoskedastisitas, maka model regresinya berada dalam kondisi heteroskedastisitas. Heteroskedastisitas adalah kondisi dimana varian dari galat model regresi tidak konstan atau varian dari galat model regresi berbeda.

2.2. *Mendeteksi Heteroskedastisitas*

Untuk menguji apakah terjadi heteroskedastisitas pada suatu data, maka dapat dilakukan beberapa uji statistik. Salah satunya adalah uji *Breusch Pagan Godfrey* (BPG). Adapun statistik uji *Breusch Pagan Godfrey* dapat dituliskan sebagai berikut :

$$BPG = \frac{1}{2}\left(\sum_{i=1}^{n} x_i f_i\right)^T \left(\sum_{i=1}^{n} x_i x_i^T\right)\left(\sum_{i=1}^{n} x_i f_i\right)$$

Uji ini menyebar χ^2_{k-1}, dimana k adalah banyaknya parameter regresi. Jika nilai $BPG > \chi^2_{k-1}$, maka tolak H_0 artinya heteroskedastisitas. Dan sebaliknya jika $BPG < \chi^2_{k-1}$, maka tidak tolak H_0.

2.3. *Estimasi Parameter Model dengan Metode Regresi Kuantil*

Regresi kuantil merupakan suatu pendekatan dalam analisis regresi yang dikenalkan oleh Roger Koenker dan Gilbert Basset. Persamaan regresi kuantil ke-θ, dapat ditulis dengan bentuk berikut [2] :

$$y_i = x_i^T \beta_\theta + \varepsilon_i,\ dengan\ i = 1, 2, ..., n, 0 < \theta < 1$$

Secara umum, estimasi parameter dengan menggunakan Metode Regresi Kuantil dapat dituliskan dalam bentuk persamaan berikut ini [4] :

$$\hat{\beta}(\theta) = min \sum_{i=1}^{n} \rho_\theta \left(y_i - x_i^T \beta_\theta\right), 0 < \theta < 1$$

dimana $\rho_\theta(y_i - x_i^T\beta_\theta)$ merupakan *loss function asymmetrics* yang dapat dinyatakan sebagai berikut ini [4] :

$$\rho_\theta(y_i - x_i^T\beta_\theta) = [\theta - I(y_i - x_i^T\beta_\theta < 0)](y_i - x_i^T\beta_\theta)$$

2.4. *Estimasi Parameter Model dengan Metode Bayes*

Jika $y_i \sim N(\beta_0 + \beta_1 x_i, \sigma_i^2)$ dan $\sigma_i^2 = \dfrac{1}{\tau_i}$ adalah parameter presisi, maka fungsi *likelihood* dari y dinyatakan sebagai berikut :

$$\begin{aligned} f(y|\beta_0,\beta_1,\tau_i) &= \prod_{i=1}^{n}(y;\beta_0,\beta_1,\tau_i) \\ &= \prod_{i=1}^{n}\left(\frac{\sqrt{\tau_i}}{\sqrt{2\pi}} exp\left\{-\frac{\tau_i}{2}(y_i-\beta_0-\beta_1 x_i)^2\right\}\right) \\ &= \left(\frac{\sqrt{\tau_i}}{\sqrt{2\pi}}\right)^n exp\left\{-\frac{1}{2}\sum_{i=1}^{n}\tau_i(y_i-\beta_0-\beta_1 x_i)^2\right\} \\ &\propto \left(\sqrt{\tau_i}\right)^n exp\left\{-\frac{1}{2}\sum_{i=1}^{n}\tau_i(y_i-\beta_0-\beta_1 x_i)^2\right\} \end{aligned}$$

Distribusi *prior* yang digunakan adalah distribusi *prior* konjugat. *Prior* konjugat adalah *prior* yang ditentukan berdasarkan pada pola *likelihood* data [1]. Adapun distribusi *prior* yang digunakan adalah sebagai berikut :

a. $\beta_0 \sim N(\mu_{\beta_0}, \sigma_{\beta_0}^2)$
b. $\beta_1 \sim N(\mu_{\beta_1}, \sigma_{\beta_1}^2)$
c. $\tau_i \sim gamma(a,b)$

Kombinasi antara *likelihood* data dan *prior* gabungan akan membentuk distribusi *posterior* gabungan dari seluruh parameter yang akan diestimasi. Distribusi *posterior* gabungan dapat dinyatakan dalam bentuk proporsional sebagai berikut :

$$\begin{aligned} f(\beta_0,\beta_1,\tau_i|y) &\propto f(y|\beta_0,\beta_1,\tau_i)f(\beta_0,\beta_1,\tau_i) \\ &\propto \left(\sqrt{\tau_i}\right)^n exp\left\{-\frac{1}{2}\sum_{i=1}^{n}\tau_i(y_i-\beta_0-\beta_1 x_i)^2\right\} \times \sqrt{\tau_{\beta_0}} \\ &\quad exp\left\{-\frac{\tau_{\beta_0}}{2}\left(\beta_0-\mu_{\tau_{\beta_0}}\right)^2\right\} \times \sqrt{\tau_{\beta_1}} exp\left\{-\frac{\tau_{\beta_1}}{2}\left(\beta_1-\mu_{\tau_{\beta_1}}\right)^2\right\} \\ &\quad \times \tau_i^{a-1} exp\left\{-b\tau_i\right\} \end{aligned}$$

Proses distribusi *posterior marginal* setiap parameter mengandung integral dengan dimensi yang cukup tinggi dan kompleks, sehingga untuk mendapatkan solusinya digunakan pendekatan numerik melalui MCMC dan *Gibbs Sampling*. Proses tersebut dilakukan melalui pengambilan sampel secara berulang melalui bentuk *full conditional posterior* [3]. Distribusi *full conditional posterior* untuk setiap parameter adalah bentuk proporsional dari distribusi *posterior* gabungan seluruh parameter, dengan mengeliminasi komponen dari distribusi *posterior* gabungan yang tidak mengandung parameter yang akan diestimasi, karena parameter lain selain parameter yang akan diestimasi tersebut dianggap bernilai tetap [1].

2.5. *Markov Chain Monte Carlo (MCMC)*

Pendekatan MCMC sangat efektif untuk mengurangi beban komputasi dalam menyelesaikan persamaan integrasi yang kompleks. Ide dasar dari MCMC adalah membangkitkan data sampel dari distribusi *posterior* sesuai proses *Markov Chain* dengan menggunakan simulasi *Monte Carlo* secara iteratif, sehingga diperoleh kondisi yang konvergen terhadap *posterior* [3]. Sampel parameter dalam *Markov Chain* diambil setelah kondisi stasioner (*equilibrium*) tercapai.

Markov Chain merupakan proses iterasi sekumpulan variabel acak β sedemikian sehingga $f(\beta^{\,t+1}|\, \beta^t, ..., \beta^1) = f(\beta^{t+1}|\beta^t)$, yang artinya proses untuk mendapatkan estimasi parameter β pada saat iterasi ke-$(t+1)$ hanya dipengaruhi oleh nilai pada saat iterasi ke- t saja. Hasil estimasi *posterior* parameter dikatakan baik, jika terpenuhinya sifat dari *Markov Chain* yang *strongly ergodic* berupa :

(1) *Irreducible*, artinya sampel parameter yang dibangkitkan melalui proses MCMC adalah bersifat acak.
(2) *Aperiodic*, artinya sampel parameter yang dibangkitkan tersebut tidak memiliki pola yang periodik dalam domain nilai tertentu.
(3) *Reccurent*, artinya perubahan sampel parameter terjadi secara stabil dalam domain nilai tertentu.

2.6. *Gibbs Sampling*

Proses *Gibbs Sampling* dilakukan dengan mengambil sampel dengan cara membangkitkan rangkaian *Gibbs* variabel acak berdasarkan sifat-sifat dasar proses *Markov Chain* [3]. Hal ini merupakan kelebihan dari *Gibbs Sampling* karena variabel acak tersebut dibangkitkan dengan menggunakan konsep distribusi yang terstruktur sebagai bentuk distribusi *full conditional.*

Proses ini adalah proses iterasi yang mengikuti proses *Markov Chain* secara iteratif dalam melakukan estimasi parameter, dengan diberikan informasi nilai parameter ditahap iterasi sebelumnya. Proses ini akan berlangsung secara *full conditional* untuk setiap parameter yang selanjutnya disusun bergantian sebagai tahapan iteratif simulasi stokastik.

3. Data dan Hasil

Data yang akan digunakan dalam penelitian ini merupakan data simulasi yang dibangkitkan secara acak menggunakan *software R*. Data yang dibangkitkan mengalami heteroskedastisitas pada variabel galatnya (ϵ_i) .

Langkah-langkah dalam simulasi data adalah sebagai berikut :

(1) Membangkitkan data sebanyak n untuk variabel prediktor $(x_i), i = 1, 2, ..., n$, dari distribusi $uniform(0, 1)$ dengan $n = 300$ dan 600.
(2) Membangkitkan galat dari distribusi normal $(0, \sigma_j^2 x_i)$, dimana $i = 1, 2, ..., n$ dan $j = 1, 2, 3$. Nilai dari masing-masing varian adalah $\sigma_1^2 = 1$, $\sigma_2^2 = 4$ dan $\sigma_3^2 = 9$.
(3) Membangkitkan data sebanyak n untuk setiap variabel respon y_i, berdasarkan model regresi linier sederhana, yaitu $y_i = \beta_0 + \beta_1 x_i + \epsilon_i$ dengan $\beta_0 = 1$ dan $\beta_1 = 2$.

(4) Melakukan uji *Breusch Pagan Godfrey* (BPG) untuk mendeteksi adanya kasus heteroskedastisitas.

Langkah 1 sampai 4 dilakukan pengulangan sebanyak 30 kali. Setelah itu akan dilakukan estimasi parameter regresi β_0 dan β_1 menggunakan Metode Regresi Kuantil dan Metode Bayes. Untuk Metode Regresi Kuantil akan dilakukan estimasi parameter pada kuantil ke- $\theta = 0.10, 0.25, 0.50, 0.75, 0.90$ dan untuk Metode Bayes akan dilakukan estimasi parameter dengan iterasi 10000.

3.1. *Mendeteksi Heteroskedastisitas*

Pada bagian ini, akan diuji apakah data yang telah dibangkitkan mengalami heteroskedastisitas. Berdasarkan data yang dibangkitkan dengan ukuran sampel $n = 300$ dan varian $\sigma_1^2 = 1$. Metode yang digunakan untuk mendeteksi adanya heteroskedastisitas adalah metode uji *Breusch Pagan Godfrey* (BPG) dengan taraf nyata $\alpha = 0.05$ menggunakan bantuan *software R*. Hal yang sama juga dilakukan untuk ukuran sampel dan varian yang berbeda pada data .

Berdasarkan uji BPG yang telah dilakukan, diperoleh hasil uji BPG dengan nilai statistik BP adalah 48,2, dimana nilai statistik BP lebih dari nilai tabel $\chi_1^2 = 3.841$. Karena nilai statistik BP lebih dari nilai tabel χ_1^2, maka tolak H_0. Hal ini menunjukkan bahwa adanya masalah heteroskedastisitas pada sebaran data.

3.2. *Perbandingan Hasil Estimasi Parameter Model dengan Metode Regresi Kuantil dan Metode Bayes*

Setelah dilakukan simulasi data sebanyak 30 kali, selanjutnya akan dibandingkan nilai dugaan dan nilai absolut bias pada hasil estimasi parameter model regresi linier sederhana yang diperoleh dari Metode Regresi Kuantil dan Metode Bayes.

Pada tabel berikut ini dapat dilihat nilai dugaan dan nilai absolut bias yang diperoleh dari Metode Regresi Kuantil pada kuantil ke $\theta = 0{,}50$ dan Metode Bayes pada iterasi 10000 kali dengan ukuran sampel ($n = 300$ dan 600) dan dengan varian ($\sigma_1^2 = 1, \sigma_2^2 = 4, \sigma_3^2 = 9$).

Hasil estimasi parameter model pada varian $\sigma_1^2 = 1$ disajikan pada tabel berikut ini :

	$\sigma_1^2 = 1$							
	Nilai Dugaan				Nilai Absolut Bias			
N	Kuantil		Bayes		Kuantil		Bayes	
	β_0	β_1	β_0	β_1	β_0	β_1	β_0	β_1
300	1,0018	1,9882	1,0185	1,9565	0,0090	0,0813	0,0448	0,1164
600	0,9993	2,0144	0,9911	2,0276	0,0069	0,0498	0,0398	0,1070

Hasil estimasi parameter model pada varian $\sigma_2^2 = 4$ disajikan pada tabel berikut ini :

N	$\sigma_2^2 = 4$							
	Nilai Dugaan				Nilai Absolut Bias			
	Kuantil		Bayes		Kuantil		Bayes	
	β_0	β_1	β_0	β_1	β_0	β_1	β_0	β_1
300	1,0100	1,9623	1,0982	1,9616	0,0312	0,2832	0,1646	0,3048
600	0,9935	2,0484	0,9656	2,0500	0,0311	0,2311	0,1279	0,4148

Hasil estimasi parameter model pada varian $\sigma_3^2 = 9$ disajikan pada tabel berikut ini :

N	$\sigma_3^2 = 9$							
	Nilai Dugaan				Nilai Absolut Bias			
	Kuantil		Bayes		Kuantil		Bayes	
	β_0	β_1	β_0	β_1	β_0	β_1	β_0	β_1
300	0,9933	1,9484	0,9728	2,0761	0,0740	0,6847	0,3128	0,7748
600	1,0163	1,9586	0,9876	1,9261	0,0540	0,7354	0,3591	0,8763

Hasil estimasi parameter yang disajikan pada tabel di atas menunjukkan bahwa Metode Regresi Kuantil memiliki nilai absolut bias lebih kecil dibandingkan dengan Metode Bayes pada ukuran sampel $n = 300$ dan 600.

3.3. *Perbandingan MSE dengan Metode Regresi Kuantil dan Metode Bayes*

Pada tabel berikut ini dapat dilihat nilai *Mean Square Error* (MSE) yang diperoleh dari Metode Regresi Kuantil pada kuantil ke- $\theta = 0.50$ dan Metode Bayes pada iterasi 10000 kali dengan ukuran sampel ($n = 300$ dan 600) dan dengan varian ($\sigma_1^2 = 1, \sigma_2^2 = 4, \sigma_3^2 = 9$).

Nilai MSE pada varian $\sigma_1^2 = 1$ disajikan pada tabel berikut ini :

N	$\sigma_1^2 = 1$			
	Nilai MSE			
	Kuantil		Bayes	
	β_0	β_1	β_0	β_1
300	0,0002	0,0174	0,0051	0,0358
600	0,0001	0,0063	0,0038	0,0274

Nilai MSE pada varian $\sigma_2^2 = 4$ disajikan pada tabel berikut ini :
Nilai MSE pada varian $\sigma_3^2 = 9$ disajikan pada tabel berikut ini : Nilai MSE

	$\sigma_2^2 = 4$			
	Nilai MSE			
N	Kuantil		Bayes	
	β_0	β_1	β_0	β_1
300	0,0028	0,2279	0,0598	0,2456
600	0,0028	0,1339	0,0413	0,4279

	$\sigma_3^2 = 9$			
	Nilai MSE			
N	Kuantil		Bayes	
	β_0	β_1	β_0	β_1
300	0,0170	1,2123	0,3293	1,5333
600	0,0076	1,3257	0,3094	1,8534

yang disajikan pada tabel di atas menunjukkan bahwa Metode Regresi Kuantil memiliki nilai MSE lebih kecil dibandingkan dengan Metode Bayes pada ukuran sampel $n = 300$ dan 600.

4. Kesimpulan

Berdasarkan hasil perbandingan estimasi parameter dan MSE yang dilakukan dengan Metode Regresi Kuantil dan Metode Bayes, diambil kesimpulan sebagai berikut.

(1) Dari simulasi data yang dilakukan sebanyak 30 kali, nilai absolut bias dari parameter β_0 dan β_1 dengan Metode Regresi Kuantil memiliki nilai yang lebih kecil daripada Metode Bayes pada ukuran sampel $n = 300$ dan 600. Hal yang sama juga ditunjukkan pada nilai MSE.

(2) Pengaruh perubahan varian pada hasil estimasi parameter menunjukkan bahwa varian $\sigma_1^2 = 1$ memiliki nilai absolut bias dan nilai MSE lebih kecil dibandingkan dengan varian $\sigma_2^2 = 4$ dan $\sigma_3^2 = 9$.

5. Ucapan Terima Kasih

Penulis mengucapkan terima kasih kepada Bapak Dr. Dodi Devianto, Ibu Monika Rianti Helmi, M.Si dan Ibu Dr. Arrival Rince Putri yang telah memberikan bimbingan, masukan, kritik dan saran sehingga tugas akhir ini dapat diselesaikan dengan baik.

Daftar Pustaka

[1] Bolstad, W.M. 2007. *Introduction to Bayesian Statistical. Second Edition.* A John Wiley and Sons. Inc, America

[2] Koenker, R. 2005. *Quantile Regression.* Cambridge University Press, New York

[3] Ntzoufras, I. 2009. *Bayesian Modelling Using Winbugs.* A John Wiley and Sons Inc, America

[4] Schuzle, N. 2004. *Applied Quantile Regression : Microeconometric, Financial and Enviromental Analysis*. Tubingen : Universitat Tubingen