

PERBANDINGAN METODE KUADRAT TERKECIL DAN BAYES PADA MODEL REGRESI DENGAN GALAT YANG TIDAK NORMAL

NURHAKIKI, MAIYASTRI, RITA DIANA

*Program Studi Matematika,
Fakultas Matematika dan Ilmu Pengetahuan Alam, Universitas Andalas,
Kampus UNAND Limau Manis Padang, Indonesia,
email : kikitanjung27@gmail.com*

Abstrak. Metode Kuadrat Terkecil (MKT) merupakan metode yang sering digunakan untuk menduga parameter model regresi. Pendugaan parameter dengan menggunakan MKT harus memenuhi asumsi-asumsi tertentu terhadap galatnya yang dinamakan dengan asumsi klasik. Salah satu asumsinya adalah asumsi kenormalan, dimana asumsi residual berdistribusi normal. Jika asumsi kenormalan tidak dipenuhi maka penduga MKT menjadi tidak efisien, oleh karena itu diperlukan metode pendugaan lain, salah satunya adalah metode bayes. Pada metode bayes, parameter model diasumsikan sebagai suatu peubah acak yang memiliki distribusi tertentu yang disebut distribusi prior. Distribusi posterior diperoleh dengan menggabungkan informasi prior dan informasi yang diperoleh dari sampel melalui teorema Bayes. Rataan dari distribusi posterior ini merupakan penduga bagi parameter model. Hasil penelitian ini menunjukkan bahwa metode Bayes menghasilkan nilai absolut bias dan MSE lebih kecil dibandingkan dengan metode MKT dalam mengestimasi parameter regresi linier sederhana untuk galat berdistribusi tidak normal.

Kata Kunci: Analisis Regresi Linier, Metode Kuadrat Terkecil, Metode Bayes, Galat Tidak Normal.

1. Pendahuluan

Analisis regresi merupakan suatu teknik untuk memodelkan hubungan antara variabel *dependent* (terikat) dengan variabel *independent* (bebas). Untuk mendapatkan model regresi dapat diperoleh dengan melakukan pendugaan terhadap parameternya menggunakan metode tertentu. Metode Kuadrat Terkecil (MKT) merupakan metode yang sering digunakan untuk mengestimasi parameter model regresi. Pendugaan parameter yang diduga dengan menggunakan MKT ini harus memenuhi asumsi-asumsi tertentu terhadap galatnya yang dinamakan dengan asumsi klasik. Salah satu asumsi yang harus dipenuhi adalah asumsi normalitas, dimana asumsi residual berdistribusi normal. Jika asumsi kenormalan tidak dipenuhi maka penduga MKT menjadi tidak efisien. Selain menggunakan Metode Kuadrat Terkecil, Metode Bayes juga dapat digunakan untuk mengestimasi parameter model regresi.

Pada Metode Bayes, parameter model diasumsikan sebagai suatu peubah acak yang memiliki suatu sebaran tertentu. Informasi parameter ini dinamakan informasi prior, sehingga sebaran yang akan dimiliki oleh parameter ini disebut distribusi

prior. Penentuan distribusi prior ini dapat diperoleh dari data penelitian sebelumnya atau berdasarkan intuisi peneliti. Dengan menggabungkan informasi prior dan informasi yang kita peroleh dari sampel, kita dapat menentukan distribusi posterior melalui teorema Bayes ini. Rataan dari distribusi posterior ini dapat digunakan sebagai penduga bagi parameter model [1].

2. Kajian Pustaka

Kajian pustaka yang digunakan dalam tulisan ini adalah sebagai berikut:

2.1. Metode Kuadrat Terkecil

Analisis regresi linier sederhana merupakan suatu teknik statistika yang digunakan untuk menarik kesimpulan tentang hubungan antara satu variabel *dependent* (terikat) dan satu variabel *independent* (bebas). Secara umum, persamaan regresi linier sederhana adalah [4]

$$Y_i = \beta_0 + \beta_1 x_i + \varepsilon_i, \quad i = 1, 2, 3, \dots, n,$$

dimana

Y_i : Variabel dependen pada model

x_i : Variabel independen

β_0, β_1 : parameter regresi.

Salah satu metode untuk mengestimasi parameter regresi adalah metode kuadrat terkecil. Metode kuadrat terkecil digunakan untuk menduga nilai β_0 dan β_1 sehingga akan didapat suatu dugaan model regresi, yaitu [1]

$$\hat{Y}_i = \hat{\beta}_0 + \hat{\beta}_1 x_i, \quad i = 1, 2, 3, \dots, n,$$

dimana

\hat{Y}_i : Variabel dependen pada model

x_i : Variabel independen

$\hat{\beta}_0, \hat{\beta}_1$: parameter regresi.

Prinsip metode kuadrat terkecil memilih β_0 dan β_1 yang meminimalkan Jumlah Kuadrat Sisaan (JKS). Sehingga diperoleh

$$\hat{\beta}_0 = \bar{y} - b_1 \bar{x} \quad \text{dan} \quad \hat{\beta}_1 = \frac{n \sum x_i y_i - \sum x_i \sum y_i}{n \sum x_i^2 - (\sum x_i)^2}.$$

2.2. Metode Bayes

Metode Bayes merupakan suatu metode dalam statistika inferensia yang menggabungkan informasi dari data pengamatan dengan informasi data sampel yang telah diperoleh sebelumnya.

Misalkan data observasi $\mathbf{y} = (y_1, y_2, \dots, y_n)'$ mempunyai distribusi tertentu dengan himpunan parameter $\theta = (\theta_1, \theta_2, \dots, \theta_n)'$ yang merupakan variabel random,

maka distribusi *posterior* dari parameter θ atau $f(\theta|\mathbf{y})$ dapat dinyatakan dalam bentuk:

$$f(\theta|\mathbf{y}) = \frac{f(\mathbf{y}|\theta)f(\theta)}{f(\mathbf{y})},$$

dengan $f(\mathbf{y}) = \int_{\Omega_{\theta_1}} \cdots \int_{\Omega_{\theta_p}} f(\mathbf{y}|\theta)f(\theta)d\theta_1 \cdots d\theta_n$ adalah *normalizing constant*, $f(\theta)$ adalah distribusi *prior* dari parameter θ dan $f(\mathbf{y}|\theta) = \prod f(y_i|\mathbf{y})$ adalah fungsi *Likelihood* data. Nilai tengah dari distribusi posterior digunakan untuk menentukan pendugaan Bayes bagi parameter θ yang tidak diketahui.

Markov Chain merupakan proses iterasi sekumpulan variabel acak θ sedemikian sehingga $f(\theta^{t+1}|\theta^t, \dots, \theta^1) = f(\theta^{t+1}|\theta^t)$, yang berarti proses untuk mendapatkan estimasi parameter θ pada saat iterasi ke $(i+1)$, hanya dipengaruhi nilai pada saat iterasi ke- i . Hasil estimasi parameter posterior dikatakan baik jika terpenuhi sifat Markov Chain yang *strongly ergodic* berupa: [3]

- (1) *Irreducible*, artinya pergerakan θ^i selama iterasi tidak ada kecenderungan tetap pada range nilai tertentu sehingga tidak mampu berubah sesuai domain θ yang seharusnya.
- (2) *Aperiodic*, artinya nilai suatu parameter θ yang dibangkitkan tidak memiliki pola yang periodik untuk muncul kembali dengan nilai yang sama pada domain tertentu.
- (3) *Recurrent*, artinya kemunculan nilai θ dalam iterasi proses MC akan muncul lagi diiterasi berikutnya adalah suatu kepastian sehingga perubahan sample parameter terjadi secara stabil dalam domain nilai tertentu.

Penggunaan metode MCMC dalam analisis Bayesian memerlukan algoritma sampling yang sesuai untuk mendapatkan sampel dari suatu distribusi. Salah satu algoritma yang sering digunakan adalah *Gibbs Sampling* [2]. Proses Gibbs Sampling dilakukan dengan cara membangkitkan rangkaian *Gibbs* variabel random berdasarkan sifat-sifat dasar proses *Markov Chain* yang menggunakan konsep distribusi unidimensional yang terstruktur dalam bentuk *full conditional*.

3. Data dan Hasil

Data yang digunakan adalah data simulasi yang dibangkitkan secara acak dengan menggunakan *software* R. Data yang dibangkitkan adalah galat yang tidak normal ($\varepsilon_i \sim \text{Logis}(\lambda, \xi)$) dimana $\lambda = 0$ dan $\xi = 0.5, 1, 1.5$ dan 2 kemudian nilai x_i ditetapkan dengan $x_i \sim \text{Unif}(0, 1)$ dan parameter $\beta_0 = 2$ dan $\beta_1 = 3$ sehingga nilai y_i diperoleh dengan persamaan

$$y_i = 2 + 3x_i + \varepsilon_i, i = 1, 2, \dots, n.$$

Banyak data direncanakan sebanyak 10, 20, 30, 40 dan 50 data. Percobaan dilakukan dengan simulasi Monte Carlo dengan ulangan sebanyak 1000, 2000 dan 3000 kali. Selanjutnya dilakukan pendugaan parameter β_0 dan β_1 berdasarkan Metode Kuadrat Terkecil dan Metode Bayes.

3.1. Estimasi Parameter Model Regresi Linier Sederhana dengan Metode Bayes

Jika y_i adalah variabel random yang merupakan respon dengan galat tidak normal $\varepsilon_i \sim \text{Logis}(0, \xi)$ maka $y_i \sim \text{Logis}(\beta_0 + \beta_1 x_i, \xi)$ dan $\tau_y = \frac{1}{\xi}$ adalah parameter presisi, maka fkp dari y_i adalah

$$f(y_i) = \frac{\tau_y (\exp(\tau_y (y_i - (\beta_0 + \beta_1 x_i))))}{(1 + \exp(\tau_y (y_i - (\beta_0 + \beta_1 x_i))))^2}, i = 1, 2, 3, \dots, n.$$

Dari fkp diatas, diperoleh fungsi likelihoodnya yaitu:

$$L(\beta_0 + \beta_1 x_i, \frac{1}{\tau_y}) = \frac{\tau_y^n (\exp(\tau_y (\sum y_i - (n\beta_0 + \beta_1 \sum x_i))))}{\prod (1 + \exp(\tau_y (y_i - (\beta_0 + \beta_1 x_i))))^2}.$$

Tipe distribusi prior yang digunakan adalah *informative prior*, dengan distribusi prior yang digunakan untuk parameter β_0 dan β_1 adalah distribusi normal yang dinyatakan sebagai berikut.

$$\beta_j \sim N(\mu_j, \tau_j), j = 0, 1.$$

Dan distribusi prior yang digunakan untuk parameter τ_y adalah distribusi gamma yang dinyatakan sebagai berikut.

$$\tau_y \sim \gamma(a, b).$$

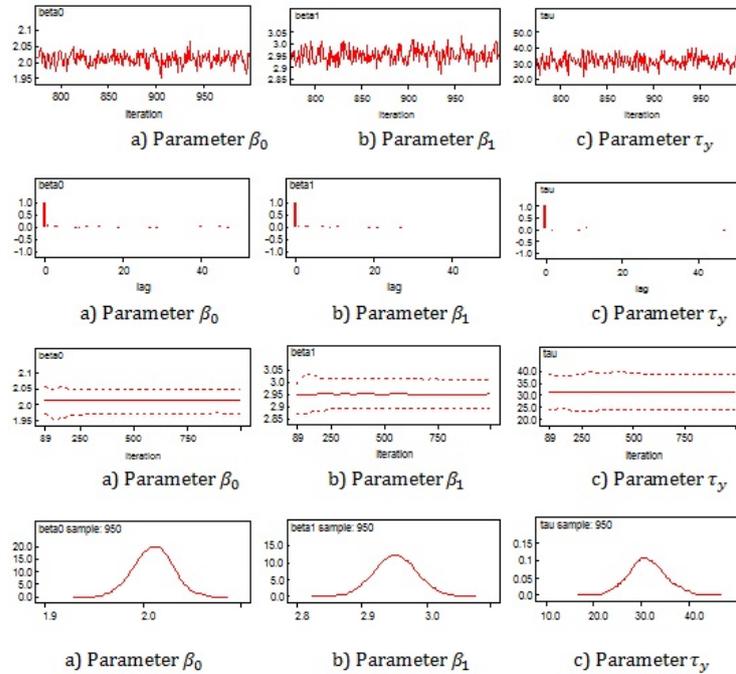
Sehingga diperoleh bentuk proposional distribusi posterior sebagai berikut:

$$f(\beta_0, \beta_1, \tau_y) \propto \frac{\tau_y^n (\exp(\tau_y (\sum y_i - (n\beta_0 + \beta_1 \sum x_i))))}{\prod (1 + \exp(\tau_y (y_i - (\beta_0 + \beta_1 x_i))))^2} \times \tau_0^{\frac{1}{2}} \exp(-\frac{\tau_0}{2} (\beta_0 - \mu_0)^2) \\ \times \tau_1^{\frac{1}{2}} \exp(-\frac{\tau_1}{2} (\beta_1 - \mu_1)^2) \times \tau_y^{a-1} \exp(-b\tau_y).$$

Dari distribusi posterior diatas, dengan menggunakan data yang dibangkitkan dengan sampel $n = 30$ dengan $\varepsilon_i \sim \text{Logis}(0, 2)$ dan iterasi sebanyak 3000 dilakukan proses MCMC dengan algoritma *Gibbs Sampling* menggunakan bantuan *software* WinBUGS dan diperoleh hasil pada Gambar 1.

Trace Plot parameter menunjukkan nilai acak dengan pola yang stasioner dan bersifat random. Kemudian, *Autocorrelation plot* memperlihatkan hanya autokorelasi pertama sampel yang tinggi, ini menunjukkan bahwa sampler baik dilakukan, dan nilai autokorelasi selanjutnya mendekati nol ini yang menunjukkan bahwa antar iterasi satu dengan yang lainnya saling bebas. Hal ini menunjukkan bahwa sampel yang dibangkitkan dalam proses MCMC memiliki sifat yang random. Selanjutnya, *Quantiles plot* menunjukkan *ergodic mean* hasil estimasi parameter β_0 , β_1 dan τ_y yang dihasilkan telah mencapai nilai yang stabil dan berada dalam *credible interval*. Kondisi ini merupakan salah satu indikasi bahwa proses iterasi estimasi parameter β_0 , β_1 dan τ_y telah mencapai kondisi equilibrium atau konvergen. Dan, *Density plot*, menunjukkan *density plot* untuk β_0 dan β_1 berdistribusi normal, dan parameter τ_y menunjukkan *density plot* berdistribusi gamma.

Ringkasan posterior parameter yang diperoleh dari proses estimasi parameter model regresi linier sederhana dapat dilihat pada Tabel 1.



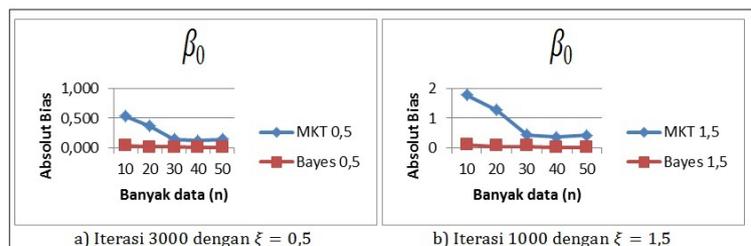
Gambar 1. Diagnostic plot pada proses *Markov Chain Monte Carlo*

Tabel 1. Nilai Koefisien Regresi Linier Sederhana dengan Pendekatan Bayes

No	Parameter	Posterior Mean	Standar Deviasi
1	β_0	2,011	0,01889
2	β_1	2,954	0,03085
3	τ_y	31,45	3,673

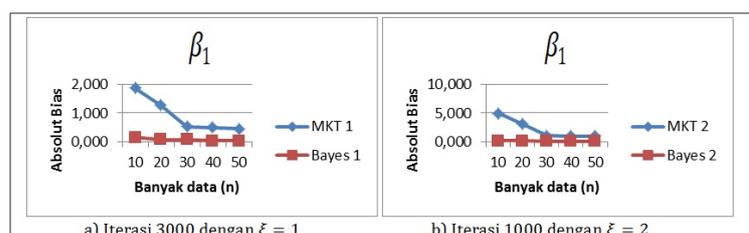
3.2. Perbandingan Penduga Paramater Regresi β_0 dan β_1 dengan Menggunakan Metode Kuadrat Terkecil (MKT) dan Metode Bayes

a. Absolut Bias



Gambar 2. Nilai Absolut Bias β_0 MKT dan Bayes

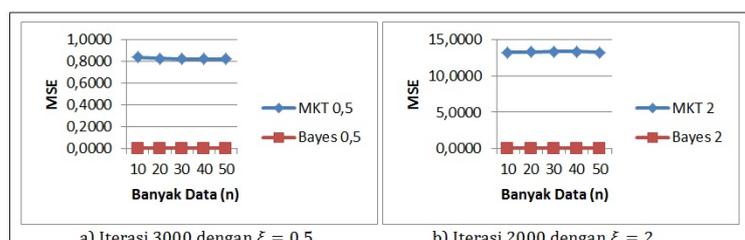
Pada Gambar 2 dapat dilihat bahwa nilai absolut bias β_0 MKT lebih besar daripada nilai absolut bias β_0 bayes. Nilai absolut bias β_0 MKT bergerak turun pada $n = 10$ dan $n = 20$, kemudian nilai absolut bias β_0 MKT bergerak di sekitar nilai 0,14 dengan $\xi = 0,5$ dan bergerak turun disekitar nilai 0,40 dengan $\xi = 1,5$. Sedangkan nilai absolut bias β_0 Bayes tetap bergerak turun menuju nilai 0 baik dengan $\xi = 0,5$ maupun dengan $\xi = 1,5$.



Gambar 3. Nilai Absolut Bias β_1 MKT dan Bayes

Pada Gambar 3 dapat dilihat bahwa nilai absolut bias β_1 MKT lebih besar daripada nilai absolut bias β_1 bayes. Nilai absolut bias β_1 MKT bergerak turun pada $n = 10$ dan $n = 20$, kemudian nilai absolut bias β_1 MKT bergerak di sekitar nilai 0,5 dengan $\xi = 1$ dan bergerak turun disekitar nilai 1 dengan $\xi = 2$. Sedangkan nilai absolut bias β_1 Bayes tetap bergerak turun menuju nilai 0 baik dengan $\xi = 1$ maupun dengan $\xi = 2$.

b. Mean Square Error (MSE)



Gambar 4. Nilai MSE Metode Kuadrat Terkecil dan Metode Bayes

Perhitungan nilai *Mean Square Error* (MSE) diperoleh dari penjumlahan seluruh nilai kuadrat galat dibagi dengan banyaknya data. Sementara nilai galat diperoleh dari hasil pengurangan atau selisih dari data aktual terhadap data hasil dugaan.

Pada Gambar 4 dapat terlihat bahwa nilai yang dihasilkan oleh Metode Kuadrat Terkecil dan Metode Bayes berbeda. Pada Gambar 4(a) dapat dilihat bahwa semakin besar sampel yang diberikan, dengan iterasi 3000 dan $\xi = 0,5$ nilai MSE pada Metode Kuadrat Terkecil berada di sekitar nilai 0,8 sedangkan nilai MSE pada Metode Bayes berada di sekitar 0, begitu pula dengan Gambar

4(b), nilai MSE pada Metode Kuadrat Terkecil berada di sekitar nilai 13,3, sedangkan nilai MSE pada metode Bayes berada pada sekitar 0 dengan iterasi 2000 dan $\xi = 2$.

Dari kedua metode diatas dapat dilihat bahwa pada Metode Bayes menghasilkan nilai MSE lebih kecil dibandingkan dengan Metode Kuadrat Terkecil. Ini berarti Metode Bayes menghasilkan pendugaan parameter yang lebih baik daripada Metode Kuadrat Terkecil.

4. Kesimpulan

- (1) Pada percobaan simulasi data menunjukkan bahwa semakin besar ukuran sampel (n) dan semakin besar iterasi (M) yang dilakukan maka nilai Absolut Bias MKT dan Absolut Bias Bayes akan semakin kecil, namun nilai Absolut Bias Bayes lebih kecil daripada Absolut Bias MKT baik Absolut Bias β_0 maupun Absolut Bias β_1 . Sehingga, pada pendugaan parameter model untuk penduga $\hat{\beta}_0$ dan $\hat{\beta}_1$, pendugaan parameter dengan menggunakan Bayes lebih efisien dan stabil dibandingkan dengan penduga MKT.
- (2) Dari percobaan simulasi data juga menunjukkan bahwa semakin besar ukuran sampel (n) dan semakin besar iterasi (M) yang dilakukan maka nilai MSE MKT dan MSE Bayes juga akan semakin kecil, namun nilai MSE Bayes jauh lebih kecil daripada nilai MSE MKT, sehingga pendugaan parameter dengan menggunakan metode Bayes lebih baik daripada metode MKT.

5. Ucapan Terima Kasih

Penulis mengucapkan terima kasih kepada ibu Dr. Ferra Yanuar, Bapak Dr. Admi Nazra dan Bapak Yudiantri Asdi, M.Sc yang telah memberikan bimbingan, masukan dan saran sehingga makalah ini dapat diselesaikan dengan baik.

Daftar Pustaka

- [1] Casella, George and Berger RL. 2002. *Statistical Inference*. Boston:Duxbury Press
- [2] Gelman, Andrew, dkk. 2004. *Bayesian Data Analysis Second Edition*. Florida:CRC Press LLC
- [3] Ntzoufras, Ioannis. 2009. *Bayesian Modeling Using WinBUGS*. Canada:John Wiley & Sons, Inc
- [4] Rawlings, John O. 1988. *Applied Regression Analysis (A Research Tool)*. California:Brooks/Cole Publishing.