JOMTA
Journal of Mathematics: Theory and Applications
Vol.3, No.2, 2021, P-ISSN 2685-9653 e-ISSN 2722-2705

Estimasi Parameter Model Regresi Linier dengan Pendekatan Metode Bayesian

Surianti[1], Hikmah[2], Rahmawati[3]
[1,3]Program Studi Matematika, [2]Program Studi Statistika, Universitas Sulawesi Barat
e-mail: [1]suriantiputrimamasa@gmail.com , [2]hikmah@unsulbar.ac.id., [3]rahmah@unsulbar.ac.id.

Abstrak. Model Regresi merupakan model yang digunakan dalam menganalisis data yang melibatkan variabel terikat dan variabel bebas yang dipengaruhi oleh beberapa parameter regresi yang belum diketahui nilainya sehingga diperlukan untuk mengestimasi. Suatu model regresi mempunyai dua bentuk yaitu linier atau nonlinear, yang tidak terlepas dari permasalahan sisaan. Sisaan (Residual) sering disebut Outlier. Tujuan pada penelitian ini adalah untuk mengestimasi parameter model regresi linier dengan pendekatan metode Bayesian. Langkah-langkah yang dilakukan yaitu menjelaskan estimasi parameter dan uji signifikansi parameter regresi linier dengan menggunakan metode bayesian, serta melakukan analisis error untuk membandingkan model estimasi terbaik yang dihasilkan Metode Ordinary Least Square (OLS) dan Metode Bayesian dengan menggunakan Prior Non-Informatif pada contoh kasus. Dari model regresi yang dihasilkan metode Ordinary Least Square (OLS) teridentifikasi terdapat satu data Outlier. Berdasarkan Pembahasan diperoleh kesimpulan bahwa estimasi parameter Bayesian yang dihasilkan dengan menggunakan WinBUGS dan metode Ordinary Least Square (OLS) menggunakan R. Dapat dilihat bahwa nilai MSE yang diperoleh dari estimasi Metode Bayesian lebih kecil dibandingkan dengan nilai MSE yang diperoleh dari estimasi Metode OLS. Nilai Mean Square Error (MSE) yang diperoleh dari estimasi Metode OLS yaitu 0,8469 sedangkan Nilai Mean Square Error (MSE) yang diperoleh dari estimasi Metode Bayesian MCMC yaitu 0,3723. Hal ini menunjukkan bahwa Metode Bayesian MCMC jauh lebih baik dibandingkan Metode OLS.

Kata Kunci: Metode Bayesian, outlier, MCMC, etimasi parameter, WinBUGS

Abstract. *The method used to estimate the regression parameters in this study is the Bayes method. The steps in estimating the parameters of the linear regression model using the Bayes method are to determine the Likelihood function from the normal distribution density function, determine the Prior Non-Informative distribution from a normal distribution, then look for the Posterior distribution by switching the Prior distribution with the Likelihood function. Thesis aims to determine the estimation of the parameters of the linear regression model with the Bayesian method approach. From the regression model generated by the OLS method, it was identified that there was one Outlier data. A model will be shown in an example data case by comparing the results of the Ordinary Least Square (OLS) method using R and the results of the Bayesian method using WinBUGS. It can be seen that the MSE value obtained from the Bayesian Method estimation is smaller than the MSE value obtained from the OLS Method estimate. The Mean Square Error (MSE) value obtained from the estimation of the OLS Method is 0,8469 while the Mean Square Error (MSE) value obtained from the estimation of the Bayesian MCMC Method is 0,3723. This shows that the Bayesian MCMC method is much better than the OLS method.*

Keywords: *Bayesian Linear Regression, outliers, MCMC, parameter estimation, WinBUGS.*

I. PENDAHULUAN

Model Regresi merupakan model yang digunakan dalam menganalisis data penelitian yang melibatkan variabel terikat dan variabel bebas yang dipengaruhi oleh beberapa parameter regresi yang belum diketahui nilainya, sehingga diperlukan untuk mengestimasi Pembentukan model regresi dilakukan dengan mengestimasi model regresi, sehingga akan menghasilkan koefisien regresi setiap variabel independen [1]. Suatu model regresi mempunyai dua bentuk yaitu berbentuk linier atau nonlinier, yang tidak terlepas dari permasalahan sisaan. Sisaan (*Residual*) sering disebut *outlier*.

Terdapat dua metode yang digunakan untuk mengestimasi parameter. Metode yang pertama adalah Metode Klasik (Metode OLS dan Metode *Maximum Likelihood*). Metode Klasik menggunakan pendekatan statistika klasik dimana pendugaan parameter berdasarkan informasi dari data sampel dan mengabaikan informasi awal (*Prior*). Metode yang kedua adalah Metode Bayesian [2].

Keunggulan utama penggunaan Metode Bayesian adalah penyederhanaan dari cara Metode Klasik dengan integral untuk memperoleh model Estimasi. Dalam metode Bayesian dapat memberikan hasil pendugaan yang lebih baik daripada pendugaan Metode Klasik, Karena Metode Klasik parameternya hanya berdasarkan informasi dari data sampel, dimana ukuran sampel sangat berpengaruh terhadap hasil pendugaan [3]. Metode Bayesian dikatakan terbaik karena sebelum menarik sampel dari populasi terkadang diperoleh informasi mengenai parameter yang akan di estimasi jika informasi tersebut ingin dimasukkan data dalam analisis data, maka estimasi parameter regresi dengan Metode Klasik tidak memungkinkan untuk memasukkan data informasi, sehingga diperlukan Metode Bayes untuk menyelesaikan permasalahan.

Penelitian sebelumnya menyatakan dalam Metode Bayesian, seseorang dapat memberikan kepercayaan awal (*prior*) terhadap suatu parameter karena adanya asumsi bahwa parameter merupakan suatu variabel acak [4]. Parameter Regresi diperlukan informasi tambahan parameter populasi yang berasal dari data sebelumnya, jika informasi menjadi faktor pertimbangan pada analisis data, maka estimasi dengan Metode Bayesian untuk mengestimasi parameter regresinya. Tujuan penelitian ini adalah untuk mengestimasi parameter model regresi linier dengan pendekatan Metode Bayesian Kasus Nilai Tukar Petani di Provinsi Sulawesi Barat pada masa pandemi Covid-19 pada Bulan Januari-Juli 2021.

II. LANDASAN TEORI

Model Regresi merupakan model yang digunakan dalam menganalisis data penelitian yang melibatkan variabel terikat dan variabel bebas yang dipengaruhi oleh beberapa parameter regresi yang belum diketahui nilainya sehingga diperlukan untuk mengestimasi. adapun batasan-batasan masalah yang digunakan yaitu, Distribusi sampel yang digunakan adalah distribusi normal univariat, Distribusi *Prior* yang digunakan adalah distribusi *Prior Non-informatif*, Uji siginifkansi parameter regresi linier berdasarkan distribusi *Prior Non-Informatif* dengan Metode Bayesian, Algoritma yang digunakan pada MCMC adalah *Gibbs Sampler.*

III. METODE PENLITIAN

Jenis penelitian yang digunakan adalah penelitian kuantitatif yaitu metode penelitian yang bertujuan untuk memverifikasi suatu teori atau kebenaran, membangun fakta, menunjukkan deskripsi, serta menganalisa hasilnya dengan prosedur yang sistematis dengan data berupa numerical, angka, atau grafik. Data yang digunakan pada penelitian ini adalah data sekunder. Variabel Respon yang digunakan adalah Nilai Tukar Petani Tanaman Pangan, Sedangkan variabel bebas yaitu Nilai Tukar Petani Hortikultura. Data tersebut merupakan data bulanan pada Masa Pandemi Covid-19 dari bulan Januari-Juli 2021.

IV. HASIL DAN PEMBAHASAN

Pada pembahasan ini diuraikan tentang model regresi linier dengan metode *Bayesian*, dan estimasi parameter model tersebut. Adapun langkah-langkah dalam mengestimasi model regresi linier dengan pendekatan Bayesian adalah sebagai berikut:

4.1 Menentukan Fungsi Model Regresi Linier

Maka bentuk regresi linier dapat dituliskan sebagai:

$$y = x\beta + \varepsilon \tag{1}$$

4.2 Menentukan Estimasi Model Regresi dengan Metode OLS

Prinsipnya Metode OLS digunakan untuk estimasi parameter yaitu meminimumkan JKG.

$$JKG = \sum \varepsilon_i^2 = \sum \left(y_i - \hat{y}_i\right)^2 = \sum \left(y_i - \beta_0 - \beta_1 x_i\right)^2 \tag{2}$$

Turunkan persamaan (2) terhadap β_0 dan β_1, maka diperoleh:

$$\begin{aligned} \frac{\partial JKG}{\partial \beta_0} &= -2\sum\left(y_i - \beta_0 - \beta_1 x_i\right) \\ &= \sum y_i - n\beta_0 - \beta_1 \sum x_i \\ &= 0 \end{aligned} \tag{3}$$

dan

$$\begin{aligned} \frac{\partial JKG}{\partial \beta_1} &= -2\sum\left(y_i - \beta_0 - \beta_1 x_i\right)x_i \\ &= \sum y_i x_i - \beta_0 \sum x_i - \beta_1 \sum x_i x_i \\ &= 0 \end{aligned} \tag{4}$$

Sehingga dari persamaan (3) dan (4) diperoleh

$$\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x}_i \tag{5}$$

dan

$$\hat{\beta}_1 = \frac{\sum_{i=1}^{n} x_i y_i - \frac{1}{n}\sum_{i=1}^{n} x_i \sum_{i=1}^{n} y_i}{\left[\sum_{i=1}^{n} x_i^2 - \frac{1}{n}\left(\sum_{i=1}^{n} x_i\right)^2\right]} \tag{6}$$

Dengan demikian persamaan (6) diperoleh:

$$= \frac{\sum_{i=1}^{n}\left(x_i - \bar{x}\right)\left(y_i - \bar{y}\right)}{\sum_{i=1}^{n}\left(x_i - \bar{x}\right)^2}$$

4.3 Menentukan Model Regresi Bayesian Fungsi Likelihood

Fungsi *likelihood* adalah fungsi densitas probabilitas bersama dari n variabel random. Misal $T_1, T_2, ..., T_n$ adalah variabel random dengan pengamatannya adalah $t_1, t_{2,} ..., t_n$ fungsi *Likelihood* dapat ditulis dengan $F\left(t_1, t_2, ..., t_n; \theta\right)$. Untuk nilai $t_1, t_{2,} ..., t_n$ fungsi *likelihood* merupakan fungsi θ ditulis dengan $L(\theta)$. Jika $T_1, T_2, ..., T_n$ adalah variabel random yang independen,

$$\begin{aligned} L(\theta) &= \prod_{i=1}^{n} f\left(x_i; \theta\right) \\ &= f\left(x_1; \theta\right) f\left(x_2; \theta\right) ..... f\left(x_n; \theta\right) \end{aligned}$$

Selanjutnya yaitu menentukan fungsi *likelihood*nya, dimana akan diasumsikan vektor data $y_1, ..., y_n$ berdistribusi normal dengan rataan μ dan variansi σ^2 yang memiliki probabilitas yaitu

$$p\left(y_i \mid \beta_0, \beta_1, \sigma^2\right) = -\frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[-\frac{1}{2\sigma^2}\left(y - X\beta\right)^2\right]$$

Sehingga fungsi *Likelihood* n pengamatan berdasarkan pdf vektor data yaitu,

$$L\left(\beta_0, \beta_1, \sigma^2 \mid y\right) \propto -\frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[-\frac{1}{2\sigma^2}\sum\left(y_i - \left(\beta_0 + \beta_1 x_i\right)\right)^2\right] \tag{7}$$

4.4 Menentukan Estimasi Regresi Linier Bayesian Prior Non-Informatif

Distribusi *Prior Non-Informatif* diperoleh dengan cara mengalihkan densitas skala parameter β_0 dan β_1 atau

$p(\beta_0)$ dan $p(\beta_1)$ dengan densitas lokasi σ^2 atau $p(\sigma^2)$ [5] Sehingga distribusi *Prior Non-Informatif* gabungan untuk parameter β_0, β_1 dan σ^2 yaitu

$$\begin{aligned} p(\beta_0,\beta_1,\sigma^2) &= p(\beta_0)\,p(\beta_1)\,p(\sigma^2) \\ &= 1\times 1\times \frac{1}{\sigma^2} \\ &= \frac{1}{\sigma^2} \end{aligned}$$

Langkah selanjutnya yaitu menentukan distribusi *Posterior*nya dengan cara mengalihkan *Prior* dengan fungsi *likelihood.*

$$\begin{aligned} p(\beta_0,\beta_1,\sigma^2 \mid y) &\propto L(\beta_0,\beta_1,\sigma^2 \mid y)\, p(\beta_0,\beta_1,\sigma^2) \\ &\propto c(-\sigma)^{-\frac{n}{2}-2} \exp\left[-\frac{1}{2\sigma^2}\sum\left(y_i-(\beta_0+\beta_1 x_i)\right)^2\right] \end{aligned} \quad (8)$$

Berdasarkan hasil estimasi parameter yang dihasilkan pada regresi linear sederhana *Prior Non-Informatif* menggunakan Metode Bayes sama dengan hasil estimasi parameter dengan menggunakan Metode OLS.

$$\hat{\beta}_0 = \frac{\sum y_i - \hat{\beta}_1 \sum x_i}{n} \quad (9)$$

$$\hat{\beta}_1 = \frac{\sum (x_i-\hat{x})\sum (y_i-\hat{y})}{\sum (x_i-\hat{x})^2} \quad (10)$$

4.5 *Menerapkan Studi Kasus Nilai Tukar Petani di Provinsi Sulawesi Barat pada Masa Pandemi Covid-19 pada Bulan Januari-Juli 2021 yang Diasumsikan Menggunakan Bantuan Paket Program Windows Bayesian Inference Using Gibbs Sampling (WinBUGS) versi 1.4*

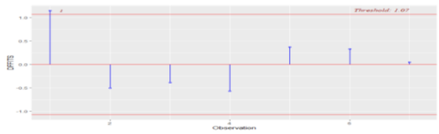
4.5.1 Mengecek Outlier pada Data untuk Metode OLS

Salah satu metode yang digunakan unutk mengecek adanya *Outlier* yaitu Metode DFFITS. Suatu data yang mempunyai nilai $|DFFITS| > 2\sqrt{\frac{k}{n}} = 2\sqrt{\frac{1}{7}} = 0{,}75$, Sehingga di defenisikan sebagai *Outlier*. Berikut hasil Metode DFFITS menggunakan Software R:

Tabel 1. Pendeteksian *outlier* dengan metode DFFITS

Data ke-i	|DFFITS|	Ket
1	1,14839	*Outlier*
2	-0,50247	Bukan *Outlier*
3	-0,38496	Bukan *Outlier*
4	-0,56934	Bukan *Outlier*
5	0,37150	Bukan *Outlier*
6	0,32952	Bukan *Outlier*
7	0,04687	Bukan *Outlier*

Berdasarkan Tabel 1. Diperoleh bahwa terdapat *outlier*, yaitu data ke-1. Dalam perhitungan DFFITS tersebut menunjukkan bahwa persamaan model regresi dari metode OLS memperlihatkan data *outlier*, sehingga estimasi parameter dengan metode OLS tersebut tidak dapat dijadikan sebagai acuaan utama.



Gambar 1. Plot Data yang Mengandung *Outlier*

Pada Gambar 1. menunjukkan grafik yang tidak beraturan atau terdapat data yang menyimpang dari data lain yaitu pada data ke-1, dimana disebut sebagai *outlier.* Data tersebut bisa saja berpengaruh terhadap estimasi parameter regresinya.

4.5.2. Estimasi Parameter dengan Metode OLS

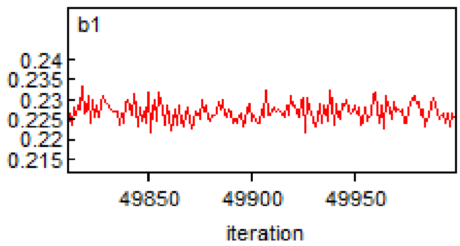
Hasil estimasi parameter menggunakan *Ordinary Least Square* (OLS) diperoleh dalam tabel 2

Tabel 2. Hasil estimasi paramete dengan OLS menggunakan R

Node	***Mean***	***SD***
β_0	77,9288	68,9867
β_1	0,2267	0,6320

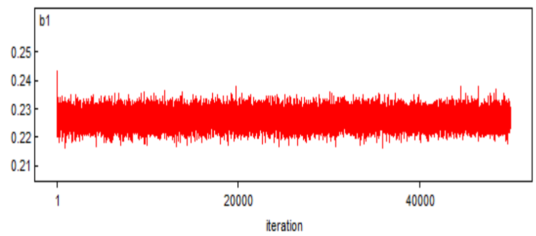
4.5.3 Estimasi Parameter dengan Metode Bayesian MCMC

Pada simulasi MCMC, iterasi yang dilakukan sampai sampel acak mencapai kekonvergenan, sampel dapat dilihat dari pola yang teratur pada grafik *Dynamic Trace.* Dalam penelitian ini, jumlah iterasi yang lakukan yaitu dengan jumlah *burn in* sebanyak 50000 iterasi. Berikut merupakan tampilan *Dynamic Trace Plot* dari salah satu parameter-parameter model regresi:



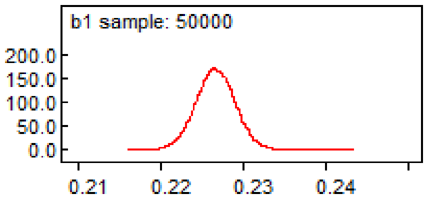
Gambar 2. *Dynamic Trace Plot* untuk parameter

Gambar 2. menunjukkan bahwa rantai markov yang diperoleh dari iterasi *Gibbs Sampler* untuk parameter β_1 membentuk pola yang teratur, sehingga dapat dikatakan konvergen. Begitu pula demikian dapat dilihat pada Gambar 3 berikut.U



Gambar 3. *Time Series*

Gambar 3 di atas menunjukkan bahwa grafik *time series* dimana menggambarkan keseluruhan grafik *trace plot* untuk semua data yang tersimpan. Data pada grafik *time series* lebih banyak dari data pada *dynamic trace*. Pola pada grafik *time series* terlihat tidak teratur, maka dapat dikatakan peubah acak tersebut telah konvergen, sedangkan distribusi *Posterior* dapat dilihat pada *Kernel Density*. Berikut tampilan *Kernel Density* dari salah satu parameter model regresi yaitu: Berikut tampilan *Kernel Density* dari salah satu parameter model regresi yaitu:



Gambar 4. *Kernel Density*

Pada Gambar 4. terlihat bahwa distribusi *Posterior* pada parameter β_1 menyerupai distribusi normal. Adapun hasil *mean dan standar device* dan *MC Error* yang dihasilkan dari estimasi parameter dengan Bayesian MCMC menggunakan WinBUGS versi 1.4 ditunjukkan pada tabel 3

Tabel 3. Hasil estimasi parameter dengan menggunakan WinBGUS versi 1.4

Node	*Mean*	*SD*	*MC Error*
β_0	77,93	0,1204	5,212E-4
β_1	0,2267	0,002387	1,17E-5

Tabel 3. menunjukkan bahwa nilai *mean* dan *standar deviation* adalah rata-rata dan simpangan baku dari beberapa sampel iterasi untuk masing-masing parameter. Nilai ini digunakan sebagai pendugaan dalam model regresi linear sederhana, sedangkan simpangan baku yang dihasilkan adalah simpangan baku beberapa sampel iterasi dan *MC Error* adalah ukuran variabilitas setiap estimasi dalam simulasi atau simpangan baku dari proses *Markov Chain* (Rantai Markov). Pada tabel 4.4 terlihat nilai *MC Error* kurang dari selang kepercayaan 5%, sehingga dapat disimpulkan bahwa nilai dari parameter $\beta_{0,}\beta_1$ yang diperoleh dari metode Bayesian pada WinBUGS sudah baik.

4.6 Uji signifikansi Parameter

Uji signifikansi parameter pada Bayesian yaitu melihat interval kepercayaan pada taraf signifikasi 95% yang ditandai dengan percentile $2{,}5\%$ dan $97{,}5\%$ dari *Posterior* yang tidak memuat nol. Nilai-nilai percentil dari *Posterior* disajikan dalam tabel berikut.

Tabel 4. Nilai percentil dari *posterior* parameter

Node	Interval Kepercayaan	
	Percentil $2{,}5\%$	**Percentil** $97{,}5\%$
β_0	77,69	78,17
β_1	0,2219	0,2314

Berdasarkan tabel 4 menunjukkan bahwa parameter β_0 β_1 dan σ tidak memuat nol pada $2{,}5\%$ dan $97{,}5\%$, artinya bahwa variabel Nilai Tukar Petani Tanaman Pangan (NTP-P) berpengaruhi signifikansi terhadap Nilai Tukar Petani Hortikultura (NTP-H)

4.7 *Perbandingan Model OLS dan Bayesian*

Hasil Model yang diperoleh menggunakan Metode OLS dan Metode Bayesian MCMC yaitu membandingkan nilai taksiran varian model. Dalam estimasi parameter yang perlu diperhatikan adalah tingkat kesalahannya semakin kecil tingkat kesalahannya suatu pendugaan maka semakin baik estimasinya.

Salah satu kriteria yang digunakan untuk melihat tingkat kesalahannya adalah MSE (*Mean Square Error*).

Tabel 5. Perbandingan model

Metode	*Mean Square Error* (MSE)
Metode OLS	0,8469
Metode Bayesian	0,3723

Berdasarkan Tabel 5, Menunjukkan bahwa nilai MSE dalam metode OLS lebih besar yaitu 0,8469 daripada MSE dalam metode Bayesian yaitu 0,3723, maka dapat dikatakan bahwa metode Bayesian lebih baik daripada metode OLS.

V. KESIMPULAN

Berdasarkan hasil estimasi parameter regresi linier dengan pendekatan metode Bayesian yaitu:

$$\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x}_i$$

$$\hat{\beta}_1 = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^{n}(x_i - \bar{x})^2}$$

Uji Signifikansi parameter β_0 dan β_1 pada metode Bayesian selang kepercayaan 95% untuk masing-masing hasil estimasi parameter MCMC adalah:

$$\beta_0\,(77{,}69; 78{,}17)$$

$$\beta_1\,(0{,}2219; 0{,}2314)$$

Dapat dilihat bahwa nilai MSE yang diperoleh dari estimasi Metode Bayesian lebih kecil dibandingkan dengan nilai MSE yang diperoleh dari estimasi Metode OLS. Nilai MSE yang diperoleh dari Metode OLS yaitu 0,8469 sedangkan Nilai MSE yang diperoleh dari Metode Bayesian MCMC yaitu 0,3723. Karena nilai MSE Metode Bayesian lebih kecil daripada nilai Metode OLS, Hal ini menunjukkan bahwa Metode Bayesian MCMC jauh lebih baik dibandingkan Metode OLS.

REFERENSI

[1] Diana SP, Tanty H. 2018. Linear Regression Model Using Bayesian Approach.For Energy Perfomance Of Residental Building. Journal Procedia. Computer Science 672-677. Bina Nusantara University.

[2] Farahdiba, D. 2020, *Estimasi parameter pada model regresi linier menggunakan metode Bayesian dengan distribusi prior konjugat dan non-informatif jeffreys* (Bachelor’s thesis, Fakultas Sains dan Teknologi Universitas Islam Negeri Syarif Hidayatullah Jakarta.

[3] Katianda, K. R., dkk., 2021, Estimasi Parameter Model Regresi Linier dengan Pendekatan Bayes. *JURNAL EKSPONESIAL*, *11*(2), 127-132.

[4] Ainul, A.M., dkk. 2018. Penerapan Model Analisis Regresi Linier Berganda dengan Pendekatan Bayesian pada Data Aset Bank di Indonesia. Jurnal Keteknikan dan Sains (JUTEKS). 1(1). 41-47.

[5] Br Damanik, F.A. 2020. Analsis Model Regresi Linier Dengan Menggunakan Metode Bayes.