Jurnal Matematika UNAND
Vol. **VII** No. **1** Hal. 115 – 124
ISSN : 2303–291X


# PERBANDINGAN METODE KUADRAT TERKECIL DAN METODE BAYES PADA MODEL REGRESI LINIER DENGAN GALAT YANG AUTOKORELASI

RIDHA KHAIRIYAH, MAIYASTRI, RITA DIANA

*Jurusan Matematika,*
*Fakultas Matematika dan Ilmu Pengetahuan Alam, Universitas Andalas,*
*Kampus UNAND Limau Manis Padang, Indonesia,*
*email : ridhakhairiyah6@gmail.com*

**Abstrak.** Analisis regresi merupakan metode dalam statistik yang digunakan untuk melihat hubungan antara variabel bebas dengan variabel tak bebas. Model regresi linier sederhana melibatkan satu variabel tak bebas dan satu variabel bebas. Dalam regresi linier sederhana, metode yang biasa digunakan dalam menduga parameter regresi adalah Metode Kuadrat Terkecil (MKT). Pendugaan parameter dengan menggunakan MKT harus memenuhi asumsi-asumsi tertentu terhadap galatnya yang dinamakan dengan asumsi klasik. Jika salah satu asumsi tidak terpenuhi seperti terjadinya autokorelasi maka pendugaan dengan MKT tidak efisien. Oleh karena itu diperlukan metode pendugaan lain yaitu metode Bayes. Penelitian ini menunjukkan bahwa metode Bayes menghasilkan MSE lebih kecil dibandingkan dengan MKT, sehingga metode Bayes dapat mengatasi kasus galat autokorelasi dari penduga metode OLS.

*Kata Kunci*: Analisis Regresi Linier, Metode Kuadrat Terkecil, Metode Bayes, Galat Autokorelasi

## 1. Pendahuluan

Regresi secara umum adalah metode statistik yang memberikan penjelasan tentang pola hubungan (model) antara dua variabel atau lebih. Dalam analisis regresi dikenal dua jenis variabel yaitu, variabel bebas dan variabel terikat. Dalam regresi linier sederhana, metode yang biasa digunakan dalam mengestimasi parameter regresi adalah Metode Kuadrat Terkecil atau *Ordinary Least Squares.* Estimasi parameter model regresi yang diperoleh dengan Metode Kuadrat Terkecil ini merupakan estimator yang baik jika model regresi memenuhi asumsi-asumsi tertentu yang disebut asumsi model regresi linier klasik. Salah satu asumsi dari model regresi linier klasik adalah tidak adanya serial korelasi antara galat yang dinamakan autokorelasi.

Pendekatan Metode Kuadrat Terkecil mengandalkan proses inferensia pada data sampel yang diambil dari populasi sedangkan pendekatan Bayes, disamping memanfaatkan data sampel yang diperoleh dari populasi juga memperhitungkan suatu distribusi awal yang disebut *prior.* Inferensia statistik dengan pendekatan Bayes berbeda dengan pendekatan Metode Kuadrat Terkecil. Pendekatan Metode Kuadrat Terkecil, parameter yang akan diestimasi mempunyai ukuran yang bernilai tetap sedangkan pendekatan Bayes memandang semua parameter yang tidak diketahui sebagai peubah acak yang memiliki distribusi disebut distribusi *prior.*

## 2. Kajian Pustaka

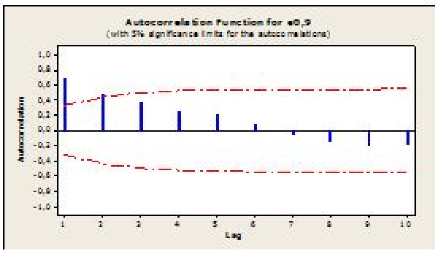
Kajian pustaka yang digunakan dalam tulisan ini adalah sebagai berikut.

### 2.1. *Autokorelasi*

Autokorelasi berarti adanya korelasi anggota observasi satu dengan observasi yang lain yang berlainan waktu[2]. Dalam kaitannya dengan asumsi metode OLS, autokorelasi merupakan korelasi antara satu variabel galat dengan variabel galat yang lain. Sedangkan salah satu asumsi penting metode OLS berkaitan dengan variabel galat adalah tidak adanya hubungan antara galat yang satu dengan galat yang lain.

### 2.2. *Mendeteksi Autokorelasi*

Sebelum melakukan sesuatu penting untuk mengetahui apakah autokorelasi ada dalam suatu situasi tertentu. Uji *Autocorrelation Function* (ACF) adalah uji yang biasa digunakan untuk mengetahui ada tidaknya autokorelasi [2]. Uji ACF menguji adanya autokorelasi pada lag-1, lag-2, lag-3, dan seterusnya. Langkah pertama dalam uji ini adalah regresikan antara y dengan x sehinggan diperoleh residual. Pada uji ACF, kasus autokorelasi terjadi ketika ada lag pada plot ACF yang keluar batas signifikansi.




Gambar 1. Plot ACF yang mengandung Autokorelasi

### 2.3. *Metode Kuadrat Terkecil*

Analisis regresi adalah suatu analisis yang bertujuan untuk menunjukkan hubungan matematis antara variabel bebas dengan variabel terikat. Bila model antara variabel terikat dan variabel bebas adalah linier maka model disebut model regresi linier. Model regresi linier yang terdiri dari satu variabel terikat dengan satu variabel bebas disebut model regresi linier sederhana. Persamaan umum regresi linier sederhana adalah

$$y_i = \beta_0 + \beta_1 x_i + \varepsilon_i, \; i = 1, 2, 3, \cdots, n,$$

dengan:

$$\begin{aligned} y_i &: \text{ nilai variabel terikat (tidak bebas)}, \\ x_i &: \text{ nilai variabel bebas}, \\ \beta_0, \beta_1 &: \text{ parameter regresi}, \\ \varepsilon_i &: \text{ nilai galat.} \end{aligned}$$

Salah satu metode untuk mengestimasi parameter regresi adalah Metode Kuadrat Terkecil. Metode Kuadrat Terkecil digunakan untuk menduga nilai $\beta_0$ dan $\beta_1$ sehingga akan didapat suatu dugaan model regresi, yaitu [2]:

$$\hat{y_i}\hat{\beta_0} + \hat{\beta_1}x_i, \ i = 1, 2, 3, \cdots, n,$$

dimana :

$$\begin{aligned} \widehat{y_i} &: \text{ nilai variabel terikat (tidak bebas)}, \\ x_i &: \text{ nilai variabel bebas}, \\ \widehat{\beta_0}, \ \widehat{\beta_1} &: \text{ parameter regresi.} \end{aligned}$$

Metode Kuadrat Terkecil merupakan metode yang digunakan untuk menentukan parameter model regresi dengan meminimumkan jumlah kuadrat sisa (JKS). Sehingga diperoleh

$$\widehat{\beta_0} = \bar{y} - b_1\bar{x} \text{ dan } \widehat{\beta_1} = \frac{n\sum x_iy_i - \sum x_i \sum y_i}{n\sum x_i^2 - (\sum x_i)^2}.$$

## 2.4. *Metode Bayes*

Metode Bayes merupakan suatu metode dalam statistika inferensia yang menggabungkan informasi dari data pengamatan dengan informasi data sampel yang telah diperoleh sebelumnya. Metode Bayes memiliki dua jenis distribusi yaitu distribusi *Prior* dan distribusi *Posterior*.

(1) Distribusi Prior.
Dalam membentuk distribusi posterior parameter model diperlukan informasi sampel yang dinyatakan dengan fungsi likelihood dan informasi awal yang dinyatakan sebagai distribusi prior. Ada beberapa distribusi prior yang dikenal dalam metode Bayes, yaitu :
   (a) *Conjugate prior* atau *non conjugate prior* merupakan prior yang ditentukan berdasarkan pada pola likelihood dari data atau tidak [1].
   (b) *Informative prior* atau *non-informative prior* merupakan *prior* yang ditentukan berdasarkan pada ketersediaan informasi sebelumnya mengenai pola distribusi data. Informasi tersebut dapat berasal dari penelitian sebelumnya [1].

(2) Distribusi Posterior.
Misalkan data observasi $\mathbf{y} = (y_1, y_2, \cdots, y_n)'$ mempunyai distribusi tertentu dengan himpunan parameter $\theta = (\theta_1, \theta_2, \cdots, \theta_n)'$ yang merupakan peubah

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

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

dengan

$$f(\mathbf{y}) = \int_{\Omega\theta 1} \cdots \int_{\Omega\theta p} f(\mathbf{y}|\theta)f(\theta)d\theta_1 \cdots d\theta_n \text{ adalah normalizing constant,}$$

$$f(\theta) \quad \text{adalah distribusi prior dari parameter } \theta,$$

$$f(\mathbf{y}|\theta) = \prod f(y_i|\mathbf{y}) \text{ adalah fungsi Likelihood data.}$$

### 2.5. *Markov Chain Monte Carlo*

Metode MCMC merupakan metode estimasi parameter model dengan menggunakan teknik simulasi numerik yang digunakan untuk menyelesaikan masalah pemodelan yang kompleks. Untuk model yang cukup kompleks, proses integrasi tersebut sangat sulit sehingga untuk mendapatkan distribusi posterior marginal dilakukan teknik simulasi numerik berupa MCMC [3]. Ide dasar dari MCMC adalah membangkitkan data sampel dari ditribusi posterior setiap parameter model sesuai proses Markov Chain dengan menggunakan simulasi Monte Carlo secara iteratif hingga diperoleh kondisi yang konvergen. Kondisi tersebut dinamakan kondisi stasioner atau *equilibrium*.

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

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

### 2.6. *Gibbs Sampling*

Penggunaan metode MCMC dalam analisis Bayes memerlukan algoritma sampling yang sesuai untuk mendapatkan sampel dari suatu distribusi. Salah satu algoritma yang sering digunakan adalah *Gibbs Sampling* [3]. *Gibbs sampling* merupakan teknik pengambilan sampel untuk membangkitkan peubah acak dari suatu distribusi marginal secara langsung tanpa harus menghitung fungsi densitasnya. Proses *Gibbs Sampling* dilakukan dengan cara membangkitkan rangkaian Gibbs peubah acak 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 berautokorelasi kemudian nilai $x$ ditetapkan dengan $x_i \sim unif(0, n)$, dan parameter $\beta_0 = 2$ dan $\beta_1 = 3$ sehingga nilai $y_i$ dimana $i = 1, 2, \cdots, n$ diperoleh dengan persamaan

$$y_i = \beta_0 + \beta_1 x_i + \varepsilon_i, i = 1, 2, \cdots, n.$$

Untuk melihat pengaruh autokorelasi, nilai sisaan dibangkitkan melalui model autoregresif tingkat satu *(first-order autoregresive)* yaitu:

$$\varepsilon_i = \rho\varepsilon_{i-1} + \delta_i,$$

untuk $i = 1, 2, \cdots, n$. Prosedur yang dilakukan adalah sebagai berikut.

(a) Tetapkan $\rho$.
(b) Bangkitkan secara acak n nilai $\delta_i$ dari sebaran normal dengan nilai tengah nol dan ragam satu, $\delta_i \sim N(0, 1)$
(c) Tetapkan $\varepsilon_0 = 5$.
(d) Hitung

$$\varepsilon_i = \rho\varepsilon_{i-1} + \delta_i, \ i = 1, 2, \cdots, n.$$

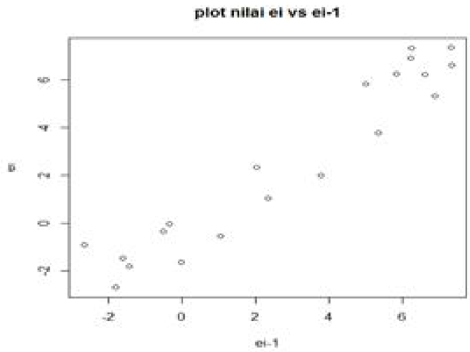
Jadi diperoleh galat yang memiliki sifat autokorelasi tingkat satu. Hitung data :

$$y_i = \beta_0 + \beta_1 x_i + \varepsilon_i.$$

Banyak data direncanakan sebanyak 20, 30 dan 40 data. Percobaan dilakukan dengan simulasi Monte Carlo dengan ulangan sebanyak 2500, 5000, dan 10000 kali untuk nilai koefisien autokorelasi 0.9, 0.7, 0.5, $-0.5$, $-0.7$ dan $-0.9$. Selanjutnya dilakukan pendugaan parameter $\beta_0$ dan $\beta_1$ berdasarkan Metode Kuadrat Terkecil dan Metode Bayes.

### 3.1. *Mendeteksi Autokorelasi*

Pada subbab ini akan diuji apakah galat yang dibangkitkan pada data sebelumnya telah mengalami autokorelasi, yaitu untuk pengamatan ini menggunakan uji ACF. Berdasarkan data yang dibangkitkan dengan jumlah sampel $n = 20$, $= 0,9$ dan iterasi sebanyak 2500 dilakukan uji ACF menggunakan software minitab 16 dan diperoleh data yang dibangkitkan mengalami autokorelasi. Pada tiap perlakuan berikutnya lakukan hal yang sama dengan ukuran sampel dan ragam serta iterasi pada data yang dibangkitkan dan diperoleh bahwa data yang dibangkitkan mengalami autokorelasi.




Gambar 2. Plot Galat yang mengandung Autokorelasi

### 3.2. *Estimasi Paramater Model Regresi Linier Sederhana dengan Metode Bayes*

Jika $y_i \sim N(\beta_0+\beta_1 x_i, \sigma^2)$ dan $\sigma^2 = \frac{1}{\tau}$ adalah parameter presisi yang nilainya tetap, maka bentuk fungsi kepekatan peluang dari $y_i$ sebagai berikut :

$$f(y_i|\beta_0, \beta_1, \tau) = \frac{\sqrt{\tau}}{\sqrt{2\pi}} \exp(-\frac{\tau}{2}(y_i - (\beta_0 + \beta_1 x_i)))^2.$$

Dari fkp diatas, diperoleh fungsi likelihoodnya yaitu:

$$\begin{aligned} L(\beta_0, \beta_1, \tau) &= \prod_{i=1}^{n}(f(y_i|\beta_0, \beta_1, \tau), \\ &= \prod_{i=1}^{n} \frac{\sqrt{\tau}}{\sqrt{2\pi}} \exp(-\frac{\tau}{2}(y_i - (\beta_0 + \beta_1 x_i))^2), \\ &= \frac{\sqrt{\tau}}{\sqrt{2\pi}} \exp(-\frac{\tau}{2}(y_1 - (\beta_0 + \beta_1 x_1))^2) \frac{\sqrt{\tau}}{\sqrt{2\pi}} \exp(-\frac{\tau}{2}(y_2 - (\beta_0 + \beta_1 x_2))^2) \cdots \\ &\quad \frac{\sqrt{\tau}}{\sqrt{2\pi}} \exp(-\frac{\tau}{2}(y_n - (\beta_0 + \beta_1 x_n))^2) \\ &\propto \ (\frac{\sqrt{\tau}}{\sqrt{2\pi}})^n \exp(-\frac{\tau}{2}\sum_{i=1}^{n}(y_i - (\beta_0 + \beta_1 x_i))^2). \end{aligned}$$

Tipe distribusi *prior* yang digunakan adalah *conjugate*, dengan distribusi *prior* yang digunakan untuk parameter $\beta_0$ dan $\beta_1$ adalah distribusi normal yang dinyatakan sebagai berikut.

$$\beta_j \sim N(\mu_j, \sigma_j^2), \ j = 0, 1.$$

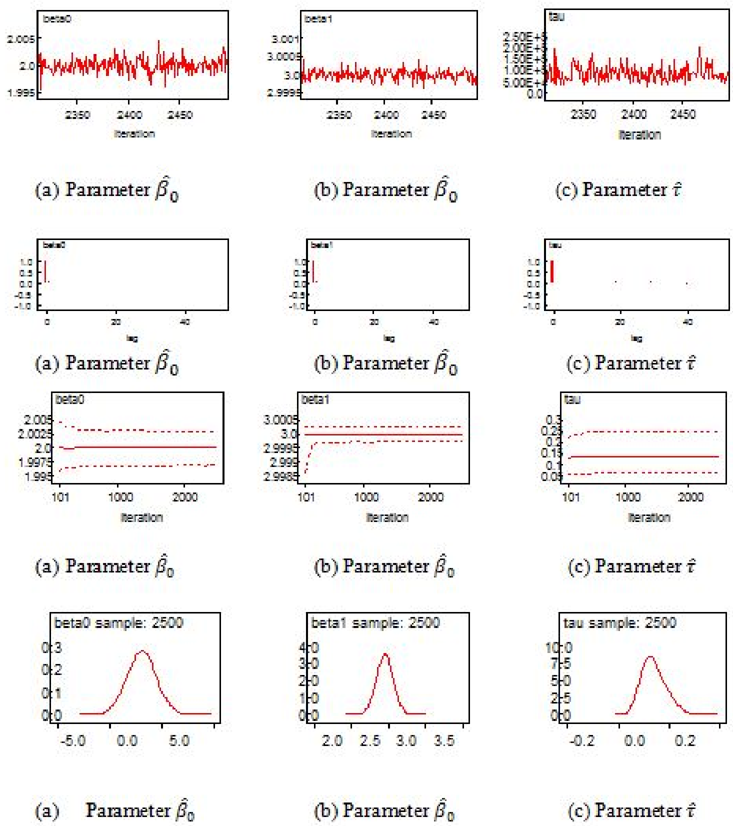
Dan distribusi *prior* yang digunakan untuk parameter $\tau$ adalah distribusi gamma yang dinyatakan sebagai berikut.

$$\tau \sim Gamma(a, b).$$

Sehingga diperoleh bentuk proposional distribusi *posterior* untuk sebagai berikut:

$$
\begin{aligned}
f(\beta_0, \beta_1, \tau | y_i) &\propto f(y_i|\beta_0, \beta_1, \tau) f(\beta_0, \beta_1, \tau), \\
&\propto (\frac{\sqrt{\tau}}{\sqrt{2\pi}})^n \exp(-\frac{\tau}{2}\sum_{i=1}^{n}(y_i - (\beta_0 + \beta_1 x_i))^2). \\
&\propto \ \frac{\sqrt{\tau}_{\beta_0}}{\sqrt{2\pi}} \exp(-\frac{\tau_{\beta_0}}{2}(\beta_0 - \mu_{\beta_0})^2). \\
&\propto \ \frac{\sqrt{\tau}_{\beta_1}}{\sqrt{2\pi}} \exp(-\frac{\tau_{\beta_1}}{2}(\beta_1 - \mu_{\beta_1})^2). \\
&\propto \ \frac{1}{\Gamma(a)b^a}\tau^{a-1} \exp(\frac{-\tau}{b}).
\end{aligned}
$$

Dari distribusi *posterior* diatas, dengan menggunakan data yang dibangkitkan dengan sampel $n = 20$, iterasi sebanyak 2500 kali dilakukan proses MCMC dengan algritma *Gibbs Sampling* menggunakan bantuan software WinBugs dan diperoleh hasil pada Gambar 3.




Gambar 3. Diagnostic plot pada proses *Markov Chain Monte Carlo*

*Trace plot* parameter menunjukkan nilai $\widehat{\beta_0}$ , $\widehat{\beta_1}$, $\widehat{\tau}$ acak dengan pola yang stasioner dan bersifat random. Kemudian*Autocorrelation plot* parameter $\widehat{\beta_0}$ , $\widehat{\beta_1}$, $\widehat{\tau}$ yang menunjukkan terpenuhinya sifat*irreducible*, *aperiodic*, dan*recurrent*. *Quantiles plot* menunjukkan *ergodic mean* hasil estimasi parameter $\widehat{\beta_0}$ , $\widehat{\beta_1}$, $\widehat{\tau}$ yang dihasilkan telah mencapai nilai yang stabil. Plot fungsi kepadatan posterior koefisien regresi model regresi linier sederhana dengan pendekatan Bayes yang diperoleh sesuai dengan distribusi *prior* yang digunakan untuk parameter $\widehat{\beta_0}$ , $\widehat{\beta_1}$ yaitu distribusi normal dan untuk parameter $\widehat{\tau}$ yaitu distribusi gamma.

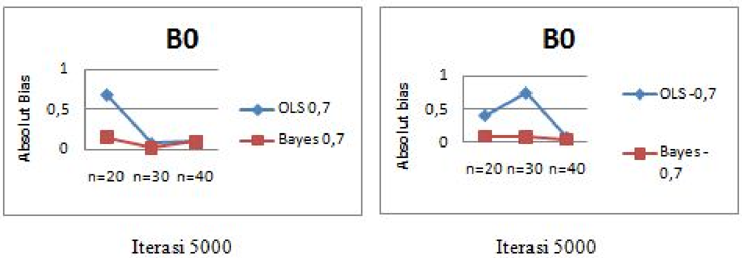
Ringkasan *posterior* parameter yang diperoleh dari proses estimasi paramater model regresi linier sederhana dapat dilihat pada tabel pada Gambar 4 berikut.

| No | Parameter | Nilai Estimasi | Absolut bias |
|---|---|---|---|
| 1 | $\hat{\beta}_0$ | 1.7817 | 0.2183 |
| 2 | $\hat{\beta}_1$ | 3.0103 | 0.0103 |

Gambar 4. Ringkasan *posterior* parameter

### 3.3. *Perbandingan Penduga Paramater Regresi $\beta_0$ dan $\beta_1$ dengan Menggunakan Metode Kuadrat Terkecil (MKT) dan Metode Bayes*

(a) Absolut Bias.
Grafik perbandingan nilai absolut bias dari $\beta_0$ penduga Metode Kuadrat Terkecil dan Bayes dengan iterasi sebanyak 5000 kali untuk ukuran sampel dan nilai $\rho$ akan disajikan dalam Gambar 5.

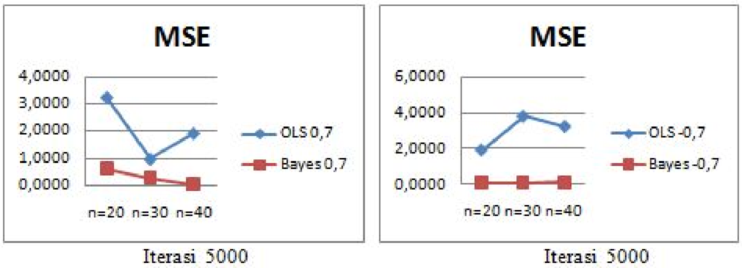



Gambar 5. Nilai Absolut Bias $\beta_0$ MKT dan Bayes

Gambar 5 menunjukkan nilai absolut bias $\beta_0$ penduga Metode Kuadrat Terkecil dan Bayes untuk koefisien autokorelasi $\rho$ yang berbeda-beda. Pada grafik diatas juga terlihat bahwa untuk kedua koefisien yaitu $\rho = 0,7$ dan

$\rho = -0,7$ nilai absolut bias $\beta_0$ Bayes lebih kecil dari pada nilai absolut bias $\beta_0$ Metode Kuadrat Terkecil.

(b) *Mean Square Error* (MSE).



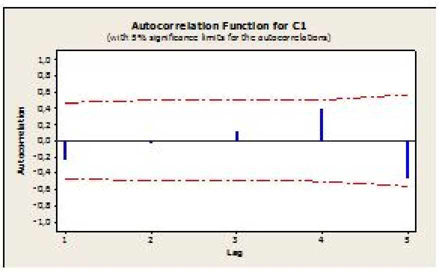

Gambar 6. Nilai MSE Metode Kuadrat Terkecil dan Metode Bayes

Gambar 6 menunjukkan nilai MSE dari penduga Metode Kuadrat Terkecil dan Bayes untuk $\rho = 0,7$ dan $\rho = -0,7$ pada iterasi 5000. Pada grafik diatas terlihat bahwa nilai MSE Bayes bergerak lebih kecil dan stabil dari pada nilai MSE Metode Kuadrat Terkecil.

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

### 3.4. *Mendeteksi Galat pada Metode Bayes*

Pada subbab ini akan dibahas apakah galat pada metode Bayes masih mengandung autokorelasi atau tidak. Berdasarkan data yang dibangkitkan dengan jumlah sampel n=20, $\rho = -0,9$, dengan iterasi sebanyak 2500 dilakukan uji ACF menggunakan software minitab 16. Terlihat tidak ada lag yang keluar dari batas signifikansi sehingga galat bebas dari adanya autokorelasi. Hal yang sama untuk ukuran sampel dan koefisien autokorelasi yang berbeda.




Gambar 7. Plot ACF galat yang tidak mengandung autokorelasi

## 4. Kesimpulan

(1) Dari percobaan simulasi pada Metode Kuadrat Terkecil semakin tinggi nilai koefisien autokorelasi $\rho$ maka nilai absolut bias dan MSE semakin besar yang berarti semakin tinggi nilai koefisie nautokorelasi $\rho$ pengaruh autokorelasi semakin substansial.
(2) Penduga dengan metode Bayes memiliki nilai MSE yang lebih kecil dari penduga Metode Kuadrat Terkecil.
(3) Penduga dengan metode Bayes dapat mengatasi kasus galat autokorelasi dari penduga Metode Kuadrat Terkecil.

## 5. Ucapan Terima Kasih

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

## Daftar Pustaka

[1] Box, G.E.P and Tiao, G.C. 1973. *Bayesian Inference in Statistical Analysis.* Addision-Wesley Publishing Company, Inc; Philippines.
[2] Gujarati, Damodar. 1997. *Ekonometrika Dasar*. Jakarta. Erlangga.
[3] Ntzoufras, Ioannis. 2009. *Bayesian Modeling Using WinBUGS.*Canada. John Wiley & Sons, Inc.