HASIL CEK_60110621

by Suparman 60110621

Bayesian model selection in Moving Average using reversible jump MCMC: A comparison between Laplacian and Gaussian Noises

Suparman[1*], Abdellah Salhi[2]

1Department of Mathematics Education, Universitas Ahmad Dahlan, Yogyakarta, Indonesia

2Department of Mathematical Sciences, University of Essex, Colchester, United Kingdom

*E-mail:suparman@pmat.uad.ac.id

Abstract

Moving average (MA) is among the time series models that is often used for pattern forecasting and recognition. It contains a noise that is often assumed to have a Gaussian distribution. However, in various applications, noise often does not have this distribution. Therefore, this paper proposed the use of Laplacian noise in the MA model. Meanwhile, the comparison of Gaussian and Laplacian were also investigated to ascertain the right noise for the model. Also, the Bayesian method was used to estimate the parameters, such as the order and coefficient of the model, as well as noise variance. The posterior distribution has a complex form because the parameters are in combination of spaces with different dimensions. Therefore, to overcome this problem, the Markov chain Monte Carlo (MCMC) reversible jump algorithm was adopted. A simulation study was conducted to evaluate its performance. After it worked properly, it was applied to model the human heart rate data. Therefore, the results showed that the MCMC algorithm can estimate the parameters of MA model. This was developed using Laplace distributed noise. Meanwhile, when compared with Gaussian, the Laplacian resulted in a higher order model and produced a smaller variance.

Keywords: Bayesian, Laplacian, moving average, reversible jump MCMC

1. Introduction

Moving average (MA) is often used to model data in various fields of life. Studies on its application is seen in several literatures, such as (Reghunath, 2005), (Akram, 2013), (Adin, 2017), (Silva de Souza, 2018), and (Gautam, 2019). In fact, Reghunath et al. (2005) used MA to analyze water resource data. The model was used to determine the long-term trends in groundwater fluctuations. To eliminate seasonal effects on this fluctuation, the 12th order MA was used to model water source data. Furthermore,

Akrami et al. (2014) used it to analyze rainfall data. To observe long and short-term trends, several MA models with different orders were examined. Silva de Souza et al. (2018) used it for profitability checks in technical analysis on the stock market. An automated trading system was developed based on moving averages from previous prices. Meanwhile, Gautam and Abhishekh (2019) used the model to analyze the data of fuzzy time series by reducing fluctuations.

Researchers like (Middleton, 2009) and (Safi, 2013) often use the Gaussian distributed noise. Meanwhile, Middleton et. al. (2009) used the type that contained Gaussian to model the noise channel. Safi (2013) examined the heteroscedastic autocorrelation function of residue in MA. Therefore, in determining this functions, the model was assumed to have Gaussian noise.

In various applications, mathematical models often have non-Gaussian noise, such as (Jureckova, 2009), (Zhang, 2015), (Suparman, 2018), and (Suparman, 2019). Jureckova et. al. (2009) examined non-parametric tests in linear autoregressive models that use Pareto noise. Also, Zhang (2015) examined the autoregressive type that uses G-GARCH as noise. Suparman (2018) developed a stationary model that has exponentially distributed noise. In addition, an inversible moving range type was developed, which has exponentially distributed noise. In this autoregressive and moving average model, the order is assumed to be unknown, hence the posterior distribution has a complicated shape. To determine the Bayesian estimator, the research adopted the MCMC reversible jump algorithm.

Conversely, Laplace distributed noise was also investigated by several researchers, such as (Minchole, 2014) and (Miertoi, 2019). Minchole et. al. (2014) detected that changes in body position from the ECG use Laplacian noise. Also, Miertoiu and Dumitrescu (2019) investigated the signal representation that rarely uses Laplacian noise.

However, MA with Laplacian have not been studied. Therefore, this research proposed the development of the model with Laplacian, which is then used for the human heart rate data. The Laplacian was also compared with Gaussian to determine the effect of noise on MA models.

2. Method

The Bayesian method was used to estimate the MA model parameters, such as its order, coefficients, and noise. In this method, the parameters were treated as random variables that are assumed to have a certain distribution, known as "prior." This was combined with the likelihood function for data, which produced a posterior distribution. In this case, the order of the MA model is also a parameter. Therefore, the shape of posterior distribution becomes very complex, which makes it difficult to explicitly determine the Bayes estimator. To determine this, the MCMC reversible jump algorithm was adopted (Green, 1995). The basic idea of this was to create a Markov chain. This chain was designed, hence its limit distribution approached the posterior for the parameters. Also, the algorithm uses 3 transformations, such as birth and death of the order, as well as coefficient change. The performance of this algorithm was evaluated using simulation studies and it was applied to model human heart rate data.

3. Results and Discussion

This section explained the functions of data, posterior and prior distribution, as well as MCMC reversible jump in determining Bayes estimators.

3.1. Likelihood function

For example, $x_1, \ldots, x_n$ are data and n is the amount. Therefore, it is said to be modeled by the q order MA, written as $MA(q)$, when it satisfies the following equation

$$x_t = z_t + \sum_{j=1}^{q} \theta_j z_{t-j} \qquad (1)$$

The magnitudes of $\theta_1, \dots, \theta_q$ express the coefficients of the $MA(q)$ model. Meanwhile, the variables of z_t ($t = 1, \dots, n$) express noise. This was assumed to be a Laplace distribution with β parameter. Therefore, the probability function for z_t is written as follows

$$g(z_t|\beta) = \frac{1}{2\beta} exp - \frac{|z_t|}{\beta} \tag{2}$$

Furthermore, the probability function for x_t can be determined using the variable transformation between x_t and z_t. The function of x_t can be stated as follows

$$g(x_t|\beta) = \frac{1}{2\beta} exp - \frac{|\sum_{j=1}^{q} \theta_j z_{t-j} + z_t|}{\beta} \tag{3}$$

For example, $x = (x_1, \dots, x_n)$ and $\theta^{(q)} = (\theta_1, \dots, \theta_q)$. Hence, the function can be stated as follows

$$f(x|q, \theta^{(q)}, \beta)$$

$$= \prod_{t=q+1}^{n} \frac{1}{2\beta} exp - \frac{|\sum_{j=1}^{q} \theta_j z_{t-j} + z_t|}{\beta}$$

$$= \left(\frac{1}{2\beta}\right)^{n-q} exp \tag{4}$$

$$- \frac{1}{\beta} \sum_{t=q+1}^{n} \left| \sum_{j=1}^{q} \theta_j z_{t-j} + z_t \right|$$

The amount of z_t is calculated using the following equation

$$z_t = x_t - \sum_{j=1}^{q} \theta_j z_{t-j} \tag{5}$$

for $t = q + 1, \dots, n$ and $z_t = 0$ for t = 1, 2, ... ,q.

The $MA(q)$ model is inversible when the coefficients of $\theta_1, \dots, \theta_q$ fulfill the I_q inversibility area where

$$I_q = \left\{ \theta^{(q)} \in R^q \middle| 1 + \theta_1 a + \theta_2 a^2 + \cdots + \theta_q a^q \neq 0, |a| \leq 1 \right\}.$$

For large order values, the area is very difficult to identify. Therefore, a transformation is needed to overcome this problem. For example, $\rho_1, \rho_2, \cdots, \rho_q$ expresses the inverse partial autocorrelation functions that correspond to the $MA(q)$ model. G is a transformation from $(\theta_1, \dots, \theta_q) \in I_q$ to

$\left(\rho_1, \rho_2, \cdots, \rho_q\right) \in (-1,1)^q$ where I_q is the inversibility area (Bhansali, 1983). Therefore, this transformation can easily be used to identify areas of inversibility even for large orders. The $MA(q)$ model is inversible only when $\left(\rho_1, \rho_2, \cdots, \rho_q\right) \in (-1,1)^q$. Also, the function for x can be rewritten as follows

$$f\left(x|q, \theta^{(q)}, \beta\right)$$
$$= \prod_{t=q+1}^n \frac{1}{2\beta} exp$$
$$-\frac{\left|\sum_{j=1}^q G^{-1}(\rho_j)z_{t-j} + z_t\right|}{\beta}$$
$$= \left(\frac{1}{2\beta}\right)^{n-q} exp \qquad (6)$$
$$-\frac{1}{\beta} \sum_{t=q+1}^n \left|\sum_{j=1}^q G^{-1}(\rho_j)z_{t-j} + z_t\right|$$

where G^{-1} is the inverse transformation of G.

3.2. Prior distribution

The MA parameter is seen as a variable with prior distribution. The distribution for this model order are Binomial with q_{max} and λ parameters.

$$\pi(q|\lambda) = C_q^{q max} \lambda^q (1-\lambda)^{q max - q} \qquad (7)$$

The value of q_{max} is limited to order 10. Meanwhile, the binomial distribution was chosen because it is a conjugate. For example, $\rho^{(q)} = \left(\rho_1, \rho_2, \cdots, \rho_q\right)$. The prior or $\rho^{(q)}$ given the value of order q is a uniform distribution at intervals $(-1, 1)^q$

$$\pi\left(\rho^{(q)}|q\right) = \frac{1}{2^q} \qquad (8)$$

Finally, the prior distribution for β is the Gamma inverse with parameters u and v

$$\pi(\beta|u, v) = \frac{v^u}{\Gamma(u)} \beta^{-(u+1)} exp - \frac{v}{\beta} \qquad (9)$$

Where $u = 1$. The prior distribution for MA model order contains hyperparameter λ and that for β contains v. The hyperprior distribution for λ is uniform at intervals (0.1). Meanwhile, v is of Jeffreys

$\pi(v) \propto \frac{1}{v}$. Therefore, the combined distribution of prior and hyperprior can be written as follows

$$\pi\left(q, \rho^{(q)}, \lambda, \beta, v\right)$$
$$= C_q^{q\,max}\lambda^q(1 \tag{10}$$
$$- \lambda)^{q\,max-q}\frac{1}{2^q}\frac{v^u}{\Gamma(u)}\beta^{-(u+1)}exp - \frac{v}{\beta}\frac{1}{v}$$

3.3. Posterior distribution

By using the Bayes theorem, the posterior distribution for $\left(q, \rho^{(q)}, \lambda, \beta, v\right)$ can be stated as follows

$$\pi\left(q, \rho^{(q)}, \lambda, \beta, v\middle|x\right)$$
$$= \left(\frac{1}{2\beta}\right)^{n-q} exp$$
$$-\frac{1}{\beta}\sum_{t=q+1}^{n}\left|\sum_{j=1}^{q}G^{-1}(\rho_j)z_{t-j} + x_t\right|$$
$$C_q^q\lambda^q(1-\lambda)^{q\,max-q}\frac{1}{2^q}\frac{v^u}{\Gamma(u)}\beta^{-(u+1)}exp$$
$$-\frac{v}{\beta}\frac{1}{v}$$

$$= \left(\frac{1}{2}\right)^{n-q}\left(\frac{1}{\beta}\right)^{n-q-1} exp$$
$$-\frac{1}{\beta}\sum_{t=q+1}^{n}\left|\sum_{i=1}^{q}G^{-1}(\rho_j)z_{t-j} + x_t\right|$$
$$C_q^{q\,max}\lambda^q(1 \tag{11}$$
$$- \lambda)^{q\,max-q}\frac{1}{2^q}\frac{v^{u-1}}{\Gamma(u)}\beta^{-(u+1)}exp - \frac{v}{\beta}$$

Parameter space is a combination of those with different dimensions. This makes it difficult to analytically determine the Bayes estimator. Therefore, the MCMC reversible jump algorithm was used to achieve this.

3.4. Reversible Jump MCMC

Simulation of posterior distribution for parameters was performed in 2 stages, which are conditional simulation for (λ, β, v) when given $\left(q, \rho^{(q)}\right)$ and for $\left(q, \rho^{(q)}\right)$ when given (λ, β, v). Meanwhile, the conditional distribution for (λ, β, v) when given $\left(q, \rho^{(q)}\right)$ is a multiplication results of the Gamma inversion, Binomial, and the Gamma distribution. Therefore, the conditional simulation for (λ, β, v) when given $\left(q, \rho^{(q)}\right)$ can be performed as follows:

$$\beta \sim IG\left(n-q, v + \sum_{t=q+1}^{n}\left|\sum_{i=1}^{q}G^{-1}(\rho_j)z_{t-i} + x_t\right|\right), \lambda \sim B(q+1, q_{max}-q+1), v \sim G(u, \frac{1}{\beta})$$

However, conditional distribution for $(q, \rho^{(q)})$ with (λ, β, v) has a complex form. Therefore, the simulation for $(q, \rho^{(q)})$ with (λ, β, v) is performed using the MCMC algorithm. This uses 3 types of transformation, such as coefficient change, birth, and death of the order (Suparman, 2018).

3.4.1. Coefficient change

For example, $w = (q, \rho^{(q)})$ is the old Markov Chain and $w^* = (q^*, \rho^{*(q^*)})$ is the new. The change in the coefficient of MA model does not alter the order value. Meanwhile, the alteration from w to w^* was performed in two steps. Firstly, value $q^* = q$ was taken, secondly, $i \in \{1, \dots, q\}$ was chosen and $\rho_i^* = b$ where $b \sim U(-1,1)$ were defined. The ratio between the likelihood function $f(x|w^*)$ and $f(x|w)$ can be stated as follows:

$$\frac{f(x|w^*)}{f(x|w)}$$

$$= \frac{exp - \frac{1}{\beta} \sum_{t=q+1}^{n} |\sum_{i=1}^{q} G^{-1}(\rho_i^*) z_{t-i} + x_t|}{exp - \frac{1}{\beta} \sum_{t=q+1}^{n} |\sum_{i=1}^{q} G^{-1}(\rho_i) z_{t-i} + x_t|} \quad (12)$$

The ratio between prior distribution for q^* and for q is $\frac{\pi(q^*)}{\pi(q)} = 1$. Meanwhile, the ratio between conditional prior for ρ^* when given q^* and conditional prior distribution for ρ with q is $\frac{\pi(\rho^*|q^*)}{\pi(\rho|q)} = 1$.

Also, the ratio between instrument distribution $q(w^*, w)$ and $q(w, w^*)$ is:

$$\frac{q(w^*, w)}{q(w, w^*)} = \left(\frac{(1 + u_i)(1 - u_i)}{(1 + r_i)(1 - r_i)} \right)^{1/2}. \quad (13)$$

$\alpha_c(w, w^*)$ is the acceptance probability for the coefficient change, which is written as follows:

$$\alpha_c(w, w^*) = min \left\{ 1, \frac{f(x|w^*)}{f(x|w)} \frac{q(w^*, w)}{q(w, w^*)} \right\} \quad (14)$$

3.4.2. Birth of the order

For example, $w = (q, \rho^{(q)})$ is the old Markov chain and $w^* = (q^*, \rho^{*(q^*)})$ is the new. The birth of MA alters the order and coefficient value of the model. Meanwhile, change from w to w^* was performed in

two steps. Firstly, $q^* = q + 1$ was taken, secondly, $w^* = (w, b)$ where $b \sim U(-1,1)$ was defined. The ratio between the function $f\left(x|w^*\right)$ and $f(x|w)$ can be stated as follows:

$$\frac{f\left(x|q^*,\rho^*\right)}{f(x|q,\rho)}$$

$$= \frac{exp -\frac{1}{\beta}\sum_{t=q+1}^{n}\left|\sum_{i=1}^{q}G^{-1}(\rho_i^*)z_{t-i} + x_t\right|}{exp -\frac{1}{\beta}\sum_{t=q+1}^{n}\left|\sum_{i=1}^{q}G^{-1}(\rho_i)z_{t-i} + x_t\right|}\left(\frac{1}{2\beta}\right) \qquad (15)$$

The ratio between the prior distribution for ρ^* and p for q is

$$\frac{\pi(q^*)}{\pi(q)} = \frac{q_{max} - q}{q + 1}\frac{\lambda}{1 - \lambda} \qquad (16)$$

Also, the ratio between conditional prior distribution for ρ^* when given q^*, and for ρ with q is $\frac{\pi(q^*)}{\pi(q)} = 1$. This can be written as:

$$\frac{\pi\left(\rho^*|q^*\right)}{\pi(\rho|q)} = \frac{1}{2} \qquad (17)$$

The ratio between the distribution of the instrument $q\left(w^*,w\right)$ and $q\left(w,w^*\right)$ depends on the value of when $a < 0$ is:

$$\frac{q\left(w^*,w\right)}{q\left(w,w^*\right)} = \frac{1}{u + 1} \qquad (18)$$

Meanwhile, when $a > 0$, the ratio between the distribution of instrument $q\left(w^*,w\right)$ and $q\left(w,w^*\right)$ is:

$$\frac{q\left(w^*,w\right)}{q\left(w,w^*\right)} = \frac{1}{1 - u} \qquad (19)$$

$\alpha_b\left(w, w^*\right)$ is the acceptance probability for birth of order, which can be written as follows:

$$\alpha_b\left(w, w^*\right)$$
$$= min\left\{1, \frac{f\left(x\middle|q^*, \rho^*\right)}{f(x|q, \rho)} \frac{\pi(q^*)}{\pi(q)} \frac{\pi\left(\rho^*\middle|q^*\right)}{\pi(\rho|q)} \frac{q\left(w^*, w\right)}{q\left(w, w^*\right)}\right\} \quad (20)$$

3.4.3. Death of order

The death of the MA model order was different from the transformation. For example, $w = \left(q + 1, \rho^{(q+1)}\right)$ is the old markov chain and $w^* = \left(q, \rho^q\right)$ is the new. The birth changes the order and coefficient value of the model. Also, alterations from w to w^* are made in two steps. Firstly, $q^* = q$ was taken, and secondly, $w^* = w \setminus \left\{\rho_{q+1}\right\}$ was defined. $\alpha_d\left(w, w^*\right)$ is the acceptance probability for the death of the order, which can be written as follows:

$$\alpha_d\left(w, w^*\right) = min\left\{1, \frac{1}{\alpha_b\left(w, w^*\right)}\right\}. \quad (21)$$

3.5. Simulation

The performance of MCMC reversible jump algorithm was tested using simulation study. This addresses two MA models. The first assumes that noise is Laplacian distributed. Meanwhile, the second model assumes that noise is Gaussian.

3.5.1. Laplacian Noise

For the first simulation, 250 synthesis time series were made according to equation (1) with Laplacian distributed noise. This was given in equation (2). Meanwhile, the MA model parameters are stated in Table 1. The maximum order is limited to order 10, $q_{maks} = 10$.

Table 1: Synthesis time series parameter

q	$\left(\theta_1, \theta_2, \theta_3\right)$	β
3	$(0.57, -0.55, -0.99)$	2

The resulting time series is presented in Figure 1.
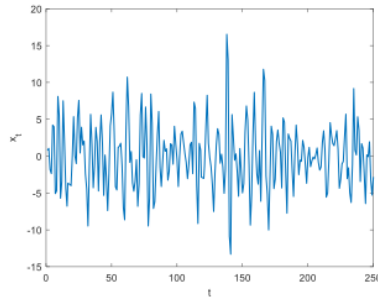
Figure 1: Synthesis time series

Furthermore, the parameters were estimated based on the synthesis time series. The MCMC reversible [1] jump algorithm was used to estimate the parameters. Also, it was run in 20,000 iterations with a burn-in period of 5000. The histogram for the order is presented in Figure 2.
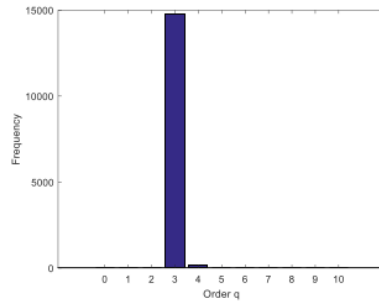
Figure 2: Histogram for order q

Figure 2 showed that the maximum frequency was reached in order 3. Therefore, the estimator for q is $\hat{q} = 3$. Given $\hat{q} = 3$, the results of the model coefficients are presented in Table 2.

Table 2: Estimation results of MA parameters (3)

$\hat{q}$	$\left(\hat{\theta}_1, \hat{\theta}_2, \hat{\theta}_3\right)$	$\hat{\beta}$
3	$(0.59, -0.51, -0.93)$	2.56

Comparing Table 1 and 2, the estimated values of the parameters are close to the actual values. The simulation showed that the algorithm can estimate the parameters that has Laplacian distributed noise.

3.5.2. Noise Gaussian

For the second simulation, 250 synthesis time series were made according to equation (1) but the noise is Gaussian distributed, which is obtained by the following equation:
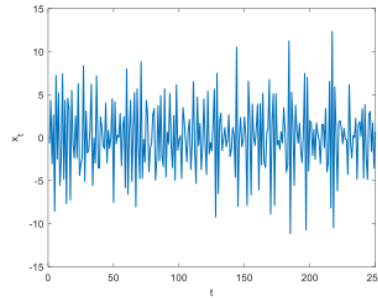
$$g(z_t|\beta) = \frac{1}{\sqrt{2\pi\sigma^2}} \, exp - \frac{1}{2\sigma^2} z_t^2$$

The MA model parameters are shown in Table 3. As in the first simulation, the order is limited to order 10, $q_{maks} = 10$.

Table 3: The synthesis time series parameter

q	$\left(\theta_1, \theta_2\right)$	σ
2	$(-1,53, 0.57)$	4

The resulting time series is presented in Figure 3.

Figure 3: Synthesis time series

Furthermore, the parameters are also estimated based on the synthesis time series. The MCMC was also used to estimate the parameters. As in the first simulation, the algorithm was also run in 20,000 iterations with a burn-in period of 5000. The histogram for the order is presented in Figure 4.
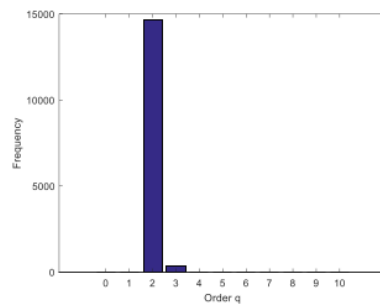
Figure 4: Histogram for order q

Figure 4 showed that the maximum frequency was reached in order 2. Therefore, the estimator for q is $\hat{q} = 2$. Given $\hat{q} = 2$, the results of the model coefficients are presented in Table 4.

Table 4: Estimation results of MA model parameters (2)

$\hat{q}$	$\left(\hat{\theta}_1, \hat{\theta}_2\right)$	$\hat{\sigma}$
2	$(0.55, -0.59)$	5.08

As in the first simulation, comparing the Table 3 and 4, the estimated values of the parameters are close to the actual. This simulation study showed that the MCMC algorithm can estimate the parameters of the MA model that has Gaussian distributed noise.

3.6. Heart Rate Data

Heart rate is the number of cardiac cycles per minute. The heart is an organ that responds in the same manner as other excitable tissues. In fact, it hypertrophy and becomes stronger when a person exercises. In the absence of exercise, this organ continues to pump blood throughout the body to facilitate tissue repair and recovery. Studies on heart rate can be found in a variety of literatures, such as (Mezentseva, 2011), (Mazzoleni, 2016), Suparman (2020).

In this paper, the authors use MA to model human heart rate data. The number of beats per minute was recorded for 100 minutes. The recording data are presented in Figure 5.
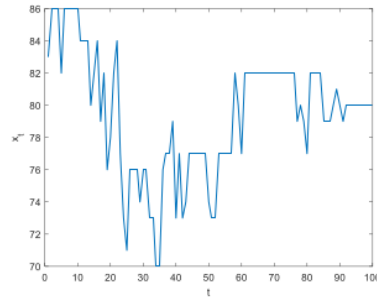
Figure 5: Heart rate data

3.5.2. Laplacian Noise

In this first case, this rate was modeled by MA with Laplacian distributed noise. Meanwhile, the MCMC algorithm was implemented to estimate the parameters with Laplacian noise. The algorithm was run in 100000 iterations with a burn-in period of 10000. Figure 6 showed a histogram for the order.
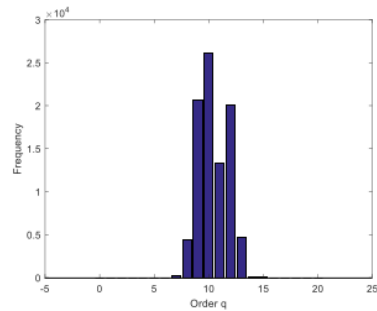
Figure 6 showed that the maximum frequency was reached in order 10. Therefore, the estimator for q is $\hat{q} = 2$. Given q = 2, the results of the model coefficients are presented in Table 5.

Table 5: Estimation results of MA parameters (10)

$\hat{q}$	$\left(\hat{\theta}_1, \hat{\theta}_2, \hat{\theta}_3, \hat{\theta}_4, \hat{\theta}_5, \hat{\theta}_6, \hat{\theta}_7, \hat{\theta}_8, \hat{\theta}_9, \hat{\theta}_{10}\right)$	$\hat{\beta}$
10	$\left(1.88, 2.65, 3.20, 3.38, 3.26, 2.84, 2.2, 1.49. 0.8, 0.27\right)$	7.897

3.5.2. Gaussian Noise

In this second case, the rate was modeled with the Gaussian distributed noise. The MCMC was implemented to estimate the parameters. Furthermore, the algorithm was run in 100000 iterations with a burn-in period of 10000. Figure 6 showed a histogram for the order.
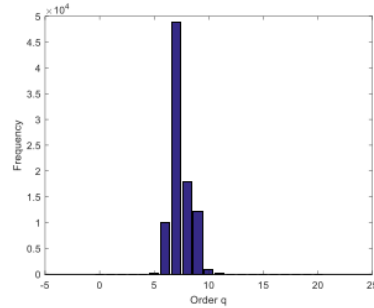

Figure 6 showed that the maximum frequency was reached in order 10. Therefore, the estimator for q is $\hat{q} = 7$. Given $\hat{q} = 7$, the results of the model coefficient is presented in Table 6.

Table 6: Estimation results of MA parameters (7)

$\hat{q}$	$\left(\hat{\theta}_1, \hat{\theta}_2, \hat{\theta}_3, \hat{\theta}_4, \hat{\theta}_5, \hat{\theta}_6, \hat{\theta}_7\right)$	$\hat{\beta}$
7	$(1.51, 1.86, 1.97, 1.76, 1.38, 0.85, 0.36)$	166.02

When the model with Laplacian noise was compared to the MA with Gaussian, the Laplacian noise resulted in a higher order model. Conversely, it produced a model with smaller noise variance.

4. Conclusion

This research resulted in the development of MA, which is the model with Laplacian noise. Also, the Bayesian approach was used to estimate the model, and the order was assumed as a parameter. Furthermore, the MCMC reversible Jump algorithm was applied to generate a Markov chain. This chain was designed, and the limit distribution approached the posterior for the parameter. This was then used to determine the Bayes estimator.

Simulation study showed that the MCMC algorithm can estimate the order and coefficient of the MA model, and noise variance. Also, the human heart rate was modeled by MA with Gaussian and Laplacian noises. This was implemented to ascertain a suitable MA for the heart rate. Laplacian noise produced a model that has a higher order than Gaussian noise. The advantage of this is that those with Laplacian has a smaller noise variance than a model with Gaussian.

HASIL CEK_60110621
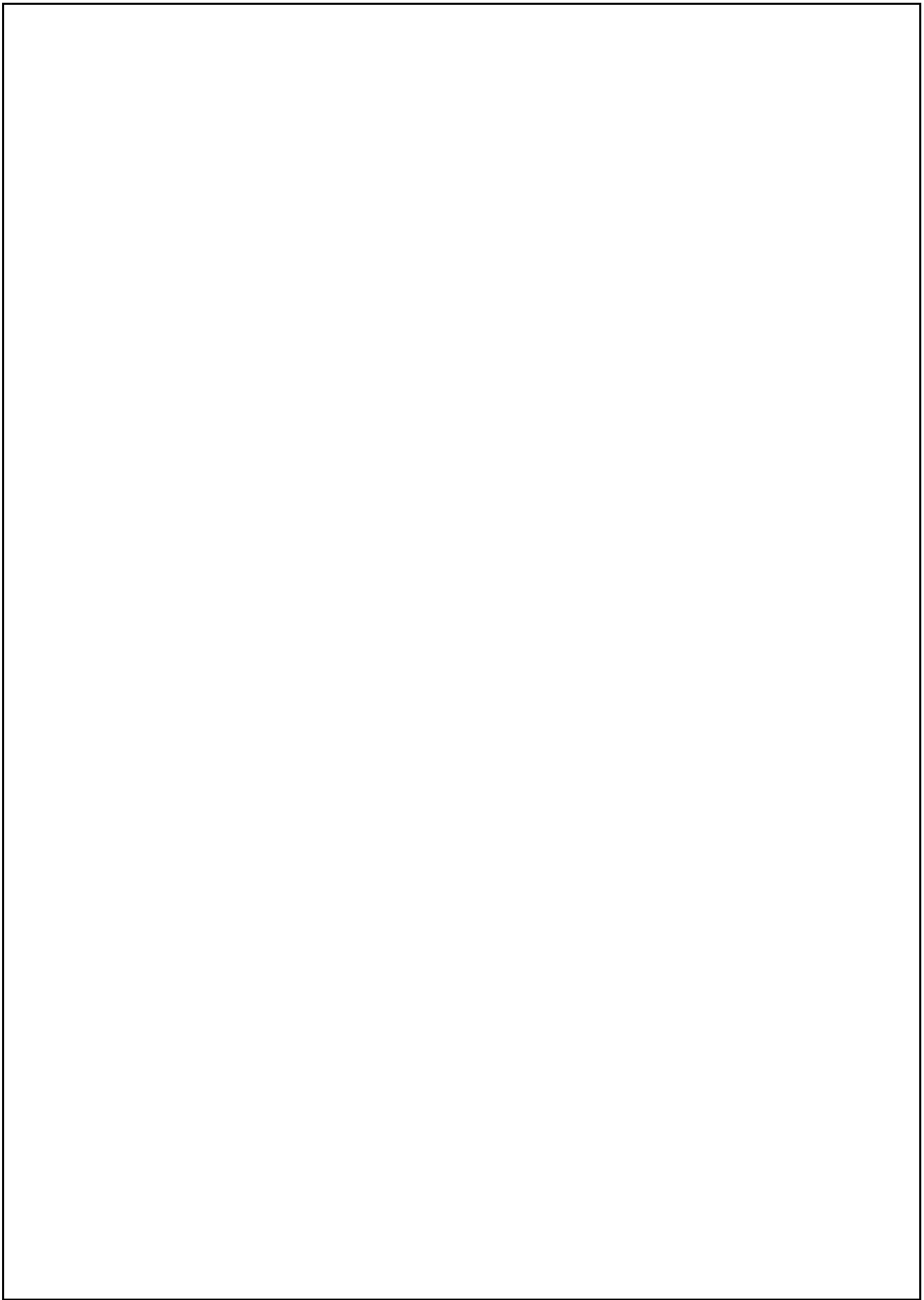
4% SIMILARITY INDEX

2% INTERNET SOURCES

1% PUBLICATIONS

2% STUDENT PAPERS

PRIMARY SOURCES

1 Submitted to University Tun Hussein Onn Malaysia
   Student Paper
   1%

2 indianrealtylaw.com
   Internet Source
   1%

3 www.skateaustria.com
   Internet Source
   1%

4 "The Bayesian Choice", Springer Nature, 2007
   Publication
   1%

Exclude quotes	On	Exclude matches	< 1%
Exclude bibliography	On		