Parameter Estimation Methods in Multiple Linear Regression Analysis with Autocorrelation and Heavy-Tailed Distributed Data

Katechan Jampachaisri* and Khanittha Tinochai

Department of Mathematics, Faculty of Science, Naresuan University, Phitsanulok, 65000, Thailand

* Corresponding author. E-mail address: katechanj@nu.ac.th

Received: 2 May 2019; Revised: 10 June 2019; Accepted: 17 June 2019

Abstract

Regression analysis is extensively used in a wide variety of fields, especially for predictive purpose. Its assumptions play a crucial role in parameter estimation. This paper focuses on parameter estimation in multiple linear regression when the assumptions are violated with simultaneous presence of autocorrelated random errors of AR(1) structure and heavy-tailed distribution, using hierarchical Bayes approach, which prior information about parameters, both noninformative and informative priors, is incorporated into the model. The result is also compared with frequently used method, maximum likelihood estimation, using the mean square error (MSE) as a criterion for comparison. The result reveals that hierarchical Bayes with informative priors outperform the maximum likelihood method, yielding the smallest MSEs for all sample sizes and correlation coefficients.

Keywords: Autocorrelation, Heavy-tailed distribution, Maximum likelihood, Hierarchical Bayes

Introduction

The relationship of explanatory variables on the interest outcome is usually performed using the well-known, traditional approach, so called multiple linear regression. The postulated assumptions on random error about independent and normally distributed with constant variance are noteworthy in inferential statistics. For predictive purpose, the reliability of parameter estimation is of concern to most researchers. However, data that are collected over time usually violate the independency among observations, often called autocorrelation or serial correlation. In addition, the outliers or influential points hidden in data can damage the assumptions about normality and constant variance, leading to data not conforming to the imposed method. As a consequence, parameter estimation based on the usual ordinary least square method (OLS) is no longer efficient and does not have the best linear unbiased estimator (BLUE) property.

Various alternatives to the least square method have been proposed to improve the estimated values of parameters when data do not follow the underlying assumptions. Beach and MacKinnon (1978) estimated parameters in simple linear regression model with autocorrelated errors of order 1, or AR(1), using the maximum likelihood method. Pires and Rodrigues (2007) compared two methods of parameter estimation in multiple linear regression; the classical method, OLS , and maximum likelihood (ML) estimation. Tanizaki (2003) studied two parameter estimation methods; maximum likelihood estimation and Bayesian method, on regression model in small sample when errors were assumed to be related. Farrell and Ludwig (2008) presented the estimation of parameters in hierarchical response time models, using Bayesian and maximum likelihood estimations. Liu and Dey (2007) estimated parameters in hierarchical overdispersed Poisson model with the presence of autocorrelation. Besides the problem of dependent random error, data may possess heavier tail than normal, such as Student-t, Cauchy or logistic distributions, which usually appeared in finance and economic fields (Ravi and Butar, 2010). Lange, Little, and Taylor (1989) studied robust statistical models, both linear and non-linear regression, when random error having multivariate-t distribution. Fernandez and Steel (1999) performed multivariate regression analysis with multivariate-t distributed errors. Rahman and Khan (2007) focused on predictive distribution using Bayesian approach in linear regression model when errors distributed as multivariate-t. Nadarajah and Kotz (2008) illustrated the maximum likelihood method using EM algorithm to estimate parameters in multivariate-t distribution. Roy and Hobert (2010) utilized Monte Carlo method in Bayesian multiple linear regression analysis with heavy-tailed error. Li and Zhao (2015) proposed a robust coefficient estimation and variable selection method based on Bayesian adaptive Lasso *t*-regression, along with an application to real data. Zellner and Ando (2010) used Bayesian and non-Bayesian approaches in the seemingly unrelated regression model with Student-t errors, accompanying by the application for forecasting.

As aforementioned, several methods have been proposed to improve parameter estimation in regression analysis when the underlying assumptions about constant variance and independent, normally distributed random errors are violated. However, most studies focused only on one problem at a time. Although some transformation techniques can abate the present problem, another problem can occur. For example, the method to solve for non-constant variance may result in heavy-tailed data instead. This paper is thus aimed to propose the use of hierarchical Bayes to estimate parameters in multiple linear regression model when two problems of correlated random errors with heavy-tailed distribution occur simultaneously. In hierarchical Bayes, prior information about parameters is incorporated into consideration. The results are then compared with the commonly used approach, maximum likelihood estimation.

The remaining of this paper is organized as follow. In the next section, multiple linear regression with correlated and heavy-tailed distributed random error, parameter estimation methods following by Gibbs sampler procedure are expressed. The result of simulation study is demonstrated in the third section. Conclusion, discussion and suggestions are then provided in the last section.

Methods and Materials

1. Multiple Linear Regression with Autocorrelation and Heavy-tailed Distributed Errors

The standard form of multiple linear regression model is written as

$$\underset{\sim}{y} = X\underset{\sim}{\beta} + \underset{\sim}{\varepsilon}, \tag{1}$$

where $\underset{\sim}{y}$ is an $n \times 1$ vector of response variable, X is an $n \times p$ matrix of regressors, $\underset{\sim}{\beta}$ is a $p \times 1$ vector of the regression coefficients and $\underset{\sim}{\varepsilon}$ is an $n \times 1$ vector of random errors. The underlying assumption about $\underset{\sim}{\varepsilon}$ is assumed to be independent and normally distributed with zero mean vector and constant covariance matrix, $\underset{\sim}{\varepsilon} \sim N(\mathbf{0},\ \sigma^2\mathbf{I})$, where $\mathbf{I}$ is an $n \times n$ matrix of identity.

As previously mentioned, data collected over time usually violate the assumption about independent random errors. At the same time, a few anomaly observations lurking in data may damage the normal assumption, causing inefficient least square estimators in model building. In this paper, we focus on parameter estimation in multiple linear regression model when random errors are correlated in the form of autoregressive of order 1, or AR(1), with heavy-tailed distribution, assumed to be multivariate t-distribution, $\varepsilon_i = \rho\varepsilon_{i-1} + u_i,\ \ i = 1, 2, ..., n$, where $u_i \sim t\left(0,\ \sigma^2,\ n\right)$. The response vector $\underset{\sim}{y}$ is accordingly distributed as a multivariate Student-t distribution,

$\underset{\sim}{Y} \sim t\left(X\underset{\sim}{\beta},\ \Sigma,\ \nu\right)$, where $\mathbf{\Sigma} = \sigma^2\mathbf{\Omega}$ is an $n \times n$ positive definite matrix and ν is the degrees of freedom. Let $\mathbf{\Omega}$ be an $n \times n$ correlation coefficient matrix, illustrated as

$$\Sigma \;=\; \sigma^2\Omega \;=\; \frac{\sigma^2}{1-\rho^2}\begin{bmatrix} 1 & \rho & \rho^2 & \dots & \rho^{n-1} \\ \rho & 1 & \rho & \dots & \rho^{n-2} \\ \rho^2 & \rho & 1 & \dots & \rho^{n-3} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \rho^{n-1} & \rho^{n-2} & \rho^{n-3} & \dots & 1 \end{bmatrix}, -1 < \rho < 1, \tag{2}$$

where ρ denotes population correlation coefficient, representing relationship between pairs of measurement. The joint probability density function of $\underset{\sim}{y} \mid \underset{\sim}{\beta},\sigma^2$ is defined as

$$f\left(\underset{\sim}{y} \mid \underset{\sim}{\beta},\sigma^2\right) \;=\; \frac{\Gamma\left(\frac{\nu+n}{2}\right)}{\left(\nu\pi\sigma^2\right)^{\frac{n}{2}}\Gamma\left(\frac{\nu}{2}\right)|\Omega|^{\frac{1}{2}}}\left[1+\frac{1}{\nu\sigma^2}\left(\underset{\sim}{y}-X\underset{\sim}{\beta}\right)'\Omega^{-1}\left(\underset{\sim}{y}-X\underset{\sim}{\beta}\right)\right]^{-\frac{(\nu+n)}{2}}. \tag{3}$$

After rearranging the term $\left(\underset{\sim}{y}-X\underset{\sim}{\beta}\right)'\Omega^{-1}\left(\underset{\sim}{y}-X\underset{\sim}{\beta}\right)$, we obtain

$$\left(\underset{\sim}{y}-X\underset{\sim}{\beta}\right)'\Omega^{-1}\left(\underset{\sim}{y}-X\underset{\sim}{\beta}\right) \;=\; \left(\underset{\sim}{y}-X\hat{\underset{\sim}{\beta}}\right)'\Omega^{-1}\left(\underset{\sim}{y}-X\hat{\underset{\sim}{\beta}}\right)+\left(\underset{\sim}{\beta}-\hat{\underset{\sim}{\beta}}\right)'X'\Omega^{-1}X\left(\underset{\sim}{\beta}-\hat{\underset{\sim}{\beta}}\right), \tag{4}$$

where $\hat{\underset{\sim}{\beta}} \;=\; \left(X'\Omega^{-1}X\right)^{-1}X'\Omega^{-1}\underset{\sim}{y}$.

Substituting (4) into (3), the joint probability function of $\underset{\sim}{y} \mid \underset{\sim}{\beta},\sigma^2$ is rewritten as

$$f\left(\underset{\sim}{y} \mid \underset{\sim}{\beta},\sigma^2\right) = \frac{\Gamma\left(\frac{\nu+n}{2}\right)}{\left(\nu\pi\sigma^2\right)^{\frac{n}{2}}\Gamma\left(\frac{\nu}{2}\right)|\Omega|^{\frac{1}{2}}}$$

$$\times\left\{1+\frac{1}{\nu\sigma^2}\left[\left(\underset{\sim}{y}-X\hat{\underset{\sim}{\beta}}\right)'\Omega^{-1}\left(\underset{\sim}{y}-X\hat{\underset{\sim}{\beta}}\right)+\left(\underset{\sim}{\beta}-\hat{\underset{\sim}{\beta}}\right)'X'\Omega^{-1}X\left(\underset{\sim}{\beta}-\hat{\underset{\sim}{\beta}}\right)\right]\right\}^{-\frac{(\nu+n)}{2}} \tag{5}$$

2. Methodology for Parameter Estimation

2.1 Maximum Likelihood Method (ML)

The parameter estimation in regression model is usually performed using a well-known method of maximum likelihood. For convenience, the likelihood function is replaced by the log-likelihood function. Values of parameters that maximize the log-likelihood function are then calculated as following:

Consider the likelihood function based on multivariate t-distribution

$$
\begin{aligned}
L\left(\underset{\sim}{\beta},\sigma^2 \mid \underset{\sim}{y}\right) &= \prod_{i=1}^{n} f\left(y_i \mid \underset{\sim}{\beta},\sigma^2\right) \\
&= \frac{\Gamma\left(\frac{\nu+n}{2}\right)}{\left(\nu\pi\sigma^2\right)^{\frac{n}{2}}\Gamma\left(\frac{\nu}{2}\right)\left|\Omega\right|^{\frac{1}{2}}}\left[1+\frac{1}{\nu\sigma^2}\left(\underset{\sim}{y}-X\underset{\sim}{\beta}\right)'\Omega^{-1}\left(\underset{\sim}{y}-X\underset{\sim}{\beta}\right)\right]^{-\frac{(\nu+n)}{2}},
\end{aligned}
$$

and corresponding log-likelihood function

$$
\begin{aligned}
\ln L\left(\underset{\sim}{\beta},\sigma^2\right) &= \ln\Gamma\left(\frac{\nu+n}{2}\right)-\frac{n}{2}\ln\nu\pi\sigma^2-\ln\Gamma\left(\frac{\nu}{2}\right)-\frac{1}{2}\ln\left|\Omega\right| \\
&\quad -\frac{(\nu+n)}{2}\ln\left[1+\frac{1}{\nu\sigma^2}\left(\underset{\sim}{y}-X\underset{\sim}{\beta}\right)'\Omega^{-1}\left(\underset{\sim}{y}-X\underset{\sim}{\beta}\right)\right].
\end{aligned} \tag{6}
$$

Differentiate (6) with respect to $\underset{\sim}{\beta}$ and σ^2, we obtain

$$
\frac{\partial\ln L\left(\underset{\sim}{\beta},\sigma^2\right)}{\partial\underset{\sim}{\beta}} = -\frac{(\nu+n)}{2}\cdot\frac{1}{\nu\sigma^2 Q}\left(-2X'\Omega^{-1}\underset{\sim}{y}+2X'\Omega^{-1}X\underset{\sim}{\beta}\right), \tag{7}
$$

$$
\frac{\partial\ln L\left(\underset{\sim}{\beta},\sigma^2\right)}{\partial\sigma^2} = -\frac{n}{2\sigma^2}-\frac{(\nu+n)}{2}\cdot\frac{1}{\nu\sigma^4 Q}\left(\underset{\sim}{y}-X\underset{\sim}{\beta}\right)'\Omega^{-1}\left(\underset{\sim}{y}-X\underset{\sim}{\beta}\right), \tag{8}
$$

where $Q=\left[1+\frac{1}{\nu\sigma^2}\left(\underset{\sim}{y}-X\underset{\sim}{\beta}\right)'\Omega^{-1}\left(\underset{\sim}{y}-X\underset{\sim}{\beta}\right)\right]$.

Setting (7) and (8) to zero and solving both equations for $\underset{\sim}{\beta}$ and σ^2, the maximum likelihood estimators of $\underset{\sim}{\beta}$ and σ^2 are resulted as

$$
\begin{aligned}
\hat{\underset{\sim}{\beta}}_{MLE} &= \left(X'\Omega^{-1}X\right)^{-1}X'\Omega^{-1}\underset{\sim}{y}, \\
\hat{\sigma}^2_{MLE} &= \frac{1}{n}\left(\underset{\sim}{y}-X\hat{\underset{\sim}{\beta}}\right)'\Omega^{-1}\left(\underset{\sim}{y}-X\hat{\underset{\sim}{\beta}}\right).
\end{aligned}
$$

2.2 Hierarchical Bayes Method (HB)

Parameter estimation based on classical approach, such as least squares method or maximum likelihood method, regards the parameters of a probability density function as unknown, but constant, while Bayesian approach consider parameters as random variables with some probability density function, known as prior distribution. The prior distribution form is characterized by its own parameters, called hyperparameters. Hierarchical Bayes approach arises when the form of probability density function of hyperparameters are known (Gill, 2008). Thus, parameter estimation can be derived from the joint posterior distribution functions of all parameters.

In this paper, noninformative and informative priors on parameters are considered in calculating the estimated values of parameter, as follows.

2.2.1. Noninformative priors for $\underset{\sim}{\beta}$ and σ^2

We assume vague priors for both $\underset{\sim}{\beta}$ and σ^2 as $\pi\left(\underset{\sim}{\beta}\right) \propto c$, where c is constant, and $\pi\left(\sigma^2\right) \propto \dfrac{1}{\sigma^2}$. Consider the joint posterior distribution functions of $\underset{\sim}{\beta}$ and σ^2

$$h\left(\underset{\sim}{\beta},\, \sigma^2 \mid \underset{\sim}{y}\right) = \frac{L\left(\underset{\sim}{y} \mid \underset{\sim}{\beta}, \sigma^2\right) \cdot \pi\left(\underset{\sim}{\beta}\right) \cdot \pi\left(\sigma^2\right)}{\int\limits_{\sigma^2}\int\limits_{\beta} L\left(\underset{\sim}{y} \mid \underset{\sim}{\beta}, \sigma^2\right) \cdot \pi\left(\underset{\sim}{\beta}\right) \cdot \pi\left(\sigma^2\right) d\underset{\sim}{\beta}\, d\sigma^2} \propto L\left(\underset{\sim}{y} \mid \underset{\sim}{\beta}, \sigma^2\right) \cdot \pi\left(\underset{\sim}{\beta}\right) \cdot \pi\left(\sigma^2\right)$$

$$\propto \frac{\Gamma\left(\frac{v+n}{2}\right)}{\left(v\pi\sigma^2\right)^{\frac{n}{2}} \Gamma\left(\frac{v}{2}\right) |\Omega|^{\frac{1}{2}}} \times \left\{1 + \frac{1}{v\sigma^2}\left[\left(\underset{\sim}{y} - X\hat{\underset{\sim}{\beta}}\right)' \Omega^{-1}\left(\underset{\sim}{y} - X\hat{\underset{\sim}{\beta}}\right) + \left(\underset{\sim}{\beta} - \hat{\underset{\sim}{\beta}}\right)' X'\Omega^{-1}X\left(\underset{\sim}{\beta} - \hat{\underset{\sim}{\beta}}\right)\right]\right\}^{-\frac{(v+n)}{2}}. \tag{9}$$

The full conditional posterior distribution function of $\underset{\sim}{\beta}$ with squared error loss function can be derived as

$$h_1\left(\underset{\sim}{\beta} \mid \sigma^2,\, \underset{\sim}{y}\right) \propto \left\{1 + \frac{1}{v\sigma^2}\left(\underset{\sim}{y} - X\hat{\underset{\sim}{\beta}}\right)' \Omega^{-1}\left(\underset{\sim}{y} - X\hat{\underset{\sim}{\beta}}\right) + \frac{1}{v}\left(\underset{\sim}{\beta} - \hat{\underset{\sim}{\beta}}\right)' \left(\left(\frac{X'\Omega^{-1}X}{\sigma^2}\right)^{-1}\right)^{-1} \left(\underset{\sim}{\beta} - \hat{\underset{\sim}{\beta}}\right)\right\}^{-\frac{(v+n)}{2}}.$$

Thus, the posterior distribution of parameter $\underset{\sim}{\beta}$ is obtained as multivariate *t* distribution, given as

$$\underset{\sim}{\beta} \mid \sigma^2,\, \underset{\sim}{Y} \sim t\left(v,\, \hat{\beta},\, \left(\frac{X'\Omega^{-1}X}{\sigma^2}\right)^{-1}\right). \tag{10}$$

Solving the joint posterior distribution functions of $\underset{\sim}{\beta}$ and σ^2 in equation (9), then we obtain the full conditional posterior distribution function of σ^2 as

$$h_2\left(\sigma^2 \mid \underset{\sim}{\beta},\, \underset{\sim}{y}\right) \propto \frac{1}{\left(\sigma^2\right)^{\frac{n}{2}}} \times \left\{1 + \frac{1}{v\sigma^2}\left[\left(\underset{\sim}{y} - X\hat{\underset{\sim}{\beta}}\right)' \Omega^{-1}\left(\underset{\sim}{y} - X\hat{\underset{\sim}{\beta}}\right) + \left(\underset{\sim}{\beta} - \hat{\underset{\sim}{\beta}}\right)' X'\Omega^{-1}X\left(\underset{\sim}{\beta} - \hat{\underset{\sim}{\beta}}\right)\right]\right\}^{-\frac{(v+n)}{2}}. \tag{11}$$

Because the full conditional posterior distribution function of σ^2 is complicated and cannot be solved into an explicit form. Numerical procedure, such as Gibbs sampling, is implemented to find the estimate of parameter σ^2.

2.2.2. Informative priors for $\underset{\sim}{\beta}$ and σ^2

For informative priors, we assume $\underset{\sim}{\beta} \sim N\left(\underset{\sim}{\mu},\, m\sigma^2 I_p\right)$, where m is a known constant, and the hyperparameter $\underset{\sim}{\mu}$ is normally distributed, that is $\underset{\sim}{\mu} \sim N\left(\mu_0, v_0 I_p\right)$, where μ_0 and v_0 are known values. Also, assuming $\sigma^2 \sim IG(\alpha, \gamma)$ and hyperparameters α and γ are distributed as inverted gamma, $\alpha \sim IG\left(f_0, e_0\right)$ and $\gamma \sim IG\left(r_0, m_0\right)$, where f_0, e_0, r_0 and m_0 are known values.

Similarly, $\underset{\sim}{\beta}$ and σ^2 are assumed to be independent for simplicity. The joint posterior distribution functions of $\underset{\sim}{\beta}$, σ^2, $\underset{\sim}{\mu}$, α and γ are obtained as follow.

$$h\left(\underset{\sim}{\beta},\sigma^2,\underset{\sim}{\mu},\alpha,\gamma \mid \underset{\sim}{y}\right) \propto L\left(\underset{\sim}{y} \mid \underset{\sim}{\beta},\sigma^2\right)\cdot\pi\left(\underset{\sim}{\beta} \mid \underset{\sim}{\mu}\right)\cdot\pi\left(\underset{\sim}{\mu} \mid \mu_0,\, v_0\right)\cdot\pi\left(\sigma^2 \mid \alpha,\gamma\right)\cdot\pi\left(\alpha \mid f_0,e_0\right)\cdot\pi\left(\gamma \mid r_0,m_0\right), \tag{12}$$

where $L\left(\underset{\sim}{y} \mid \underset{\sim}{\beta},\sigma^2\right)$ is likelihood function, $\pi\left(\underset{\sim}{\beta} \mid \underset{\sim}{\mu}\right)$ and $\pi\left(\sigma^2 \mid \alpha,\gamma\right)$ are prior distribution functions of $\underset{\sim}{\beta}$ and σ^2, $\pi\left(\underset{\sim}{\mu} \mid \mu_0,\, v_0\right)$, $\pi\left(\alpha \mid f_0,e_0\right)$ and $\pi\left(\gamma \mid r_0,m_0\right)$ are distribution functions of hyperparameters $\underset{\sim}{\mu}$, α and γ.

Substituting all distributional forms into (12), the joint posterior distribution functions of $\underset{\sim}{\beta}$, σ^2, $\underset{\sim}{\mu}$, α and γ can be expressed as

$$\begin{aligned}
h\left(\underset{\sim}{\beta},\sigma^2,\underset{\sim}{\mu},\alpha,\gamma \mid \underset{\sim}{y}\right) \propto\ & \frac{\Gamma\left(\frac{\nu+n}{2}\right)}{\left(\nu\pi\sigma^2\right)^{\frac{n}{2}}\Gamma\left(\frac{\nu}{2}\right)|\Omega|^{\frac{1}{2}}} \\
& \times \left\{1+\frac{1}{\nu\sigma^2}\left[\left(\underset{\sim}{y}-X\hat{\underset{\sim}{\beta}}\right)'\Omega^{-1}\left(\underset{\sim}{y}-X\hat{\underset{\sim}{\beta}}\right)+\left(\underset{\sim}{\beta}-\hat{\underset{\sim}{\beta}}\right)'X'\Omega^{-1}X\left(\underset{\sim}{\beta}-\hat{\underset{\sim}{\beta}}\right)\right]\right\}^{-\frac{(\nu+n)}{2}} \\
& \times \frac{1}{\left(2\pi m\sigma^2\right)^{\frac{p}{2}}}e^{-\frac{1}{2m\sigma^2}\left(\underset{\sim}{\beta}-\underset{\sim}{\mu}\right)'\left(\underset{\sim}{\beta}-\underset{\sim}{\mu}\right)} \times \frac{1}{\left(2\pi v_0\right)^{\frac{p}{2}}}e^{-\frac{1}{2v_0}\left(\underset{\sim}{\mu}-\mu_0\right)'\left(\underset{\sim}{\mu}-\mu_0\right)} \\
& \times \frac{\gamma^{\alpha}}{\Gamma(\alpha)}\left(\sigma^2\right)^{-(\alpha+1)}e^{-\frac{\gamma}{\sigma^2}} \times \frac{e_0^{f_0}}{\Gamma\left(f_0\right)}(\alpha)^{-(f_0+1)}e^{-\frac{e_0}{\alpha}} \times \frac{m_0^{r_0}}{\Gamma\left(r_0\right)}(\gamma)^{-(r_0+1)}e^{-\frac{m_0}{\gamma}}.
\end{aligned} \tag{13}$$

After integrating (13) with respect to σ^2, $\underset{\sim}{\mu}$, α and γ, the conditional posterior distribution function of $\underset{\sim}{\beta}$ is obtained up to proportionality as

$$\begin{aligned}
h_1\left(\underset{\sim}{\beta} \mid \sigma^2,\underset{\sim}{\mu},\, \alpha,\, \gamma,\, \underset{\sim}{y}\right) \propto\ & \left\{1+\frac{1}{\nu\sigma^2}\left[\left(\underset{\sim}{y}-X\hat{\underset{\sim}{\beta}}\right)'\Omega^{-1}\left(\underset{\sim}{y}-X\hat{\underset{\sim}{\beta}}\right)+\left(\underset{\sim}{\beta}-\hat{\underset{\sim}{\beta}}\right)'X'\Omega^{-1}X\left(\underset{\sim}{\beta}-\hat{\underset{\sim}{\beta}}\right)\right]\right\}^{-\frac{(\nu+n)}{2}} \\
& \times e^{-\frac{1}{2m\sigma^2}\left(\underset{\sim}{\beta}-\underset{\sim}{\mu}\right)'\left(\underset{\sim}{\beta}-\underset{\sim}{\mu}\right)}.
\end{aligned} \tag{14}$$

Likewise, the conditional posterior distribution functions of σ^2 can be derived by integrating (13) with respect to $\underset{\sim}{\beta}$, $\underset{\sim}{\mu}$, α and γ, yielding

$$h_2\left(\sigma^2 \mid \underset{\sim}{\beta},\, \underset{\sim}{\mu},\, \alpha,\, \gamma,\, \underset{\sim}{y}\right) \propto \left\{1+\frac{1}{v\sigma^2}\left[\left(\underset{\sim}{y}-X\hat{\underset{\sim}{\beta}}\right)'\Omega^{-1}\left(\underset{\sim}{y}-X\hat{\underset{\sim}{\beta}}\right)+\left(\underset{\sim}{\beta}-\hat{\underset{\sim}{\beta}}\right)'X'\Omega^{-1}X\left(\underset{\sim}{\beta}-\hat{\underset{\sim}{\beta}}\right)\right]\right\}^{-\frac{(v+n)}{2}}$$

$$\times\frac{1}{\left(\sigma^2\right)^{\frac{n+p}{2}+(\alpha+1)}}e^{-\frac{1}{\sigma^2}\left(\underset{\sim}{\beta}-\underset{\sim}{\mu}\right)'\left(\underset{\sim}{\beta}-\underset{\sim}{\mu}\right)-\frac{\gamma}{\sigma^2}}. \tag{15}$$

Similarly, the conditional posterior distribution functions of $\underset{\sim}{\mu}$ after integrating out (13) with respect to $\underset{\sim}{\beta},\ \sigma^2,\ \underset{\sim}{\mu},\ \alpha$ and γ is resulted as

$$h_3\left(\underset{\sim}{\mu} \mid \underset{\sim}{\beta},\ \sigma^2, \alpha,\ \gamma,\ \underset{\sim}{y}\right) \propto e^{-\frac{1}{2}\left(\underset{\sim}{\mu}-\tilde{\mu}\right)'A^{-1}\left(\underset{\sim}{\mu}-\tilde{\mu}\right)}, \tag{16}$$

where $A^{-1} = \left(\frac{1}{m\sigma^2}+\frac{1}{v_0}\right)$ and $\tilde{\mu} = A\left(\frac{\underset{\sim}{\beta}}{m\sigma^2}+\frac{\mu_0}{v_0}\right)$.

Alternatively, $(\underset{\sim}{\mu} \mid \underset{\sim}{\beta},\ \sigma^2, \alpha,\ \gamma,\ \underset{\sim}{Y}) \sim N\left(\tilde{\mu},\ A\right)$,

Finally, the conditional posterior distribution functions of α and γ are calculated similarly, illustrated as

$$h_3\left(\alpha \mid \underset{\sim}{\beta},\ \sigma^2,\ \gamma,\ \underset{\sim}{y}\right) \propto \frac{1}{\Gamma(\alpha)}(\alpha)^{-(f_0+1)}e^{-\frac{e_0}{\alpha}}\gamma^{\alpha}\left(\sigma^2\right)^{-\alpha}, \tag{17}$$

$$h_4\left(\gamma \mid \underset{\sim}{\beta},\ \sigma^2,\ \alpha,\ \underset{\sim}{y}\right) \propto (\gamma)^{-(r_0+1-\alpha)}e^{-\frac{m_0}{\gamma}-\frac{\gamma}{\sigma^2}}. \tag{18}$$

Alternatively, $\left(\alpha \mid \underset{\sim}{\beta},\ \sigma^2,\ \gamma,\ \underset{\sim}{y}\right) \sim IG\left(f_0, e_0\right)$ and $\left(\gamma \mid \underset{\sim}{\beta},\ \sigma^2,\ \alpha,\ \underset{\sim}{y}\right) \sim IG\left(r_0, m_0\right)$, respectively.

Since the full conditional distribution function of $\underset{\sim}{\beta}$ and σ^2 are quite complicated and cannot be derived into an analytic- closed form. The Gibbs sampler is accordingly adopted to find the estimators of $\underset{\sim}{\beta}$ and σ^2.

2.3. Gibbs Sampler Procedure

With hierarchical Bayes approach, the estimate values of parameters in regression model cannot be obtained directly due to the complicated distributional forms, as mentioned in the previous section. Numerical method, such as Gibbs sampler, is implemented by sequentially drawing from these conditional distributions to create the realization of parameter values.

2.3.1. Gibbs sampler for noninformative priors

The estimates of parameter can be obtained according to the following steps.

(1) Drawing $\underset{\sim}{\beta}^{(t+1)}$ from $\underset{\sim}{\beta}^{(t+1)} \mid \sigma^{2(t)}, \underset{\sim}{y} \sim t\left(v,\ \hat{\beta},\ \left(\frac{X'\Omega^{-1}X}{\sigma^{2(t)}}\right)^{-1}\right)$,

(2) Drawing $\sigma^{2(t+1)}$ from $\sigma^{2(t+1)} \mid \underset{\sim}{\beta}^{(t+1)}, \underset{\sim}{y} \sim IG\left(0.01,\ 0.01\right)$.

2.3.2. Gibbs sampler for informative priors

The conditional posterior distributions of $\underset{\sim}{\beta}$ and σ^2 are constructed according to the following steps.

(3) Drawing $\underset{\sim}{\beta}^{(t+1)}$ from $\underset{\sim}{\beta}^{(t+1)} \mid \sigma^{2(t)},\ \alpha^{(t)},\ \gamma^{(t)},\ \underset{\sim}{y}\ \sim t\left(\nu,\ \mu,\ \sigma^{2\,(t)}\right)$, where $\mu \sim N\left(\mu_0, v_0\right)$ and $\sigma^{2\,(t)} \sim IG\left(\alpha,\ \gamma\right)$,

(4) Drawing $\sigma^{2(t+1)}$ from $\sigma^{2(t+1)} \mid \underset{\sim}{y} \sim IG\left(\alpha,\ \gamma\right)$, where $\alpha \sim IG\left(f_0, e_0\right)$ and $\gamma \sim IG\left(r_0, m_0\right)$,

(5) Drawing $\underset{\sim}{\mu}^{(t+1)}$ from $\underset{\sim}{\mu}^{(t+1)} \mid \underset{\sim}{\beta}^{(t+1)}, \sigma^{2(t+1)}, \underset{\sim}{y} \sim N\left(\tilde{\mu},\ A\right)$, where $A^{-1} = \left(\dfrac{1}{m\sigma^{2(t+1)}} + \dfrac{1}{v_0}\right)$ and $\tilde{\mu} = A\left(\dfrac{\underset{\sim}{\beta}^{(t+1)}}{m\sigma^{2(t+1)}} + \dfrac{\mu_0}{v_0}\right)$,

(6) Drawing $\alpha^{(t+1)}$ from $\alpha^{(t+1)} \mid \underset{\sim}{\beta}^{(t+1)}, \sigma^{2(t+1)}, \underset{\sim}{\mu}^{(t+1)},\ \gamma^{(t)}, \underset{\sim}{y}\ \sim IG\left(f_0, e_0\right)$, where $f_0 = 0.01,\ e_0 = 0.01$,

(7) Drawing $\gamma^{(t+1)}$ from $\gamma^{(t+1)} \mid \underset{\sim}{\beta}^{(t+1)},\ \sigma^{2(t+1)},\ \underset{\sim}{\mu}^{(t+1)},\ \alpha^{(t+1)},\ \underset{\sim}{y}\ \sim IG\left(r_0, m_0\right)$, where $r_0 = 0.01,\ m_0 = 0.01$.

Results

In this paper, the multiple linear regression model with two predictors are considered under the situation of autoregressive of order 1 (AR(1)) random error with heavy-tailed distribution. All predictors are generated from standard normal distributions. The study is performed on the sample size of 50, 200 and 500 with 3 levels of correlation coefficient: low$\left(\rho = 0.1\right)$, moderate $\left(\rho = 0.5\right)$ and high $\left(\rho = 0.9\right)$. Data are simulated and repeated 1,000 times and Gibbs sampler is implemented for parameter estimation. Three estimators obtained from maximum likelihood (*ML*), hierarchical Bayes using vague priors (*HBV*) and informative priors (*HBI*) are compared. The mean square error is utilized as a criterion for comparison.

Table 1 Mean square error (MSE) of regression coefficient β_0.

n	ρ	MSE		
		ML	***HBV***	***HBI***
50	0.1	1.2051x10^{-5}	0.7514	0.0271
	0.5	1.7785x10^{-4}	0.7774	0.0876
	0.9	4.7038x10^{-4}	0.9006	1.6646
200	0.1	2.2304x10^{-5}	0.2050	0.0066
	0.5	1.0250x10^{-5}	0.2720	0.0215
	0.9	1.6175x10^{-3}	0.6334	0.4503
500	0.1	2.1427x10^{-5}	0.0346	0.0025
	0.5	2.0564x10^{-5}	0.0595	0.0078
	0.9	4.2012x10^{-8}	0.3407	0.1954

Table 2 Mean square error (MSE) of regression coefficient β_1.

n	ρ	MSE		
		ML	*HBV*	*HBI*
50	0.1	1.0814	0.7592	0.0218
	0.5	1.0805	0.7755	0.0309
	0.9	1.0747	0.8688	0.0848
200	0.1	1.0169	0.2089	0.0056
	0.5	1.0177	0.2739	0.0069
	0.9	1.0124	0.5864	0.0242
500	0.1	1.0083	0.0343	0.0020
	0.5	1.0033	0.0560	0.0029
	0.9	0.9965	0.3040	0.0099

Table 3 Mean square error (MSE) of regression coefficient β_2.

n	ρ	MSE		
		ML	*HBV*	*HBI*
50	0.1	1.0655	0.7699	0.0230
	0.5	1.0782	0.7878	0.0280
	0.9	1.0745	0.8790	0.0859
200	0.1	1.0102	0.2063	0.0048
	0.5	1.0281	0.2723	0.0060
	0.9	1.0084	0.5950	0.0272
500	0.1	1.0066	0.0344	0.0021
	0.5	1.0055	0.0551	0.0025
	0.9	0.9954	0.3019	0.0098

Table 4 Mean square error (MSE) of all regression coefficients $(\beta_0, \beta_1, \beta_2)$ and variance (σ^2).

n	ρ	MSE					
		$(\beta_0, \beta_1, \beta_2)$			σ^2		
		ML	*HBV*	*HBI*	*ML*	*HBV*	*HBI*
50	0.1	2.2118	2.2805	0.0719	3.7830	5.0174	0.0817
	0.5	2.3163	2.3408	0.1464	6.0553	3.0398	0.5313
	0.9	4.6925	2.6484	1.8353	8.2235	5.4041	0.7047
200	0.1	2.0436	0.6201	0.0170	4.0368	0.4606	0.0206
	0.5	2.0761	0.8182	0.0345	6.2097	2.9244	0.4648
	0.9	3.1740	1.8148	0.5017	7.0706	4.8055	0.8199
500	0.1	2.0217	0.1032	0.0066	4.0825	0.0249	0.0086
	0.5	2.0211	0.1706	0.0132	6.2350	0.7916	0.4599
	0.9	2.3793	0.9465	0.2151	7.1226	3.8812	0.9316

The result shown in Table 1 reveals that the estimate of parameter β_0 which represents an intercept term in regression model using *ML* yields the least MSEs for all sample sizes and three levels of correlation coefficient, mostly followed by *HBI* and *HBV*, respectively. On the contrary to the parameters β_1 and β_2 estimates, *HBI* mostly results in the smallest MSEs, successively followed by *HBV* and *ML*, as illustrated in Table 2 and Table 3 respectively.

On overall, the MSEs of each method for estimating β_0, β_1 and β_2 are summed together, as displayed in Table 4 and disclosed that the MSEs of estimates obtained from *HBI* are smallest for all sample sizes at low, moderate and high correlation coefficients. In addition, the MSEs of *ML* tends to be smaller than those of *HBV* as n =50, and vice versa for n =200 and 500. Similar results are also obtained for σ^2 with the smallest MSEs of the estimate using *HBI*, followed by *ML* and *HB*V as n = 50 and vice versa for n = 200 and 500. It is also observed that the MSEs of all three methods tend to be lower when the sample sizes increase but tends to higher with increasing the levels of correlation coefficient.

Conclusion Discussion and Suggestions

The Validity of assumptions in regression model makes a major contribution to the quality of parameter estimators. Alternative to searching for technique to solving the problem of assumption violation which is difficult in practice, especially when a few assumptions are violated simultaneously, this study instead focuses on incorporating information into the estimation in order to mitigate the problem using hierarchical Bayes approach and comparing the result with classical approach, maximum likelihood. Based on simulated study, the results indicate that hierarchical Bayes approach using both noninformative and informative priors to estimate parameters in multiple linear regression model with correlated and heavy-tailed random error performs rather well, especially when correlation between observations are large. As larger sample size, the hierarchical Bayes also outperformed the widely used method, maximum likelihood.

This study indicates that hierarchical Bayes approach performs superior to the maximum likelihood method when data do not follow the independent and normal distribution assumptions. The result is also evident when data are highly correlated with large sample sizes. As a consequence, prior information incorporating into the model with the hierarchical Bayes approach is able to mitigate the problem of dependent and nonnormal data. This finding is benefit for practical use of regression model, especially in predictive purpose. For further study, the comparison of parameter estimation can perform on other circumstances of assumption violation.

Acknowledgments

This research is supported by Naresuan University, Thailand. The authors would like to express their deep thanks to all referees and the editor for valuable comments and suggestions to improve this work.

References

Beach, C. M., & MacKinnon, J. G. (1978). A Maxim Procedure for Regression with Autocorrelated Errors. *Econometrica*l, *46*(1), 51-58.

Farrell, S., & Ludwig, C. G. S. (2008). Bayesian and Maximum Likelihood Estimation of Hierarchical Response Time Models. *The Psychonomic Society*, *15*(6), 1209-1217.

Fernandez, C., & Steel, M. F. J. (1999). Multivariate Student-t Regression Model: Pitfalls and Inference. *Biometrika*, *86*(1), 53-167.

Gill, J. (2008). *Bayesian Methods: A Social and Behavioral Sciences Approach*(2nd ed.). USA: CRC.

Lange, K. L., Little, R. J. A. & Taylor, J. M. G. (1989). Robust Statistical Modeling Using the t Distribution. *Journal of the American Statistical Association*, *84*, 881-896.

Li, Z., & Zhao, W. (2015). *Robust Bayesian Regularized Estimation based on t Regression Model.* Retrieved from http://dx.doi.org/10.1155/2015/989412.

Liu, J., & Dey, D. K. (2007). Hierarchical Overdispersed Poisson Model with Macrolevel Autocorrelation. *Statistical Methodology*, *4*, 354-370.

Nadarajah, S., & Kotz, S. (2008). Estimations for the Multivariate t Distribution. *Acta Applicandae Mathematicae*, *102*, 99-118.

Pires, A. M., & Rodrigues, I. M. (2007). Multiple Linear Regression with Some Correlated Errors: Classical and Robust Methods. *Statistics in Medicine*, *26*, 2901-2918.

Press, S. J. (2005). *Applied Multivariate Analysis: Using Bayesian and Frequentist Methods of Inference* (2nd ed.). Malabar, Florida, USA: Robert E. Krieger Publishing Company.

Rahman, A., & Khan, S. (2007). Prediction Distribution for Linear Regression Model with Multivariate Student-t Errors under the Bayesian Approach. *Proceedings in the 3rd International Conferences on Research and Education in Mathematics (ICREM3).* 10-12 Apr 2007 (pp. 188-193). Kuala Lumpur: Malaysia.

Ravi, A., & Butar, F. (2010). An Insight into Heavy-Tailed Distribution. *Journal of Mathematical Sciences & Education*, *5*(1), 1-15.

Roy, V., & Hobert, J. P. (2010). On Monte Carlo Methods for Bayesian Multivariate Regression Models with Heavy-Tailed Errors. *Journal of Multivariate Analysis*, *101*, 1190-1202.

Tanizaki, H. (2003). On Regression Models with Autocorrelated Error: Small Sample Properties. *International journal of pure and applied mathematics*, *5*(3), 247-252.

Zellner, A., & Ando, T. (2010). Bayesian and Non-Bayesian Analysis of the Seemingly Unrelated Regression Model with Student-t Errors and Its Application for Forecasting. *International Journal of Forecasting*, *26*, 413-434.