The Journal of Applied Science
วารสารวิทยาศาสตร์ประยุกต์
ISSN 1513-7805 Printed in Thailand

Vol. 18 No. 2: 11-21 [2019]
doi: 10.14416/j.appsci.2019.07.002
ISSN 2586-9663 (Online)

Research Article

Parameter estimation methods in multiple linear regression analysis with intraclass correlation and heavy-tailed distributed data

Katechan Jampachaisri[1]* and Khanittha Tinochai[1]

Department of Mathematics, Faculty of Science, Naresuan University, Phitsanulok 65000, Thailand
***E-mail**: katechanj@nu.ac.th
Received: 11/05/2019; Revised: 02/07/2019; Accepted: 26/07/2019

Abstract

The underlying assumptions play an important role in the linear regression analysis. Non-validity of assumptions can cause the estimators of regression coefficients no longer possessing the properties of best linear unbiased estimator (BLUE). This research focuses on parameter estimation in regression model when random errors are correlated with intraclass correlation and followed heavy-tailed distribution simultaneously. Alternative to searching for techniques to remedy the problem of violation, hierarchical Bayes approach is implemented to estimate model parameter, in which prior knowledge about parameters is incorporated to reduce the effect of violated assumptions. Its performance is then compared to the classical approach, maximum likelihood (ML), through Monte Carlo simulation. The study indicates that hierarchical Bayes with informative priors yields the estimators more efficient than ML.

Keywords: intraclass correlation, maximum likelihood, hierarchical Bayes, heavy- tailed distribution

Introduction

Multiple linear regression analysis is extensively used in many fields. Its assumptions about random error, independent and normally distributed with constant variance, are of importance for statistical inference, especially in estimation and hypothesis testing. The accuracy of parameter estimation plays a significant role in prediction. In practice, we occasionally encounter various problems when dealing with real world situations. Outliers or influential observations may lead to nonnormal distributed data or heteroscedasticity or both. Data, especially in science, business and economic area, are usually collected over time, causing the violation of independence among observations, called autocorrelation or serial correlation. Consequently, statistical inference based on the traditional least square estimators is not efficient and does not possess the property of best linear unbiased estimator (BLUE) (Montgomery et al., 2012). Furthermore, violating the underlying assumptions of a model also has the effect on established relationship. One structure of correlated random errors is called intraclass correlation. Intraclass correlation describes the same degree of relationship between observations. If this occurs, traditional statistical methods that assumes independence cannot be utilized, due to incorrect estimates of variance for parameter estimators and hence p-values.

Several techniques have been proposed to account for between-observation correlation in model parameter estimation. Paul (1990) introduced the estimate of intraclass correlation in

doi: 10.14416/j.appsci.2019.07.002

generalized linear model using the estimating equation approach which can be solved iteratively and then the model parameters were estimated using maximum likelihood method. The proposed method was also applied to familial data with multivariate normal distribution. Tanizaki (2003) compared the maximum likelihood estimator with Bayes estimator in linear regression with the presence of first-order autocorrelated errors in small samples. The results showed that Bayes approach provides less biased and more efficient estimator than maximum likelihood approach. Farrell & Ludwig (2008) considered parameter estimation in multilevel response time model based on maximum likelihood and Bayesian approaches, which assume to be ex-Gaussian error distribution. Babatunde et al. (2014) performed bootstrapping experiments to estimate parameters in regression model with autocorrelated errors. The study also revealed that the estimated parameters were affected by the levels of autocorrelation, leading to biased and inefficient estimators.

In addition to correlated error components, lots of studies have contributed to tackling non-normal data, especially in longer and heavier tail than normal distribution, such as t distribution. Ravi & Butar (2010) utilized the maximum likelihood method to estimate parameters in stock-market data based on heavy-tailed distribution. Nadarajah & Kotz (2008) provided a review of methods for estimation and simulation in multivariate t distributions. Lange et al. (1989) illustrated robust statistical inferences in general linear model with multivariate t distributed errors and applied to various data, using both linear and nonlinear regression. Fernandez & Steel (1999) considered the likelihood-based parameter estimations in multivariate linear regression model with independent Student-t distributed errors. Rahman & Khan (2007) demonstrated the use of Bayesian method to derive predictive distribution for multiple linear regression models when error components followed multivariate Student-t distribution.

Various techniques have been proposed to solve for violation of the underlying assumptions under formulated model in an attempts to improve the properties of parameter estimators. However, most studies usually focus on one problem at a time. Unfortunately, when one problem is remedied, another problem possibly occurs. Only a few of them have tackled several problems simultaneously. Ayinde et al. (2014) examined performances of four estimators; Ordinary Least Square (OLS), Cochrane-Orcutt (COR), Maximum Likelihood (ML) and Principal Component (PC) based estimators in linear regression model under related regressors and error terms for prediction. Instead of seeking remedial approaches, this research focuses on parameter estimation techniques that can mitigate non-validity of the underlying assumptions when random errors in regression model are simultaneously correlated and follow heavy-tailed distribution. We incorporate prior knowledge about parameters into estimation process using hierarchical Bayes approach. The result is then compared to the classical approach, maximum likelihood (ML). Bias, variance and mean square error of maximum likelihood and hierarchical Bayes estimators are examined through Monte Carlo simulations. Gibbs sampling is performed to obtain the parameter estimates due to computer-based process in hierarchical Bayes.

The outline of this paper is as follows. Multiple linear regression model with intraclass correlation is described in the next section, accompanying with parameter estimation methods, including maximum likelihood (ML) and hierarchical Bayes (HB), and Gibbs sampling procedure. Simulation result is provided in the third section. The conclusion is drawn in the final section.

Methods

1. Multiple linear regression model with intraclass correlation

Consider the multiple linear regression model

doi: 10.14416/j.appsci.2019.07.002

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

where $\underset{\sim}{y}$ is an $n \times 1$ vector of response variables, 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 and constant variance, $\underset{\sim}{\varepsilon} \sim N(\mathbf{0},\ \sigma^2\mathbf{I})$, where $\mathbf{I}$ is the $n \times n$ identity matrix.

Consider the form of intraclass correlation. Let $\Sigma = \sigma^2\Omega$ be an $n \times n$ positive definite matrix, where Ω denotes an $n \times n$ intraclass correlation coefficient matrix (Donner & Bull, 1983; Press, 2005) and ρ denotes intraclass correlation coefficient, written as

$$\Sigma \ = \ \sigma^2\Omega \ = \ \sigma^2 \begin{bmatrix} 1 & \rho & \cdots & \rho \\ \rho & 1 & \cdots & \rho \\ \vdots & \vdots & \ddots & \vdots \\ \rho & \cdots & \cdots & 1 \end{bmatrix}, \tag{2}$$

where $-\dfrac{1}{n-1} < \rho \le 1.$

Alternatively,

$$\Sigma \ = \ \sigma^2\Omega \ = \ \sigma^2\left[(1-\rho)I_n + \rho\underset{\sim}{e}\underset{\sim}{e}'\right],$$

where $\underset{\sim}{e}$ is an $n \times 1$ vector one and I_n is the $n \times n$ identity matrix. The determinants of Ω and Σ are then obtained as following:

$$|\Omega| \ = \ (1-\rho)^{n-1}\left[1+(n-1)\rho\right],$$

and

$$|\Sigma| \ = \ \sigma^{2n}(1-\rho)^{n-1}\left[1+(n-1)\rho\right], \quad |\Sigma| > 0.$$

Also, the inverse matrices of Ω and Σ are

$$\Omega^{-1} \ = \ \frac{I_n}{(1-\rho)} \ - \ \frac{\rho\underset{\sim}{e}\underset{\sim}{e}'}{(1-\rho)\left[1+(n-1)\rho\right]},$$

$$\Sigma^{-1} \ = \ \frac{I_n}{\sigma^2(1-\rho)} \ - \ \frac{\rho\underset{\sim}{e}\underset{\sim}{e}'}{\sigma^2(1-\rho)\left[1+(n-1)\rho\right]}.$$

In this paper, we consider regression model when random errors are correlated and have heavy-tailed distribution. The errors in this case are assumed to follow multivariate *t* distribution, denoted as $\underset{\sim}{\varepsilon} \sim t(\nu,\ 0,\ \Sigma)$, where $\Sigma = \sigma^2\Omega$. Thus, the random variable $\underset{\sim}{Y}$ also follows multivariate *t* distribution, $\underset{\sim}{Y} \sim t(\nu,\ X\underset{\sim}{\beta},\ \Sigma)$.

 doi: 10.14416/j.appsci.2019.07.002

Joint probability distribution function of $\underset{\sim}{Y}$ can be expressed as

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

Since $\left(\underset{\sim}{y}-X\underset{\sim}{\beta}\right)' \Omega^{-1}\left(\underset{\sim}{y}-X\underset{\sim}{\beta}\right) = \left(\underset{\sim}{y}-X\hat{\underset{\sim}{\beta}}\right)' \Omega^{-1}\left(\underset{\sim}{y}-X\hat{\underset{\sim}{\beta}}\right)+\left(\underset{\sim}{\beta}-\hat{\underset{\sim}{\beta}}\right)' X'\Omega^{-1}X\left(\underset{\sim}{\beta}-\hat{\underset{\sim}{\beta}}\right)$, equation (3) can be rewritten as

$$f\left(\underset{\sim}{y} \mid \underset{\sim}{\beta}, \sigma^2\right)=\frac{\Gamma\left(\frac{\nu+n}{2}\right)}{\left(\nu\pi\sigma^2\right)^{\frac{n}{2}} \Gamma\left(\frac{\nu}{2}\right)|\Omega|^{\frac{1}{2}}}\left\{1+\frac{1}{\nu\sigma^2}\left[\left(\underset{\sim}{y}-X\hat{\underset{\sim}{\beta}}\right)' \Omega^{-1}\left(\underset{\sim}{y}-X\hat{\underset{\sim}{\beta}}\right)+\left(\underset{\sim}{\beta}-\hat{\underset{\sim}{\beta}}\right)' X'\Omega^{-1}X\left(\underset{\sim}{\beta}-\hat{\underset{\sim}{\beta}}\right)\right]\right\}^{-\frac{(\nu+n)}{2}}. \quad (4)$$

2. Maximum likelihood estimation (ML)

The principle of maximum likelihood method is to find the estimate values of parameters that most likely to occur by maximizing the likelihood function, which is a joint function of all parameters and observed values. In practice, the likelihood function is usually replaced by the log-likelihood function for convenient in calculation. The estimations of parameters $\underset{\sim}{\beta}$ and σ^2 can be demonstrated as follow.

2.1 Estimation of parameter $\underset{\sim}{\beta}$

Consider the likelihood function

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

Then, the log-likelihood function is

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

Next, differentiate the log-likelihood function with respect to the parameter $\underset{\sim}{\beta}$ and equate to zero, the maximum likelihood estimator of $\underset{\sim}{\beta}$ is thus derived as

 doi: 10.14416/j.appsci.2019.07.002

$$\hat{\underset{\sim}{\beta}} = \left(X'\Omega^{-1}X \right)^{-1} X'\Omega^{-1}\underset{\sim}{y}. \quad (5)$$

2.2 Estimation of parameter σ^2

Similarly, differentiate the log-likelihood function with respect to the parameter σ^2 and equate to zero, the maximum likelihood estimator of σ^2, after replacing $\underset{\sim}{\beta}$ with $\hat{\underset{\sim}{\beta}}$, is resulted as.

$$\hat{\sigma}^2 = \frac{1}{n}\left(\underset{\sim}{y} - X\hat{\underset{\sim}{\beta}}\right)' \Omega^{-1}\left(\underset{\sim}{y} - X\hat{\underset{\sim}{\beta}}\right). \quad (6)$$

3. Hierarchical Bayes estimation (HB)

Parameter estimation with classical approach, such as least squares method or maximum likelihood method, considers a known probability density function consisting of constant, but unknown parameters, while Bayesian approach consider parameters as random variables with some probability density function, called prior distribution. The form of prior distribution is defined by its own parameters, called hyperparameters. Hierarchical Bayes approach (Gill, 2008) is involved when the form of probability density function of hyperparameters are known. Let $Y_1, Y_2, ..., Y_n$ be random samples from a population with probability density function $f(y;\theta) = f(y \mid \theta)$, where parameter θ is a value of random variable Θ and $f(y_i \mid \theta)$ is a conditional density function of random variable Y when defined $\Theta = \theta$. Let $h(\theta \mid Y_1, Y_2, ..., Y_n)$ be a conditional density function of random variable Θ given the observed values $Y_1, Y_2, ..., Y_n$. Suppose $Y_1 = y_1, Y_2 = y_2, ..., Y_n = y_n$ then $h(\theta \mid \underline{y})$ is the posterior density function. $\pi(\theta)$ is a prior density function of Θ. For continuous random variable Θ, the posterior density function can be derived as following.

$$h\left(\theta \mid \underset{\sim}{y}\right) = \frac{L(\theta)\cdot\pi(\theta)}{\int_{\theta} L(\theta)\cdot\pi(\theta)\, d\theta} \propto L(\theta)\cdot\pi(\theta). \quad (7)$$

Likewise, the integral sign in equation (7) can be replaced by summation to obtain the posterior density function for discrete random variable Θ.

In this research, the derivation of parameter estimators is divided into two cases as follows.

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

Assuming independent noninformative priors for both $\underset{\sim}{\beta}$ and σ^2 as $\pi\left(\underset{\sim}{\beta}\right) \propto c,$ where c is constant, and $\pi\left(\sigma^2\right) \propto \frac{1}{\sigma^2}.$

Consider the joint posterior density functions of $\underset{\sim}{\beta}$ and σ^2

Vol. 18 No. 2: 11-21 [2019]
doi: 10.14416/j.appsci.2019.07.002

$$h\left(\underset{\sim}{\beta},\ \sigma^2 \mid \underset{\sim}{y}\right) \propto \frac{\Gamma\left(\frac{\nu+n}{2}\right)}{\left(\nu\pi\sigma^2\right)^{\frac{n}{2}}\Gamma\left(\frac{\nu}{2}\right)\left|\Omega\right|^{\frac{1}{2}}} \times \left\{1+\frac{1}{\nu\sigma^2}\left[\left(\underset{\sim}{y}-X\hat{\underset{\sim}{\beta}}\right)'\Omega^{-1}\left(\underset{\sim}{y}-X\hat{\underset{\sim}{\beta}}\right)+\left(\underset{\sim}{\beta}-\hat{\underset{\sim}{\beta}}\right)'X'\Omega^{-1}X\left(\underset{\sim}{\beta}-\hat{\underset{\sim}{\beta}}\right)\right]\right\}^{-\frac{(\nu+n)}{2}}. \quad (8)$$

Then, the full conditional posterior density function of $\underset{\sim}{\beta}$ with squared error loss function is obtained as

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

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

$$\underset{\sim}{\beta} \mid \sigma^2,\ \underset{\sim}{Y}\ \sim\ t\left(\nu,\ \hat{\beta},\ \left(\frac{X'\Omega^{-1}X}{\sigma^2}\right)^{-1}\right). \quad (9)$$

Solving the joint posterior density functions of $\underset{\sim}{\beta}$ and σ^2 in equation (8), then we obtain the full conditional posterior density function of σ^2 as

$$h_2\left(\sigma^2 \mid \underset{\sim}{\beta},\ \underset{\sim}{y}\right) \propto \frac{1}{\left(\sigma^2\right)^{\frac{n}{2}}} \times \left\{1+\frac{1}{\nu\sigma^2}\left[\left(\underset{\sim}{y}-X\hat{\underset{\sim}{\beta}}\right)'\Omega^{-1}\left(\underset{\sim}{y}-X\hat{\underset{\sim}{\beta}}\right)+\left(\underset{\sim}{\beta}-\hat{\underset{\sim}{\beta}}\right)'X'\Omega^{-1}X\left(\underset{\sim}{\beta}-\hat{\underset{\sim}{\beta}}\right)\right]\right\}^{-\frac{(\nu+n)}{2}}. \quad (10)$$

Since, the full conditional posterior density function of σ^2 is complicated and cannot rearrange into a known closed form. Hence, numerical procedure, such as Gibbs sampling, is applied to find the estimate of parameter σ^2.

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

Suppose parameters $\underset{\sim}{\beta}$ and σ^2 are independent and assume informative priors for $\underset{\sim}{\beta}$ and σ^2 as follows. $\underset{\sim}{\beta} \sim N\left(\underset{\sim}{\mu},\ m\sigma^2 I_p\right)$, and hyperparameter $\underset{\sim}{\mu} \sim N\left(\mu_0, v_0 I_p\right)$, where m, μ_0 and v_0 are known values. $\sigma^2 \sim IG(\alpha,\gamma)$ and hyperparameters α and γ are distributed as inverted gamma, $\alpha \sim IG(f_0, e_0)$ and $\gamma \sim IG(r_0, m_0)$, where $f_0,\ e_0,\ r_0$ and m_0 are known values.

The joint posterior density function of $\underset{\sim}{\beta}, \sigma^2, \underset{\sim}{\mu}, \alpha$ and γ are expressed as

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

$$\cdot\frac{1}{\left(2\pi m\sigma^2\right)^{\frac{p}{2}}}e^{-\frac{1}{2m\sigma^2}\left(\underset{\sim}{\beta}-\underset{\sim}{\mu}\right)'\left(\underset{\sim}{\beta}-\underset{\sim}{\mu}\right)}\times\frac{1}{\left(2\pi v_0\right)^{\frac{p}{2}}}e^{-\frac{1}{2v_0}\left(\underset{\sim}{\mu}-\mu_0\right)'\left(\underset{\sim}{\mu}-\mu_0\right)}$$

$$\cdot\frac{\gamma^{\alpha}}{\Gamma(\alpha)}\left(\sigma^2\right)^{-(\alpha+1)}e^{-\frac{\gamma}{\sigma^2}}\times\frac{e_0^{f_0}}{\Gamma\left(f_0\right)}(\alpha)^{-(f_0+1)}e^{-\frac{e_0}{\alpha}}\times\frac{m_0^{r_0}}{\Gamma\left(r_0\right)}(\gamma)^{-(r_0+1)}e^{-\frac{m_0}{\gamma}}.$$

Then, the full conditional posterior density function of $\underset{\sim}{\beta}$ is obtained as

$$h_1\left(\underset{\sim}{\beta}\mid\sigma^2,\underset{\sim}{\mu},\ \alpha,\ \gamma,\ \underset{\sim}{y}\right)\propto\left\{1+\frac{1}{\nu\sigma^2}\left[\left(\underset{\sim}{y}-X\hat{\underset{\sim}{\beta}}\right)'\Omega^{-1}\left(\underset{\sim}{y}-X\hat{\underset{\sim}{\beta}}\right)+\left(\underset{\sim}{\beta}-\hat{\underset{\sim}{\beta}}\right)'X'\Omega^{-1}X\left(\underset{\sim}{\beta}-\hat{\underset{\sim}{\beta}}\right)\right]\right\}^{-\frac{(\nu+n)}{2}}$$
$$\cdot e^{-\frac{1}{2m\sigma^2}\left(\underset{\sim}{\beta}-\underset{\sim}{\mu}\right)'\left(\underset{\sim}{\beta}-\underset{\sim}{\mu}\right)},\tag{11}$$

and the full conditional posterior density function of σ^2 is

$$h_2\left(\sigma^2\mid\underset{\sim}{\beta},\ \underset{\sim}{\mu},\ \alpha,\ \gamma,\ \underset{\sim}{y}\right)\propto\left\{1+\frac{1}{\nu\sigma^2}\left[\left(\underset{\sim}{y}-X\hat{\underset{\sim}{\beta}}\right)'\Omega^{-1}\left(\underset{\sim}{y}-X\hat{\underset{\sim}{\beta}}\right)+\left(\underset{\sim}{\beta}-\hat{\underset{\sim}{\beta}}\right)'X'\Omega^{-1}X\left(\underset{\sim}{\beta}-\hat{\underset{\sim}{\beta}}\right)\right]\right\}^{-\frac{(\nu+n)}{2}}$$
$$\cdot\frac{1}{\left(\sigma^2\right)^{\frac{n+p}{2}+(\alpha+1)}}e^{-\frac{1}{\sigma^2}\left(\underset{\sim}{\beta}-\underset{\sim}{\mu}\right)'\left(\underset{\sim}{\beta}-\underset{\sim}{\mu}\right)-\frac{\gamma}{\sigma^2}}.\tag{12}$$

Next, the full conditional posterior density function of $\underset{\sim}{\mu}$ is yielded as

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

 doi: 10.14416/j.appsci.2019.07.002

where $A^{-1}=\left(\frac{1}{m\sigma^2}+\frac{1}{\nu_0}\right)$ and $\tilde{\mu} = A\left(\frac{\underset{\sim}{\beta}}{m\sigma^2}+\frac{\mu_0}{\nu_0}\right)$.

Thus, the posterior distribution of $\underset{\sim}{\mu}$ follows normal distribution,

$$\underset{\sim}{\mu} \mid \underset{\sim}{\beta},\ \sigma^2, \alpha,\ \gamma,\ \underset{\sim}{Y}\ \sim\ N\left(\tilde{\mu},\ A\right). \tag{13}$$

Finally, full conditional posterior density functions of $\alpha,\ \gamma$ are resulted as

$$h_3\left(\alpha \mid \underset{\sim}{\beta}, \sigma^2, \gamma, \underset{\sim}{y}\right) \propto \frac{1}{\Gamma(\alpha)}(\alpha)^{-(f_0+1)} e^{-\frac{e_0}{\alpha}} \gamma^{\alpha}\left(\sigma^2\right)^{-\alpha}, \tag{14}$$

$$h_4\left(\gamma \mid \underset{\sim}{\beta}, \sigma^2, \alpha, \underset{\sim}{y}\right) \propto (\gamma)^{-(r_0+1-\alpha)} e^{-\frac{m_0}{\gamma}-\frac{\gamma}{\sigma^2}}. \tag{15}$$

Similar to noninformative priors, full conditional posterior density functions of $\underset{\sim}{\beta}$ and σ^2 are complicated and cannot be simplified to an explicit form. To ease computation, Gibbs sampler is used to obtain the estimates of parameters by alternately generating parameters from marginal posterior distributions. With large enough sample, a sequence of simulated parameter which approximately follows a specified probability distribution is obtained and therefore its characteristics, such as mean and variance, can be estimated.

4 Gibbs Sampling Procedure

4.1 Gibbs sampler for noninformative priors

The estimates of parameters using hierarchical Bayes method with vague priors can obtain via Gibbs sampler as following steps.

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

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

4.2 Gibbs sampler for informative priors

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

1) 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 $\mu \sim N\left(\mu_0, \nu_0\right)$ and $\sigma^{2\,(t)} \sim IG(\alpha,\ \gamma)$,

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

3) 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

 doi: 10.14416/j.appsci.2019.07.002

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

4) 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 $f_0 = 0.01,\ e_0 = 0.01$,

5) 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 and Discussion

In this paper, the multiple linear regression model with two regressors is considered with the presence of correlated random errors with intraclass correlation structure and multivariate *t* distribution, Both regressors are constructed from standard normal distribution. Data are constructed on three correlation coefficient levels as low $(\rho = 0.1)$, moderate $(\rho = 0.5)$ and high $(\rho = 0.9)$ with the sample size of 60, 100, 250 and 350. The simulation is repeated 1,000 times and Gibbs sampler is implemented for parameter estimation. For each situation, three estimates of parameter obtained from maximum likelihood (*ML*), hierarchical Bayes using noninformative priors (*HBN*) and informative priors (*HBI*) are compared. The bias, variance and mean square error (MSE) are considered as criteria for comparison.

Table 1. The MSE, bias and variance of three regression coefficients $(\beta_0, \beta_1, \beta_2)$.

n	ρ	MSE			Bias			Variance		
		ML	*HBN*	*HBI*	*ML*	*HBN*	*HBI*	*ML*	*HBN*	*HBI*
60	0.1	4.705	6.089	0.572	2.005	2.507	0.011	0.260	0.006	0.156
	0.5	5.131	5.331	0.916	2.013	2.464	0.031	0.652	0.019	0.571
	0.9	4.225	4.596	0.138	2.022	2.293	0.034	1.044	0.072	0.915
100	0.1	4.617	4.094	0.540	2.004	2.141	0.021	0.210	0.013	0.138
	0.5	5.165	2.093	0.870	2.004	2.009	0.001	0.601	0.058	0.540
	0.9	4.178	1.087	0.119	2.057	1.326	0.055	0.934	0.334	0.867
250	0.1	4.543	0.773	0.512	2.007	1.019	0.009	0.150	0.048	0.119
	0.5	5.137	0.910	0.971	2.001	0.723	0.016	0.537	0.250	0.511
	0.9	4.180	0.603	0.110	2.035	0.173	0.042	0.998	0.880	0.970
350	0.1	4.550	0.603	0.485	2.013	0.736	0.016	0.128	0.062	0.110
	0.5	5.064	0.865	0.905	2.014	0.736	0.024	0.495	0.062	0.484
	0.9	4.705	6.089	0.572	2.035	0.124	0.038	0.922	0.849	0.904

The MSE, bias and variance of three regression coefficients obtained from *ML*, *HBN* and *HBI* can be found in Table 1. On overall, MSEs for *HBI* are smallest in most cases, except for $\rho = 0.5$ and $n = 250,\ 350.$ In all situations, *HBI* also yields the least bias while *HBN* yields the least variance. However, the bias of *HBN* seems to reduce for larger sample sizes. It is also observed that MSEs of three methods tend to decrease with increasing *n*.

 doi: 10.14416/j.appsci.2019.07.002

Table 2. The MSE, bias and variance of the estimated error variance $\left(\sigma^2\right)$

n	ρ	MSE			Bias			Variance		
		ML	*HBN*	*HBI*	*ML*	*HBN*	*HBI*	*ML*	*HBN*	*HBI*
60	0.1	4.664	5.744	0.068	2.069	2.231	0.067	0.383	0.767	0.063
	0.5	15.146	34.943	0.066	3.775	4.767	0.038	0.896	12.217	0.064
	0.9	380.125	1221.248	0.070	19.114	24.738	0.048	14.769	609.294	0.067
100	0.1	4.825	3.408	0.042	2.141	1.675	0.033	0.241	0.603	0.005
	0.5	15.294	20.921	0.040	3.842	3.322	0.036	0.532	9.887	0.038
	0.9	387.047	736.790	0.041	19.451	13.107	0.034	8.724	564.987	0.040
250	0.1	4.865	0.235	0.015	2.184	0.401	0.017	0.094	0.074	0.015
	0.5	15.637	1.400	0.015	3.925	0.506	0.017	0.234	1.144	0.014
	0.9	388.775	5.716	0.016	19.622	0.145	0.015	3.765	5.695	0.015
350	0.1	4.902	0.067	0.012	2.199	0.202	0.002	0.068	0.026	0.012
	0.5	15.867	0.066	0.011	3.964	0.202	0.013	0.158	0.026	0.011
	0.9	384.836	0.015	0.011	19.794	0.033	0.008	3.053	0.014	0.011

Table 2 displays MSE, bias and variance of the estimated error variance obtained from *ML*, *HBN* and *HBI*. The results reveal that *HBI* produces the smallest bias, variance and hence MSEs in all situations. The variance of three estimators tends to decrease as the sample size increases. In addition, the bias and MSEs for *HBN* and *HBI* are likely to be lower as an increase of *n*.

Conclusion

Several applications of regression model involve prediction which requires good estimates of model parameters. When the assumption of regression analysis is violated, it can influence the quality of parameter estimates and then prediction performance. This research focuses on parameter estimation method when random errors in regression model are correlated with intraclass structure and follow heavy-tailed distribution. Instead of seeking for techniques to resolve the problems, we choose to include more information, both informative and noninformative priors, into the estimation method, called hierarchical Bayes approach, in order to lessen the problem and then compare the result with the classical approach, maximum likelihood.

For regression coefficient estimation, the simulated study indicates that *HBN* yields the smallest variance while *HBI* yields the smallest bias and MSEs. For error variance estimation, *HBI* yields the smallest bias, variance and thus MSEs in all situations under study. As a result, *HBI* tends to produce the estimators for regression coefficients and error variance more efficient than *ML* and *HBN*, which is benefit for prediction performance.

As seen that the parameter estimation based on hierarchical Bayes approach performs rather well when error components violate the model assumptions, specifically correlated errors with intraclass structure and heavy-tailed distribution. Hence, informative priors incorporating into the parameter estimation method can abate the mentioned problem of violation which is useful in practice.

Acknowledgement

This research is financially 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

Ayinde, K., Apata, E. O. & Alaba, O. O. (2012). Estimators of linear regression model and prediction under some assumptions violation. *Open Journal of Statistics*, *2*, 534-546.

Babatunde, O. S., Ikughur, A. J., Ogunmola, A. O. & Oguntunde, P. E. (2014). On the effect of autocorrelation in regression model due to specification error. *International Journal of Modern Mathematical Science*, *10*(3), 239-246.

Donner, A. & Bull, S. (1983). Inferences concerning a common intraclass correlation coefficient. *Biometrics*, *39*(3), 771-775.

Farrell, S. & Ludwig, C. (2008). Bayesian and maximum likelihood estimation of hierarchical response time model. *Psychonomic Bulletin & Review*, *15*(6), 1209-1217.

Fernandez, C. & Steel, M. F. J. (1999). Multivariate Student-*t* regression model: pitfalls and inference. *Biometrika*, *86*(1), 153-167.

Gill, J. (2008). *Bayesian Methods: a Social and Behavioral Sciences Approach* (2nd ed.). USA: Chapman & Hall.

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*(408), 881-896.

Montgomery, D. C., Peck, E. A. & Vining, G. G. (2012). *Introduction to Linear Regression* (5th ed.). New Jersey, USA: John Wiley & Sons.

Nadarajah, S. & Kotz, S. (2008). Estimations for the multivariate *t* distribution. *Acta Applicandae Mathematicae*, *102*(1), 99-118.

Paul, S. R. (1990). Maximum likelihood estimation of intraclass correlation in the analysis of familial data: estimating equation approach. *Biometrika*, *77*(3), 549-555.

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

Rahman, A. & Khan, S. (2007). *Prediction distribution for linear regression model with multivariate Student-t errors under the Bayesian approach*. pp. 188-193. In the 3rd International Conference on Research and Education in Mathematics (ICREM3), 10-12 April 2007, Kuala Lumpur, Malaysia.

Ravi, A. & Butar, F. (2010). An insight into heavy-tailed distribution. *Journal of Mathematical Sciences & Mathematics Education*, *5*(1), 19-31.

Tanizaki, H. (2003). On regression models with autocorrelated error: small sample properties. *International Journal of Pure and Applied Science*, *5*(3), 247-252.