Communications for Statistical Applications and Methods
2014, Vol. 21, No. 2, 193–200

DOI: http://dx.doi.org/10.5351/CSAM.2014.21.2.193
ISSN 2287-7843

Bayesian Inference for Censored Panel Regression Model

Seung-Chun Lee[1,a], Byongsu Choi[b]

[a]Department of Applied Statistics, Hanshin University, Korea
[b]Department of Multimedia Engineering, Hansung University, Korea

Abstract

It was recognized by some researchers that the disturbance variance in a censored regression model is frequently underestimated by the maximum likelihood method. This underestimation has implications for the estimation of marginal effects and asymptotic standard errors. For instance, the actual coverage probability of the confidence interval based on a maximum likelihood estimate can be significantly smaller than the nominal confidence level; consequently, a Bayesian estimation is considered to overcome this difficulty. The behaviors of the maximum likelihood and Bayesian estimators of disturbance variance are examined in a fixed effects panel regression model with a limited dependent variable, which is known to have the *incidental parameter problem.* Behavior under random effect assumption is also investigated.

Keywords: Censored panel regression, Gibbs sampling, incidental parameter problem.

1. Introduction

In many statistical data analysis, a dependent variable could be subject to censoring due to a variety of reasons. In particular, the censored data is common in survival analysis. The ordinary least squares method fails to provide consistent estimates for a conventional regression model, if the fraction of censored data is significant. This leads to discuss the estimation methods in a censored regression model (see, *e.g.*, Tobin, 1958; Maddala, 1983; Amemiya, 1984). Currently, the maximum likelihood estimation is a dominating method and is implemented in various **R** packages such as **AER** (Kleiber and Zeileis, 2009), **censReg** (Henningsen, 2011), and **NADA** (Lee, 2013).

A Bayesian estimation of the censored regression model was considered first by Chib (1992). In his pioneering work, several Bayesian computations were examined under the assumption of normally distributed errors. He concluded that Bayesian estimates achieve the improvement in small samples over maximum likelihood estimates, and the Gibbs sampling method provides close estimates to exact Bayesian estimates. Bayesian estimation using Bayesian Markov Chain Monte Carlo (MCMC) was also implemented in **R** package, **MCMCpack** (Martin *et al.*, 2013).

Lee and Choi (2013) also studied a Bayesian estimation for the same problem, and observed that the maximum likelihood estimate tends to have smaller standard errors than the Bayesian estimate when the sample size is small. This might contradict to the conclusion of Chib (1992). The contradiction originates from the tendency of the maximum likelihood estimation to underestimate the disturbance variance, which has an implication for the estimation of standard errors.

The underestimation of disturbance variance was also recognized by Green (2004a, b). He examined the behavior of a maximum likelihood estimator in a fixed effects panel regression model with

This work was supported by Hanshin University research grant.
This research was financially supported by Hansung University.
[1] Corresponding author: Professor, Department of Applied Statistics, Hanshin University, Hanshin Dae-Gil 137, Osan, Gyunggi-Do 447-791, Korea. E-mail: seung@hs.ac.kr

Published online 31 March 2014 / journal homepage: http://csam.or.kr

censored dependent variable. The model is known to have *incidental parameters problem* (see, *e.g.* Lancaster, 2000). That is, the maximum likelihood estimator of the time-invariant individual specific effect is inconsistent when T(the length of panel) is fixed. He concluded that the location coefficients were unaffected by the *incidental parameters problem,* but the finite sample bias appeared in the estimation of disturbance variance. This has an implication for the estimation of the marginal effect as well as the asymptotic standard error. For instance, the coverage probability of the confidence interval based on the maximum likelihood estimate and its asymptotic standard error is significantly lower than nominal level. It would be dangerous to use the maximum likelihood method for statistical problems such as the interval estimation and the hypothesis test when the sample size is small.

It seems that the finite sample bias of the maximum likelihood estimate in the estimation of disturbance variance would not be necessarily due to the *incidental parameters problem* because the same problem occurs under the plain censored regression model considered by Lee and Choi (2013). In their paper, an Wald-like confidence interval based on Bayesian estimate and its standard error was shown to have a proper coverage probability. Thus, it is believed that the finite sample bias in the estimation of disturbance variance was less significant in the Bayesian estimation. In this note, we consider the Bayesian estimation for the panel regression model with censored dependent variable, and demonstrate that the Bayesian estimation can provide the proper estimate of the disturbance variance. This shows that the Bayesian method has an advantage over the maximum likelihood method in the inference of the censored panel regression model with small sample.

Several authors have considered the Bayesian inference for the panel regression model. For instance, Hamilton (1999) analyzed a survival data using a Bayesian panel Tobit model. Bruno (2004) compared the Bayesian estimation with classical maximum likelihood estimates assuming random effect and limited dependent variable. Morawetz (2006) build a Bayesian model for original panel regression. See also Martin *et al.* (2013). All of these works employed conjugate priors, and it requires specifying the value of hyper-parameters. In this paper, we use noninformative priors to build a Bayesian model, in which there is an advantage of no specifying hyper-parameters.

2. Censored Panel Regression Model

We consider a standard censored regression model for panel data with individual specific effects:

$$y_{it}^* = \boldsymbol{x}_{it}'\boldsymbol{\beta} + \mu_i + \epsilon_{it}, \tag{2.1}$$

$$y_{it} = \begin{cases} a, & \text{if } y_{it}^* \le a, \\ y_{it}^*, & \text{if } a < y_{it}^* < b, \\ b, & \text{if } y_{it}^* \ge b. \end{cases}$$

Here the subscripts $i = 1, \ldots, N$, and $t = 1, \ldots, T_i$ indicate the individual and the time period, respectively. T_i is the number of time periods observed for i^{th} individual, μ_i is a time-invariant individual specific effect, and ϵ_{it} is the remaining disturbance. In addition, y_{it} is observed output, $\boldsymbol{x}_{it}$ is the vector of inputs for the i^{th} individual in the t^{th} period, and $\boldsymbol{\beta}$ is a vector of slope parameters.

When μ_i is assumed to be constant over time, the model is referred to as the "fixed effects" model, which is known to have *incidental parameter problem,* while "random effects" model treats μ_i as a random variable just like ϵ_{it}. Green (2004a) argued that the random effects model requires an unpalatable orthogonality assumption that the effects should be uncorrelated with dependent variables, which can be relaxed by the fixed effects model, but Maddala (1983) gave several reasons to employ a random effects model. For instance, the individual specific effect should be treated as random if one wants to make inferences about the population from which cross-section data came from. Thus, choice between the fixed and the random effects models should depend upon the statistical properties of implied estimators. As McCulloch (1996) stated, a frequentist decision to regard an effect as fixed or random is indeed complicated one.

In what follows, we will assume that disturbance terms are independently and normally distributed with mean 0 and variance σ^2, and we will focus on the Bayesian estimation. You might refer to Henningsen (2011) for the maximum likelihood estimation.

2.1. Random effects model

Let $NT = \sum_{i=1}^{N} T_i$ and write (2.1) as a matrix form,

$$\boldsymbol{y}^* = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{W}\boldsymbol{\mu} + \boldsymbol{\epsilon}, \tag{2.2}$$

where $\boldsymbol{y}^*$ is a $NT \times 1$ vector of latent values, $\boldsymbol{X}$ is a $NT \times k$ full column rank design matrix of input variables, $\boldsymbol{W} = \oplus_{i=1}^{N} \mathbf{1}_{T_i}$, $\boldsymbol{\mu}$ is a $N \times 1$ vector of individual specific effects, and $\boldsymbol{\epsilon}$ is a $NT \times 1$ vector of disturbance terms. Here $\oplus$ denotes the direct sum of matrices, and $\mathbf{1}_{T_i}$ is a $T_i \times 1$ vector of 1's.

For (2.2), a hierarchical Bayesian model can be setup as $\boldsymbol{y}^*|\boldsymbol{\beta}, \boldsymbol{\mu}, \sigma^2 \sim N(\boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{W}\boldsymbol{\mu}, \mathbf{I}\,\sigma^2)$, and $\boldsymbol{\beta}, \boldsymbol{\mu}$ and σ^2 have a certain proper or improper joint prior distribution. In the absence of prior information about the fixed effects $\boldsymbol{\beta}$ and the disturbance variance σ^2, $p(\boldsymbol{\beta}, \sigma^2) \propto 1/\sigma^2$ is customarily chosen for noninformative prior (see, *e.g.*, Zeger and Karim, 1991; Kass and Natarajan, 2006; Gelman *et al.*, 2013). From a Bayesian standpoint, all effects are random. Thus, if a frequentist decides that an effect is indeed random and provides a distribution, then most Bayesian might be willing to use that distribution as a prior. In this point of view, $N(\mathbf{0}, \mathbf{I}\,\tau^2)$ is chosen for the random effects, because frequentist believes that μ_i's are independently and identically distributed normal random variables with mean 0 and common variance τ^2; however, this prior requires a second stage prior for τ^2. A conventional diffuse noninformative prior for τ^2 is $p(\tau^2) \propto 1/\tau$ (Gelman *et al.*, 2013).

Note that certain improper priors lead to improper posterior distributions, and the Gibbs chains corresponding to improper posterior are quite ill behaved. One must be careful in choosing the improper prior since Gibbs output does not inform the user that the posterior is improper, However, it can be shown that the priors described above yield proper posterior distributions. See Theorem 1 of Hobert and Casella (1996) for further details.

The full conditional posterior distributions for the Gibbs sampling can be obtained from the joint posterior distribution,

$$p\left(\boldsymbol{\beta}, \boldsymbol{\mu}, \tau^2, \sigma^2 \,|\, \boldsymbol{y}^*\right) \propto p\left(\boldsymbol{y}^*|\boldsymbol{\beta}, \boldsymbol{\mu}, \sigma^2\right) p\left(\boldsymbol{\beta}\right) p\left(\boldsymbol{\mu} \,|\, \tau^2\right) p\left(\tau^2\right) p\left(\sigma^2\right). \tag{2.3}$$

From (2.3), we have the following full conditional distributions.

1. $\boldsymbol{\beta} \,|\, \boldsymbol{\mu}, \sigma^2, \boldsymbol{y}^* \sim N((\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'(\boldsymbol{y}^* - \boldsymbol{W}\boldsymbol{\mu}), \sigma^2(\boldsymbol{X}'\boldsymbol{X})^{-1})$.
2. $\boldsymbol{\mu} \,|\, \boldsymbol{\beta}, \tau^2, \sigma^2, \boldsymbol{y}^* \sim N((\boldsymbol{W}'\boldsymbol{W} + \sigma^2/\tau^2\mathbf{I})^{-1}\boldsymbol{W}'(\boldsymbol{y}^* - \boldsymbol{X}\boldsymbol{\beta}), \sigma^2(\boldsymbol{W}'\boldsymbol{W} + \sigma^2/\tau^2\mathbf{I})^{-1})$.
3. $\tau^2 \,|\, \boldsymbol{\mu} \sim \mathrm{IG}((N-1)/2, \boldsymbol{\mu}'\boldsymbol{\mu}/2)$.
4. $\sigma^2 \,|\, \boldsymbol{\beta}, \boldsymbol{\mu}, \boldsymbol{y}^* \sim \mathrm{IG}(NT/2, 1/2(\boldsymbol{y}^* - \boldsymbol{X}\boldsymbol{\beta} - \boldsymbol{W}\boldsymbol{\mu})'(\boldsymbol{y}^* - \boldsymbol{X}\boldsymbol{\beta} - \boldsymbol{W}\boldsymbol{\mu}))$.

where $\mathrm{IG}(\alpha, \beta)$ represents the inverse-gamma distribution with shape parameter α and scale parameter β.

We need additional full conditional distributions of latent variables, which will be employed at data-augment step (Tanner and Wong, 1987). Note that, given $\boldsymbol{\beta}, \boldsymbol{\mu}, \sigma^2$ and $\boldsymbol{y}$, each y_{it}^* is independent, and has a truncated normal distribution on $(-\infty, a]$, if $y_{it} = a$, or a truncated normal distribution on $[b, \infty)$, if $y_{it} = b$. Otherwise, it has a degenerating distribution at y_{it}.

Let $\theta^{(t)}$ represent the sample value from the full conditional distribution of θ at t^{th} iteration. Then, the Gibbs algorithm is defined to draw samples from the full conditional distributions at each iteration, where the sampling is carried out by the following procedure:

1. draw $\boldsymbol{y}^{*(t)}$ from the distribution $p(\boldsymbol{y}^*|\boldsymbol{\beta}^{(t-1)}, \boldsymbol{\mu}^{(t-1)}, \sigma^{2^{(t-1)}}, \boldsymbol{y})$.
2. draw $\boldsymbol{\beta}^{(t)}$ from the distribution $p(\boldsymbol{\beta}\,|\,\boldsymbol{\mu}^{(t-1)}, \sigma^{2^{(t-1)}}, \boldsymbol{y}^{*(t)})$.
3. draw $\boldsymbol{\mu}^{(t)}$ from the distribution $P(\boldsymbol{\mu}\,|\,\boldsymbol{\beta}^{(t)}, \tau^{2^{(t-1)}}, \sigma^{2^{(t-1)}}, \boldsymbol{y}^{*(t)})$.
4. draw $\tau^{2^{(t)}}$ from the distribution $p(\tau^2\,|\,\boldsymbol{\mu}^{(t)})$.
5. draw $\sigma^{2^{(t)}}$ from the distribution $p(\sigma^2\,|\,\boldsymbol{\beta}^{(t)}, \boldsymbol{\mu}^{(t)}, \boldsymbol{y}^{*(t)})$.

2.2. Fixed effects model

When we consider fixed effects model, the intercept term should not be allowed in $\boldsymbol{X}$, and hence, by the nature of panel data, the column spaces of $\boldsymbol{X}$ and $\boldsymbol{W}$ in (2.2) are disjoint. Thus, if we consider $\boldsymbol{\mu}$ as fixed effects, (2.2) is essentially identical to the plain censored regression model described in Lee and Choi (2013), and the same Bayesian method is applicable. Likewise, **censReg** does not have a specific routine designed for the maximum likelihood estimation of the fixed effects model.

Using a uniform prior for $\boldsymbol{\mu}$ just like $\boldsymbol{\beta}$, the full conditional distributions of $\boldsymbol{y}^*, \boldsymbol{\beta}$ and σ^2 implementing the Gibbs sampling are the same as the random effects model, but the full conditional distribution of $\boldsymbol{\mu}$ is changed to $\boldsymbol{\mu}\,|\,\boldsymbol{\beta}, \sigma^2, \boldsymbol{y}^* \sim N((\boldsymbol{W}'\boldsymbol{W})^{-1}\boldsymbol{W}'(\boldsymbol{y}^* - \boldsymbol{X\beta}), \sigma^2(\boldsymbol{W}'\boldsymbol{W})^{-1})$. The sampling step for τ^2 is not necessary in the Gibbs sampling procedure for the fixed effects model.

The only concern in the fixed effects model is the behavior of the Bayesian estimation under the *incidental parameter problem*. We will address this topic in Section 3 by comparing the behaviors of Bayesian and maximum likelihood estimators through simulation studies.

3. Monte Carlo Results

To compare our results with previous studies, the experiments were done on the designs of Bruno (2004) and Green (2004b) for each random and fixed effects models.

For the random effects model, the latent variable was generated according to

$$y_{it}^* = \beta_0 + \beta_1 x1_{it} + \beta_2 x2_i + \mu_i + \epsilon_{it}, \quad \mu_i \sim \text{NID}(0,1),\ \epsilon_{it} \sim \text{NID}(0,1)$$

and the first regressor $x1_{it}$ is time-dependent to mimic a time series and was generated by an autoregressive process, $x1_{it} = 0.1t + 0.5x1_{i,t-1} + u_{it}$, $u_{it} \sim U(-0.5, 0.5)$ with initial value generated by $x1_{i0} = 5 + 10u_{i0}$. The time invariant variable $x2_i$ was given to 0 or 1 according to $x2_i^*$ being less or greater than 0.5, where $x2_i^*$'s are uniform random numbers. The observed value was mapped from the latent variable by

$$y_{it} = \begin{cases} y_{it}^*, & \text{if } y_{it}^* > 0, \\ 0, & \text{otherwise.} \end{cases}$$

The values for β's were 1.5, −1, and 1, respectively, which resulted in about 30% ∼ 40% censoring rate.

Table 1: Monte Carlo Simulations for random effects model

T	Method	Category	$N = 30$					$N = 50$				
			β_0	β_1	β_2	$\log(\tau)$	$\log(\sigma)$	β_0	β_1	β_2	$\log(\tau)$	$\log(\sigma)$
5	MLE	Estimate	1.5050	1.0085	0.9839	−0.0478	−0.0187	1.5167	−1.0132	0.9905	−0.0299	−0.0087
	Bayes		1.5111	−1.0308	1.0014	0.0160	0.0082	1.5154	−1.0271	1.0047	0.0131	0.0051
	MLE	SE	0.3581	0.1833	0.4205	0.2063	0.0940	0.2662	0.1347	0.3138	0.1447	0.0687
	Bayes		0.3679	0.1616	0.4506	0.1968	0.0844	0.2763	0.1237	0.3369	0.1431	0.0646
	MLE	MSE	0.1431	0.0259	0.2273	0.0530	0.0076	0.0820	0.0162	0.1225	0.0267	0.0044
	Bayes		0.1285	0.0271	0.1961	0.0418	0.0070	0.0730	0.0167	0.1040	0.0222	0.0042
	MLE	Coverage	0.9270	0.9600	0.9045	0.9215	0.9495	0.9260	0.9575	0.9225	0.9200	0.9470
	Bayes		0.9540	0.9495	0.9450	0.9655	0.9545	0.9495	0.9450	0.9610	0.9505	0.9510
10	MLE	Estimate	1.5309	−1.0067	0.9726	−0.0529	−0.0071	1.5232	−1.0030	0.9787	−0.0493	−0.0022
	Bayes		1.5077	−1.0168	1.0102	0.0210	0.0064	1.5127	−1.0096	0.9961	0.0181	0.0025
	MLE	SE	0.2664	0.1378	0.2808	0.1394	0.0622	0.1938	0.0996	0.2023	0.0959	0.0455
	Bayes		0.3359	0.1220	0.4213	0.1655	0.0559	0.2531	0.0936	0.3170	0.1237	0.0431
	MLE	MSE	0.1518	0.0147	0.2522	0.0318	0.0032	0.0912	0.0091	0.1590	0.0188	0.0018
	Bayes		0.1081	0.0144	0.1666	0.0293	0.0032	0.0581	0.0089	0.0951	0.0163	0.0018
	MLE	Coverage	0.8100	0.9615	0.7200	0.8900	0.9535	0.7870	0.9545	0.6750	0.8550	0.9500
	Bayes		0.9515	0.9535	0.9460	0.9450	0.9470	0.9545	0.9505	0.9535	0.9460	0.9480
15	MLE	Estimate	1.5582	−1.0007	0.9428	−0.0943	−0.0023	1.5481	−0.9999	0.9298	−0.0957	0.0015
	Bayes		1.4894	−1.0089	1.0256	0.0248	0.0056	1.5034	−1.0049	0.9945	0.0165	0.0027
	MLE	SE	0.2117	0.0925	0.2091	0.1105	0.0548	0.1520	0.0673	0.1492	0.0758	0.0398
	Bayes		0.3190	0.0818	0.4156	0.1604	0.0482	0.2401	0.0629	0.3117	0.1203	0.0372
	MLE	MSE	0.1593	0.0064	0.2462	0.0312	0.0023	0.0966	0.0039	0.1715	0.0239	0.0014
	Bayes		0.0978	0.0063	0.1567	0.0262	0.0024	0.0538	0.0038	0.0859	0.0145	0.0014
	MLE	Coverage	0.6855	0.9695	0.5800	0.8140	0.9680	0.6665	0.9545	0.5290	0.6485	0.9580
	Bayes		0.9515	0.9540	0.9550	0.9500	0.9465	0.9545	0.9530	0.9585	0.9535	0.9485

The second study design for the fixed effects model was

$$y^*_{it} = \beta_1(x1_{it} + D) + \beta_2 x2_{it} + \mu_i + \epsilon_{it}, \quad \epsilon_{it} \sim \text{NID}(0, 1),$$

where $x1_{it} \sim N(0, 1)$, $x2_{it} = \mathbf{1}[(x_{it} + v_{it}) > 0]$, $v_{it} \sim N(0, 1)$, and $\mu_i = \omega \times (\sqrt{T}\bar{x}_i) + 0.3 \times u_i$, $u_i \sim N(0, 1)$. Note that ω is used to control the amount of correlation between the effects and the regressor. In our study, $\beta_1 = \beta_2 = 1$ and $\omega = 0.5$ were used. The constant D is used to control the rate of censoring. We gave several values of D to maintain the rate of censored data at around 30% – 40%.

We examine the behavior of two estimators on a small sample size because the large sample would guarantee the efficiency of the Bayesian and the maximum likelihood estimators; in addition, the *incidental parameter problem* originates from too many parameters to be estimated with a relatively small sample. That is, the goal of this study is to investigate the small sample properties of the Bayesian and the maximum likelihood estimators.

The numbers of periods and individuals analyzed were $T = 5, 10$ and 15, and $N = 30, 50$. In each experiment, the averages of Bayesian and maximum likelihood estimates and the averages of standard errors were calculated. The average of estimates can be used to measure the bias. These experimental results were based on 2000 replications. In addition, using

$$\frac{1}{2000}\sum_{i=1}^{2000}(\text{estimate} - \text{true value})^2,$$

the mean squared error was estimated to compare the efficiency. Finally, we estimated the coverage probability of 95% Wald-like confidence interval. We believe that the coverage probability would be close to 0.95, if the standard error of estimate was given appropriately.

Table 2: Monte Carlo Simulations for fixed effects model

T	Method	Category	$N = 30$			$N = 50$		
			β_1	β_2	$\log(\sigma)$	β_1	β_2	$\log(\sigma)$
5	MLE	Estimate	1.0054	1.0072	−0.1558	1.0060	0.9986	−0.1463
	Bayes		1.0295	1.0207	0.0188	1.0612	1.0678	0.0436
	MLE	SE	0.1164	0.2103	0.0716	0.0901	0.1632	0.0553
	Bayes		0.1248	0.2350	0.0742	0.1098	0.1996	0.0667
	MLE	MSE	0.0179	0.0575	0.0307	0.0106	0.0333	0.0253
	Bayes		0.0166	0.0542	0.0057	0.0157	0.0429	0.0062
	MLE	Coverage	0.9100	0.9150	0.4390	0.9015	0.9200	0.2735
	Bayes		0.9430	0.9530	0.9540	0.9170	0.9450	0.9115
10	MLE	Estimate	1.0005	1.0040	−0.0668	1.0007	1.0024	−0.0665
	Bayes		1.0087	1.0077	0.0087	1.0245	1.0310	0.0188
	MLE	SE	0.0813	0.1478	0.0498	0.0628	0.1143	0.0385
	Bayes		0.0790	0.1512	0.0468	0.0685	0.1251	0.0419
	MLE	MSE	0.0071	0.0240	0.0071	0.0046	0.0144	0.0061
	Bayes		0.0060	0.0215	0.0022	0.0054	0.0163	0.0021
	MLE	Coverage	0.9425	0.9385	0.7225	0.9280	0.9405	0.5925
	Bayes		0.9490	0.9540	0.9465	0.9330	0.9460	0.9305
15	MLE	Estimate	0.9987	1.0021	−0.0449	0.9993	1.0070	−0.0417
	Bayes		1.0141	1.0198	0.0115	1.0139	1.0239	0.0125
	MLE	SE	0.0655	0.1197	0.0402	0.0509	0.0928	0.0311
	Bayes		0.0695	0.1271	0.0425	0.0538	0.0982	0.0328
	MLE	MSE	0.0044	0.0145	0.0037	0.0028	0.0089	0.0028
	Bayes		0.0047	0.0155	0.0018	0.0031	0.0097	0.0012
	MLE	Coverage	0.9440	0.9485	0.7915	0.9410	0.9445	0.7195
	Bayes		0.9500	0.9565	0.9440	0.9425	0.9485	0.9295

Gibbs sampling was implemented using 51,000 iterations with the first 1,000 samples ignored, commonly called burn-in periods. After the burn-in periods, we used only every fifth draw to migrate the impact of autocorrelation. Thus, all the Bayes estimates were based on 10,000 samples. The statistics related to maximum likelihood estimation were obtained from **censReg**.

Table 1 shows the experiment results for the random effects model. In the estimation of regression coefficients, the maximum likelihood estimate is slightly less biased, but the bias of the Bayesian estimate seems not to be significant as well. However, the maximum likelihood estimate tends to have significantly smaller standard errors than the Bayesian estimate. This might lead to thoughtless conclusion that the maximum likelihood estimation has better efficiency. Note that the Bayesian estimate has smaller numerical mean squared errors than the maximum likelihood estimate in most cases. It should be noted that the maximum likelihood method underestimates the variance components, which has an implication for the standard error of estimates. To understand the underestimation of standard errors, we calculated the coverage probability of 95% confidence interval. As shown in Table 1, the coverage probability of maximum likelihood confidence interval is getting close to the nominal level as sample size is increasing, but it is significantly small when sample size is small. Thus, one must be careful in using the maximum likelihood method with small sample.

Similar results might be concluded as well for the fixed effects model. As shown in Table 2, and stated by Green (2004a, b), the maximum likelihood estimation of slope parameters are not affected greatly by the *incidental parameter problem*, but we could conclude that the asymptotic standard errors or asymptotic variances are generally underestimated, because the empirical coverage probabilities of the maximum likelihood estimation are significantly smaller than the nominal level. In contrast to the maximum likelihood estimation, the coverage probabilities of the Bayesian estimation are close to the nominal level. The Bayesian method can provide the proper estimate of standard error despite a small sample size.

4. Conclusions

The Bayesian method usually requires intensive computer works, but advances in contemporary computing power enable an intensive Bayesian method to solve various complicated statistical problems. We believe that the censored regression model is one of challenging statistical problem to which Bayesian should pay attention, because the frequentist method seems inefficient when sample size is small. For instance, the maximum likelihood estimation is quite reliable, but other statistical inference such as the hypothesis test or the interval estimation, the maximum likelihood method may be inadquate. In this paper, we have shown that a Bayesian model could be a better alternative when the sample size is small. Hence, it is believed that the Bayesian method can meet the various demands in the inference of a censored regression model. In addition, it would be desirable to compare the property of Bayesian models that employ various priors such as the reference prior and the matching probability prior for further study.

References

Amemiya, T. (1984). Tobit models: A survey, *Journal of Econometrics*, **24**, 3–61.

Bruno, G. (2004). Limited dependent panel models: A comparative analysis of classical and Bayesian inference among econometrics packages, *Computing in Economics and Finance, Society for Computational Economics*, `http://editorialexpress.com/cgi-bin/conference/download.cgi?db_name=SCE2004%&paper_id=41.`

Chib, S. (1992). Bayesian inference in the Tobit censored regression model, *Journal of Econometrics*, **51**, 77–99.

Cowles, M. K., Carlin, B. P. and Connett, J. E. (1996). Bayesian Tobit modeling of longitudinal ordinal clinical trial compliance data with nonignorable missingness, *Journal of American Statistical Association*, **91**, 86–98.

Gelman, A., Carlin, J. B., Stern, H. S., Dunson, D. B. and Vehtari, A. (2013). *Bayesian Data Analysis*, 3rd edition, Chapman & Hall/CRC, New York.

Green W. (2004a). The behavior of the maximum likelihood estimator of limited dependent variable models in the presence of fixed effects, *Econometrics Journal*, **7**, 98–119.

Green W. (2004b). Fixed effects and bias due to the incidental parameters problem in the Tobit model, *Econometrics Reviews*, **23**, 125-147.

Hamilton, B. H. (1999). HMO selection and medicare costs: Bayesian MCMC estimation of a robust panel data Tobit model with survival, *Health Economics and Econometrics*, **8**, 403–414.

Henningsen, A. (2011). *censReg: Censored Regression (Tobit) Models*, R package version 0.5, `http://CRAN.R-project.org/package=censReg.`

Hobert, J. P. and Casella, G. (1996). The effect of improper priors on Gibbs sampling in hierarchical linear mixed models, *Journal of the American Statistical Association*, **91**, 1461–1473.

Kass, R. E. and Natarajan, R. (2006). A default conjugate prior for variance components in generalized linear mixed models (Comment on article by Browne and Draper), *Bayesian Analysis*, **1**, 535–542.

Kleiber, C. and Zeileis, A. (2009). *AER: Applied Econometrics with R*, R package version 1.1, `http://CRAN.R-project.org/package=AER.`

Lancaster, T. (2000). The incidental parameter problem since 1948, *Journal of Econometrics*, **95**, 391–413.

Lee, L. (2013). *Nondetects And Data Analysis for Environmental Data*, R package version 1.5–6, `http://CRAN.R-project.org/package=NADA.`

Lee, S.-C. and Choi, B. (2013). Bayesian interval estimation of Tobit regression model, *The Korean Journal of Applied Statistics*, **26**, 737–746.

Maddala, G. S. (1983). *Limited-Dependent and Qualitative Variables in Econometrics*, Cambridge University Press, New York.

Martin, A. D., Quinn, K. M. and Park, J. H. (2013). *Markov chain Monte Carlo (MCMC) Package*, R package version 1.3-3, `http://CRAN.R-project.org/package=MCMCpack.`

McCulloch, C. E. (1996). Fixed and random effects and best prediction, In *Proceedings of the Kansas State Conference on Applied Statistics in Agriculture.*

Morawetz, U. (2006). Bayesian modelling of panel data with individual effects applied to simulated data, *Institut für Nachhaltige Wirtschaftsentwicklung*.

Tanner, M. A. and Wong, W.-H. (1987). The calculation of posterior distributions by data augmentation (with discussion), *Journal of the American Statistical Association*, **82**, 528–550.

Tobin, J. (1958). Estimation of relationships for limited dependent variables, *Econometrica,* **26**, 24–36.

Zeger, S. L. and Karim, M. R. (1991). Generalized linear models with random effects: A Gibbs sampling approach, *Journal of the American Statistical Association*, **86**, 79–86.

Received February 13, 2014; Revised March 8, 2014; Accepted March 10, 2014