Communications for Statistical Applications and Methods
2017, Vol. 24, No. 3, 291–301

https://doi.org/10.5351/CSAM.2017.24.3.291
Print ISSN 2287-7843 / Online ISSN 2383-4757

The restricted maximum likelihood estimation of a censored regression model

Seung-Chun Lee[1,a]

[a]Department of Applied Statistics, Hanshin University, Korea

Abstract

It is well known in a small sample that the maximum likelihood (ML) approach for variance components in the general linear model yields estimates that are biased downward. The ML estimate of residual variance tends to be downwardly biased. The underestimation of residual variance, which has implications for the estimation of marginal effects and asymptotic standard error of estimates, seems to be more serious in some limited dependent variable models, as shown by some researchers. An alternative frequentist's approach may be restricted or residual maximum likelihood (REML), which accounts for the loss in degrees of freedom and gives an unbiased estimate of residual variance. In this situation, the REML estimator is derived in a censored regression model. A small sample the REML is shown to provide proper inference on regression coefficients.

Keywords: censored regression, REML, limited dependent variable, observed Fisher information

1. Introduction

Consider a linear regression model:

$$w_i = \boldsymbol{x}_i'\boldsymbol{\beta} + \epsilon_i, \qquad i = 1, 2, \ldots, n, \tag{1.1}$$

where w_i is the dependent variable, $\boldsymbol{x}_i$ is the $p \times 1$ vector of predictors, $\boldsymbol{\beta}$ is the $p \times 1$ vector of regression coefficients. The standard assumption of the error term, viz. ϵ_i's are independent and identically distributed normal random variables, implies that the dependent variable can be any real number; however, in many statistical analysis the dependent variable can only have limited ranges. For example, the variable of interest is constrained to lie between zero and one (as in the case of a probability) or is constrained to be positive (as in the case of wages or hours worked). In an econometrics context, a dependent variable whose range of possible values is "restricted in some important way" is defined to be a limited dependent variable (LDV). The probit or logistic regression, the Poisson regression, the truncated regression and the censored regression are examples of and LDV model.

Currently, the maximum likelihood (ML) is a dominating method for LVD models; however, the ML gives biased estimates for variance components in a general mixed linear model. In particular, the ML estimate of disturbance variance is known to be biased downward. Therefore, the inferences on the regression coefficients will have inflated type I error rates because their precision is overstated. The underestimation is because the fixed effects are assumed to be known without error in the ML approach. The downward bias would be more severe in the LDV models because the dependent

[1] Department of Statistics, Hanshin University, 137, Hanshindae-gil, Osan-si, Gyeonggi-Do 18101, Korea. E-mail: seung@hs.ac.kr

Published 31 May 2017 / journal homepage: http://csam.or.kr

variable is observable only in a limited range as shown by Green (2004), and Lee and Choi (2013, 2014). This leads to a discussion of the estimation method for the LDV models in small samples.

The restricted maximum likelihood (REML) could be an alternative frequentist's method, which also has a Bayesian justification. It might be expected that the REML can eliminate the bias of ML since the REML divides the mean squared deviation by the degrees of freedom instead of by the sample size. Consequently, we investigate the behavior of the REML estimator.

Hughes (1999) provided the REML estimate in a general mixed effects linear model with censored data using a Monte Carlo EM algorithm and claimed that the approach can be used with an arbitrarily complex design matrix. A Monte Carlo EM algorithm employing the Gibbs sampler is adequate to provide an estimate in the general mixed effects model. Many numerical methods have been employed to get the REML estimate in the LDV models, see Noh and Lee (2007) for further details. However, the numerical method lacks the capacity to provide standard error estimates. For instance, Hughes (1999) gave an asymptotic approximation for the variances of fixed effects only, but the asymptotic variance is obtained by not including the estimation of variance components. Thus, the approximation may be too rough to evaluate the performance of estimation methods. A sharp asymptotic approximation requires the computation of the second order derivative of log-likelihood functions. This paper shows that second order derivatives are tractable in a censored regression model.

Censored regression models commonly arise in econometrics. For instance, suppose a labor supply model estimates the relationship between hours worked and employee characteristics. The estimates undertaken using linear regression will be biased since the hours of work would be zeros for the people who are unemployed. The Tobit model, proposed by Tobin (1958), describes the relationship between a non-negative dependent variable and independent variables. It is a special case of a censored regression that is essentially equivalent to general censored regression called Tobit type I, see Amemiya (1985). Section 2 provides the REML for the Tobit regression model.

2. The restricted maximum likelihood (REML) estimation of Tobit model

The standard Tobit model assumes that observed dependent variable y_i is given by:

$$y_i = \begin{cases} w_i, & \text{if } w_i > 0, \\ 0, & \text{if } w_i \leq 0, \end{cases} \tag{2.1}$$

where w_i is a latent variable determine by equation (1.1). A common variation of the Tobit model is censoring from below at a value y_L different from zero, or is censoring above at a value y_U, or is simultaneously censoring from above and below. The variations classified into the Tobit type I, and the REML method of the Tobit type I models are essentially equal to the original Tobit model.

For model (2.1), Tobin (1958) provided the ML estimate. Later, Amemiya (1984) proved that the maximum likelihood estimation (MLE) suggested by Tobin is consistent. The ML method is implemented in various R packages that include applied econometrics with R (Kleiber and Zeileis, 2009), censored regression model (censReg) (Henningsen, 2017), and nondetects and data analysis (Lee, 2017).

The REML introduced by Patterson and Thomson (1971) transforms the dependent variable, $z_1 = \boldsymbol{L}_1'\boldsymbol{w}$ and $z_2 = \boldsymbol{L}_2'\boldsymbol{w}$ to partition the likelihood into two independent parts, where $\boldsymbol{L}_1 = \boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1}$ and $\boldsymbol{L}_2$ be such that $\boldsymbol{L}_2'\boldsymbol{X} = \boldsymbol{0}, \boldsymbol{L}_2'\boldsymbol{L}_2 = \boldsymbol{I}$ and $\boldsymbol{L}_2\boldsymbol{L}_2' = \boldsymbol{I} - \boldsymbol{H}$, where $\boldsymbol{H} = \boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'$ and $\boldsymbol{X} = (\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n)'$. The first transformed variable relates to the fixed effects and the second one relates to the residual contrast. Then, under the normality assumption of error terms, we have

$$\begin{pmatrix} z_1 \\ z_2 \end{pmatrix} \sim \mathcal{MVN}\left(\begin{pmatrix} \boldsymbol{\beta} \\ \mathbf{0} \end{pmatrix}, \sigma^2 \begin{pmatrix} (\boldsymbol{X}'\boldsymbol{X})^{-1} & \mathbf{0} \\ \mathbf{0} & \boldsymbol{I} \end{pmatrix}\right).$$

Thus, $L(\boldsymbol{\beta}, \sigma^2; z) = L_1(\boldsymbol{\beta}, \sigma^2; z_1) \times L_2(\sigma^2; z_2)$, where

$$L_1\left(\boldsymbol{\beta}, \sigma^2; z_1\right) \propto \left(\sigma^2\right)^{p/2} \exp\left[-\frac{1}{2\sigma^2}(z_1 - \boldsymbol{\beta})' \boldsymbol{X}'\boldsymbol{X}(z_1 - \boldsymbol{\beta})\right]$$

and

$$L_2\left(\sigma^2; z_2\right) \propto \left(\sigma^2\right)^{(n-p)/2} \exp\left[-\frac{1}{2\sigma^2} z_2' z_2\right].$$

Let $\mathcal{R} = \{\boldsymbol{w} : \boldsymbol{y}(\boldsymbol{w}) = \boldsymbol{y}\}$ be the set of latent variables given the observed data $\boldsymbol{y}$, then the marginal log-likelihood function of the observed data $\boldsymbol{y}$ is

$$\log L\left(\boldsymbol{\beta}, \sigma^2; \boldsymbol{y}\right) = \log \int_{\mathcal{R}} L_1\left(\boldsymbol{\beta}, \sigma^2; z_1\right) d\mu + \log \int_{\mathcal{R}} L_2\left(\sigma^2; z_2\right) d\mu, \tag{2.2}$$

where μ is the Lebesgue measure. The REML of $\boldsymbol{\theta} = (\boldsymbol{\beta}, \sigma^2)$ is obtained by maximizing the two terms in equation (2.2) separately, and are the solutions of equations:

$$\begin{aligned}
S_1(\boldsymbol{\theta}) = \frac{\partial \ell_1}{\partial \boldsymbol{\beta}} &= \frac{\partial}{\partial \boldsymbol{\beta}} \log \int_{\mathcal{R}} L_1\left(\boldsymbol{\beta}, \sigma^2; z_1\right) d\mu \\
&= \int_{\mathcal{R}} -\frac{\boldsymbol{X}'\boldsymbol{X}}{\sigma^2}(\boldsymbol{\beta} - z_1) f(z_1)\, d\mu \Big/ \int_{\mathcal{R}} f(z_1)\, d\mu \\
&= -\frac{\boldsymbol{X}'\boldsymbol{X}}{\sigma^2}(\boldsymbol{\beta} - E(z_1 | \boldsymbol{y})) = \mathbf{0}, \qquad (2.3)\\
S_2(\boldsymbol{\theta}) = \frac{\partial \ell_2}{\partial \sigma^2} &= \frac{\partial}{\partial \sigma^2} \log \int_{\mathcal{R}} L_2\left(\sigma^2; z_2\right) d\mu \\
&= -\frac{n-p}{2\sigma^2} + \int_{\mathcal{R}} \frac{z_2' z_2}{2\sigma^4} f(z_2)\, d\mu \Big/ \int_{\mathcal{R}} f(z_2)\, d\mu \\
&= -\frac{n-p}{2\sigma^2} + \frac{1}{2\sigma^4} E(z_2' z_2 | \boldsymbol{y}) = 0. \qquad (2.4)
\end{aligned}$$

If $y_i = 0$, then $w_i \sim TN_{(-\infty,0]}(\boldsymbol{x}_i'\boldsymbol{\beta}, \sigma^2)$, a truncated normal distribution on $(-\infty, 0]$ with mean $\boldsymbol{x}_i'\boldsymbol{\beta}$ and variance σ^2, while w_i is degenerating at y_i, if $y_i > 0$. Thus, the moments of a truncated normal random variable are useful in what follow: Let ϕ and Φ are the pdf and cdf of a standard normal distribution, and suppose $Y \sim TN_{(-\infty,0]}(\mu, \sigma^2)$. Then, using

$$M_Y(t) = \exp\left(\mu t + \frac{1}{2}\sigma^2 t^2\right)\left[\Phi\left(-\frac{\mu}{\sigma} - \sigma t\right) \Big/ \Phi\left(-\frac{\mu}{\sigma}\right)\right],$$

we have

$$E(Y) = \mu - \sigma\lambda\left(\frac{\mu}{\sigma}\right), \qquad E\left(Y^2\right) = \mu^2 + \sigma^2 - \mu\sigma\lambda\left(\frac{\mu}{\sigma}\right),$$
$$E\left(Y^3\right) = 2\sigma^2 E(Y) + \mu E\left(Y^2\right), \qquad E\left(Y^4\right) = 3\sigma^2 E\left(Y^2\right) + \mu E\left(Y^3\right),$$

where $\lambda(x) = \phi(x)/\Phi(-x)$, the inverse of Mill's ratio. In particular, if $y_i = 0$,

$$E_{\boldsymbol{\theta}}\left(w_i|\, \boldsymbol{y}\right) = \boldsymbol{x}_i'\boldsymbol{\beta} - \sigma\lambda\left(\frac{\boldsymbol{x}_i'\boldsymbol{\beta}}{\sigma}\right),$$
$$E_{\boldsymbol{\theta}}\left(w_i^2\middle|\,\boldsymbol{y}\right) = \boldsymbol{\beta}'\,\boldsymbol{x}_i\,\boldsymbol{x}_i'\,\boldsymbol{\beta} + \sigma^2 - \sigma\,\boldsymbol{x}_i'\,\boldsymbol{\beta}\,\lambda\left(\frac{\boldsymbol{x}_i'\boldsymbol{\beta}}{\sigma}\right),$$
$$\mathrm{Var}_{\boldsymbol{\theta}}\left(w_i|\,\boldsymbol{y}\right) = \sigma^2 + \sigma\,\boldsymbol{x}_i'\boldsymbol{\beta}\,\lambda\left(\frac{\boldsymbol{x}_i'\boldsymbol{\beta}}{\sigma}\right) - \sigma^2\lambda^2\left(\frac{\boldsymbol{x}_i'\boldsymbol{\beta}}{\sigma}\right).$$

Also,

$$\left(\frac{\partial}{\partial\boldsymbol{\beta}}\right)E\left(w_i|\,\boldsymbol{y}\right) = \left(\frac{1}{\sigma^2}\right)\mathrm{Var}\left(w_i|\boldsymbol{y}\right)\,\boldsymbol{x}_i,$$
$$\left(\frac{\partial}{\partial\boldsymbol{\beta}}\right)E\left(w_i^2\middle|\,\boldsymbol{y}\right) = \left[E\left(w_i|\,\boldsymbol{y}\right) + \left(\frac{1}{\sigma^2}\right)\boldsymbol{x}_i'\,\boldsymbol{\beta}\,\mathrm{Var}\left(w_i|\,\boldsymbol{y}\right)\right]\boldsymbol{x}_i,$$
$$\left(\frac{\partial}{\partial\sigma^2}\right)E\left(w_i|\,\boldsymbol{y}\right) = -\left(\frac{1}{2\sigma^3}\right)\lambda\left(\frac{\boldsymbol{x}_i'\boldsymbol{\beta}}{\sigma}\right)E\left(w_i^2\middle|\,\boldsymbol{y}\right),$$
$$\left(\frac{\partial}{\partial\sigma^2}\right)E\left(w_i^2\middle|\,\boldsymbol{y}\right) = 1 - \left(\frac{1}{2\sigma^3}\right)\boldsymbol{x}_i'\boldsymbol{\beta}\lambda\left(\frac{\boldsymbol{x}_i'\boldsymbol{\beta}}{\sigma}\right)E\left(w_i^2\middle|\,\boldsymbol{y}\right),$$
$$\left(\frac{\partial}{\partial\gamma}\right)\mathrm{Var}\left(w_i|\,\boldsymbol{y}\right) = \left(\frac{\partial}{\partial\gamma}\right)E\left(w_i^2\middle|\,\boldsymbol{y}\right) - 2E\left(w_i|\,\boldsymbol{y}\right)\left(\frac{\partial}{\partial\gamma}\right)E\left(w_i|\,\boldsymbol{y}\right),$$

where $\gamma = \boldsymbol{\beta}$ or $\gamma = \sigma^2$.

For notational convenience, we assume that the last m observations of dependent variable are zero. Then, the vector of observed dependent variable is $\boldsymbol{y}' = (\boldsymbol{y}_1', \boldsymbol{0}')$. Write $\boldsymbol{w}' = (\boldsymbol{w}_1', \boldsymbol{w}_2')$ and $\boldsymbol{X}' = (\boldsymbol{X}_1', \mathbf{X_2}')$, where $\boldsymbol{w}_2 = (w_{21}, \ldots, w_{2m})'$ and $\boldsymbol{X}_2$ are the vector of latent variables and the design matrix, respectively, corresponding to zero observations. Let $\boldsymbol{H}_{ij} = \boldsymbol{X}_i(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}_j$. With these notations, the conditional expectations in equations (2.3) and (2.4) can be written as

$$E_{\boldsymbol{\theta}}\left(z_1|\boldsymbol{y}\right) = \left(\boldsymbol{X}'\boldsymbol{X}\right)^{-1}\left(\boldsymbol{X}_1'\boldsymbol{y}_1 + \boldsymbol{X}_2'E_{\boldsymbol{\theta}}\left(\boldsymbol{w}_2|\boldsymbol{y}\right)\right)$$

and

$$\begin{aligned} E_{\boldsymbol{\theta}}\left(z_2'z_2|\boldsymbol{y}\right) =& \boldsymbol{y}_1'\left(\boldsymbol{I} - \boldsymbol{H}_{11}\right)\boldsymbol{y}_1 - 2\boldsymbol{y}_1'\boldsymbol{H}_{12}E_{\boldsymbol{\theta}}\left(\boldsymbol{w}_2|\boldsymbol{y}\right) \\ &+ \sum_{i=1}^{m}\left(1 - h_{ii}^{22}\right)\mathrm{Var}_{\boldsymbol{\theta}}\left(w_{2i}|\boldsymbol{y}\right) + E_{\boldsymbol{\theta}}\left(\boldsymbol{w}_2'|\boldsymbol{y}\right)\left(\boldsymbol{I} - \boldsymbol{H}_{22}\right)E_{\boldsymbol{\theta}}\left(\boldsymbol{w}_2|\boldsymbol{y}\right), \end{aligned}$$

where h_{ij}^{22} is (i, j) element of $\boldsymbol{H}_{22}$.

Let $\boldsymbol{S}'(\boldsymbol{\theta}) = (S_1'(\boldsymbol{\theta}), S_2(\boldsymbol{\theta}))$. The REML estimate of $\boldsymbol{\theta}$ can be found by applying the Newton-Raphson method,

$$\hat{\boldsymbol{\theta}}_{(i+1)} = \hat{\boldsymbol{\theta}}_{(i)} - \boldsymbol{J}^{-1}(\hat{\boldsymbol{\theta}}_{(i)})\boldsymbol{S}(\hat{\boldsymbol{\theta}}_{(i)})$$

with an arbitrary initial value $\hat{\boldsymbol{\theta}}_{(0)}$, where $\boldsymbol{J}(\boldsymbol{\theta})$ is the Jacobian of $\boldsymbol{S}(\boldsymbol{\theta})$,

$$\boldsymbol{J}(\boldsymbol{\theta}) = \frac{\partial}{\partial \boldsymbol{\theta}}\boldsymbol{S}(\boldsymbol{\theta}) = \begin{bmatrix} \frac{\partial}{\partial \boldsymbol{\beta}'} S_1(\boldsymbol{\theta}) & \frac{\partial}{\partial \sigma^2} S_1(\boldsymbol{\theta}) \\ \frac{\partial}{\partial \boldsymbol{\beta}'} S_2(\boldsymbol{\theta}) & \frac{\partial}{\partial \sigma^2} S_2(\boldsymbol{\theta}) \end{bmatrix}.$$

The elements of Jacobian matrix are:

$$\frac{\partial S_1(\boldsymbol{\theta})}{\partial \boldsymbol{\beta}'} = -\frac{1}{\sigma^2}\boldsymbol{X}'\boldsymbol{X} + \frac{1}{\sigma^4}\boldsymbol{X}_2'\boldsymbol{V}_2\boldsymbol{X}_2, \tag{2.5}$$

$$\frac{\partial S_1(\boldsymbol{\theta})}{\partial \sigma^2} = \frac{\boldsymbol{X}'\boldsymbol{X}}{\sigma^4}\left[\boldsymbol{\beta} - E(\boldsymbol{z}_1|\,\boldsymbol{y})\right] + \frac{1}{\sigma^2}\boldsymbol{X}_2'\frac{\partial}{\partial \sigma^2}E(\boldsymbol{w}_2|\,\boldsymbol{y}),$$

$$\frac{\partial S_2(\boldsymbol{\theta})}{\partial \boldsymbol{\beta}'} = \frac{1}{\sigma^6}\left[\left\{E\left(\boldsymbol{w}_2'|\,\boldsymbol{y}\right)(\boldsymbol{I} - \boldsymbol{H}_{22}) - \boldsymbol{y}_1'\boldsymbol{H}_{12}\right\}\boldsymbol{V}_2\boldsymbol{X}_2 + \frac{1}{2}\sum_{i=1}^{m}\left(1 - h_{ii}^{22}\right)\frac{\partial}{\partial \boldsymbol{\beta}'}\mathrm{Var}\left(w_{2i}|\,\boldsymbol{y}\right)\right],$$

$$\begin{aligned}\frac{\partial S_2(\boldsymbol{\theta})}{\partial \sigma^2} = \frac{n-p}{2\sigma^4} - \frac{1}{\sigma^6}E(\boldsymbol{z}_2'\boldsymbol{z}_2|\boldsymbol{y}) + \frac{1}{\sigma^4}&\left[\frac{1}{2}\sum_{i=1}^{m}\left(1 - h_{ii}^{22}\right)\frac{\partial}{\partial \sigma^2}\mathrm{Var}(w_{2i}|\boldsymbol{y})\right.\\ &\left. + \left\{E(\boldsymbol{w}_2'|\boldsymbol{y})(\boldsymbol{I} - \boldsymbol{H}_{22}) - \boldsymbol{y}_1'\boldsymbol{H}_{12}\right\}\frac{\partial}{\partial \sigma^2}E(\boldsymbol{w}_2|\boldsymbol{y})\right].\end{aligned}$$

The computation of REML estimate is relatively easy. In fact, we can easily get the REML estimate, $\hat{\boldsymbol{\theta}}$ by applying the EM algorithm. The real problem in REML is the computation of the standard error of estimate, because we cannot use standard ML theory. For instance, second-order derivatives of $\ell_2(\boldsymbol{\theta})$ are useless. However, if the variance of $\boldsymbol{S}(\boldsymbol{\theta})$ is available, the asymptotic variance of $\hat{\boldsymbol{\theta}}$ can be obtained by using Taylor series expansion,

$$\boldsymbol{S}(\boldsymbol{\theta}) \approx \boldsymbol{S}\left(\hat{\boldsymbol{\theta}}\right) + \boldsymbol{J}\left(\hat{\boldsymbol{\theta}}\right)\left(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}\right).$$

Since, $\boldsymbol{S}(\hat{\boldsymbol{\theta}}) = \boldsymbol{0}$, an asymptotic variance of $\boldsymbol{\theta}$ is

$$\mathrm{Var}\left(\hat{\boldsymbol{\theta}}\right) \approx \boldsymbol{J}^{-1}\left(\hat{\boldsymbol{\theta}}\right)\left[\mathrm{Var}\left(\boldsymbol{S}\left(\boldsymbol{\theta}\right)\right)\right]\left[\boldsymbol{J}^{-1}\left(\hat{\boldsymbol{\theta}}\right)\right]'. \tag{2.6}$$

To compute equation (2.6), we first note that $\partial S_1(\boldsymbol{\theta})/\partial \boldsymbol{\beta}' = \partial^2 \ell_1/\partial \boldsymbol{\beta}\partial \boldsymbol{\beta}' = \partial^2 \log L/\partial \boldsymbol{\beta}\partial \boldsymbol{\beta}'$. That is, the score functions of ML and REML are the same. Thus, using the ML theory, we have

$$\mathrm{Var}\left(S_1\left(\boldsymbol{\theta}\right)\right) = -E\left[\frac{\partial^2 \ell_1}{\partial \boldsymbol{\beta}\partial \boldsymbol{\beta}'}\right] = \frac{1}{\sigma^2}\boldsymbol{X}'\boldsymbol{X} - \frac{1}{\sigma^4}\boldsymbol{X}_2'\boldsymbol{V}_2\boldsymbol{X}_2.$$

However, $\partial S_2(\boldsymbol{\theta})/\partial \sigma^2$ cannot be related to the variance of $S_2(\boldsymbol{\theta})$. To compute $\mathrm{Var}(S_2(\boldsymbol{\theta}))$, we can use the relationship, $\mathrm{Var}(E(X|Y)) = \mathrm{Var}(X) - E(\mathrm{Var}(X|Y))$. Consequently,

$$\mathrm{Var}\left(S_2\left(\boldsymbol{\theta}\right)\right) = \frac{1}{4\sigma^8}\left[\mathrm{Var}\left(\boldsymbol{w}'(\boldsymbol{I} - \boldsymbol{H})\boldsymbol{w}\right) - E\left(\mathrm{Var}(\boldsymbol{w}'(\boldsymbol{I} - \boldsymbol{H})\boldsymbol{w}|\,\boldsymbol{y})\right)\right]. \tag{2.7}$$

Since, $\boldsymbol{w}'(\boldsymbol{I}-\boldsymbol{H})\boldsymbol{w}/\sigma^2 \sim \chi^2_{n-p}$, the first term of right-hand side of equation (2.7) is equal to $(n-p)/2\sigma^4$, and in the second terms,

$$\begin{aligned}\mathrm{Var}\left(\boldsymbol{w}\left(\boldsymbol{I}-\boldsymbol{H}\right)\boldsymbol{w}|\,\boldsymbol{y}\right) &= \mathrm{Var}\left(\boldsymbol{w}_2'\left(\boldsymbol{I}-\boldsymbol{H}_{22}\right)\boldsymbol{w}_2 - 2\boldsymbol{y}_1'\boldsymbol{H}_{12}\boldsymbol{w}_2|\,\boldsymbol{y}\right)\\ &= \mathrm{Var}\left(\boldsymbol{w}_2'\left(\boldsymbol{I}-\boldsymbol{H}_{22}\right)\boldsymbol{w}_2|\,\boldsymbol{y}\right) + 4\boldsymbol{y}_1'\boldsymbol{H}_{12}\boldsymbol{V}_2\boldsymbol{H}_{12}'\boldsymbol{y}_1 - 4\boldsymbol{y}_1'\boldsymbol{H}_{12}\mathrm{Cov}(\boldsymbol{w}_2, \boldsymbol{w}_2'(\boldsymbol{I}-\boldsymbol{H}_{22})\,\boldsymbol{w}_2|\,\boldsymbol{y}).\end{aligned}$$

It can be shown that

$$\begin{aligned}\mathrm{Var}\left(\boldsymbol{w}_2\left(\boldsymbol{I}-\boldsymbol{H}_{22}\right)\boldsymbol{w}_2|\,\boldsymbol{y}\right) = &-4\sum\sum_{i\neq j}\left(1-h_{ii}^{22}\right)h_{ij}^{22}\left[E\left(w_{2i}^3\middle|\,\boldsymbol{y}\right) - E\left(w_{2i}^2\middle|\,\boldsymbol{y}\right)E\left(w_{2i}|\boldsymbol{y}\right)\right]E\left(w_{2j}|\boldsymbol{y}\right)\\ &+ \sum_{i=1}^{m}\left(1-h_{ii}^{22}\right)^2\left[E\left(w_{2i}^4\middle|\,\boldsymbol{y}\right) - \left(E\left(w_{2i}^2\middle|\,\boldsymbol{y}\right)\right)^2\right]\\ &+ 4\sum\sum\sum_{i\neq j\neq k} h_{ij}^{22}h_{ik}^{22}\mathrm{Var}\left(w_{2i}|\,\boldsymbol{y}\right)E\left(w_{2j}|\boldsymbol{y}\right)E\left(w_{2k}|\,\boldsymbol{y}\right)\\ &+ 2\sum\sum_{i\neq j}\left(h_{ij}^{22}\right)^2\left[E\left(w_{2i}^2\middle|\,\boldsymbol{y}\right)E\left(w_{2j}^2\middle|\,\boldsymbol{y}\right) - \left(E\left(w_{2i}|\,\boldsymbol{y}\right)\right)^2\left(E\left(w_{2j}|\,\boldsymbol{y}\right)\right)^2\right]\end{aligned}$$

and

$$\begin{aligned}\mathrm{Cov}\left(\boldsymbol{w}_2, \boldsymbol{w}_2'\left(\boldsymbol{I}-\boldsymbol{H}_{22}\right)\boldsymbol{w}_2|\,\boldsymbol{y}\right) = \Bigg\{&\left(1-h_{ii}^{22}\right)\left[E\left(w_{2i}^3\middle|\,\boldsymbol{y}\right) - E\left(w_{2i}|\,\boldsymbol{y}\right)E\left(w_{2i}^2\middle|\,\boldsymbol{y}\right)\right]\\ &- 2\sum_{i\neq j}h_{ij}^{22}\mathrm{Var}\left(w_{2i}|\,\boldsymbol{y}\right)E\left(w_{2j}|\,\boldsymbol{y}\right)\Bigg\}_{i=1}^{m}.\end{aligned}$$

$\mathrm{Var}(S_2(\boldsymbol{\theta}))$ in equation (2.7) plays a role of the expected information; however, Efron and Hinkley (1978) showed that (in most cases) observed information is a more appropriate measure of information than the expected information. Thus, we substitute the non-expected version of equation (2.7) for $\mathrm{Var}(S_2(\boldsymbol{\theta}))$, i.e., we replace

$$\frac{n-p}{2\sigma^4} + \frac{1}{\sigma^8}\left[\boldsymbol{y}_1'\boldsymbol{H}_{12}\left\{\boldsymbol{V}_2\boldsymbol{H}_{12}'\boldsymbol{y}_1 - \mathrm{Cov}\left(\boldsymbol{w}_2, \boldsymbol{w}_2'\left(\boldsymbol{I}-\boldsymbol{H}_{22}\right)\boldsymbol{w}_2|\,\boldsymbol{y}\right)\right\} + \frac{1}{4}\mathrm{Var}\left(\boldsymbol{w}_2\left(\boldsymbol{I}-\boldsymbol{H}_{22}\right)\boldsymbol{w}_2|\,\boldsymbol{y}\right)\right]$$

for $\mathrm{Var}(S_2(\boldsymbol{\theta}))$.

Similarly, the covariance of $S_1(\boldsymbol{\theta})$ and $S_2(\boldsymbol{\theta})$ is

$$\begin{aligned}\mathrm{Cov}\left(S_1\left(\boldsymbol{\theta}\right), S_2\left(\boldsymbol{\theta}\right)\right) &= \frac{1}{2\sigma^6}\left[\mathrm{Cov}\left(\boldsymbol{z}_1, \boldsymbol{z}_2'\boldsymbol{z}_2\right) - \boldsymbol{X}'\boldsymbol{X}\,E\left(\mathrm{Cov}\left(\boldsymbol{z}_1, \boldsymbol{z}_2'\boldsymbol{z}_2|\,\boldsymbol{y}\right)\right)\right]\\ &= \frac{1}{\sigma^6}E\left[\boldsymbol{X}_2'\left\{\frac{1}{2}\mathrm{Cov}\left(\boldsymbol{w}_2, \boldsymbol{w}_2'\left(\boldsymbol{I}-\boldsymbol{H}_{22}\right)\boldsymbol{w}_2|\,\boldsymbol{y}\right) - \boldsymbol{V}_2\boldsymbol{H}_{12}'\boldsymbol{y}_1\right\}\right], \qquad (2.8)\end{aligned}$$

but we use the non-expected value of equation (2.8) for $\mathrm{Cov}(S_1(\boldsymbol{\theta}), S_2(\boldsymbol{\theta}))$.

Remark 1. When compute the estimate, the EM algorithm is preferable to the Newton-Raphson method, because it does not require second-order derivatives and a matrix inversion. The number of iterations seems to be almost the same in our experiment. Thus, we compute the second-order derivatives to compute the standard error of estimate after obtaining the estimate by EM algorithm. The $(i+1)^{th}$ iteration of EM algorithm can be done by

$$\begin{aligned}\hat{\boldsymbol{\beta}}_{(i+1)} &= E_{(\hat{\boldsymbol{\theta}}_{(i)}, \hat{\sigma}^2_{(i)})}(\boldsymbol{z}_1|\boldsymbol{y}),\\ \hat{\sigma}^2_{(i+1)} &= \frac{1}{n-p}E_{(\hat{\boldsymbol{\beta}}_{(i+1)}, \hat{\sigma}^2_{(i)})}(\boldsymbol{z}_2'\boldsymbol{z}_2|\,\boldsymbol{y}).\end{aligned}$$

Table 1: Estimates of Tobit model for Tobin data

Parameter	Maximum likelihood				Restricted maximum likelihood			
	Estimate	Std err	t value	Pr (> t)	Estimate	Std err	t value	Pr (> t)
(Intercept)	15.1449	16.0795	0.942	0.346	15.7773	20.4878	0.770	0.441
age	−0.1291	0.2186	−0.590	0.555	−0.1639	0.2842	−0.577	0.564
quant	−0.0455	0.0583	−0.782	0.434	−0.0458	0.0741	−0.617	0.537
logSigma	1.7179	0.3103	5.536	3.1e−8	1.9527	0.4061	4.808	1.5e−6
Sigma	5.5725	1.7293			7.0478	2.8624		

3. Examples

3.1. Tobin data

Tobin data, introduced by Tobin (1958) to illustrate the Tobit model, consists of 20 observations on 3 variables, age, quant (liquidity ratio) and durable (durable goods purchase), and 13 observations of dependent variable, durable are censored from below at 0. This data may be adequate to demonstrate the small sample properties of the ML and REML.

The ordinary least squares method provides biased estimates of regression coefficients due to the large fraction of censored dependent variable; however, the ML method provides a downwardly biased estimate of residual variance. The estimates of ML and REML are shown in Table 1. We use censReg (Henningsen, 2017) to compute MLEs.

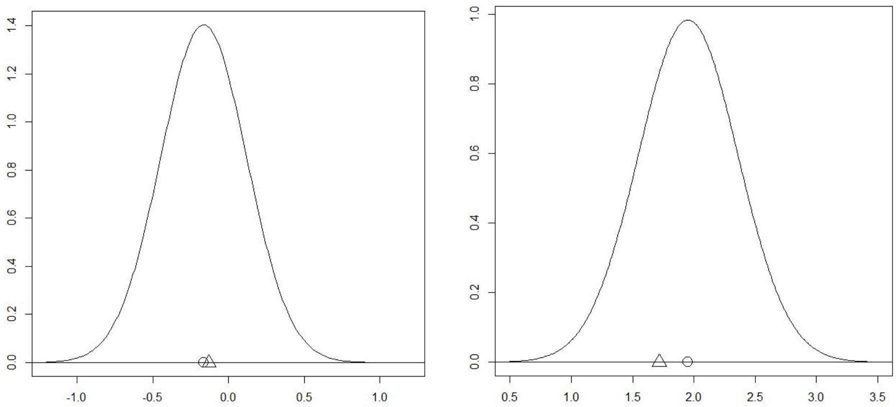
Regression coefficients are estimated to nearly the same values for all variables. The largest absolute difference in the estimation of the regression coefficient, occurs at the coefficient of "age" is 0.0348, which may be negligible. To see this more clearly, the two estimates are plotted in $N(-0.1639, 0.2842^2)$ by assuming the REML provides correct values so that the asymptotic distribution of REML estimator would be a normal distribution with mean −0.1639 and standard deviation 0.2842. Figure 1(a) indicates that it may be concluded that two estimates are quite close. Unlike the regression coefficients, the ML and REML estimates of residual variance are different. The ML gives a smaller estimate of the residual variance than REML. Figure 1(b) shows the distance between estimates in $N(1.9527, 0.4061^2)$, to demonstrate that the difference seems to be significant.

The MLE of residual variance is often biased downward when the sample size is small. The underestimation of residual variance would lead to the underestimated standard error of regression coefficients. REML gives larger standard errors for all slope parameters as well variance estimate than the ML; therefore, REML is believed to somewhat remedy the underestimation of MLE in Tobin data.

3.2. Mroz data

The second data set comes from the 1976 Panel Study of Income Dynamics (Mroz , 1987). It consists 753 observations on 19 variables, but we take 8 variables, hour (working hours of married white women, dependent variable), age, wage, educ (educational attainment in years), kidslt6 (number of children less than 6 years old in household), kidge6 (number of children between ages 6 and 18 in household), nwifeinc (net wife income), exper (actual years of wife's previous labor market experience), and exper2 (exper^2) to fit the model given in Wooldridge (2009).

Of the 753 observations, the first 428 are for women with positive hours worked in 1975, while the remaining 325 observations are for women who did not work for pay in 1975. This data was cited in various literatures as an example of probit and logistics regression models by making dependent variable dichotomous. It is believed that the Mroz data is suitable to show a large sample property of REML.



(a) Age (b) $\log \sigma$

Figure 1: *Distance between maximum likelihood ("△") and restricted maximum likelihood ("◯") estimates.*

Table 2: Estimates of Tobit model for Mroz data

Parameter	Maximum likelihood				Restricted maximum likelihood			
	Estimate	Std err	*t* value	Pr (> *t*)	Estimate	Std err	*t* value	Pr (> *t*)
(Intercept)	965.3053	446.4360	2.162	0.0306	960.7502	450.4161	2.133	0.0329
nwifeinc	−8.8142	4.4591	−1.977	0.0481	−8.8597	4.4988	−1.969	0.0489
educ	80.6456	21.5832	3.736	0.0002	81.1313	21.7766	3.726	0.0002
exper	131.5643	17.2794	7.614	2.e−14	132.1379	17.4375	7.578	3.e−14
exper2	−1.8642	0.5377	−3.467	0.0005	−1.8738	0.5425	−3.454	0.0005
age	−54.4050	7.4185	−7.334	2.e−13	−54.6125	7.4854	−7.296	2e−13
kidslt6	−894.0217	111.8780	−7.991	1.e−15	−897.7255	112.8874	−7.952	1.e−15
kidsge6	−16.2180	38.6414	−0.420	0.6747	−15.9904	38.9846	−0.410	0.6817
logSigma	7.0229	0.0371	189.514	0.0000	7.0315	0.0374	187.916	0.0000

Table 2 shows the ML and REML estimates. As the case of Tobin data, the estimates of regression coefficients are nearly equal. The estimation of residual variance indicates that the estimate of REML is slightly larger than ML. REML often has slightly larger standard errors due to the slightly larger estimate of residual variance; however, the differences are statistically insignificant. Both methods report almost identical t and p-values due to the similar values; consequently, the inference based on the ML and REML estimates would get a similar conclusion on the regression model. This result is predictable since the downward bias of MLE is a small sample property. Based upon these observations, we may say that two methods are statistically identical in Mroz data.

4. Simulation study and conclusion

REML is distinguished from ML in the estimations of the residual variance and the standard error of regression coefficient when the sample size is small. REML is believed to be a proper method when the sample size is small. To see this, a simulation study has been done on the study design used in Bilias *et al.* (2000) and Yu and Stander (2007). That is, for a Tobit regression model, the latent variable was generated according to

$$w_i = \max\{\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \epsilon, 0\}, \qquad \epsilon \sim N(0, \sigma^2).$$

Table 3: Simulation result with $(\beta_0, \beta_1, \beta_2, \log\sigma) = (1, 1, 1, 0)$

n	θ	Maximum likelihood					Restricted maximum likelihood				
		Estimate	Std err	Coverage	MSE	MAE	Estimate	Std err	Coverage	MSE	MAE
30	β_0	0.9894	0.2187	0.9396	0.0531	0.1810	0.9704	0.2336	0.9514	0.0564	0.1856
	β_1	0.9989	0.1849	0.9288	0.0380	0.1552	1.0086	0.1970	0.9424	0.0390	0.1573
	β_2	1.0067	0.1565	0.9346	0.0274	0.1312	1.0163	0.1669	0.9472	0.0284	0.1333
	$\log\sigma$	−0.0792	0.1476	0.8918	0.0311	0.1391	−0.0172	0.1579	0.9474	0.0257	0.1262
50	β_0	0.9914	0.1724	0.9386	0.0325	0.1421	0.9822	0.1787	0.9448	0.0335	0.1440
	β_1	1.0047	0.1512	0.9420	0.0245	0.1256	1.0106	0.1567	0.9486	0.0249	0.1266
	β_2	1.0075	0.1487	0.9392	0.0244	0.1252	1.0133	0.1540	0.9470	0.0248	0.1263
	$\log\sigma$	−0.0431	0.1123	0.9232	0.0155	0.0990	−0.0083	0.1166	0.9488	0.0139	0.0943
100	β_0	0.9979	0.1158	0.9414	0.0138	0.0934	0.9931	0.1179	0.9434	0.0140	0.0941
	β_1	1.0007	0.1117	0.9398	0.0133	0.0912	1.0044	0.1137	0.9436	0.0134	0.0917
	β_2	1.0036	0.1150	0.9404	0.0139	0.0940	1.0065	0.1170	0.9432	0.0140	0.0943
	$\log\sigma$	−0.0237	0.0820	0.9352	0.0075	0.0689	−0.0063	0.0836	0.9482	0.0070	0.0669
200	β_0	0.9984	0.0801	0.9484	0.0065	0.0638	0.9959	0.0809	0.9488	0.0065	0.0641
	β_1	1.0012	0.0805	0.9558	0.0064	0.0643	1.0033	0.0813	0.9578	0.0065	0.0645
	β_2	1.0020	0.0822	0.9532	0.0066	0.0652	1.0039	0.0829	0.9558	0.0066	0.0654
	$\log\sigma$	−0.0114	0.0598	0.9406	0.0037	0.0485	−0.0024	0.0605	0.9490	0.0036	0.0480

MSE = mean squared error; MAE = mean absolute error.

Regressors x_1 and x_2 are generated by a Bernoulli random number taking -1 and 1 each with probability 1/2, and a standard normal random number, respectively. The parameter $(\beta_0, \beta_1, \beta_2, \log\sigma)$ is set to $(1, 1, 1, 0)$. With this setting, the censoring level is approximately 30%.

Based upon 5,000 replications, the averages of estimates and standard errors are presented in Table 3 for $n = 30, 50, 100$, and 200. The empirical mean squared error and mean absolute error are computed. In addition, to measure an adequacy of standard error, a 95% Wald confidence interval is constructed for each parameter that then calculates the empirical coverage probability. The coverage probability should be close to 0.95 if the standard error were correct.

The regression parameter estimates are shown to be quite close for all sample sizes. However, the estimates of $\log\sigma$ are not similar when the sample size is small (Figure 2). In particular, when $n = 30$, the differences in the standard errors and the estimates of $\log\sigma$ seem to be significant. Note that the coverage probabilities of REML are closer to the nominal level than the ML for all cases, particularly for $\log\sigma$, it may conclude that the REML can provide a better estimate of the residual variance and the standard error. However, the ML has a uniformly smaller mean squared error and mean absolute error for the estimation of the regression coefficient. The ML theoretically appears when we estimate the slope parameter itself. However, the inference such as the interval estimation or the hypothesis testing on the slope parameters based on the MLE may be inadequate when the sample size is small, say $n \leq 50$. Perhaps, this is well known in the general linear model.

Acknowledgements

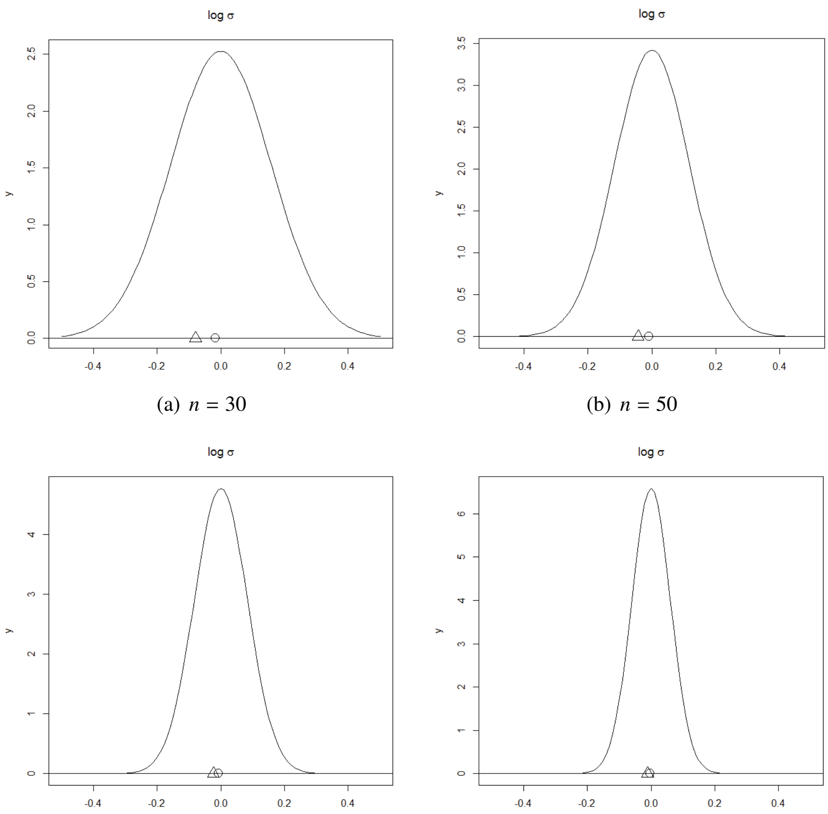
This work was supported by Hanshin University research grant.

References

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

Amemiya T (1985). Tobit models. In T Amemiya (Ed), *Advanced Econometrics* (pp. 360–411), Basil Blackwell, Oxford.

Bilias Y, Chen S, and Ying Z (2000). Simple resampling methods for censored regression quantiles,



(a) $n = 30$ (b) $n = 50$

(c) $n = 100$ (d) $n = 200$

Figure 2: *Distance between maximum likelihood ("△") and restricted maximum likelihood ("◯") estimates of* $\log \sigma$.

Journal of Econometrics, **99**, 373–386.
Efron B and Hinkley DV (1978). Assessing the accuracy of the maximum likelihood estimator: observed versus expected information, *Biometrika*, **65**, 457–482.
Green W (2004). The behaviour of maximum likelihood estimator of limited dependent variable models in the presence of fixed effects, *Econometric Journal*, **7**, 98–119.
Henningsen A (2017). Estimating censored regression models in R using the censReg packages, Retrieved May 4, 2017, from: http://cran.r-project.org/package=censReg
Hughes JP (1999). Mixed effects models with censored data with application to HIV RNA levels, *Biometrics*, **55**, 625–629.
Kleiber C and Zeileis A (2009). AER: applied econometrics with R, R package version 1.1, Retrieved May 4, 2017, from: http://CRAN.R-project.org/package=AER
Lee SC and Choi BS (2013). Bayesian interval estimation of Tobit regression model, *The Korean Journal of Applied Statistics*, **26**, 737–746.
Lee SC and Choi B (2014). A comparison of Bayesian and maximum likelihood estimations in a SUR Tobit regression model, *Korean Journal of Applied Statistics*, **27**, 991–1002.
Lee L (2017). NADA: nondetects and data analysis for environmental data, Retrieved May 4, 2017,

from: http://cran.r-project.org/package=NADA
Mroz TA (1987). The sensitivity of an empirical model of married women's hours of work to economic and statistical assumptions, *Econometrica*, **55**, 765–799.
Noh M and Lee Y (2007). REML estimation for binary data in GLMMs, *Journal of Multivariate Analysis*, **98**, 896–915.
Patterson H and Thomson R (1971). Recovery of inter-block information when block sizes are unequal, *Biometrika*, **58**, 545–554.
Tobin J (1958). Estimation of relationships for limited dependent variables, *Econometrica*, **26**, 24–36.
Yu K and Stander J (2007). Bayesian analysis of a Tobit quantile regression model, *Journal of Econometrics*, **137**, 260–276.
Wooldridge JM (2009). *Introductory Econometrics: A Modern Approach*, South-Western Cengage Learning, Mason.

Received February 18, 2017; Revised March 25, 2017; Accepted April 2, 2017