Communications for Statistical Applications and Methods
2019, Vol. 26, No. 4, 371–383

https://doi.org/10.29220/CSAM.2019.26.4.371
Print ISSN 2287-7843 / Online ISSN 2383-4757

Restricted maximum likelihood estimation of a censored random effects panel regression model

Minah Lee[a], Seung-Chun Lee[1,b]

[a]Data Analysis Team, Samsung SDS, Korea;
[b]Department of Applied Statistics, Hanshin University, Korea

Abstract

Panel data sets have been developed in various areas, and many recent studies have analyzed panel, or longitudinal data sets. Maximum likelihood (ML) may be the most common statistical method for analyzing panel data models; however, the inference based on the ML estimate will have an inflated Type I error because the ML method tends to give a downwardly biased estimate of variance components when the sample size is small. The under estimation could be severe when data is incomplete. This paper proposes the restricted maximum likelihood (REML) method for a random effects panel data model with a censored dependent variable. Note that the likelihood function of the model is complex in that it includes a multidimensional integral. Many authors proposed to use integral approximation methods for the computation of likelihood function; however, it is well known that integral approximation methods are inadequate for high dimensional integrals in practice. This paper introduces to use the moments of truncated multivariate normal random vector for the calculation of multidimensional integral. In addition, a proper asymptotic standard error of REML estimate is given.

Keywords: restricted maximum likelihood, censored dependent variable, panel regression

1. Introduction

The panel regression model with individual specific effects has the following specification:

$$w_{it} = \mathbf{x}_{it}'\boldsymbol{\beta} + \mu_i + \epsilon_{it}, \quad i = 1, 2, \ldots, n; \ t = 1, 2, \ldots, T_i, \tag{1.1}$$

where w_{it} is the dependent variable, $\mathbf{x}_{it}$ is the $p\times1$ vector of predictors, $\boldsymbol{\beta}$ is the $p\times1$ vector of regression coefficients, μ_i is the time-invariant individual specific effect, and ϵ_{it} is the remaining disturbance term. Here, subscripts i and t represent the individual and the time period, respectively. 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 an incidental parameter problem (Lancaster, 2000), while the "random effects" model treats μ_i as a random variable. As McCulloch (1996) stated, a frequentists decision to regard an effect as fixed or random is a complicated one, but we will assume that the individual specific effect μ_i's are random and independent of the regressors $\mathbf{x}_{it}$.

The standard assumption of the error term, viz. ϵ_{it}'s are independent and identically distributed normal random variable. This implies that the dependent variable can be any real number; however, in many statistical analyses the dependent variable can only be observable on limited ranges. For

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

Published 31 July 2019 / journal homepage: http://csam.or.kr

example, a dependent variable is constrained to be positive, as in the case of wage or hours worked. Unemployed people are left-censored at zero since the wage or working hours would be zero. To incorporate the unemployed in the model, the dependent variable is commonly treated as an unobservable latent variable, which leads to the standard Tobit model (Tobin, 1958).

A generalization of the standard Tobit model is that the dependent variable can be censored in either left, right or both directions of a Type I Tobit model, where the observed values of the dependent variable are defined as

$$y_i = \begin{cases} \ell, & \text{if } w_i \le \ell, \\ w_i, & \text{if } \ell < w_i < u, \\ u, & \text{if } w_i \ge u, \end{cases} \tag{1.2}$$

where ℓ and u are known lower and upper censored points. Thus, the standard Tobit model is a special case of Type I Tobit model with $\ell = 0$ and $u = \infty$.

Such a censored dependent variable commonly occurs in survival analysis, biomedical and epidemiological studies. The usual estimation method fails to provide consistent estimates for a conventional regression model. This leads to discussing the estimation methods in the censored regression model (Tobin, 1958; Maddala, 1983; and Amemiya, 1984). Currently, the dominating method may be the maximum likelihood (ML), which is implemented in most statistical packages dealing with a censored regression model. For example, R packages such as **AER** (Kleiber and Zeileis, 2009), and **NADA** (Lee, 2017) give ML estimates for the usual linear regression model. See also “qlim” procedure in SAS (2011). In particular, **censReg** (Henningsen, 2017) and “xttobit” of Stata (2017) provide ML estimates for the random effects panel regression model.

The ML estimator in nonlinear panel data model with fixed effects is widely understood to be biased and inconsistent when the length of panel T is small; however, Green (2004) found in simulation studies, that the finite sample bias of the ML method appears in the disturbance variance rather than in the slope parameters. Since, it is known that when the sample size is small, the ML estimate of disturbance variance is biased downward, and inferences on the regression coefficients will have an inflated Type I error rate because their precision is overstated. It is desirable to consider other estimation methods when the sample size is small. We believe that Bayesian estimation could be an alternative (Lee, 2016); however, a natural frequentist substitute may be the restricted maximum likelihood (REML) method (Patterson and Thomson, 1971).

Note that REML estimates variance components on the basis of residuals resulting after eliminating the fixed effects contained in a model. This makes REML divide the mean squared deviation by degrees of freedom instead of by sample size, which can remedy the downward bias of ML. It also has a Bayesian justification. Today, REML is widely used for the estimation of variance components in various mixed effects models with complete data, but surprisingly it is not well established for limited dependent variable models such as the binary or the censored dependent variable model. Even the definition of REML procedure is unclear. For example, Lee and Nelder (2001) regarded the REML as an adjusted profile likelihood method, but Drum and McCullagh (1993) considered it as an unbiased estimation equation method. See, Noh and Lee (2007) for further details.

The difficulty of an ML based approach for the limited dependent variable model lies mainly in computational problem rather than theoretical aspect. For instance, Hughes (1999) provided ML and REML estimates 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; however, such a EM based method lacks the capability of providing standard errors of variance components. He gave asymptotic standard errors for only fixed effects relying on the maximum likelihood theory. Since the asymptotic standard error does not take account of the estimation of variance components, the approximation may not be theoretically appealing. Like Hughes (1999), most works dealing with REML focus on the parameter estimation itself and did not mention the standard error of REML estimates. It is believed that the calculation of the asymptotic standard error of REML estimate is another challenging work.

The problem lies in the likelihood based methods with a limited dependent variable is the computation likelihood function. It is difficult to maximize the function directly because the likelihood function contains multidimensional integrals. Many authors proposed to use an integral approximation method for the computation of a likelihood function, and then maximize the approximated likelihood function. Various integral approximation methods such as Newton and Gauss-Hermite quadratures, Monte Carlo integration and Markov Chain Monte Carlo have been employed for this purpose. The approximation method enables the likelihood based estimation in the limited dependent variable model; however, it is well known that such integral approximation methods are inadequate for high dimensional integrals. It may be difficult to get a good approximation when the number of censored observation is large under the censored random effects panel model. Indeed, it is observed that statistical packages using different methods give quite different results. See Zhang *et al.* (2011) for further details.

The main object of this paper is to present a REML procedure for a censored dependent variable model. Many authors have considered REML in binary dependent variable models, but we can only find limited literature on REML estimation with censored data. The censored dependent variable model resembles the normal-probit model for binary data; however, the methodology used in the normal-probit model cannot be directly applicable to the model considered here in that we do not use an approximation method. It is also demonstrated through a simulation study that when the sample size is small, REML is a proper method in the sense that inferences based on it have the Type I error rate close to a nominal level.

2. Restricted maximum likelihood method

2.1. Estimation

Let $\boldsymbol{\mu} \sim \mathcal{N}(\mathbf{0}, \sigma_\mu^2 \mathbf{I})$ and $\boldsymbol{\epsilon} \sim \mathcal{N}(\mathbf{0}, \sigma_\mu^2 \mathbf{I})$ where $\boldsymbol{\mu}$ and $\boldsymbol{\epsilon}$ are independent random vectors of μ_i's and ϵ_{it}'s, respectively, and $N = \sum_{i=1}^n T_i$. Writing (1.1) as $\mathbf{w} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\boldsymbol{\mu} + \boldsymbol{\epsilon}$, where $\mathbf{X}$ is the $N \times p$ matrix of regressors and $\mathbf{Z}$ is the $N \times n$ incident matrix for μ_i's, the likelihood can be written as

$$L\left(\boldsymbol{\beta}, \sigma_\mu^2, \sigma_\epsilon^2; \mathbf{y}\right) = \int_{\mathcal{R}} f(\mathbf{w}) du, \tag{2.1}$$

where $f(\mathbf{w})$ is the probability density function of a multivariate normal random vector with mean $\mathbf{X}\boldsymbol{\beta}$, variance-covariance matrix $\mathbf{V} = \mathbf{Z}\mathbf{Z}'\sigma_\mu^2 + \mathbf{I}\sigma_\epsilon^2$, $\mathcal{R} = \{\mathbf{w} : \mathbf{y}(\mathbf{w}) = \mathbf{y}\}$ is the set of latent variables given the observed data $\mathbf{y}$, and u is the Lebesgue measure.

Note that when data is complete, i.e., no censored observations, the REML equations for variance components are shown to be

$$\mathbf{w}'\mathbf{P}\frac{\partial \mathbf{V}}{\partial \sigma_i^2}\mathbf{P}\mathbf{w} = \operatorname{tr}\left(\mathbf{P}\frac{\partial \mathbf{V}}{\partial \sigma_i^2}\right), \quad i = \mu, \epsilon,$$

where $\mathbf{P} = \mathbf{V}^{-1} - \mathbf{V}^{-1}\mathbf{X}\left(\mathbf{X}'\mathbf{V}^{-1}\mathbf{X}\right)^{-}\mathbf{X}'\mathbf{V}^{-1}$ (Searle *et al.*, 2006, p.251). For a censored data, we take conditional expectation on the REML equations, and hence estimates are obtained by solving the following equations.

$$S_{\sigma_\mu^2} = \mathrm{E}\left(\mathbf{w}'\mathbf{PZZ}'\mathbf{Pw} \mid \mathbf{y}\right) - \mathrm{tr}\left(\mathbf{PZZ}'\right) = 0, \tag{2.2}$$

$$S_{\sigma_\epsilon^2} = \mathrm{E}\left(\mathbf{w}'\mathbf{PPw} \mid \mathbf{y}\right) - \mathrm{tr}\left(\mathbf{P}\right) = 0, \tag{2.3}$$

where tr($\mathbf{A}$) denotes the trace of a square matrix $\mathbf{A}$. As we noted before, REML is not uniquely defined when data is incomplete. For instance, McCulloch (1994) gave an EM algorithm for REML estimation in a probit-normal model by treating the fixed effects as random effects whose variance tend to infinity. The EM algorithm solve the above estimation equations; therefore, our definition of REML coincides with the approach of McCulloch (1994). See also, Hughes (1999).

In general, the REML estimation includes no procedure to estimate fixed effects. The fixed-effects are estimated with the estimated random components in a complete data case; however, the conditional expectations are determined by the variance components as well as by the slope parameter $\boldsymbol{\beta}$. The random components and fixed effects should therefore be estimated simultaneously. As Searle *et al.* (2006) suggested, the log-likelihood equation for the ML of $\boldsymbol{\beta}$ will be used.

$$S_{\boldsymbol{\beta}} = \frac{\partial}{\partial \boldsymbol{\beta}} \log \int_{\mathcal{R}} f(\mathbf{w}) du = \frac{\int_{\mathcal{R}} \mathbf{X}'\mathbf{V}^{-1}(\mathbf{w} - \mathbf{X}\boldsymbol{\beta}) f(\mathbf{w}) du}{\int_{\mathcal{R}} f(\mathbf{w}) du} = \mathbf{X}'\mathbf{V}^{-1}(\mathrm{E}(\mathbf{w}|\mathbf{y}) - \mathbf{X}\boldsymbol{\beta}). \tag{2.4}$$

To solve the equations, it needs to compute the conditional expectations related to the moments of a multivariate truncated normal random vector. Thus, the calculation of the moments is essential and is the main problem of ML based approaches. Many researchers employed the Gibbs sampler or the Gauss-Hermit numerical integration method to approximate the moments and then used the EM or variations of EM algorithms. However, such numerical methods are generally not recommended for high dimensional integrals. Since, the dimensionality is increasing with the number of censored observations, it could not give proper approximation when the number of censored observations is large. An EM based method also lacks the capability of providing the standard error of REML estimate. To compute the standard error, it requires to compute up to the 4^{th} order moments, which may be hard to approximate by numerical methods.

There is a long history of the moment calculation for a multivariate truncated normal random vector. Using the moment generating function or recurrence relationships, many moment calculation methods have been proposed under various conditions, see Arismendi (2013). Among them, Kan and Robotti (2017) gave a method meet our demand. In what follows, we assume safely that necessary moments of truncated variables could be obtainable. It would be worth mentioning that their algorithm requires to compute 5^m conditional expectations where m is the number of censored observations that may be huge in some applications; however, once we notice that in the panel regression model, observations from different individuals are independent, the computational burden could be reduced greatly.

To compute the conditional expectations in (2.2), (2.3), and (2.4), we assume that the last m of $\mathbf{y}$ are censored observations. The vectors of uncensored and censored observations will be denoted by $\mathbf{y}_1 = (y_{1i}, \ldots, y_{1N-m})'$ and $\mathbf{y}_2 = (y_{21}, \ldots, y_{2m})'$, respectively. Likewise the vector of latent variables and the matrix of regressors are partitioned as $\mathbf{w}' = (\mathbf{w}_1', \mathbf{w}_2')$ and $\mathbf{X}' = (\mathbf{X}_1', \mathbf{X}_2')$. Then, $\mathbf{w}_2|\mathbf{w}_1 = \mathbf{y}_1$ is a multivariate normal random vector with mean $\boldsymbol{\mu}_{\mathbf{w}_2|\mathbf{y}_1} = \mathbf{X}_2\boldsymbol{\beta} + \mathbf{V}_{21}\mathbf{V}_{11}^{-1}(\mathbf{y}_1 - \mathbf{X}_1\boldsymbol{\beta})$ and variance-covariance matrix $\mathbf{V}_{\mathbf{w}_2|\mathbf{y}_1} = \mathbf{V}_{22} - \mathbf{V}_{21}\mathbf{V}_{11}^{-1}\mathbf{V}_{12}$ where $\mathbf{V}_{ij}$, $i, j = 1, 2$ are the partitioned matrices of $\mathbf{V}$ according to $\mathbf{w}_1$ and $\mathbf{w}_1$. This shows that $\mathbf{w}_2|\mathbf{y}$ is a multivariate truncated normal random vector. To be precise, notations related to a multivariate truncated normal distribution are defined as follows.

Suppose $\mathbf{z}_{n\times 1} \sim \mathcal{N}(\boldsymbol{\mu}, \mathbf{V})$ and $R = \{(z_1, \ldots, z_n) : a_i \le z_i \le b_i, i = 1, \ldots, n\}$, then the distribution of $\mathbf{z}|\mathbf{z} \in R$ is a multivariate truncated normal on R, and it will be represented by $\mathbf{z}|\mathbf{z} \in R \sim TN_{(\mathbf{a},\mathbf{b})}(\boldsymbol{\mu}, \mathbf{V})$, where $\mathbf{a} = \{a_i\}_{i=1}^n$ and $\mathbf{b} = \{b_i\}_{i=1}^n$.

Note that each element of $\mathbf{y}_2$ is either ℓ or u. For each $i = 1, \ldots, m$, $y_{2i} = \ell$ indicates that w_{2i} is left-truncated, and then define $a_i^* = -\infty$ and $b_i^* = \ell$. Similarly, if $y_{2i} = u$, let $a_i^* = u$ and $b_i^* = \infty$. Then, we have $\mathbf{w}_2|\mathbf{y} \sim TN_{(\mathbf{a}^*,\mathbf{b}^*)}(\boldsymbol{\mu}_{\mathbf{w}_2|\mathbf{y}_1}, \mathbf{V}_{\mathbf{w}_2|\mathbf{y}_1})$.

The conditional expectation of a quadratic form of $\mathbf{w}$ is equal to

$$\mathrm{E}\left(\mathbf{w}'\mathbf{A}\mathbf{w} \mid \mathbf{y}\right) = \mathbf{y}_1'\mathbf{A}_{11}\mathbf{y}_1 + 2\mathbf{y}_1'\mathbf{A}_{12}\mathrm{E}(\mathbf{w}_2|\mathbf{y}) + \mathrm{E}\left(\mathbf{w}_2'\mathbf{A}_{22}\mathbf{w}_2 \mid \mathbf{y}\right). \tag{2.5}$$

Here, a symmetric matrix $\mathbf{A}$ is partitioned in an obvious manner. Thus, it needs to compute up to the second order moments of a multivariate truncated normal random variable to evaluate the conditional expectation. In fact, we need to compute every quantities $\mathrm{E}(w_{2i_1}^{j_1} w_{2i_2}^{j_2} w_{2i_3}^{j_3} w_{2i_4}^{j_4}|\mathbf{y})$ where $i_k \in (1, \ldots, m)$, $k = 1, \ldots, 4$ and j_k's are nonnegative integers satisfying $\sum_{k=1}^4 j_k \le 4$. Then, the conditional expectation of a quadratic form of the latent variables shown in (2.2) or (2.3) can be calculated by (2.5) and $\mathrm{E}(\mathbf{w}_2'\mathbf{A}_{22}\mathbf{w}_2|\,\mathbf{y}) = \mathrm{tr}(\mathbf{A}_{22}\mathrm{E}(\mathbf{w}_2\mathbf{w}_2'|\mathbf{y}))$.

The Newton-Raphson method is applicable to solve estimation equations that require the derivative of the equations to form a Jacobian matrix. One may consider a numerical derivative method, which may reduce some computational burden, and hence may speed up getting the Jacobian matrix. However, we found that a numerical method is quite unstable. Besides, the Jacobian can be obtained analytically.

Using a well-known ML theory, the partial derivatives of $S_{\boldsymbol{\beta}}$ with respect to $\boldsymbol{\beta}'$ and σ_i^2, $i = \mu, \epsilon$ are shown to be

$$\frac{\partial S_{\boldsymbol{\beta}}}{\partial \boldsymbol{\beta}'} = -\mathbf{X}'\mathbf{V}^{-1}\mathbf{X} + \mathbf{X}'\mathbf{V}^{-1}\mathbf{V}_{\mathbf{w}|\mathbf{y}}\mathbf{V}^{-1}\mathbf{X}$$

and

$$\begin{aligned}\frac{\partial}{\partial \sigma_i^2} S_{\boldsymbol{\beta}} = {} & -\mathbf{X}'\mathbf{V}^{-1}\frac{\partial \mathbf{V}}{\partial \sigma_i^2}\mathbf{V}^{-1}\Big[\mathrm{E}(\mathbf{w}|\mathbf{y}) - \mathbf{X}\boldsymbol{\beta}\Big] - \mathbf{X}'\mathbf{V}^{-1}\mathbf{V}_{\mathbf{w}|\mathbf{y}}\mathbf{V}^{-1}\frac{\partial \mathbf{V}}{\partial \sigma_i^2}\mathbf{V}^{-1}\mathbf{X}\boldsymbol{\beta} \\ & + \frac{1}{2}\mathbf{X}'\mathbf{V}^{-1}\mathrm{Cov}\left(\mathbf{w}, \mathbf{w}'\mathbf{V}^{-1}\frac{\partial \mathbf{V}}{\partial \sigma_i^2}\mathbf{V}^{-1}\mathbf{w}\,\middle|\, \mathbf{y}\right),\end{aligned}$$

where $\mathbf{V}_{\mathbf{w}|\mathbf{y}}$ is the conditional variance of $\mathbf{W}$ given observed $\mathbf{y}$, $\partial\mathbf{V}/\partial\sigma_\mu^2 = \mathbf{Z}\mathbf{Z}'$ and $\partial\mathbf{V}/\partial\sigma_\epsilon^2 = \mathbf{I}$. For the differentiation of $S_{\sigma_\mu^2}$ and $S_{\sigma_\epsilon^2}$, the following theorem is applicable.

Theorem 1. *Let* $\mathbf{A}$ *be a symmetric matrix which depends on* σ_μ^2 *and* σ_ϵ^2*, but not* $\boldsymbol{\beta}$*, then for* $i = \mu$ *or* ϵ*, we have*

$$\frac{\partial}{\partial \boldsymbol{\beta}'}\mathrm{E}\left(\mathbf{w}'\mathbf{A}\mathbf{w}|\mathbf{y}\right) = \mathrm{Cov}(\mathbf{w}'\mathbf{A}\mathbf{w}, \mathbf{w}|\mathbf{y})\mathbf{V}^{-1}\mathbf{X},$$

and

$$\begin{aligned}\frac{\partial}{\partial \sigma_i^2}\mathrm{E}(\mathbf{w}'\mathbf{A}\mathbf{w}|\mathbf{y}) = {} & \mathrm{E}\left(\mathbf{w}'\frac{\partial \mathbf{A}}{\partial \sigma_i^2}\mathbf{w}\,\middle|\,\mathbf{y}\right) + \frac{1}{2}\mathrm{Cov}\left(\mathbf{w}'\mathbf{A}\mathbf{w}, \mathbf{w}'\mathbf{V}^{-1}\frac{\partial \mathbf{V}}{\partial \sigma_i^2}\mathbf{V}^{-1}\mathbf{w}|\mathbf{y}\right) \\ & - \mathrm{Cov}\left(\mathbf{w}'\mathbf{A}\mathbf{w}, \mathbf{w}|\mathbf{y}\right)\mathbf{V}^{-1}\frac{\partial \mathbf{V}}{\partial \sigma_i^2}\mathbf{V}^{-1}\mathbf{X}\boldsymbol{\beta}.\end{aligned}$$

Proof: The conditional expectation can be written as

$$\mathrm{E}(\mathbf{w}'\mathbf{A}\mathbf{w}|\mathbf{y}) = \int_{\mathcal{R}} \mathbf{w}'\mathbf{A}\mathbf{w} f(\mathbf{w}) du \Big/ \int_{\mathcal{R}} f(\mathbf{w}) du,$$

and $(\partial/\partial\boldsymbol{\beta}')f(\mathbf{w}) = (\mathbf{w} - \mathbf{X}\boldsymbol{\beta})'\mathbf{V}^{-1}\mathbf{X}f(\mathbf{w})$, we have

$$\begin{aligned}\frac{\partial \mathrm{E}(\mathbf{w}'\mathbf{A}\mathbf{w}|\mathbf{y})}{\partial\boldsymbol{\beta}'} &= \frac{\int_{\mathcal{R}} \mathbf{w}'\mathbf{A}\mathbf{w}\frac{\partial}{\partial\boldsymbol{\beta}'} f(\mathbf{w})du}{\int_{\mathcal{R}} f(\mathbf{w})du} - \frac{\int_{\mathcal{R}} \mathbf{w}'\mathbf{A}\mathbf{w} f(\mathbf{w})du \int_{\mathcal{R}} \frac{\partial}{\partial\boldsymbol{\beta}'} f(\mathbf{w})du}{\left[\int_{\mathcal{R}} f(\mathbf{w})du\right]^2} \\ &= [\mathrm{E}(\mathbf{w}'\mathbf{A}\mathbf{w}(\mathbf{w} - \mathbf{X}\boldsymbol{\beta})'|\mathbf{y}) - \mathrm{E}(\mathbf{w}'\mathbf{A}\mathbf{w}|\mathbf{y})\mathrm{E}((\mathbf{w} - \mathbf{X}\boldsymbol{\beta})'|\mathbf{y})]\,\mathbf{V}^{-1}\mathbf{X} \\ &= \mathrm{Cov}(\mathbf{w}'\mathbf{A}\mathbf{w}, \mathbf{w}|\mathbf{y})\mathbf{V}^{-1}\mathbf{X}.\end{aligned}$$

Likewise, $(\partial/\partial\sigma_i^2)f(\mathbf{w}) = (1/2)[(\mathbf{w} - \mathbf{X}\boldsymbol{\beta})'\mathbf{V}^{-1}(\partial\mathbf{V}/\partial\sigma_i^2)\mathbf{V}^{-1}(\mathbf{w} - \mathbf{X}\boldsymbol{\beta}) - \mathrm{tr}(\mathbf{V}^{-1}\partial\mathbf{V}/\partial\sigma_i^2)]f(\mathbf{w})$ gives

$$\begin{aligned}\frac{\partial \mathrm{E}(\mathbf{w}'\mathbf{A}\mathbf{w}|\mathbf{y})}{\partial\sigma_i^2} &= \frac{\int_{\mathcal{R}} \frac{\partial}{\partial\sigma_i^2}\{\mathbf{w}'\mathbf{A}\mathbf{w} f(\mathbf{w})\}\, du}{\int_{\mathcal{R}} f(\mathbf{w})du} - \frac{\int_{\mathcal{R}} \mathbf{w}'\mathbf{A}\mathbf{w} f(\mathbf{w})du \int_{\mathcal{R}} \frac{\partial}{\partial\sigma_i^2} f(\mathbf{w})du}{\left[\int_{\mathcal{R}} f(\mathbf{w})du\right]^2} \\ &= \frac{1}{2}\mathrm{E}\left[\mathbf{w}'\mathbf{A}\mathbf{w}\left\{(\mathbf{w} - \mathbf{X}\boldsymbol{\beta})'\mathbf{V}^{-1}\frac{\partial\mathbf{V}}{\partial\sigma_i^2}\mathbf{V}^{-1}(\mathbf{w} - \mathbf{X}\boldsymbol{\beta}) - \mathrm{tr}\left(\mathbf{V}^{-1}\frac{\partial\mathbf{V}}{\partial\sigma_i^2}\right)\right\}\middle|\,\mathbf{y}\right] \\ &\quad - \frac{1}{2}\mathrm{E}(\mathbf{w}'\mathbf{A}\mathbf{w}|\mathbf{y})\left\{\mathrm{E}\left((\mathbf{w}-\mathbf{X}\boldsymbol{\beta})'\mathbf{V}^{-1}\frac{\partial\mathbf{V}}{\partial\sigma_i^2}\mathbf{V}^{-1}(\mathbf{w}-\mathbf{X}\boldsymbol{\beta})\middle|\,\mathbf{y}\right) - \mathrm{tr}\left(\mathbf{V}^{-1}\frac{\partial\mathbf{V}}{\partial\sigma_i^2}\right)\right\} + \mathrm{E}\left(\mathbf{w}'\frac{\partial\mathbf{A}}{\partial\sigma_i^2}\mathbf{w}\middle|\,\mathbf{y}\right) \\ &= \mathrm{E}\left(\mathbf{w}'\frac{\partial\mathbf{A}}{\partial\sigma_i^2}\mathbf{w}\middle|\,\mathbf{y}\right) + \frac{1}{2}\mathrm{Cov}\left\{\mathbf{w}'\mathbf{A}\mathbf{w}, (\mathbf{w} - \mathbf{X}\boldsymbol{\beta})'\mathbf{V}^{-1}\frac{\partial\mathbf{V}}{\partial\sigma_i^2}\mathbf{V}^{-1}(\mathbf{w} - \mathbf{X}\boldsymbol{\beta})\middle|\,\mathbf{y}\right\} \\ &= \mathrm{E}\left(\mathbf{w}'\frac{\partial\mathbf{A}}{\partial\sigma_i^2}\mathbf{w}\middle|\,\mathbf{y}\right) + \frac{1}{2}\mathrm{Cov}\left(\mathbf{w}'\mathbf{A}\mathbf{w}, \mathbf{w}'\mathbf{V}^{-1}\frac{\partial\mathbf{V}}{\partial\sigma_i^2}\mathbf{V}^{-1}\mathbf{w}|\mathbf{y}\right) - \mathrm{Cov}(\mathbf{w}'\mathbf{A}\mathbf{w}, \mathbf{w}|\mathbf{y})\,\mathbf{V}^{-1}\frac{\partial\mathbf{V}}{\partial\sigma_i^2}\mathbf{V}^{-1}\mathbf{X}\boldsymbol{\beta}.\end{aligned}$$

□

Theorem 1 is quite standard and is useful not only for REML but also for ML estimation. Perhaps it is well known, but we could not find a statement of the second order derivatives when data is incomplete. We include the proof for completeness. The second order derivatives gives an advantage of our definition of REML as the solution of (2.2), (2.3), and (2.4) in that the asymptotic standard error can be expressed as analytic forms rather than numerical forms.

Once we note $\mathbf{PPZZ'P}$ is a symmetric matrix, $\mathbf{PPZZ'P} = \mathbf{PZZ'PP}$, and $(\partial/\partial\sigma_i^2)\mathbf{P} = -\mathbf{P}(\partial\mathbf{V}/\partial\sigma_i^2)$ $\mathbf{P}$, then Theorem 1 gives following partial derivatives

$$\begin{aligned}\frac{\partial S_{\sigma_\mu^2}}{\partial\boldsymbol{\beta}'} &= \mathrm{Cov}(\mathbf{w}'\mathbf{PZZ'Pw}, \mathbf{w}|\mathbf{y})\mathbf{V}^{-1}\mathbf{X}, \\ \frac{\partial S_{\sigma_\mu^2}}{\partial\sigma_\mu^2} &= -2\mathrm{E}\left(\mathbf{w}'\mathbf{PZZ'PZZ'Pw}\middle|\mathbf{y}\right) + \frac{1}{2}\mathrm{Cov}\left(\mathbf{w}'\mathbf{PZZ'Pw}, \mathbf{w}'\mathbf{V}^{-1}\mathbf{ZZ'}\mathbf{V}^{-1}\mathbf{w}|\mathbf{y}\right) \\ &\quad - \mathrm{Cov}\left(\mathbf{w}'\mathbf{PZZ'Pw}, \mathbf{w}|\mathbf{y}\right)\mathbf{V}^{-1}\mathbf{ZZ'}\mathbf{V}^{-1}\mathbf{X}\boldsymbol{\beta} + \mathrm{tr}\left(\mathbf{PZZ'PZZ'}\right), \\ \frac{\partial S_{\sigma_\mu^2}}{\partial\sigma_\epsilon^2} &= -2\mathrm{E}\left(\mathbf{w}'\mathbf{PPZZ'Pw}\middle|\,\mathbf{y}\right) + \frac{1}{2}\mathrm{Cov}\left(\mathbf{w}'\mathbf{PZZ'Pw}, \mathbf{w}'\mathbf{V}^{-1}\mathbf{V}^{-1}\mathbf{w}|\mathbf{y}\right) \\ &\quad - \mathrm{Cov}\left(\mathbf{w}'\mathbf{PZZ'Pw}, \mathbf{w}|\mathbf{y}\right)\mathbf{V}^{-1}\mathbf{V}^{-1}\mathbf{X}\boldsymbol{\beta} + \mathrm{tr}\left(\mathbf{PZZ'P}\right),\end{aligned}$$

$$
\begin{aligned}
\frac{\partial S_{\sigma_\epsilon^2}}{\partial \boldsymbol{\beta}'} &= \mathrm{Cov}(\mathbf{w}'\mathbf{P}\mathbf{P}\mathbf{w}, \mathbf{w}|\mathbf{y})\mathbf{V}^{-1}\mathbf{X}, \\
\frac{\partial S_{\sigma_\epsilon^2}}{\partial \sigma_\mu^2} &= -2\mathrm{E}\left(\mathbf{w}'\mathbf{P}\mathbf{P}\mathbf{Z}\mathbf{Z}'\mathbf{P}\mathbf{w}\middle|\,\mathbf{y}\right) + \frac{1}{2}\mathrm{Cov}\left(\mathbf{w}'\mathbf{P}\mathbf{P}\mathbf{w}, \mathbf{w}'\mathbf{V}^{-1}\mathbf{Z}\mathbf{Z}'\mathbf{V}^{-1}\mathbf{w}|\mathbf{y}\right) \\
&\quad - \mathrm{Cov}\left(\mathbf{w}'\mathbf{P}'\mathbf{P}\mathbf{w}, \mathbf{w}|\mathbf{y}\right)\mathbf{V}^{-1}\mathbf{Z}\mathbf{Z}'\mathbf{V}^{-1}\mathbf{X}\boldsymbol{\beta} + \mathrm{tr}\left(\mathbf{P}\mathbf{Z}\mathbf{Z}'\mathbf{P}\right), \\
\frac{\partial S_{\sigma_\epsilon^2}}{\partial \sigma_\epsilon^2} &= -2\mathrm{E}\left(\mathbf{w}'\mathbf{P}\mathbf{P}\mathbf{P}\mathbf{w}\middle|\,\mathbf{y}\right) + \frac{1}{2}\mathrm{Cov}\left(\mathbf{w}'\mathbf{P}\mathbf{P}\mathbf{w}, \mathbf{w}'\mathbf{V}^{-1}\mathbf{V}^{-1}\mathbf{w}|\mathbf{y}\right) \\
&\quad - \mathrm{Cov}\left(\mathbf{w}'\mathbf{P}\mathbf{P}\mathbf{w}, \mathbf{w}|\mathbf{y}\right)\mathbf{V}^{-1}\mathbf{V}^{-1}\mathbf{X}\boldsymbol{\beta} + \mathrm{tr}\left(\mathbf{P}\mathbf{P}\right).
\end{aligned}
$$

Let S($\boldsymbol{\theta}$) be the vector of $S_{\boldsymbol{\theta}}, S_{\sigma_\mu^2}$, and $S_{\sigma_\epsilon^2}$ where $\boldsymbol{\theta} = (\boldsymbol{\beta}', \sigma_\mu^2, \sigma_\epsilon^2)'$, then the partial derivatives and S($\boldsymbol{\theta}$) consist of conditional moments of the forms, E($\mathbf{w}|\mathbf{y}$), Var($\mathbf{w}|\mathbf{y}$), E($\mathbf{w}'\mathbf{A}\mathbf{w}|\mathbf{y}$), Cov($\mathbf{w}, \mathbf{w}'\mathbf{A}\mathbf{w}|\mathbf{y}$), and Cov($\mathbf{w}'\mathbf{A}\mathbf{w}, \mathbf{w}'\mathbf{B}\mathbf{w}|\mathbf{y}$) where $\mathbf{A}$ and $\mathbf{B}$ are some symmetric matrices. Let us consider the calculation of these quantities. The first two are

$$
\mathrm{E}(\mathbf{w}|\mathbf{y}) = \begin{pmatrix} \mathbf{y}_1 \\ \mathrm{E}(\mathbf{w}_2|\mathbf{y}) \end{pmatrix} \quad \text{and} \quad \mathbf{V}_{\mathbf{w}|\mathbf{y}} = \mathrm{Var}(\mathbf{w}|\mathbf{y}) = \begin{pmatrix} \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{V}_{\mathbf{w}_2|\mathbf{y}} \end{pmatrix},
$$

where $\mathbf{V}_{\mathbf{w}_2|\mathbf{y}} = \mathrm{E}(\mathbf{w}_2\mathbf{w}_2'|\mathbf{y}) - \mathrm{E}(\mathbf{w}_2|\mathbf{y})\mathrm{E}(\mathbf{w}_2'|\mathbf{y})$, and the third type has been seen before. Also, one may know that the first $N - m$ elements of $N \times 1$ vector Cov($\mathbf{w}, \mathbf{w}'\mathbf{A}\mathbf{w}|\mathbf{y}$) are zero. Indeed, and it can be shown

$$
\mathrm{Cov}(\mathbf{w}, \mathbf{w}'\mathbf{A}\mathbf{w}|\mathbf{y})' = \left(\mathbf{0}', \left[\mathrm{Cov}\left(\mathbf{w}_2, \mathbf{w}_2'\mathbf{A}_{22}\mathbf{w}_2|\mathbf{y}\right) + 2\mathbf{V}_{\mathbf{w}_2|\mathbf{y}}\mathbf{A}_{21}\mathbf{y}_1\right]'\right).
$$

Since, the i^{th} element of Cov($\mathbf{w}_2, \mathbf{w}_2'\mathbf{A}_{22}\mathbf{w}_2|\mathbf{y}$) is equal to

$$
\mathrm{Cov}\left(w_{2i}, \mathbf{w}_2'\mathbf{A}_{22}\mathbf{w}_2|\mathbf{y}\right) = \mathrm{E}\left(w_{2i}\mathbf{w}_2'\mathbf{A}_{22}\mathbf{w}_2|\mathbf{y}\right) - \mathrm{E}(w_{2i}|\mathbf{y})\mathrm{E}(\mathbf{w}_2'\mathbf{A}_{22}\mathbf{w}_2|\mathbf{y}),
$$

and for each $i = 1, 2, \ldots, m$,

$$
\begin{aligned}
&\mathrm{E}\left(w_{2i}\mathbf{w}_2'\mathbf{A}_{22}\mathbf{w}_2|\mathbf{y}\right) \\
&= \sum_{j=1}^{m}\sum_{k=1}^{m} a_{jk}^{22}\mathrm{E}\left(w_{2i}w_{2j}w_{2k}|\mathbf{y}\right) \\
&= a_{ii}^{22}\mathrm{E}\left(w_{2i}^3|\mathbf{y}\right) + 2\sum_{j\neq i}^{m} a_{ij}^{22}\mathrm{E}\left(w_{2i}^2 w_{2j}|\mathbf{y}\right) + \sum_{j\neq i}^{m} a_{jj}^{22}\mathrm{E}\left(w_{2i}w_{2j}^2|\mathbf{y}\right) + 2\sum_{\substack{j<k \\ j,k\neq i}}^{m}\sum^{m} a_{jk}^{22}\mathrm{E}\left(w_{2i}w_{2j}w_{2k}|\mathbf{y}\right),
\end{aligned}
$$

where a_{ij}^{22} denotes the $(i, j)^{th}$ element of $\mathbf{A}_{22}$, Cov($\mathbf{w}, \mathbf{w}'\mathbf{A}\mathbf{w}|\mathbf{y}$) is computable with up to the 3^{rd} order conditional moments of a multivariate truncated normal random vector. Finally, the conditional covariance of two quadratic forms of latent variables is shown to be

$$
\begin{aligned}
\mathrm{Cov}\left(\mathbf{w}'\mathbf{A}\mathbf{w}, \mathbf{w}'\mathbf{B}\mathbf{w}|\mathbf{y}\right) = {}& 4\mathbf{y}_1'\mathbf{A}_{12}\mathbf{V}_{\mathbf{w}_2|\mathbf{y}}\mathbf{B}_{12}'\mathbf{y}_1 + 2\mathbf{y}_1'\mathbf{A}_{12}\mathrm{Cov}\left(\mathbf{w}_2, \mathbf{w}_2'\mathbf{B}_{22}\mathbf{w}_2|\mathbf{y}\right) \\
&+ 2\mathrm{Cov}\left(\mathbf{w}_2'\mathbf{A}_{22}\mathbf{w}_2, \mathbf{w}_2|\mathbf{y}\right)\mathbf{B}_{12}'\mathbf{y}_1 + \mathrm{Cov}\left(\mathbf{w}_2'\mathbf{A}_{22}\mathbf{w}_2, \mathbf{w}_2'\mathbf{B}_{22}\mathbf{w}_2|\mathbf{y}\right),
\end{aligned}
$$

and the last term of above equation is equal to $\mathrm{E}(\mathbf{w}_2'\mathbf{A}_{22}\mathbf{w}_2\mathbf{w}_2'\mathbf{B}_{22}\mathbf{w}_2|\mathbf{y}) - \mathrm{E}(\mathbf{w}_2'\mathbf{A}_{22}\mathbf{w}_2|\mathbf{y})\mathrm{E}(\mathbf{w}_2'\mathbf{B}_{22}\mathbf{w}_2|\mathbf{y})$. Since

$$
\mathrm{E}\left(\mathbf{w}_2'\mathbf{A}_{22}\mathbf{w}_2\mathbf{w}_2'\mathbf{B}_{22}\mathbf{w}_2|\mathbf{y}\right) = \mathrm{tr}\left[\mathbf{B}_{22}\mathrm{E}\left(\mathbf{w}_2\mathbf{w}_2'\mathbf{A}_{22}\mathbf{w}_2\mathbf{w}_2'|\mathbf{y}\right)\right],
$$

the conditional covariance can be calculated if $\mathrm{E}(\mathbf{w}_2\mathbf{w}_2'\mathbf{A}_{22}\mathbf{w}_2\mathbf{w}_2'|\mathbf{y})$ is given. Let e_{ij} be the $(i,j)^{th}$ element of $\mathrm{E}(\mathbf{w}_2\mathbf{w}_2'\mathbf{A}_{22}\mathbf{w}_2\mathbf{w}_2'|\mathbf{y})$, then we have

$$e_{ii} = a_{ii}^{22}\mathrm{E}\left(w_{2i}^4|\mathbf{y}\right) + 2\sum_{k\neq i}^{m} a_{ik}^{22}\mathrm{E}\left(w_{2i}^3 w_{2k}|\mathbf{y}\right) + \sum_{k\neq i}^{m} a_{kk}^{22}\mathrm{E}\left(w_{2i}^2 w_{2k}^2|\mathbf{y}\right) + 2\sum^{m}\sum^{m}_{k<\ell,k,\ell\neq i} a_{k\ell}^{22}\mathrm{E}\left(w_{2i}^2 w_{2k} w_{2\ell}|\mathbf{y}\right),$$

for $i = 1,\ldots,m$ and

$$\begin{aligned} e_{ij} &= a_{ii}^{22}\mathrm{E}\left(w_{2i}^3 w_{2j}|\mathbf{y}\right) + a_{jj}^{22}\mathrm{E}\left(w_{2i} w_{2j}^3|\mathbf{y}\right) + 2a_{ij}^{22}\mathrm{E}\left(w_{2i}^2 w_{2j}^2|\mathbf{y}\right) + 2\sum_{k\neq(i,j)}^{m} a_{ik}^{22}\mathrm{E}\left(w_{2i}^2 w_{2j} w_{2k}|\mathbf{y}\right) \\ &\quad + 2\sum_{k\neq(i,j)}^{m} a_{jk}^{22}\mathrm{E}\left(w_{2i} w_{2j}^2 w_{2k}|\mathbf{y}\right) + \sum_{k\neq(i,j)}^{m} a_{kk}^{22}\mathrm{E}\left(w_{2i} w_{2j} w_{2k}^2|\mathbf{y}\right) + 2\sum^{m}\sum^{m}_{\substack{k<\ell \\ k,\ell\neq i,j}} a_{k\ell}^{22}\mathrm{E}\left(w_{2i} w_{2j} w_{2k} w_{2\ell}|\mathbf{y}\right), \end{aligned}$$

for $i, j = 1,\ldots,m;\ i \neq j$.

Thus, the nonlinear simultaneous equations, $\mathrm{S}(\boldsymbol{\theta}) = \mathbf{0}$ and the Jacobian consisting of

$$\mathbf{J}(\boldsymbol{\theta}) = \begin{pmatrix} \dfrac{\partial S_{\boldsymbol{\beta}}}{\partial \boldsymbol{\beta}'}, & \dfrac{\partial S_{\boldsymbol{\beta}}}{\partial \sigma_\mu^2}, & \dfrac{\partial S_{\boldsymbol{\beta}}}{\partial \sigma_\epsilon^2} \\ \dfrac{\partial S_{\sigma_\mu^2}}{\partial \boldsymbol{\beta}'}, & \dfrac{\partial S_{\sigma_\mu^2}}{\partial \sigma_\mu^2}, & \dfrac{\partial S_{\sigma_\mu^2}}{\partial \sigma_\epsilon^2} \\ \dfrac{\partial S_{\sigma_\epsilon^2}}{\partial \boldsymbol{\beta}'}, & \dfrac{\partial S_{\sigma_\epsilon^2}}{\partial \sigma_\mu^2}, & \dfrac{\partial S_{\sigma_\epsilon^2}}{\partial \sigma_\epsilon^2} \end{pmatrix}.$$

are computable with conditional moments $\mathrm{E}(w_{2i_1}^{j_1} w_{2i_2}^{j_2} w_{2i_3}^{j_3} w_{2i_4}^{j_4}|\mathbf{y})$, $\sum_{k=1}^4 j_k \leq 4$, which are assumed to be given by the algorithm of Kan and Robotti (2017).

The REML estimate of $\boldsymbol{\theta}$ can be found by applying the Newton-Raphson method,

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

with an arbitrary initial value $\hat{\boldsymbol{\theta}}_{(0)}$. If convergence is reached at the t-step, we set $\hat{\boldsymbol{\theta}}_{(t+1)}$ as the REML estimate. For the initial value, it may use the estimate of $\boldsymbol{\theta}$ based only on uncensored observations. We have used a R package, **plm** of Croissant and Millo (2008) for this purpose. With this initial value, we could get the convergence in almost all cases among 1,000 replications shown in Section 4.

2.2. Standard error

When data is incomplete, it seems that the standard error of REML is more complex than ML, because Louis (1982) provided the incomplete data Hessian in terms of the conditional expectations of the complete data Hessian, which makes it possible to work with incomplete data. Theoretically there is no difficulty in ML estimation since the Hessian is directly related to the variance of the ML estimate. Many alternatives to Louis's method have also been proposed (Meilijson, 1989; Meng and Rubin, 1991; Duan and Fulop, 2011). The problem exists in the REML estimation is that, unlike the ML case, the Hessian does not give a variance. However, it is hard to find literature mentioning the standard error despite most researchers mentioning REML estimation with incomplete data. Regardless, a slight variant of the maximum likelihood theory can give an asymptotic variance of REML estimate.

Theorem 2. *Let $\hat{\boldsymbol{\theta}}$ be a solution to* $\mathrm{S}(\boldsymbol{\theta}) = \mathbf{0}$*, then an asymptotic variance of $\hat{\boldsymbol{\theta}}$ is given by*

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

where $\mathbf{J}(\boldsymbol{\theta})$ *is the Jacobian of* $\mathrm{S}(\boldsymbol{\theta})$.

Proof: The Taylor series expansion gives $\mathbf{S}(\boldsymbol{\theta}) \approx \mathbf{S}(\hat{\boldsymbol{\theta}}) + \mathbf{J}(\hat{\boldsymbol{\theta}})(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta})$, but $\mathbf{S}(\hat{\boldsymbol{\theta}}) = \mathbf{0}$, we have $\mathbf{J}^{-1}(\hat{\boldsymbol{\theta}})\mathbf{S}(\boldsymbol{\theta}) \approx \hat{\boldsymbol{\theta}} - \boldsymbol{\theta}$. □

In the ML estimation, the Hessian is equal to the Jacobian, and the variance of score is obtained by the expected value of minus the Hessian, $-\mathrm{E}[(\partial^2 \log L)/(\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}')]$ called the Fisher information or the expected information. Thus, if we regard $\mathbf{S}(\boldsymbol{\theta})$ as the score of ML, $\mathrm{Var}(\mathbf{S}(\boldsymbol{\theta}))$ plays the role of the expected information. Note that Efron and Hinkley (1978) stated that in most cases the observed information, which is equal to minus the Hessian, is a more appropriate measure of information than the expected information, and an asymptotic variance of ML estimate is usually computed using the observed information rather than expected information. Thus, it may be desirable to replace $\mathrm{Var}(\mathbf{S}(\boldsymbol{\theta}))$ by some observed information like quantity.

To use Theorem 2, $\mathrm{Var}(\mathrm{S}(\boldsymbol{\theta}))$ should be given. Since, $\partial S_{\boldsymbol{\beta}}/\partial\boldsymbol{\beta}' = (\partial^2 \log L)/(\partial\boldsymbol{\beta}\partial\boldsymbol{\beta}')$, using the maximum likelihood theory, we have

$$\mathrm{Var}(\mathrm{S}_{\boldsymbol{\beta}}) = -\mathrm{E}\left[\frac{\partial^2 \log L}{\partial\boldsymbol{\beta}\partial\boldsymbol{\beta}'}\right] = \mathbf{X}'\mathbf{V}^{-1}\mathbf{X} - \mathbf{X}'\mathbf{V}^{-1}\mathrm{Var}(\mathbf{w}|\mathbf{y})\mathbf{V}^{-1}\mathbf{X},$$

but, $\partial \mathrm{S}_{\sigma_i^2}/\partial\sigma_i^2$ does not give the variance of $\mathrm{S}_{\sigma_i^2}$ for i= μ, ϵ. To compute $\mathrm{Var}(S_{\sigma_i^2})$, we can use a relationship $\mathrm{Var}(\mathrm{E}(X|Y)) = \mathrm{Var}(X) - \mathrm{E}(\mathrm{Var}(X|Y))$. That is, the variance of conditional expectation of a quadratic form can be obtained by

$$\begin{aligned}\mathrm{Var}\left[\mathrm{E}(\mathbf{w}'\mathbf{A}\mathbf{w}|\mathbf{y})\right] &= \mathrm{Var}(\mathbf{w}'\mathbf{A}\mathbf{w}) - \mathrm{E}\left[\mathrm{Var}(\mathbf{w}'\mathbf{A}\mathbf{w}|\mathbf{y})\right] \\ &= 2\mathrm{tr}(\mathbf{AVAV}) + 4\boldsymbol{\beta}'\mathbf{X}'\mathbf{AVAX}\boldsymbol{\beta} - \mathrm{E}\left[\mathrm{Var}\left(2\mathbf{y}_1'\mathbf{A}_{12}\mathbf{w}_2 + \mathbf{w}_2'\mathbf{A}\mathbf{w}_2|\mathbf{y}\right)\right].\end{aligned} \tag{2.7}$$

As stated before, the variance given in (2.7) plays the role of the expected information, but in view of Efron and Hinkley (1978), it would be better to use the non-expected value of (2.7). Thus, we replace the expectation part in (2.7) by

$$\mathrm{Var}\left(2\mathbf{y}_1'\mathbf{A}_{12}\mathbf{w}_2 + \mathbf{w}_2'\mathbf{A}\mathbf{w}_2|\mathbf{y}\right) = 4\mathbf{y}_1'\mathbf{A}_{12}\mathbf{V}_{\mathbf{w}_2|\mathbf{y}}\mathbf{A}_{12}'\mathbf{y}_1 + 4\mathbf{y}_1\mathbf{A}_{12}\mathrm{Cov}(\mathbf{w}_2, \mathbf{w}_2'\mathbf{A}\mathbf{w}_2|\mathbf{y}) + \mathrm{Var}(\mathbf{w}_2'\mathbf{A}\mathbf{w}_2|\mathbf{y}).$$

Since, $\mathrm{Var}(\mathbf{w}_2'\mathbf{A}\mathbf{w}_2|\mathbf{y}) = \mathrm{Cov}(\mathbf{w}_2'\mathbf{A}\mathbf{w}_2, \mathbf{w}_2'\mathbf{A}\mathbf{w}_2|\mathbf{y})$, we can use previous results for computing the conditional variance. Similarly, for the covariance of $\mathrm{S}_{\boldsymbol{\beta}}$ and $\mathrm{S}_{\sigma_i^2}$

$$\begin{aligned}\mathrm{Cov}\left[\mathbf{X}'\mathbf{V}^{-1}\left(\mathrm{E}(\mathbf{w}|\mathbf{y}) - \mathbf{X}\boldsymbol{\beta}\right), \mathrm{E}(\mathbf{w}'\mathbf{A}\mathbf{w}|\mathbf{y})\right] &= \mathbf{X}'\mathbf{V}^{-1}\mathrm{Cov}\left(\mathbf{w}, \mathbf{w}'\mathbf{A}\mathbf{w}\right) - \mathbf{X}'\mathbf{V}^{-1}\mathrm{E}\left[\mathrm{Cov}\left(\mathbf{w}, \mathbf{w}'\mathbf{A}\mathbf{w}|\mathbf{y}\right)\right] \\ &= 2\mathbf{X}'\mathbf{AX}\boldsymbol{\beta} - \mathbf{X}'\mathbf{V}^{-1}\mathrm{E}\left[\mathrm{Cov}\left(\mathbf{w}, \mathbf{w}'\mathbf{A}\mathbf{w}|\mathbf{y}\right)\right]\end{aligned} \tag{2.8}$$

is applicable. Finally, the covariance of $\mathrm{S}_{\sigma_\mu^2}$ and $\mathrm{S}_{\sigma_\epsilon^2}$ is equal to

$$\begin{aligned}\mathrm{Cov}\left(\mathrm{S}_{\sigma_\mu^2}, \mathrm{S}_{\sigma_\epsilon^2}\right) &= \mathrm{Cov}\left(\mathbf{w}'\mathbf{PZZ}'\mathbf{Pw}, \mathbf{w}'\mathbf{PPw}\right) - \mathrm{Cov}\left(\mathbf{w}'\mathbf{PZZ}'\mathbf{Pw}, \mathbf{w}'\mathbf{PPw}|\mathbf{y}\right) \\ &= 2\mathrm{tr}(\mathbf{PZZ}'\mathbf{PVPPV}) + 4\boldsymbol{\beta}'\mathbf{X}'\mathbf{PZZ}'\mathbf{PVPPX}\boldsymbol{\beta} - \mathrm{E}\left[\mathrm{Cov}\left(\mathbf{w}'\mathbf{PZZ}'\mathbf{Pw}, \mathbf{w}'\mathbf{PPw}|\mathbf{y}\right)\right].\end{aligned} \tag{2.9}$$

The expected values in (2.8) and (2.9) will be replaced by non-expected values, $\mathrm{Cov}(\mathbf{w}, \mathbf{wAw}|\mathbf{y})$ and $\mathrm{Cov}(\mathbf{w}'\mathbf{PZZ}'\mathbf{Pw}, \mathbf{w}'\mathbf{PPw}|\mathbf{y})$, respectively for calculating asymptotic standard errors.

Table 1: Estimates of EmpUK data

	REML			ML		
	Estimate	Std. err	*t*-value	Estimate	Std. err	*t*-value
(Intercept)	2.3431	0.7921	2.958	2.3423	0.7901	2.965
wage	−0.0814	0.0164	−4.970	−0.0814	0.0164	−4.977
capital	0.1245	0.0424	2.939	0.1248	0.0422	2.956
output	0.0424	0.0039	10.971	0.0424	0.0039	10.986
σ_μ^2	35.1454	4.5396		34.8675	4.4131	
σ_ϵ^2	1.1414	0.0562		1.1382	0.0558	

REML = restricted maximum likelihood; ML = maximum likelihood.

3. Examples

In this section, we demonstrate two examples to reflect some of main features of REML and ML estimates. In addition, we wish to talk about some computational issues.

Example 1. The first data “EmpUK” presented by Arellano and Bond (1991) is an unbalanced panel of 140 observations from 1976 to 1984. It consists of 1031 observations on 7 variables. The dependent variable “emp” has a heavily right-skewed distribution, and some observations are quite large compared with most observations. The values of “emp” larger than 30 are treated the same in this example. That is, we assume that the dependent variable is right-censored at 30, then 57 out of 1031 are treated as censored observations. Because of the assumption, the estimation itself may not be meaningful, but it is believed that the data is suitable to show a large sample property of REML and ML estimates.

Table 1 shows the estimates of REML and ML. All the computations are done under R (R Core Team, 2017) with RcppArmadillo (Eddelbuettel and Sanderson, 2014) and Rcpp (Eddelbuettel *et al.*, 2018). The two estimation methods do not make differences to the estimate of variance components as well as slope parameters. The ML estimate of random component is only slightly smaller and has the tendency to have a slightly smaller standard error than REML, but it is hard to find some statistical meaning. Because the downward bias is a small sample property, this result can be predictable. Since, ML is theoretically preferable method, it seems that the REML has little or no theoretical support.

Both **censReg** (Henningsen, 2017) and **xttobit** of Stata (2017) have a capability of ML estimation for a censored random-effects panel regression model. We first used **censReg** to get the ML estimates for Table 1, but found that it is quite unstable and failed to give estimates. It mainly uses numerical methods, the Gauss-Hermite quadrature (GHQ) for approximating the log-likelihood, and a numerical differential method, but GHQ is generally not recommended for high dimensional integrals. Since, the dimensionality is increasing with the number of censored observations, it could not give a proper approximation when the number of censored observations is large. Generally, increasing the number of quadratic points, nGHQ increases the accuracy of the computation, but increasing nGHQ was not helpful for the ML estimation of “EmpUK” data. Even when nGHQ is sufficiently large, small changes in nGHQ produced quite different estimates.

We did not examine **xttobit**, but it uses the same algorithm and carries an inherent possibility that **xttobit** suffers the same problem. It seems that **censReg** is only good when the number of censored observations is small. Even that case, we must be careful about the value of nGHQ. In **censReg**, the default value of nGHQ is 8, but it would not be large enough, because Lesaffre and Spiessens (2001) gave a simple example of a logistic random-intercepts model in the context of a longitudinal clinical trial where nGHQ gives valid results only for a high number of quadrature points. The author of **censReg** may recognize this point. He gave an example showing the effect of nGHQ.

Table 2: Estimates of an artificial panel data

	REML			ML		
	Estimate	Std. err	t-value	Estimate	Std. err	t-value
(Intercept)	−0.3921	0.4782	−0.820	−0.3655	0.4612	−0.792
x1	1.7020	0.2186	7.787	1.6838	0.2124	7.927
x2	2.2875	0.6919	3.306	2.2636	0.6739	3.359
σ_μ^2	0.9005	0.5109		0.7961	0.4474	
σ_ϵ^2	1.0175	0.2720		0.9734	0.2534	

REML = restricted maximum likelihood; ML = maximum likelihood.

Example 2. We borrow an artificial panel data, which include 60 observations of 15 panels, shown in Henningsen (2017) for the second example. The dependent variable y is modeled by 2 independent variables x_1 and x_2. Among the 60 values of y, 20 observations are left-censored at 0. The REML is generally distinguish from the ML when sample size is small; therefore, the data may be adequate for demonstrating the small sample property of two estimates. Table 2 shows the estimates and their standard errors.

It can be observed that the estimates are nearly the same for all slope parameters. For instance, relative distances between REML and ML estimates $(\hat{\theta}_{\text{ML}} - \hat{\theta}_{\text{REML}})/\text{std}(\hat{\theta}_{\text{ML}})$ are nearly equal to zero for all slop parameters. But two methods make a difference in the estimate of random components. The ML gives a slightly smaller estimate of variance components. Also, the standard errors of ML are smaller than the REML for all parameters. It is believed that the REML can remedy the underestimation of the ML in this small data since ML is known to underestimate variance components, which leads to the underestimated standard error.

4. Simulation study and conclusion

REML was distinguished from ML in the estimation of variance components and standard errors. When a sample size is small, an inference based on ML estimates may not be appropriate because of the inflated Type I error due to the underestimated standard error. It seems that REML is more adequate for the inference of a censored regression model since REML reports a larger standard error than ML. To see this, we have performed a simulation under the study design used in Example 2. That is, we have generated a dataset according to

$$y_{it} = \max\{\beta_0 + \beta_1 x_{1it} + \beta_2 x_{2it} + \mu_i + \epsilon_{it},\ 0\}, \quad \mu_i \overset{iid}{\sim} N\left(0, \sigma_\mu^2\right) \text{ and } \epsilon \overset{iid}{\sim} N\left(0, \sigma_\epsilon^2\right),$$

where y_{it} is the observation of the dependent variable. The regressors x_1 and x_2 are selected from a standard normal and a uniform distribution ranging 0 to 1, respectively. The parameters are given by $(\beta_0, \beta_1, \beta_2, \sigma_\mu^2, \sigma_\epsilon^2) = (-1, 2, 3, 1, 1)$.

The data generation process has been repeated 1,000 times. Some estimates of variance components may become negative. In particular, the ML method often estimates σ_μ^2 negatively. The negative estimates could be constrained to zero. We could only consider $T = 4$ and $n = 10, 15$ since the difference between two methods is statistically meaningful when sample size is small and the algorithm is computationally expensive. Based on 1,000 replications for each case, the average of estimates and the empirical mean squared error have been computed (Table 3). Also, a 95% Wald confidence interval for the slope parameter has been constructed for each replication, and then we have calculated the empirical coverage probability of the interval to measure the adequacy of the calculated standard error. If the standard error is asymptotically correct, then the coverage probability would be close to 0.95.

Table 3: Summary of simulation

n	θ	REML				ML			
		Aver. Est	Aver. std	MSE	Coverage	Aver. Est	Aver. std	MSE	Coverage
10	β_0	−1.0630	0.5933	0.3953	0.937	−1.0156	0.5604	0.3748	0.922
	β_1	2.0481	0.2918	0.0930	0.941	2.0225	0.2785	0.0891	0.936
	β_2	3.0525	0.7827	0.6436	0.937	3.0156	0.6763	0.6852	0.925
	σ_μ^2	1.0732	0.7353	0.5264		0.9133	0.6058	0.4005	
	σ_ϵ^2	1.0027	0.3657	0.1363		0.9258	0.3218	0.1194	
15	β_0	−1.0579	0.5354	0.3125	0.944	−1.0195	0.5131	0.3077	0.924
	β_1	2.0184	0.2216	0.0758	0.949	2.0122	0.2550	0.0538	0.940
	β_2	3.0500	0.7061	0.5191	0.940	3.0095	0.7608	0.5048	0.930
	σ_μ^2	1.0483	0.5884	0.3474		0.9506	0.5532	0.3316	
	σ_ϵ^2	0.9901	0.2982	0.0947		0.9408	0.2919	0.1008	

$(\beta_0,\beta_1,\beta_2,\sigma_\mu^2,\sigma_\epsilon^2) = (-1,2,3,1,1)$. REML = restricted maximum likelihood; ML = maximum likelihood; MSE = mean squared error.

Note that the coverage probabilities of ML are smaller than the nominal level for all sample sizes. In particular, when $n = 10$, we may say that the coverage probability is far below the nominal level. The coverage probabilities of REML are also below the nominal level, but REML gives coverage probabilities closer to 0.95 than ML. Thus, it may be concluded that the REML can provide a proper standard error of the estimate.

However, the ML tends to have smaller empirical mean squared error. Thus, ML is theoretically appearing when we wish to estimate the parameter itself; however, the inference, such as the interval estimation or the hypothesis testing on the slope parameter, based on ML may not be adequate when the sample size is small. Perhaps, this is well known in the general linear model.

We have considered a methodology under a censored random effect panel regression model, but the methodology can be applicable to a general mixed-effects linear model with limited dependent variables.

References

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

Arellano M and Bond S (1991). Some tests of specification for panel data: Monte Carlo evidence and an application to employment equations, *The Review of Economic Studies*, **58**, 227–297.

Arismendi JC (2013). Multivariate truncated moments, *Journal of Multivariate Analysis*, **117**, 41–75.

Croissant Y and Millo G (2008). Panel data econometrics in R: The plm package, *Journal of Statistical Software*, **27**, 1–43.

Drum ML and McCullagh P (1993). REML estimation with exact covariance in the logistic mixed model, *Biometrics*, **49**, 677–689.

Duan JC and Fulop A (2011). A stable estimator of the information matrix under EM for dependent data, *Statistics and Computing*, **21**, 83–91.

Eddelbuettel D, Francois R, Allaire J, Ushey K, Kou Q, Russell N, Bates D, and Chambers J (2018). *Seamless R and C++ Integration*. Available from: http://www.rcpp.org, http://dirk.eddelbuettel.com/code/rcpp.html

Eddelbuettel D and Sanderson C (2014). RcppArmadillo: accelerating R with high-performance C++ linear algebra, *Computational Statistics and Data Analysis*, **71**, 1054–1063.

Efron B and Hinkley DV (1978). The observed versus expected information, *Biometrika*, **65**, 457–487.

Green W (2004). Fixed effects and bias due to the incidental parameters problem in the Tobit model, *Econometric Review*, **23**, 125–147.

Henningsen A (2017). censReg: Censored Regression (Tobit) Models. R package version 0.5. Available 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.

Kan R and Robotti C (2017). On moments of folded and truncated multivariate normal distributions, Unpublished manuscript. Available from: https://sites.google.com/site/cesarerobotti/kr_JCGS.pdf

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

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

Lee L (2017) *Nondetects And Data Analysis for environmental data*. Available from: http://cran.r-project.org/ pack-age=NADA

Lee SC (2016). A Bayesian inference for fixed effects panel probit model, *Communications for statistical Applications and Methods*, **23**, 179–187.

Lesaffre E and Spiessens B (2001). On the effect of the number of quadrature points in a logistic random effects model: an example, *Journal of Royal Statistical Society, Applied Statistics, Series C*, **50**, 325–335.

Lee Y and Nelder JA (2001). Hierarchical generalized linear models: a synthesis of generalized linear models, random–effect model and structure dispersion, *Biometrika*, **88**, 987–1006.

Louis TA (1982). Finding the observed information matrix when using the EM algorithm, *Journal of the Royal Statistical Society, Series B*, **62**, 257–270.

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

Meilijson I (1989). A fast improvement to the EM algorithm on its own terms, *Journal of the Royal Statistical Society, Series B*, **51**, 127–138.

Meng XL and Rubin DB (1991). Using EM to obtain asymptotic variance-covariance matrices: the SEM algorithm, *Journal of the American Statistical Association*, **86**, 899–909.

McCulloch CE (1994). Maximum likelihood variance components estimation for binary data, *Journal of the American Statistical Association*, **89**, 330–335.

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

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*, **31**, 100–109.

R Core Team (2017). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. Available from: https://www.R-project.org/

SAS (2011). *SAS/ETS 9.3 User's Guide*.

Searle SR, Casella G, and McCulloch CE (2006). *Variance Components*, John Wiley & Sons, New York.

Stata (2017). *Finite Mixture Models Reference Manual*, Stata press.

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

Zhang H, Lu N, Feng C, Thurston SW, Xia Y, Zhu L, and Tu XM (2011). On fitting generalized linear mixed-effects models for binary responses using different statistical packages, *Statistics in Medicine*, **30**, 2562–2572.

Received January 2, 2019; Revised February 18, 2019; Accepted February 18, 2019