The Canadian Journal of Statistics
Vol. xx, No. yy, 2018, Pages 1–25
La revue canadienne de statistique

Consistent Estimator for Logistic Mixed Effect Models

Yizheng Wei[1]*, Yanyuan Ma[2], Tanya P. Garcia[3] and Samiran Sinha[4]

[1]*Department of Statistics, University of South Carolina, Columbia, SC 29208*
[2]*Department of Statistics, The Pennsylvania State University, University Park, PA 16802*
[3,4]*Department of Statistics, Texas A&M University, College Station, TX 77843*

Key words and phrases: Dependence; Distribution Free Random Slope; Logistic Mixed Effect Models; Semiparametric Models; .

MSC 2010: Primary 62J12; secondary 62G99

Abstract: We propose a consistent and locally efficient estimator to estimate the model parameters for a logistic mixed effect model with random slopes. Our approach relaxes two typical assumptions: the random effects being normally distributed, and the covariates and random effects being independent of each other. Adhering to these assumptions is particularly difficult in health studies where in many cases we have limited resources to design experiments and gather data in long-term studies, while new findings from other fields might emerge, suggesting the violation of such assumptions. So it is crucial if we could have an estimator robust to such violations and then we could make better use of current data harvested using various valuable resources. Our method generalizes the framework presented in Garcia & Ma (2016) which also deals with a logistic mixed effect model but only considers a random intercept. A simulation study reveals that our proposed estimator remains consistent even when the independence and normality assumptions are violated. This contrasts from the traditional maximum likelihood estimator which is likely to be inconsistent when there is dependence between the covariates and random effects. Application of this work to a Huntington disease study reveals that disease diagnosis can be further improved using assessments of cognitive performance. *The Canadian Journal of Statistics* xx: 1–25; 2018 © 2018 Statistical Society of Canada

Résumé: Insérer votre résumé ici. We will supply a French abstract for those authors who can't prepare it themselves. *La revue canadienne de statistique* xx: 1–25; 2018 © 2018 Société statistique du Canada

1. INTRODUCTION

A mixed effect logistic model is commonly used for analyzing clustered binary data arising in longitudinal studies of behavioral, social, health, and biomedical science. In the mixed effect logistic model, the logit of the success probability of the response is modeled as a linear function of fixed and random effect components. The observed data are $(Y_{ij}, \mathbf{X}_{ij}, \mathbf{Z}_{ij})$, $j = 1, \cdots, m_i$ and $i = 1, \cdots, n$, where Y_{ij} is the binary response variable, $\mathbf{X}_{ij}$ is a p-vector that exerts a fixed effect and $\mathbf{Z}_{ij}$ is a q-dimensional random variable that has a random effect $\mathbf{R}_i \in \mathcal{R}^q$. Here, i and j denote the index for clusters and the subject within a cluster, respectively. The random effect is completely unobserved and we assume $m_i > q$ for identifiability for all i. This identifiability requirement will become self-evident in Section 2. The mixed effect logistic model is

$$\text{pr}(Y_{ij} = 1|\mathbf{X}_{ij}, \mathbf{Z}_{ij}, \mathbf{R}_i) = \frac{\exp(\mathbf{X}_{ij}^{\mathrm{T}}\boldsymbol{\beta} + \mathbf{Z}_{ij}^{\mathrm{T}}\mathbf{R}_i)}{1 + \exp(\mathbf{X}_{ij}^{\mathrm{T}}\boldsymbol{\beta} + \mathbf{Z}_{ij}^{\mathrm{T}}\mathbf{R}_i)}, \; j = 1, \cdots, m, \; i = 1, \cdots, n, \quad (1)$$

* *Author to whom correspondence may be addressed.*
E-mail: ywei@email.sc.edu

and the main objective is to consistently estimate the p-dimensional regression coefficient $\boldsymbol{\beta}$ in the presence of the unobserved random effect.

The standard maximum likelihood approach estimates $\boldsymbol{\beta}$ assumes that $\mathbf{R}_i$ has a parametric distribution (e.g., multivariate normal with zero mean and positive definite variance-covariance matrix) and is independent of the covariates $\mathbf{X}_{ij}$ and $\mathbf{Z}_{ij}$. When the distribution for $\mathbf{R}_i$ is misspecified, however, the approach can yield biased parameter estimates and distorted type-I error rates (Heagerty & Kurland, 2001; Agresti, Ohman, & Caffo, 2004; Litiére, Alonso, & Molenberghs, 2007; Litiére, Alonso, & Molenberghs, 2008). The misspecification may occur in terms of mis-specifying the shape of the distribution, incorrectly assuming independence between the covariates and the random effect, or incorrectly assuming independence between the cluster size and the random effect. A good review on the potential bias due to misspecification of the distribution of $\mathbf{R}_i$ can be found in Neuhaus, McCulloch, & Boylan (2011).

More flexible models for the distribution of $\mathbf{R}_i$ have been considered to circumvent the misspecification bias, but under limited settings. For linear mixed models, Zhang & Davidian (2001) proposed a smooth semi-nonparametric probability density for random effect and Zhang et al. (2008) proposed a negatively skewed random effect density. However, extending either method to generalized linear models is non-trivial, and imposing smoothness constraints or a skewness condition introduces computational complexities that we can actually avoid.

In this paper, we propose estimating parameters in the mixed effect logistic model without imposing any distributional assumptions on the random effect. Taking a semiparametric approach, we treat the distribution of $\mathbf{R}_i$ as a nuisance parameter and demonstrate that consistent estimates of $\boldsymbol{\beta}$ are obtained regardless of how the distribution of $\mathbf{R}_i$ is specified. We thus avoid unnecessary assumptions, such as a particular distributional shape for $\mathbf{R}_i$ (Zhang & Davidian, 2001; Zhang et al., 2008) and the independence between covariates and the random effect $\mathbf{R}_i$. Our method generalizes the framework presented by Garcia & Ma (2016) which also deals with a logistic mixed effect model but only considered a random intercept. The presence of the random slope terms in our model means that their method no longer applies. Extending the result from random intercept to random slope is not as straightforward as it seems.

The rest of this paper is organized as follows. In Section 2, we develop semiparametric efficient estimator for $\boldsymbol{\beta}$. We demonstrate that the proposed estimator is consistent regardless of the assumed model for the distribution of $\mathbf{R}_i$, and the estimator achieves the asymptotic efficiency when the distribution for $\mathbf{R}_i$ is correctly modeled. In Section 3, we demonstrate through extensive simulation studies that the proposed estimator is robust to different distributional assumptions of $\mathbf{R}_i$, including different distributional shapes and dependence structures with covariates. The robustness property of the new estimator contrasts to the large biases of the maximum likelihood estimator when the distribution of $\mathbf{R}_i$ is misspecified. In Section 4, we apply our method to analyze a dataset from a study of Huntington disease and discover that the maximum likelihood estimator may result in misleading results about the importance of cognitive measures in relationship to diagnosis of Huntington disease. In contrast, our method detects one more cognitive measure crucial in determining the diagnostic result of Huntington disease. The paper ends with a brief discussion in Section 5. All technical details are given in an Appendix.

2. MAIN RESULTS

2.1. Notation and assumptions

Let $\mathbf{Y}_i = (Y_{i1}, \dots, Y_{im})^{\mathrm{T}}$ denote a m-dimensional vector, $\mathbf{X}_i = (\mathbf{X}_{i1}, \dots, \mathbf{X}_{im})$ denote a $p \times m$ matrix, and $\mathbf{Z}_i = (\mathbf{Z}_{i1}, \dots, \mathbf{Z}_{im})$ denote a $q \times m$ matrix. Without loss of generality, assume that the first q columns of $\mathbf{Z}_i$ form an invertible matrix. For notational simplicity and ease of presentation, we used a common m. We could also change m to m_i to account for different cluster sizes.

Let f to denote various densities described by the subindices. The likelihood for the ith cluster formed by the model in Equation (1) is

$$f_{\mathbf{Y},\mathbf{X},\mathbf{Z}}(\mathbf{y}_i, \mathbf{x}_i, \mathbf{z}_i; \boldsymbol{\beta}) = \int f_{\mathbf{Y}|\mathbf{R},\mathbf{X},\mathbf{Z}}(\mathbf{y}_i \mid \mathbf{r}_i, \mathbf{x}_i, \mathbf{z}_i, \beta) f_{\mathbf{R},\mathbf{X},\mathbf{Z}}(\mathbf{r}_i, \mathbf{x}_i, \mathbf{z}_i) d\mu(\mathbf{r}_i)$$

$$= \int \prod_{j=1}^{m} \exp[y_{ij}(\mathbf{x}_{ij}^{\mathrm{T}}\boldsymbol{\beta} + \mathbf{r}_i^{\mathrm{T}}\mathbf{z}_{ij}) - \log\{1 + \exp(\mathbf{x}_{ij}^{\mathrm{T}}\boldsymbol{\beta} + \mathbf{r}_i^{\mathrm{T}}\mathbf{z}_{ij}\}] f_{\mathbf{R},\mathbf{X},\mathbf{Z}}(\mathbf{r}_i, \mathbf{x}_i, \mathbf{z}_i) d\mu(\mathbf{r}_i),$$

where $\mu(\cdot)$ denotes the dominating measure. Throughout, we let $f_{\mathbf{R},\mathbf{X},\mathbf{Z}}(\mathbf{r}_i, \mathbf{x}_i, \mathbf{z}_i)$ be completely unspecified. To estimate $\boldsymbol{\beta}$ without needing to specify this distribution, we take a semiparametric approach as described next.

2.2. Consistent and efficient estimator

Our approach is rooted in treating $f_{\mathbf{R},\mathbf{X},\mathbf{Z}}(\mathbf{r}, \mathbf{x}, \mathbf{z})$ as an infinite dimensional nuisance parameter and using semiparametric techniques to estimate $\boldsymbol{\beta}$ (Tsiatis, 2006). The approach involves first deriving the space spanned by this infinite-dimensional nuisance parameter. This space, referred to as the nuisance tangent space, and its orthogonal complement are derived in a similar way as Section S1 of Garcia & Ma (2016). The orthogonal complement of the nuisance tangent space serves as an intermediate calculation for the estimator of interest. Specifically, the efficient score function for $\boldsymbol{\beta}$, denoted $\mathbf{S}_{\text{eff}}$, is obtained by projecting the score function with respect to $\boldsymbol{\beta}$,

$$\begin{aligned}\mathbf{S}_{\boldsymbol{\beta}}(\mathbf{Y}, \mathbf{X}, \mathbf{Z}) &\equiv \frac{\partial}{\partial \boldsymbol{\beta}} \log\{f_{\mathbf{Y},\mathbf{X},\mathbf{Z}}(\mathbf{Y}, \mathbf{X}, \mathbf{Z}, \boldsymbol{\beta})\} \\ &= E\left[\frac{\partial}{\partial \boldsymbol{\beta}} \log\{f_{\mathbf{Y}|\mathbf{R},\mathbf{X},\mathbf{Z}}(\mathbf{Y} \mid \mathbf{R}, \mathbf{X}, \mathbf{Z}, \boldsymbol{\beta})\} \mid \mathbf{Y}, \mathbf{X}, \mathbf{Z}\right], \end{aligned} \tag{2}$$

onto the orthogonal complement of the nuisance tangent space. That is,

$$\mathbf{S}_{\text{eff}}(\mathbf{Y}, \mathbf{X}, \mathbf{Z}, \boldsymbol{\beta}) = \mathbf{S}_{\boldsymbol{\beta}}(\mathbf{Y}, \mathbf{X}, \mathbf{Z}) - E\{\mathbf{h}(\mathbf{R}, \mathbf{X}, \mathbf{Z}) \mid \mathbf{Y}, \mathbf{X}, \mathbf{Z}\},$$

where $\mathbf{h}$ is a p-dimensional function that satisfies

$$E\{\mathbf{S}_{\boldsymbol{\beta}}(\mathbf{Y}, \mathbf{X}, \mathbf{Z}) \mid \mathbf{R}, \mathbf{X}, \mathbf{Z}\} = E[E\{\mathbf{h}(\mathbf{R}, \mathbf{X}, \mathbf{Z}) \mid \mathbf{Y}, \mathbf{X}, \mathbf{Z}\} \mid \mathbf{R}, \mathbf{X}, \mathbf{Z}].$$

To allow exchanging integration and differentiation in Equation (2), we assume that both $f_{\mathbf{Y}|\mathbf{R},\mathbf{X},\mathbf{Z}}(\mathbf{Y} \mid \mathbf{R}, \mathbf{X}, \mathbf{Z}, \boldsymbol{\beta})$ and its partial derivative $\partial f_{\mathbf{Y}|\mathbf{R},\mathbf{X},\mathbf{Z}}(\mathbf{Y} \mid \mathbf{R}, \mathbf{X}, \mathbf{Z}, \boldsymbol{\beta})/\partial \boldsymbol{\beta}$ are continuous functions of $\boldsymbol{\beta}$ and $\mathbf{R}$. The practical implementation of the procedure described above is however infeasible, because we are unable to perform the above computation without the true distribution form of the random effect. To this end, we adopt a working model for $f_{\mathbf{R}|\mathbf{X},\mathbf{Z}}$, denoted $f^*_{\mathbf{R}|\mathbf{X},\mathbf{Z}}$, and perform the above calculation under such a working model. We provide the detailed expressions below, with all the affected quantities marked with *. Under such a working model, the score function with respect to $\boldsymbol{\beta}$ is

$$\mathbf{S}^*_{\boldsymbol{\beta}}(\mathbf{Y}, \mathbf{X}, \mathbf{Z}) = E^*\left[\frac{\partial}{\partial \boldsymbol{\beta}} \log\{f_{\mathbf{Y}|\mathbf{R},\mathbf{X},\mathbf{Z}}(\mathbf{Y} \mid \mathbf{R}, \mathbf{X}, \mathbf{Z}, \boldsymbol{\beta})\} \mid \mathbf{Y}, \mathbf{X}, \mathbf{Z}\right], \tag{3}$$

and the locally efficient score function is

$$\mathbf{S}^*_{\text{eff}}(\mathbf{Y}, \mathbf{X}, \mathbf{Z}, \boldsymbol{\beta}) = \mathbf{S}^*_{\boldsymbol{\beta}}(\mathbf{Y}, \mathbf{X}, \mathbf{Z}) - E^*\{\mathbf{h}^*(\mathbf{R}, \mathbf{X}, \mathbf{Z}) \mid \mathbf{Y}, \mathbf{X}, \mathbf{Z}\},$$

Here, the "locally efficient score" means a function containing a working model in it. When a misspecified working model is used, the function has mean zero, and when a correct working model is used, the function is identical to the efficient score function, i.e. $\mathbf{S}_{\text{eff}}(Y, \mathbf{X}, \mathbf{Z}, \boldsymbol{\beta})$. An estimator based on solving the estimating equation formed by the locally efficient score function is named a locally efficient estimator. A locally efficient estimator subsequently has the property that if a misspecified working model is used, the estimator is consistent. When a correct working model is used, the estimator is efficient. Further, $\mathbf{h}^*$ is a p-dimensional function that satisfies

$$E\{\mathbf{S}_{\boldsymbol{\beta}}^*(\mathbf{Y}, \mathbf{X}, \mathbf{Z}) \mid \mathbf{R}, \mathbf{X}, \mathbf{Z}\} = E[E^*\{\mathbf{h}^*(\mathbf{R}, \mathbf{X}, \mathbf{Z}) \mid \mathbf{Y}, \mathbf{X}, \mathbf{Z}\} \mid \mathbf{R}, \mathbf{X}, \mathbf{Z}]. \tag{4}$$

An estimator of $\boldsymbol{\beta}$ is then obtained from solving the estimating equation

$$\sum_{i=1}^{n} \mathbf{S}_{\text{eff}}^*(\mathbf{y}_i, \mathbf{x}_i, \mathbf{z}_i, \boldsymbol{\beta}) = \mathbf{0}. \tag{5}$$

Using a working model $f^*_{\mathbf{R}|\mathbf{X},\mathbf{Z}}$ to replace the true form of $f_{\mathbf{R}|\mathbf{X},\mathbf{Z}}$ enables us to proceed with the computation. Of course, there is a cost involved with such a replacement. Fortunately, the cost is only in terms of estimation efficiency. The replacement does not affect the consistency of the resulting estimator.

Theorem 1. *The estimator $\widehat{\boldsymbol{\beta}}$ solving Equation (5) satisfies*

$$\sqrt{n}(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0) \to N\{0, \mathbf{A}^{-1}\mathbf{B}(\mathbf{A}^{-1})^{\mathrm{T}}\}$$

in distribution when $n \to \infty$. Here, $\boldsymbol{\beta}_0$ is the true value of the parameter $\boldsymbol{\beta}$, $\mathbf{A} = E\left\{\partial \mathbf{S}_{\text{eff}}^*(\mathbf{Y}, \mathbf{X}, \mathbf{Z}, \boldsymbol{\beta}_0) \ / \partial \boldsymbol{\beta}^{\mathrm{T}}\right\}$ and $\mathbf{B} = var\{\mathbf{S}_{\text{eff}}^*(\mathbf{Y}, \mathbf{X}, \mathbf{Z}, \boldsymbol{\beta}_0)\} = E\{\mathbf{S}_{\text{eff}}^*(\mathbf{Y}, \mathbf{X}, \mathbf{Z}, \boldsymbol{\beta}_0)^{\otimes 2}\}$. Additionally, if the true $f_{\mathbf{R}|\mathbf{X},\mathbf{Z}}$ is used in constructing the estimator, the resulting estimator $\widehat{\boldsymbol{\beta}}$ achieves the optimal estimation efficiency bound.

The proof of Theorem 1 is in the Appendix. Theorem 1 implies that we are free to choose the form of $f^*_{\mathbf{R}|\mathbf{X},\mathbf{Z}}$ without incurring penalties on consistency or distorted type I error rates as in (Heagerty & Kurland, 2001; Agresti, Ohman, & Caffo, 2004; Litiére, Alonso, & Molenberghs, 2007; Litiére, Alonso, & Molenberghs, 2008). If $f^*_{\mathbf{R}|\mathbf{X},\mathbf{Z}}$ happens to be the true model, then the estimator for $\boldsymbol{\beta}$ achieves the optimal efficiency bound. For computational simplicity, we therefore choose the posited model of $f^*_{\mathbf{R}|\mathbf{X},\mathbf{Z}}$ as a standard normal distribution. In the simulation results given in Tables 5, 6, 7, 8, we show that even when the true dsitribution $f_{\mathbf{R}|\mathbf{X},\mathbf{Z}}$ is not from a standard normal, our estimator for $\boldsymbol{\beta}$ is still consistent. Regarding the estimation of the covariance matrix of our estimator, we estimate the derivative $\partial \mathbf{S}_{\text{eff}}^*(\mathbf{Y}, \mathbf{X}, \mathbf{Z}, \boldsymbol{\beta}_0)/\partial \boldsymbol{\beta}^{\mathrm{T}}$ through numerical difference, and approximate the expectations via sample average.

A computational challenge in forming the estimating equation is solving Equation (4) for $\mathbf{h}^*$ as it is an ill-posed integral equation. However, as demonstrated next, a simple transformation of the response variable $\mathbf{Y}_{ij}$ and covariate $\mathbf{Z}_{ij}$ allows us to avoid solving this ill-posed problem.

2.3. Simplification of estimating equations

To circumvent the ill-posed problem in Equation (4), we transform the response variable $\mathbf{Y}_i$ and covariate $\mathbf{Z}_{ij}$ such that the transformed variables satisfy properties similar to the classical sufficiency and completeness.

Let $\mathbf{W}_i = \sum_{j=1}^{m} Y_{ij}\mathbf{Z}_{ij}$, $\mathbf{U}_i = (Y_{i(q+1)}, \ldots, Y_{im})^{\mathrm{T}}$. Write $\mathbf{Z}_i = (\mathbf{Z}_{i1}, \cdots, \mathbf{Z}_{im}) = (\mathbf{Z}_{iL}, \mathbf{Z}_{iR})$, where $\mathbf{Z}_{iL} \in \mathcal{R}^{q\times q}$ and $\mathbf{Z}_{iR} \in \mathcal{R}^{q\times(m-q)}$. That is, $\mathbf{Z}_{iL}$ is the left $q \times q$ submatrix of $\mathbf{Z}_i$ and $\mathbf{Z}_{iR}$ is the right $q \times (m-q)$ submatrix of $\mathbf{Z}_i$. Let

$$\mathbf{M}_i = \begin{pmatrix} \mathbf{Z}_{iL} & \mathbf{Z}_{iR} \\ \mathbf{0}_{(m-q)\times q} & \mathbf{I}_{(m-q)\times(m-q)} \end{pmatrix}, \quad \mathbf{M}_i^{-1} = \begin{pmatrix} \mathbf{Z}_{iL}^{-1} & -\mathbf{Z}_{iL}^{-1}\mathbf{Z}_{iR} \\ \mathbf{0}_{(m-q)\times q} & \mathbf{I}_{(m-q)\times(m-q)} \end{pmatrix}.$$

Under this notation, we transform $\mathbf{Y}_i$ as

$$\mathbf{Y}_i = \mathbf{M}_i^{-1} \begin{pmatrix} \mathbf{W}_i \\ \mathbf{U}_i \end{pmatrix}.$$

The matrix $\mathbf{M}_i$ is invertible because we assumed that the first q columns of $\mathbf{Z}_i$ form an invertible matrix. The one-to-one mapping from $(\mathbf{Z}_i,\mathbf{Y}_i)$ to $(\mathbf{W}_i,\mathbf{U}_i)$ allows us to take advantage of certain sufficiency and completeness properties of $\mathbf{W}_i$ and $\mathbf{U}_i$ as described in Theorem 2.

Theorem 2. *The variables $\mathbf{W}_i$ and $\mathbf{U}_i$ satisfy the following two properties:*

(a) Sufficiency of $\mathbf{W}$:
$f_{\mathbf{U}|\mathbf{W},\mathbf{R},\mathbf{X},\mathbf{Z}}(\mathbf{u} \mid \mathbf{w},\mathbf{r},\mathbf{x},\mathbf{z}) = f_{\mathbf{U}|\mathbf{W},\mathbf{X},\mathbf{Z}}(\mathbf{u} \mid \mathbf{w},\mathbf{x},\mathbf{z}) = f_{\mathbf{U}|\mathbf{X},\mathbf{Z}}(\mathbf{u} \mid \mathbf{x},\mathbf{z})$,
$f_{\mathbf{R}|\mathbf{U},\mathbf{W},\mathbf{X},\mathbf{Z}}(\mathbf{r} \mid \mathbf{u},\mathbf{w},\mathbf{x},\mathbf{z}) = f_{\mathbf{R}|\mathbf{W},\mathbf{X},\mathbf{Z}}(\mathbf{r} \mid \mathbf{w},\mathbf{x},\mathbf{z})$.

(b) Completeness of $\mathbf{W}$:
For any function $\mathbf{a}(\mathbf{w},\mathbf{x},\mathbf{z})$, if $E\{\mathbf{a}(\mathbf{W},\mathbf{X},\mathbf{Z}) \mid \mathbf{R},\mathbf{X},\mathbf{Z}\} = \mathbf{0}$, then $\mathbf{a}(\mathbf{W},\mathbf{X},\mathbf{Z}) = \mathbf{0}$.

The proof of Theorem 2 is in Appendix. The sufficiency and completeness properties in Theorem 2 allow us to form a statistic free of the random slope associated with $\mathbf{Z}$ and remove the component containing the random slope from the estimating equation. Indeed Theorem 2 (a) yields that $E\{\mathbf{h}^*(\mathbf{R},\mathbf{X},\mathbf{Z}) \mid \mathbf{Y},\mathbf{X},\mathbf{Z}\}$ in Equation (4) is actually equal to $E\{\mathbf{h}^*(\mathbf{R},\mathbf{X},\mathbf{Z}) \mid \mathbf{W},\mathbf{X},\mathbf{Z}\}$. The advantage of this equality is that the conditional expectation of $h^*(\mathbf{R},\mathbf{X},\mathbf{Z})$ given $(\mathbf{W},\mathbf{X},\mathbf{Z})$ satisfies

$$E^*\{\mathbf{h}^*(\mathbf{R},\mathbf{X},\mathbf{Z}) \mid \mathbf{W},\mathbf{X},\mathbf{Z}\} = E\{\mathbf{S}^*_{\boldsymbol{\beta}}(\mathbf{Y},\mathbf{X},\mathbf{Z}) \mid \mathbf{W},\mathbf{X},\mathbf{Z}\}$$

and $E\{\mathbf{S}^*_{\boldsymbol{\beta}}(\mathbf{Y},\mathbf{X},\mathbf{Z}) \mid \mathbf{W},\mathbf{X},\mathbf{Z}\}$ has a closed form, given by

$$\frac{\sum_{\mathbf{u}} \mathbf{S}^*_{\boldsymbol{\beta}}\{\mathbf{M}^{-1}(\mathbf{W}^{\mathrm{T}},\mathbf{u}^{\mathrm{T}})^{\mathrm{T}},\mathbf{X},\mathbf{Z}\} \exp\{(\mathbf{X}_1^{\mathrm{T}}\boldsymbol{\beta},\ldots,\mathbf{X}_q^{\mathrm{T}}\boldsymbol{\beta})(-\mathbf{Z}_L^{-1}\mathbf{Z}_R\mathbf{u}) + (\mathbf{X}_{(q+1)}^{\mathrm{T}}\boldsymbol{\beta},\ldots,\mathbf{X}_m^{\mathrm{T}}\boldsymbol{\beta})\mathbf{u}\}}{\sum_{\mathbf{u}} \exp\{(\mathbf{X}_1^{\mathrm{T}}\boldsymbol{\beta},\ldots,\mathbf{X}_q^{\mathrm{T}}\boldsymbol{\beta})(-\mathbf{Z}_L^{-1}\mathbf{Z}_R\mathbf{u}) + (\mathbf{X}_{(q+1)}^{\mathrm{T}}\boldsymbol{\beta},\ldots,\mathbf{X}_m^{\mathrm{T}}\boldsymbol{\beta})\mathbf{u}\}} \quad (6)$$

where the summation $\sum_{\mathbf{u}}$ is over all possible $\mathbf{u} \in \mathcal{R}^{m-p}$ such that each entry in $\mathbf{u}$ is either 0 or 1, and $\mathbf{u}$ satisfies that $\mathbf{W}_i = Z_{iL}\mathbf{Y}_{i1} + \mathbf{Z}_{iR}\mathbf{u}$. Here $\mathbf{Y}_{i1}$ is the subvector of $\mathbf{Y}_i$ formed by the first q elements.
Therefore, the estimating equation (5) which originally involved solving an ill-posed problem is now of the form

$$\sum_{i=1}^{n}\sum_{i=1}^{n}[\mathbf{S}^*_{\boldsymbol{\beta}}(\mathbf{Y}_i,\mathbf{X}_i,\mathbf{Z}_i) - E\{\mathbf{S}^*_{\boldsymbol{\beta}}(\mathbf{Y}_i,\mathbf{X}_i,\mathbf{Z}_i) \mid \mathbf{W}_i,\mathbf{X}_i,\mathbf{Z}_i\}] = \mathbf{0}. \quad (7)$$

All terms in the estimating equation can be explicitly constructed without needing to solve an ill-posed problem. The construction of $\mathbf{S}^*_{\boldsymbol{\beta}}(\mathbf{Y}_i,\mathbf{X}_i,\mathbf{Z}_i)$ does require specifying a proposed model $f^*_{\mathbf{R}|\mathbf{X},\mathbf{Z}}$, but by Theorem 1, the model does not need to be correctly specified to ensure consistency. Therefore, we have constructed a simple estimation method that does not impose stringent assumptions on the unknown random effect, nor does it involve heavy computation. All terms in the new estimating equation are easy to compute with the most difficult part being $E\{\mathbf{S}^*_{\boldsymbol\beta}(\mathbf{Y}_i, \mathbf{X}_i, \mathbf{Z}_i) \mid \mathbf{W}_i, \mathbf{X}_i, \mathbf{Z}_i\}$, which we will use the Gaussian quadrature to deal with.

In summary, our algorithm for computing $\widehat{\boldsymbol\beta}$ involves:

Step 1. Specify a working model for $f^*_{\mathbf{R}|\mathbf{X},\mathbf{Z}}$. For convenience, we suggest to model $f^*_{\mathbf{R}|\mathbf{X},\mathbf{Z}}$ using a normal distribution.

Step 2. Compute the function $\mathbf{S}^*_{\beta}(\mathbf{Y}, \mathbf{X}, \mathbf{Z})$, as in Equation (3) where the expectations are computed under $f^*_{\mathbf{R}|\mathbf{X},\mathbf{Z}}$ from Step 1.

Step 3. Compute $E\{\mathbf{S}^*_{\boldsymbol\beta}(\mathbf{Y}_i, \mathbf{X}_i, \mathbf{Z}_i) \mid \mathbf{W}_i, \mathbf{X}_i, \mathbf{Z}_i\}$ using Equation (6).

Step 4. Solve the estimating equation (7) to obtain $\widehat{\boldsymbol\beta}$.

3. SIMULATION STUDY

3.1. Design of Simulation

We compared the performance of our estimator to the traditional normal-based maximum likelihood estimator (MLE). We used the *glmer* in R package *lme4* (Bates, Maechler, & Bolker, 2016) to compute the maximum likelihood estimator. The assumption of the MLE is that the random effect is normally distributed, and that covariates and the random effects are independent. In comparison, our estimator does not assume that the random effect follows a specific distributional form, nor do we require independence between the covariates and the random effect. In this simulation study, we assess the sensitivity of our estimator and the MLE when these assumptions do not hold.

We generated 1000 data sets from the logistic random slope model in (1) with each data set having a sample size $n = 500$. We considered $m_i = 3$ covariates. We set the true parameter as $\boldsymbol\beta = (0.35, 0.6, -0.4)^{\mathrm{T}}$. To assess the distributional assumption of the random effect, we generated data according to four different distributions:

1. Standard Normal random effect. R_i is from a standard normal distribution.
2. Mixed Normal random effect: R_i is from a mixture of normal distribution with 80% of the data from Normal(3,1), and 20% of the data from Normal(6,1.5).
3. Gamma random effect. R_i is from a Gamma distribution with shape parameter 1 and scale parameter 1.25.
4. Student-t random effect. R_i is from a student-t distribution with degree of freedom 3.

Thus, the distributional shapes of the random effect include the standard bell-shaped form, bimodality, heavy tailness and skewness. The deviations from the standard bell-shaped form will allow us to assess how well our estimator performs in comparison to the MLE which assumes the random effect is indeed standard normal.

Under each of these four distributional assumptions for the random effect, we generated three different sets of covariates, first assuming their independence from the random effect:

1. Z_{ij} is from the Bernoulli distribution with success probability 0.5, and X_{ij} is from Normal(0.5,1);
2. Z_{ij} is from the Poisson distribution with parameter 0.5, and X_{ij} is from Normal(0.5,1);
3. Z_{ij} is from the Geometric distribution with success probability 0.7, and X_{ij} is from Normal(0.5,1).

Therefore, in total we considered 12 different cases: four ways of generating R_i's in combination with three ways of generating the covariates.

We further considered 12 additional cases similar to the above except that we introduced dependency between the random effect and covariates. This is aimed to assess deviations from the second assumption of the MLE in which the random effects and covariates are assumed to be independent. In the dependency case, we generated X_{ij} from Normal$(0.5R_i, 1)$ to achieve the dependency between R_i and X_{ij}'s. The generation of Z_{ij}'s were the same as before.

In summary, these settings were designed to investigate the performance of both the semiparametric estimator and the MLE when the random effect distribution is mis-specified, in combination with different covariate combinations of X and Z. For all data generation settings, we centered the generated random slopes to have zero mean to accommodate the standard normal-based MLE. In the proposed method, for all dependent and independent cases, we assumed the random effect is Normal(0,1) distribution and is independent of all the covariates. This is of course not a valid assumption in all the settings considered above.

3.2. Simulation Results

We compared the performance of the semiparametric estimator and MLE in terms of their bias, sample variance, estimated variance, and 95% coverage probabilities. The results of the independent cases are given in Tables 1 to 4 and those for the dependent cases are given in Tables 5 to 8.

Tables 1 to 4 show that when covariates and random effects are independent, the semiparametric estimator has comparable performance to that of the MLE in terms of bias and the 95% coverage probabilities meeting the nominal level. While we expected the semiparametric estimator to be consistent based on Theorem 1, we were initially surprised by the robustness of MLE to deviations from normality. However, Neuhaus, Hauck, & Kalbfleisch (1992) demonstrated that the MLE actually performs quite well for mixed effect models when the random effect is not normally distributed. In terms of estimation variability, the semiparametric estimator has somewhat larger variability compared to the normal-based MLE, although the difference in variabilities is small. This is also within our expectation since MLE adopts stronger modeling assumptions and must have smaller estimation variability.

The results in Tables 5 to 8 indicate a different phenomenon. In the case when the covariates and random effect are dependent, inconsistency of the MLE starts to manifest. Specifically, the biases of the estimates from the normal-based MLE are sufficiently large, and they cause the coverage of the 95% confidence intervals to be completely off from the nominal level. In contrast, the biases of estimates from our proposed estimator is still very small, and the coverage probability of 95% confidence intervals remain close to their nominal level. This clearly demonstrates that if we treat the random effect as independent from the covariates while in fact there is dependency between the two, the normal-based MLE loses its robustness and gives severely biased estimates with very small variability. Subsequently, inference based on MLE will be misleading. On the contrary, the semiparametric estimator continues to provide consistent estimation and valid inference results.

Summarizing the observations, the semiparametric estimator is a much more reliable method unless it is clear that the random effect and the covariates are independent of each other. Because the random effect is not observable, it is often difficult to determine its relation with the covariates. Thus, we recommend implementing the semiparametric estimator in general.

We also record the execution time of running 50 simulations using our estimator under one setting noted in Table 9. The CPU for this simulation is Intel I7-8700k@4.4GHz and the size of RAM is 32GB. From Table 9, the execution time increases as the cluster size increases or the number of parameters increases. These values show that the computation is generally sufficiently fast and we can use a single-thread R to run the entire simulation without engaging super computers with thousands of threads.

We also considered small sample performance, such as sample size $n = 25$ or $n = 50$. The algorithm does not converge in such sample sizes. As we gradually increase the sample size, we start to have some sensible results when the sample size $n = 220$. We report simulation results for $n = 220$ in Tables S.1 to S.8 in the supplement. Overall, the general conclusion based on $n = 220$ is the same as based on $n = 500$. The computing code is placed in the Supplementary Materials .

4. ANALYSIS OF A HUNTINGTON DISEASE STUDY

Huntington disease (HD) is a rare neurodegenerative disease linked to deterioration of the central nervous system. Its symptoms include unwanted choreatic movements, behavioral and psychiatric disturbances and dementia (Ross, 2010). The Cooperative Huntington Observational Research Trial (COHORT) was a large observational, longitudinal study of HD conducted from 2005 to 2011 that evaluated different cognitive and motor impairments associated with HD. The study included n=3211 participants who were annually evaluated over a four year time span. We focused on those subjects who had at least 4 consecutive visits during this study. Our main objective in analyzing COHORT is to investigate if cognitive measures are important determining the possibility of occurrence of HD. This objective stems from recent results that a major sign of HD is cognitive decline, and such decline can be observed long before motor symptoms first appear (Ross, 2010).

To assess the association between cognitive measures and occurrence of HD, we modeled the data using the mixed effect logistic model in (1). For each person $i = 1, \ldots, n$, and visit $j = 1, \ldots, m$ with $m = 6$, we set the response variable Y_{ij} as 1 if the person was diagnosed with HD, and 0 otherwise. Diagnosis of HD occurs when the participant's extrapyramidal signs are unequivocally associated with HD and the diagnosis is determined by a trained clinician. We set Z_i to be the gender for subject i. We set $\mathbf{X}_{ij}$'s to be a set of four different motor and cognitive measures. Specifically, we set X_{1ij} to be the total motor score (TMS), defined as the sum of total motor impairments as evaluated using the Unified Huntington Disease Rating Scale (Huntington Study Group, 1996). We set X_{2ij} to be the score from the Symbolic Digit Modality Test (SDMT), a test that assesses the cognitive impairment by some simple substitution tasks, such as visual scanning, attention, and motor speed. We set X_{3ij} to be the stroop color score (SCOLOR), a test that assesses the cognitive impairment by recording how many X's printed in blue, red, or green ink that a subject correctly verbally stated its color in a certain amount of time. We set X_{4ij} to be the stroop word score (SWORD), a test that assesses the cognitive impairment by recording the number of color words (blue, red, green) printed in black ink that a subject correctly verbally reads in a certain amount of time. We set X_{5ij} to be the stroop interference score (SINTER), a test that assesses the cognitive impairment by recording how many color words that were printed in colored ink (eg. BLUE printed in green ink or BLUE printed in blue ink) and correctly verbally read by a subject in a certain amount of time. Lastly, we set R_i to be the random slope associated with Z_i.

We applied our proposed estimator and the standard-normal MLE to assess the association between cognitive impairments and occurrence of HD. We suspect that the cognitive covariates and random effect are dependent based on clinical results from Downing et al. (2008). They found gender differences in cognitive function. Females tended to outperform males on tests of memorization and language skills. Males tended to outperform females on tasks involving mathematical reasoning and visuospatial ability. These results suggest that if we assess the impact of cognitive measures on HD occurrence, we may have that cognitive measures and the random effect are dependent through gender. This would imply that the MLE could yield misleading results because it assumes independence, whereas our estimator does not.

We performed our analysis in two steps. In the first step, we analyzed three subsets of the data: the 1404 subjects who had four clinical visits, the 775 subjects who had five visits, and the 132 subjects who had six visits. In each of the three sub-data sets, we implemented the semiparametric estimator to obtain estimators $\widehat{\boldsymbol\beta}_1$, $\widehat{\boldsymbol\beta}_2$, $\widehat{\boldsymbol\beta}_3$, where $\widehat{\boldsymbol\beta}_1$=($\widehat{\beta}_{1tms}$, $\widehat{\beta}_{1sdmt}$, $\widehat{\beta}_{1scolor}$, $\widehat{\beta}_{1sword}$, $\widehat{\beta}_{1sinter}$), $\widehat{\boldsymbol\beta}_2$=($\widehat{\beta}_{2tms}$, $\widehat{\beta}_{2sdmt}$, $\widehat{\beta}_{2scolor}$, $\widehat{\beta}_{2sword}$, $\widehat{\beta}_{2sinter}$), $\widehat{\boldsymbol\beta}_3$=($\widehat{\beta}_{3tms}$, $\widehat{\beta}_{3sdmt}$, $\widehat{\beta}_{3scolor}$, $\widehat{\beta}_{3sword}$, $\widehat{\beta}_{4sinter}$). We then perform the second step by taking a weighted average of the results, i.e. we set $\widehat{\boldsymbol\beta}$=($\widehat{\beta}_{tms}$, $\widehat{\beta}_{sdmt}$, $\widehat{\beta}_{scolor}$, $\widehat{\beta}_{sword}$, $\widehat{\beta}_{sinter}$).

The weighted average is denoted as $\widehat{\boldsymbol\beta} = \sum_{i=1}^{3} \mathbf{w}_i \widehat{\boldsymbol\beta}_i$, where the weights are proportional to the inverse of the variances of $\widehat{\boldsymbol\beta}_i$. That is, $\mathbf{w}_i$ is a diagonal matrix, with its jth element $w_{ij} = v_{ij}^{-1}/(\sum_{i=1}^{3} v_{ij}^{-1})$, where $v_{ij} = \text{var}(\widehat{\beta}_{ij})$. The variance of the final estimator is $\text{var}(\widehat{\beta}_j) = (\sum_{i=1}^{3} v_{ij}^{-1})^{-1}$. For comparison, we also implemented the normal-based MLE in the similar fashion.

Table 10 shows the results from both estimators. The semiparametric estimator indicates that cognitive scores from SDMT, SCOLOR, SWORD are not statistically significant, as their 95% confidence intervals contain zero. On the other hand, it detects TMS and SINTER to be significant covariates, both positively associated with the probability of developing HD. However, MLE indicates that only TMS score is statistically significant, while all the other four covariates are not statistically significant. The difference from the two analysis indicates that there is dependence between the random slope and the covariates. Based on both the theoretical results and the simulation experience, we believe the results from MLE can be misleading.

This result implies that if we adopt MLE on the data set to determine which covariates are needed for diagnosis of HD, we might neglect a vital covariate stroop interference score. The importance of stroop interference score coincided with clinical findings in Paulsen et al. (2013), where they found that prodromal HD patients have declined response shifting, and inhibition depends on efficient response shifting, while inhibition is necessary for stroop interference test. Based on these observations and our analysis results, we recommend using TMS and SINTER jointly to determine the occurrence of HD.

5. DISCUSSION

We proposed a locally efficient estimator using a semiparametric approach in a mixed effect logistic model with random slope. Locally efficient means even when we use a misspecified working model, the resulting estimator of $\boldsymbol\beta$ is still consistent. If the true model happens to be the proposed working model, then the estimator is efficient. The method does not assume independence between the random slope and the covariates, and does not estimate or model the distribution of the random slope. In fact, an important advantage of the estimator is its consistency regardless whether or not the distribution of the random effect is correctly modeled, and regardless if there is dependency between the random slopes and the covariates. Our method is developed under the mixed effect model with binary response under the logit link function. It will be interesting and valuable to investigate if the general approach can be adapted to incorporate the probit link or log-log link for the binary response, and to more general models in handling count or continuous response.

Sometimes, there is evidence that a random effect is discrete, hence it is natural to consider the treatment of a discrete random effect. In fact, if a random effect is discrete with infinitely many categories, we would recommend to ignore its discreteness and use a continuous working model for its distribution for computational purpose. In fact, our derivation has not assumed the random effect is continuous so the results derived before indeed apply. If a random effect is discrete with finitely many categories, the problem actually drastically simplifies. Indeed, in this case, treating the random effect probability masses as additional parameters, the original problem is a pure parametric model and a simple MLE will yield the efficient estimator.

ACKNOWLEDGEMENTS

Data from the COHORT study, which received support from HP Therapeutics, Inc., were used in this study. We thank the Huntington Study Group COHORT investigators and respective coordinators who collected the data, as well as participants and their families who made this work possible.

Yanyuan Ma is supported by a grant from the National Science Foundation.

Tanya P. Garcia is supported by a grant from the National Institute Of Neurological Disorders and Stroke of the National Institutes of Health.

The content is solely the responsibility of the authors and does not represent the official views of the National Institutes of Health.

BIBLIOGRAPHY

Agresti, A., Ohman, P. & Caffo, B. (2004). Examples in which misspecification of a random effects distribution reduces efficiency. *Computational Statistics & Data Analysis*, 47, 639–653.

Bates, D., Maechler, M., & Bolker, B. (2016). lme4: Linear Mixed-Effects Models using 'Eigen' and S4. https://cran.r-project.org/package=lme4.

Downing, K., Chan, S. W, Downing, W. K., Kwong, T., & Lam, T. F. (2008). Measuring gender differences in cognitive functioning. *Multicultural Education & Technology Journal*, 2, 4–18.

Garcia, T. P. & Ma, Y. (2016). Optimal Estimator for Logistic Model with Distribution-free Random Intercept. *Scandinavian Journal of Statistics*, 43, 156–171.

Heagerty, P. J. & Kurland, B. F. (2001). Misspecified Maximum Likelihood Estimates and Generalised Linear Mixed Models. *Biometrika*, 88, 973–985.

Huntington Study Group. (1996). Unified Huntington?s Disease Rating Scale: Reliability and Consistency. *Movement Disorders*, 11, 136–142.

Litiére, S., Alonso, A., & Molenberghs, G. (2007). Type I and Type II error under random-effects misspecification in generalized linear mixed models. *Biometrics*, 63, 1038–1044.

Litiére, S., Alonso, A., & Molenberghs, G. (2008). The impact of a misspecified random-effects distribution on the estimation and the performance of inferential procedures in generalized linear mixed models. *Statistics in Medicine*, 27, 3125–3144.

Neuhaus, J. M., Hauck, W. W., & Kalbfleisch, J. D. (1992). The effects of mixture distribution misspecification when fitting mixed-effects logistic models. *Biometrika*, 79, 755–762.

Neuhaus, J. M., McCulloch, C. E., & Boylan, R. (2011). A note on Type II error under random effects misspecification in generalized linear mixed models. *Biometrics*, 67, 65–660.

Paulsen, J. S., Smith, M. M, Long, J. D., & PREDICT HD investigators and Coordinators of the Huntington Study Group. (2013). Cognitive decline in prodromal Huntington Disease: Implications for Clinical Trials. *Journal of Neurology, Neurosurgery, and Psychiatry*, Nov, 1233–1239.

Ross, R. A. C. (2010). Huntington's disease: a clinical review. *Orphanet Journal of Rare Diseases*, 5, 40.

Tsiatis, A. (2006). *Semiparametric Theory and Missing Data*, Springer, New York City.

Zhang, D. & Davidian, M. (2001). Linear mixed models with flexible distributions of random effects for longitudinal data. *Biometrics*, 57, 795–802.

Zhang, P., Song, P. X., Qu, A., & Greene T. (2008). Efficient estimation for patient-specific rates of disease progression using nonnormal linear mixed models. *Biometrics*, 64, 29–38.

APPENDIX

Proof of Theorem 1.

Because $E\{\mathbf{S}^*_{\text{eff}}(\mathbf{Y},\mathbf{X},\mathbf{Z},\boldsymbol{\beta}_0)\}=\mathbf{0}$, we can expand around $\boldsymbol{\beta}_0$ to obtain

$$\begin{aligned}\mathbf{0} &= n^{-1/2}\sum_{i=1}^n \mathbf{S}^*_{\text{eff}}(\mathbf{y}_i,\mathbf{x}_i,\mathbf{z}_i,\widehat{\boldsymbol{\beta}})\\ &= n^{-1/2}\sum_{i=1}^n \mathbf{S}^*_{\text{eff}}(\mathbf{y}_i,\mathbf{x}_i,\mathbf{z}_i,\boldsymbol{\beta}_0) + n^{-1}\sum_{i=1}^n \frac{\partial \mathbf{S}^*_{\text{eff}}(\mathbf{y}_i,\mathbf{x}_i,\mathbf{z}_i,\boldsymbol{\beta}^*)}{\partial \boldsymbol{\beta}^{\mathrm{T}}} n^{1/2}(\widehat{\boldsymbol{\beta}}-\boldsymbol{\beta}_0)\\ &= n^{-1/2}\sum_{i=1}^n \mathbf{S}^*_{\text{eff}}(\mathbf{y}_i,\mathbf{x}_i,\mathbf{z}_i,\boldsymbol{\beta}_0) + E\left\{\frac{\partial \mathbf{S}^*_{\text{eff}}(\mathbf{Y},\mathbf{X},\mathbf{Z},\boldsymbol{\beta}_0)}{\partial \boldsymbol{\beta}^{\mathrm{T}}}\right\} n^{1/2}(\widehat{\boldsymbol{\beta}}-\boldsymbol{\beta}_0) + o_p(1),\end{aligned}$$

where $\boldsymbol{\beta}^*$ lies on the line connecting $\widehat{\boldsymbol{\beta}}$ and $\boldsymbol{\beta}_0$. Therefore,

$$n^{1/2}(\widehat{\boldsymbol{\beta}}-\boldsymbol{\beta}_0) = E\left\{\frac{\partial \mathbf{S}^*_{\text{eff}}(\mathbf{Y},\mathbf{X},\mathbf{Z},\boldsymbol{\beta}_0)}{\partial \boldsymbol{\beta}^{\mathrm{T}}}\right\}^{-1} n^{-1/2}\sum_{i=1}^n \mathbf{S}^*_{\text{eff}}(\mathbf{y}_i,\mathbf{x}_i,\mathbf{z}_i,\boldsymbol{\beta}_0) + o_p(1).$$

This implies that $n^{1/2}(\widehat{\boldsymbol{\beta}}-\boldsymbol{\beta}_0)\to \text{Normal}(0,\mathbf{A}^{-1}\mathbf{B}\mathbf{A}^{-T})$,$\mathbf{A}=E\{\partial \mathbf{S}^*_{\text{eff}}(\mathbf{Y},\mathbf{X},\mathbf{Z},\boldsymbol{\beta}_0)/\partial\boldsymbol{\beta}^{\mathrm{T}}\}$ and $\mathbf{B}=\text{var}\{\mathbf{S}^*_{\text{eff}}(\mathbf{Y},\mathbf{X},\mathbf{Z},\boldsymbol{\beta}_0)\}$. Finally, it is easy to check that when $f^*_{\mathbf{R},\mathbf{X},\mathbf{Z}}(\mathbf{r},\mathbf{x},\mathbf{z}) = f_{\mathbf{R},\mathbf{X},\mathbf{Z}}(\mathbf{r},\mathbf{x},\mathbf{z})$, $-\mathbf{A}=-E\{\partial\mathbf{S}_{\text{eff}}(\mathbf{Y},\mathbf{X},\mathbf{Z},\boldsymbol{\beta}_0)/\partial\boldsymbol{\beta}^{\mathrm{T}}\}=\text{var}\{\mathbf{S}_{\text{eff}}(\mathbf{Y},\mathbf{X},\mathbf{Z},\boldsymbol{\beta}_0)\}=\mathbf{B}$ since $\mathbf{S}_{\text{eff}}$ is the efficient score vector. Hence, the variance-covariance simplifies to $\mathbf{B}^{-1}$ and the estimator is efficient. □

Proof of Theorem 2.
Sufficiency:

$$\begin{aligned}& f_{\mathbf{U}\mid\mathbf{W},\mathbf{R},\mathbf{X},\mathbf{Z}}(\mathbf{u}\mid\mathbf{w},\mathbf{r},\mathbf{x},\mathbf{z})\\ &= \text{pr}(\mathbf{U}=\mathbf{u}\mid\mathbf{W}=\mathbf{w},\mathbf{R}=\mathbf{r},\mathbf{X}=\mathbf{x},\mathbf{Z}=\mathbf{z})\\ &= \frac{\exp\{\sum_{j=1}^m Y_{ij}\mathbf{X}_{ij}^{\mathrm{T}}\boldsymbol{\beta}+\mathbf{r}^{\mathrm{T}}(\sum_{j=1}^m Y_{ij}\mathbf{Z}_{ij})\}/[\prod_{j=1}^m\{1+\exp(\mathbf{X}_{ij}^{\mathrm{T}}\boldsymbol{\beta}+\mathbf{Z}_{ij}^{\mathrm{T}}\mathbf{r})\}]\Big|_{\mathbf{Z}_i\mathbf{Y}_i=(\mathbf{w}^{\mathrm{T}},\mathbf{u}^{\mathrm{T}})^{\mathrm{T}}}}{\sum_{\mathbf{Y}_i s.t. \mathbf{Z}_i\mathbf{Y}_i=\mathbf{w}}\exp\{\sum_{j=1}^m Y_{ij}\mathbf{X}_{ij}^{\mathrm{T}}\boldsymbol{\beta}+\mathbf{r}^{\mathrm{T}}(\sum_{j=1}^m Y_{ij}\mathbf{Z}_{ij})\}/[\prod_{j=1}^m\{1+\exp(\mathbf{X}_{ij}^{\mathrm{T}}\boldsymbol{\beta}+\mathbf{Z}_{ij}^{\mathrm{T}}\mathbf{r})\}]}\\ &= \frac{\exp\{(\mathbf{X}_{i1}^{\mathrm{T}}\boldsymbol{\beta},\ldots,\mathbf{X}_{iq}^{\mathrm{T}}\boldsymbol{\beta})(\mathbf{Z}_{iL}\mathbf{w}-\mathbf{Z}_{iL}^{-1}\mathbf{Z}_{iR}\mathbf{u})+(\mathbf{X}_{i(q+1)}^{\mathrm{T}}\boldsymbol{\beta},\ldots,\mathbf{X}_{im}^{\mathrm{T}}\boldsymbol{\beta})\mathbf{u}+\mathbf{r}^{\mathrm{T}}\mathbf{w}\}}{\sum_{\mathbf{u}}\exp\{(\mathbf{X}_{i1}^{\mathrm{T}}\boldsymbol{\beta},\ldots,\mathbf{X}_{iq}^{\mathrm{T}}\boldsymbol{\beta})(\mathbf{Z}_{iL}\mathbf{w}-\mathbf{Z}_{iL}^{-1}\mathbf{Z}_{iR}\mathbf{u})+(\mathbf{X}_{i(q+1)}^{\mathrm{T}}\boldsymbol{\beta},\ldots,\mathbf{X}_{im}^{\mathrm{T}}\boldsymbol{\beta})\mathbf{u}+\mathbf{r}^{\mathrm{T}}\mathbf{w}\}}\\ &= \frac{\exp\{(\mathbf{X}_{i1}^{\mathrm{T}}\boldsymbol{\beta},\ldots,\mathbf{X}_{iq}^{\mathrm{T}}\boldsymbol{\beta})(-\mathbf{Z}_{iL}^{-1}\mathbf{Z}_{iR}\mathbf{u})+(\mathbf{X}_{i(q+1)}^{\mathrm{T}}\boldsymbol{\beta},\ldots,\mathbf{X}_{im}^{\mathrm{T}}\boldsymbol{\beta})\mathbf{u}\}}{\sum_{\mathbf{u}}\exp\{(\mathbf{X}_{i1}^{\mathrm{T}}\boldsymbol{\beta},\ldots,\mathbf{X}_{iq}^{\mathrm{T}}\boldsymbol{\beta})(-\mathbf{Z}_{iL}^{-1}\mathbf{Z}_{iR}\mathbf{u})+(\mathbf{X}_{i(q+1)}^{\mathrm{T}}\boldsymbol{\beta},\ldots,\mathbf{X}_{im}^{\mathrm{T}}\boldsymbol{\beta})\mathbf{u}\}}\\ &= f_{\mathbf{U}\mid\mathbf{X},\mathbf{Z}}(\mathbf{u}\mid\mathbf{x},\mathbf{z}).\end{aligned}$$

Similarly,

$$
\begin{aligned}
& f_{\mathbf{R}\mid\mathbf{U},\mathbf{W},\mathbf{X},\mathbf{Z}}(\mathbf{r}\mid\mathbf{u},\mathbf{w},\mathbf{x},\mathbf{z}) \\
&= \mathrm{pr}(\mathbf{R}=\mathbf{r}\mid\mathbf{U}=\mathbf{u},\mathbf{W}=\mathbf{w},\mathbf{X}=\mathbf{x},\mathbf{Z}=\mathbf{z}) \\
&= \Big(f_{\mathbf{R},\mathbf{X},\mathbf{Z}}(\mathbf{r},\mathbf{x},\mathbf{z})\exp\{\sum_{j=1}^m Y_{ij}\mathbf{X}_{ij}^{\mathrm{T}}\boldsymbol{\beta}+\mathbf{r}^{\mathrm{T}}(\sum_{j=1}^m Y_{ij}\mathbf{Z}_{ij})\} \\
&\quad /[\prod_{j=1}^m\{1+\exp(\mathbf{X}_{ij}^{\mathrm{T}}\boldsymbol{\beta}+\mathbf{Z}_{ij}^{\mathrm{T}}\mathbf{r})\}]\Big|_{\mathbf{Z}_i\mathbf{Y}_i=(\mathbf{w}^{\mathrm{T}},\mathbf{u}^{\mathrm{T}})^{\mathrm{T}}}\Big) \\
&\quad /\Big(\int f_{\mathbf{R},\mathbf{X},\mathbf{Z}}(\mathbf{r},\mathbf{x},\mathbf{z})\exp\{\sum_{j=1}^m Y_{ij}\mathbf{X}_{ij}^{\mathrm{T}}\boldsymbol{\beta}+\mathbf{r}^{\mathrm{T}}(\sum_{j=1}^m Y_{ij}\mathbf{Z}_{ij})\} \\
&\quad /[\prod_{j=1}^m\{1+\exp(\mathbf{X}_{ij}^{\mathrm{T}}\boldsymbol{\beta}+\mathbf{Z}_{ij}^{\mathrm{T}}\mathbf{r})\}]\Big|_{\mathbf{Z}_i\mathbf{Y}_i=(\mathbf{w}^{\mathrm{T}},\mathbf{u}^{\mathrm{T}})^{\mathrm{T}}} d\mathbf{r}\Big) \\
&= \frac{f_{\mathbf{R},\mathbf{X},\mathbf{Z}}(\mathbf{r},\mathbf{x},\mathbf{z})\exp(\mathbf{w}^{\mathrm{T}}\mathbf{r})/[\prod_{j=1}^m\{1+\exp(\mathbf{X}_{ij}^{\mathrm{T}}\boldsymbol{\beta}+\mathbf{Z}_{ij}^{\mathrm{T}}\mathbf{r})\}]}{\int f_{\mathbf{R},\mathbf{X},\mathbf{Z}}(\mathbf{r},\mathbf{x},\mathbf{z})\exp(\mathbf{w}^{\mathrm{T}}\mathbf{r})/[\prod_{j=1}^m\{1+\exp(\mathbf{X}_{ij}^{\mathrm{T}}\boldsymbol{\beta}+\mathbf{Z}_{ij}^{\mathrm{T}}\mathbf{r})\}]d\mathbf{r}} \\
&= f_{\mathbf{R}\mid\mathbf{W},\mathbf{X},\mathbf{Z}}(\mathbf{r}\mid\mathbf{w},\mathbf{x},\mathbf{z}).
\end{aligned}
$$

Completeness:

$$
\begin{aligned}
& E\{\mathbf{a}(\mathbf{W},\mathbf{X},\mathbf{Z})\mid\mathbf{R},\mathbf{X},\mathbf{Z}\} \\
&= \int \mathbf{a}(\mathbf{w},\mathbf{X},\mathbf{Z}) f_{\mathbf{W}\mid\mathbf{R},\mathbf{X},\mathbf{Z}}(\mathbf{w}\mid\mathbf{R},\mathbf{X},\mathbf{Z})d\mu(\mathbf{w}) \\
&= \int \mathbf{a}(\mathbf{w},\mathbf{X},\mathbf{Z})\frac{\exp\{(\mathbf{X}_{i1}^{\mathrm{T}}\boldsymbol{\beta},\ldots,\mathbf{X}_{iq}^{\mathrm{T}}\boldsymbol{\beta})(\mathbf{Z}_{iL}\mathbf{w}-\mathbf{Z}_{iL}^{-1}\mathbf{Z}_{iR}\mathbf{u})+(\mathbf{X}_{i(q+1)}^{\mathrm{T}}\boldsymbol{\beta},\ldots,\mathbf{X}_{im}^{\mathrm{T}}\boldsymbol{\beta})\mathbf{u}+\mathbf{R}^{\mathrm{T}}\mathbf{w}\}}{\prod_{j=1}^m\{1+\exp(\mathbf{X}_{ij}^{\mathrm{T}}\boldsymbol{\beta}+\mathbf{Z}_{ij}^{\mathrm{T}}\mathbf{R})\}} \\
&\quad d\mu(\mathbf{u})d\mu(\mathbf{w}) \\
&= b(\mathbf{R},\mathbf{X}_i,\mathbf{Z}_i)\int \mathbf{a}(\mathbf{w},\mathbf{X},\mathbf{Z})\exp[\{(\mathbf{X}_{i1}^{\mathrm{T}}\boldsymbol{\beta},\ldots,\mathbf{X}_{iq}^{\mathrm{T}}\boldsymbol{\beta})\mathbf{Z}_{iL}+\mathbf{R}^{\mathrm{T}}\}\mathbf{w}]d\mu(\mathbf{w}),
\end{aligned}
$$

where

$$
b(\mathbf{R},\mathbf{X}_i,\mathbf{Z}_i)=\frac{\int\exp\{(\mathbf{X}_{i1}^{\mathrm{T}}\boldsymbol{\beta},\ldots,\mathbf{X}_{iq}^{\mathrm{T}}\boldsymbol{\beta})(-\mathbf{Z}_{iL}^{-1}\mathbf{Z}_{iR}\mathbf{u})+(\mathbf{X}_{i(q+1)}^{\mathrm{T}}\boldsymbol{\beta},\ldots,\mathbf{X}_{im}^{\mathrm{T}}\boldsymbol{\beta})\mathbf{u}\}d\mu(\mathbf{u})}{\prod_{j=1}^m\{1+\exp(\mathbf{X}_{ij}^{\mathrm{T}}\boldsymbol{\beta}+\mathbf{Z}_{ij}^{\mathrm{T}}\mathbf{R})\}}
$$

is a positive function. Thus, $E\{\mathbf{a}(\mathbf{W},\mathbf{X},\mathbf{Z})\mid\mathbf{R},\mathbf{X},\mathbf{Z}\}=\mathbf{0}$ implies

$$
\int \mathbf{a}(\mathbf{W},\mathbf{X},\mathbf{Z})\exp[\{(\mathbf{X}_{i1}^{\mathrm{T}}\boldsymbol{\beta},\ldots,\mathbf{X}_{iq}^{\mathrm{T}}\boldsymbol{\beta})\mathbf{Z}_{iL}+\mathbf{R}^{\mathrm{T}}\}\mathbf{w}]d\mu(\mathbf{w})=\mathbf{0}.
$$

This implies the Laplace transformation of $\mathbf{a}$ is zero, hence $\mathbf{a}(\mathbf{W},\mathbf{X},\mathbf{Z})=\mathbf{0}$. □

TABLE 1: Simulation results when random effect and covariates are independent. Bias, sample variance (var), averaged estimated variance ($\widehat{\text{var}}$), and the empirical coverage percentage of the 95% confidence interval (CI) for the semiparametric estimator and the normal-based MLE are reported. The true parameter $\boldsymbol{\beta} = (0.35, 0.6, -0.4)^{\mathrm{T}}$. Results are based on 1000 simulations with $n = 500$, $m_i = 3$. Biases are multiplied by 100, var and $\widehat{\text{var}}$ are multiplied by 1000.

	Semiparametric Estimator				Normal-based MLE			
	bias	var	$\widehat{\text{var}}$	CI	bias	var	$\widehat{\text{var}}$	CI
	$R_i \sim 0.8\text{N}(3,1) + 0.2\text{N}(6,1.5)$ and centered				$X_{ij} \sim \text{N}(0.5,1)$ $Z_i \sim \text{Bernoulli}(0.5)$			
$\widehat{\beta}_1$	2.59	3.40	3.28	94.7%	0.44	0.45	0.44	94.6%
$\widehat{\beta}_2$	3.03	3.82	3.67	96.1%	0.48	0.49	0.50	96.2%
$\widehat{\beta}_3$	−2.94	3.76	3.38	94.4%	−0.38	0.43	0.45	95.5%
	$R_i \sim 0.8\text{N}(3,1) + 0.2\text{N}(6,1.5)$ and centered				$X_{ij} \sim \text{N}(0.5,1)$ $Z_i \sim \text{Poisson}(0.5)$			
$\widehat{\beta}_1$	3.29	3.82	3.35	94.1%	0.21	0.45	0.43	94.4%
$\widehat{\beta}_2$	3.19	4.03	3.76	95.4%	0.25	0.48	0.49	95.4%
$\widehat{\beta}_3$	−2.62	3.93	3.45	94.1%	−0.30	0.44	0.44	95.9%
	$R_i \sim 0.8\text{N}(3,1) + 0.2\text{N}(6,1.5)$ and centered				$X_{ij} \sim \text{N}(0.5,1)$ $Z_i \sim \text{Geometric}(0.7)$			
$\widehat{\beta}_1$	1.16	2.53	2.49	95.3%	0.078	0.41	0.41	95.6%
$\widehat{\beta}_2$	2.39	2.93	2.88	95.3%	0.13	0.47	0.47	94.0%
$\widehat{\beta}_3$	−1.59	2.74	2.58	95.3%	0.11	0.44	0.42	95.0%

TABLE 2: Simulation results when random effect and covariates are independent. Bias, sample variance (var), averaged estimated variance ($\widehat{\text{var}}$), and the empirical coverage percentage of the 95% confidence interval (CI) for the semiparametric estimator and the normal-based MLE are reported. The true parameter $\boldsymbol{\beta} = (0.35, 0.6, -0.4)^{\mathrm{T}}$. Results are based on 1000 simulations with $n = 500$, $m_i = 3$. Biases are multiplied by 100, var and $\widehat{\text{var}}$ are multiplied by 1000.

	Semiparametric Estimator				Normal-based MLE			
	bias	var	$\widehat{\text{var}}$	CI	bias	var	$\widehat{\text{var}}$	CI
	$R_i \sim \text{Gamma}(1,1.25)$ and centered				$X_{ij} \sim \text{N}(0.5,1)$ $Z_i \sim \text{Bernoulli}(0.5)$			
$\widehat{\beta}_1$	2.17	3.34	3.09	95.4%	−2.33	0.38	0.39	92.8%
$\widehat{\beta}_2$	2.69	4.00	3.51	94.5%	−2.62	0.46	0.45	92.1%
$\widehat{\beta}_3$	−1.77	3.30	3.13	94.9%	−3.02	0.42	0.40	92.1%
	$R_i \sim \text{Gamma}(1,1.25)$ and centered				$X_{ij} \sim \text{N}(0.5,1)$ $Z_i \sim \text{Poisson}(0.5)$			
$\widehat{\beta}_1$	1.88	3.56	3.29	95.0%	−2.02	0.38	0.40	94.2%
$\widehat{\beta}_2$	1.90	4.05	3.78	94.5%	−2.61	0.45	0.45	93.1%
$\widehat{\beta}_3$	−1.24	3.46	3.38	95.1%	−2.91	0.41	0.41	93.8%
	$R_i \sim \text{Gamma}(1,1.25)$ and centered				$X_{ij} \sim \text{N}(0.5,1)$ $Z_i \sim \text{Geometric}(0.7)$			
$\widehat{\beta}_1$	1.82	2.58	2.43	95.4%	−2.00	0.37	0.38	94.2%
$\widehat{\beta}_2$	3.42	3.08	2.83	94.6%	−1.97	0.44	0.44	93.5%
$\widehat{\beta}_3$	−2.60	2.52	2.49	95.3%	−2.33	0.38	0.39	94.4%

TABLE 3: Simulation results when random effect and covariates are independent. Bias, sample variance (var), averaged estimated variance ($\widehat{\text{var}}$), and the empirical coverage percentage of the 95% confidence interval (CI) for the semiparametric estimator and the normal-based MLE are reported. The true parameter $\boldsymbol{\beta} = (0.35, 0.6, -0.4)^{\mathrm{T}}$. Results are based on 1000 simulations with $n = 500$, $m_i = 3$. Biases are multiplied by 100, var and $\widehat{\text{var}}$ are multiplied by 1000.

	Semiparametric Estimator				Normal-based MLE			
	bias	var	$\widehat{\text{var}}$	CI	bias	var	$\widehat{\text{var}}$	CI
	$R_i \sim \mathrm{t}(3)$ and centered				$X_{ij} \sim \mathrm{N}(0.5, 1)$ $Z_i \sim \mathrm{Bernoulli}(0.5)$			
$\widehat{\beta}_1$	2.47	4.12	3.80	95.3%	−1.70	0.55	0.53	94.4%
$\widehat{\beta}_2$	5.51	5.21	4.44	93.4%	−1.22	0.60	0.60	94.3%
$\widehat{\beta}_3$	−3.61	4.39	3.91	95.4%	−2.10	0.54	0.55	94.3%
	$R_i \sim \mathrm{t}(3)$ and centered				$X_{ij} \sim \mathrm{N}(0.5, 1)$ $Z_i \sim \mathrm{Poisson}(0.5)$			
$\widehat{\beta}_1$	3.05	3.78	3.58	95.1%	−1.15	0.51	0.53	95.2%
$\widehat{\beta}_2$	4.29	4.61	4.11	94.4%	−0.90	0.59	0.60	95.4%
$\widehat{\beta}_3$	−2.73	3.87	3.67	95.0%	−1.70	0.52	0.54	94.8%
	$R_i \sim \mathrm{t}(3)$ and centered				$X_{ij} \sim \mathrm{N}(0.5, 1)$ $Z_i \sim \mathrm{Geometric}(0.7)$			
$\widehat{\beta}_1$	2.37	3.38	3.02	94.3%	−0.81	0.49	0.50	95.5%
$\widehat{\beta}_2$	2.77	4.00	3.47	94.2%	−1.22	0.57	0.56	93.5%
$\widehat{\beta}_3$	−2.07	3.43	3.07	94.3%	−1.28	0.53	0.51	94.6 %

TABLE 4: Simulation results when random effect and covariates are independent. Bias, sample variance (var), averaged estimated variance ($\widehat{\text{var}}$), and the empirical coverage percentage of the 95% confidence interval (CI) for the semiparametric estimator and the normal-based MLE are reported. The true parameter $\boldsymbol{\beta} = (0.35, 0.6, -0.4)^{\mathrm{T}}$. Results are based on 1000 simulations with $n = 500$, $m_i = 3$. Biases are multiplied by 100, var and $\widehat{\text{var}}$ are multiplied by 1000.

	Semiparametric Estimator				Normal-based MLE			
	bias	var	$\widehat{\text{var}}$	CI	bias	var	$\widehat{\text{var}}$	CI
	$R_i \sim \mathrm{N}(0, 1)$				$X_{ij} \sim \mathrm{N}(0.5, 1)$ $Z_i \sim \mathrm{Bernoulli}(0.5)$			
$\widehat{\beta}_1$	2.44	3.42	3.25	95.3%	0.57	0.49	0.49	95.5%
$\widehat{\beta}_2$	3.60	4.28	3.75	94.9%	0.70	0.51	0.55	96.0%
$\widehat{\beta}_3$	−3.41	3.20	3.39	96.0 %	−0.40	0.52	0.50	94.5%
	$R_i \sim \mathrm{N}(0, 1)$				$X_{ij} \sim \mathrm{N}(0.5, 1)$ $Z_i \sim \mathrm{Poisson}(0.5)$			
$\widehat{\beta}_1$	2.56	3.42	3.35	94.2%	0.25	0.47	0.47	95.4%
$\widehat{\beta}_2$	4.76	4.48	3.88	94.1%	0.60	0.51	0.53	95.8%
$\widehat{\beta}_3$	−3.46	3.32	3.47	95.7%	−0.64	0.50	0.48	94.6%
	$R_i \sim \mathrm{N}(0, 1)$				$X_{ij} \sim \mathrm{N}(0.5, 1)$ $Z_i \sim \mathrm{Geometric}(0.7)$			
$\widehat{\beta}_1$	1.99	3.28	2.94	95.2%	0.44	0.45	0.47	96.1%
$\widehat{\beta}_2$	3.96	3.76	3.36	94.8%	1.19	0.53	0.53	95.1%
$\widehat{\beta}_3$	−1.86	3.13	2.96	95.2%	−0.71	0.47	0.48	95.7 %

TABLE 5: Simulation results when random effect and covariates are dependent: $X_{ij} \sim \mathrm{Normal}(0.5R_i, 1)$. Bias, sample variance (var), averaged estimated variance ($\widehat{\text{var}}$), and the empirical coverage percentage of the 95% confidence interval (CI) for the semiparametric estimator and the normal-based MLE are reported. The true parameter $\boldsymbol{\beta} = (0.35, 0.6, -0.4)^{\mathrm{T}}$. Results are based on 1000 simulations with $n = 500$, $m_i = 3$. Biases are multiplied by 100, var and $\widehat{\text{var}}$ are multiplied by 1000.

	Semiparametric Estimator				Normal-based MLE			
	bias	var	$\widehat{\text{var}}$	CI	bias	var	$\widehat{\text{var}}$	CI
	$R_i \sim 0.8\mathrm{N}(3,1) + 0.2\mathrm{N}(6,1.5)$ and centered				$Z_i \sim \mathrm{Bernoulli}(0.5)$			
$\widehat{\beta}_1$	1.74	3.76	3.25	94.6%	25.91	0.48	0.44	2.3%
$\widehat{\beta}_2$	2.72	3.85	3.72	96.0%	25.12	0.49	0.51	5.1%
$\widehat{\beta}_3$	−3.98	3.81	3.39	94.4%	29.41	0.42	0.41	0.9%
	$R_i \sim 0.8\mathrm{N}(3,1) + 0.2\mathrm{N}(6,1.5)$ and centered				$Z_i \sim \mathrm{Poisson}(0.5)$			
$\widehat{\beta}_1$	2.49	3.39	3.31	96.2%	25.40	0.43	0.44	1.6%
$\widehat{\beta}_2$	3.44	3.89	3.73	95.2%	24.32	0.50	0.51	4.8%
$\widehat{\beta}_3$	−3.52	3.77	3.41	94.7%	28.68	0.41	0.41	0.8%
	$R_i \sim 0.8\mathrm{N}(3,1) + 0.2\mathrm{N}(6,1.5)$ and centered				$Z_i \sim \mathrm{Geometric}(0.7)$			
$\widehat{\beta}_1$	1.86	3.12	2.76	94.1%	23.68	0.44	0.42	3.9%
$\widehat{\beta}_2$	3.15	3.70	3.18	93.2%	23.02	0.52	0.49	7.4%
$\widehat{\beta}_3$	−2.32	2.89	2.84	95.5%	27.28	0.41	0.40	2.3%

TABLE 6: Simulation results when random effect and covariates are dependent: $X_{ij} \sim \mathrm{Normal}(0.5R_i, 1)$. Bias, sample variance (var), averaged estimated variance ($\widehat{\text{var}}$), and the empirical coverage percentage of the 95% confidence interval (CI) for the semiparametric estimator and the normal-based MLE are reported. The true parameter $\boldsymbol{\beta} = (0.35, 0.6, -0.4)^{\mathrm{T}}$. Results are based on 1000 simulations with $n = 500$, $m_i = 3$. Biases are multiplied by 100, var and $\widehat{\text{var}}$ are multiplied by 1000.

	Semiparametric Estimator				Normal-based MLE			
	bias	var	$\widehat{\text{var}}$	CI	bias	var	$\widehat{\text{var}}$	CI
	$R_i \sim \mathrm{Gamma}(1, 1.25)$ and centered				$Z_i \sim \mathrm{Bernoulli}(0.5)$			
$\widehat{\beta}_1$	2.39	2.86	2.82	95.4%	20.73	0.39	0.41	8.2%
$\widehat{\beta}_2$	2.95	3.65	3.22	94.7%	20.16	0.51	0.48	16.6%
$\widehat{\beta}_3$	−1.90	3.22	2.86	95.2%	23.35	0.40	0.39	5.0%
	$R_i \sim \mathrm{Gamma}(1, 1.25)$ and centered				$Z_i \sim \mathrm{Poisson}(0.5)$			
$\widehat{\beta}_1$	2.73	3.12	3.07	95.8%	20.31	0.38	0.41	9.1%
$\widehat{\beta}_2$	3.17	4.11	3.55	94.6%	19.67	0.49	0.48	17.5%
$\widehat{\beta}_3$	−2.18	3.44	3.18	94.4%	22.72	0.39	0.40	6.2%
	$R_i \sim \mathrm{Gamma}(1, 1.25)$ and centered				$Z_i \sim \mathrm{Geometric}(0.7)$			
$\widehat{\beta}_1$	2.50	2.65	2.63	95.9%	19.80	0.40	0.41	10.8%
$\widehat{\beta}_2$	3.05	2.94	3.00	95.5%	19.05	0.49	0.47	19.6%
$\widehat{\beta}_3$	−1.29	2.85	2.67	94.8%	21.40	0.39	0.39	8.2%

TABLE 7: Simulation results when random effect and covariates are dependent: $X_{ij} \sim \mathrm{Normal}(0.5R_i, 1)$. Bias, sample variance (var), averaged estimated variance ($\widehat{\text{var}}$), and the empirical coverage percentage of the 95% confidence interval (CI) for the semiparametric estimator and the normal-based MLE are reported. The true parameter $\boldsymbol{\beta} = (0.35, 0.6, -0.4)^{\mathrm{T}}$. Results are based on 1000 simulations with $n = 500$, $m_i = 3$. Biases are multiplied by 100, var and $\widehat{\text{var}}$ are multiplied by 1000.

	Semiparametric Estimator				Normal-based MLE			
	bias	var	$\widehat{\text{var}}$	CI	bias	var	$\widehat{\text{var}}$	CI
	$R_i \sim \mathrm{t}(3)$ and centered				$Z_i \sim \mathrm{Bernoulli}(0.5)$			
$\widehat{\beta}_1$	2.16	4.84	4.33	95.2%	35.62	0.44	0.48	0.0%
$\widehat{\beta}_2$	4.30	5.79	4.94	94.4%	33.46	0.54	0.57	0.2%
$\widehat{\beta}_3$	−4.51	4.98	4.47	93.7%	42.23	0.38	0.41	0.0%
	$R_i \sim \mathrm{t}(3)$ and centered				$Z_i \sim \mathrm{Poisson}(0.5)$			
$\widehat{\beta}_1$	3.00	4.73	4.21	94.8%	34.37	0.43	0.47	0.0%
$\widehat{\beta}_2$	4.61	5.92	4.93	95.1%	32.16	0.53	0.55	0.3%
$\widehat{\beta}_3$	−3.28	4.56	4.34	95.3%	41.67	0.40	0.40	0.0%
	$R_i \sim \mathrm{t}(3)$ and centered				$Z_i \sim \mathrm{Geometric}(0.7)$			
$\widehat{\beta}_1$	1.86	3.78	3.43	94.5%	33.05	0.41	0.45	0.0%
$\widehat{\beta}_2$	3.73	4.29	3.98	93.3%	30.87	0.49	0.53	0.4%
$\widehat{\beta}_3$	−2.89	4.02	3.52	93.9%	40.16	0.36	0.39	0.0%

TABLE 8: Simulation results when random effect and covariates are dependent: $X_{ij} \sim \mathrm{Normal}(0.5R_i, 1)$. Bias, sample variance (var), averaged estimated variance ($\widehat{\text{var}}$), and the empirical coverage percentage of the 95% confidence interval (CI) for the semiparametric estimator and the normal-based MLE are reported. The true parameter $\boldsymbol{\beta} = (0.35, 0.6, -0.4)^{\mathrm{T}}$. Results are based on 1000 simulations with $n = 500$, $m_i = 3$. Biases are multiplied by 100, var and $\widehat{\text{var}}$ are multiplied by 1000.

	Semiparametric Estimator				Normal-based MLE			
	bias	var	$\widehat{\text{var}}$	CI	bias	var	$\widehat{\text{var}}$	CI
	$R_i \sim \mathrm{N}(0, 1)$				$Z_i \sim \mathrm{Bernoulli}(0.5)$			
$\widehat{\beta}_1$	3.05	4.11	3.70	95.4%	30.25	0.47	0.46	0.2%
$\widehat{\beta}_2$	5.42	4.98	4.21	94.3%	28.77	0.56	0.53	1.3%
$\widehat{\beta}_3$	−2.52	3.74	3.71	95.6%	35.39	0.40	0.41	0.1%
	$R_i \sim \mathrm{N}(0, 1)$				$Z_i \sim \mathrm{Poisson}(0.5)$			
$\widehat{\beta}_1$	2.04	4.20	3.66	92.7%	29.55	0.47	0.45	0.2%
$\widehat{\beta}_2$	4.81	4.72	4.22	94.2%	28.04	0.50	0.52	1.2%
$\widehat{\beta}_3$	−2.16	3.72	3.64	97.0%	34.73	0.39	0.41	0.1%
	$R_i \sim \mathrm{N}(0, 1)$				$Z_i \sim \mathrm{Geometric}(0.7)$			
$\widehat{\beta}_1$	1.61	3.20	2.92	94.7%	28.30	0.46	0.44	0.8%
$\widehat{\beta}_2$	3.85	3.44	3.39	94.8%	26.55	0.51	0.51	2.8%
$\widehat{\beta}_3$	−1.12	3.20	2.99	93.8%	33.22	0.460	0.40	0.2%

TABLE 9: Execution time of 50 simulations using our estimator when random effect is generated from Normal(0,1), Z_{ij} is from the Geometric distribution with success probability 0.7, Independent case means $X_{ij} \sim \mathrm{N}(0.5, 1)$ and dependent case means $X_{ij} \sim \mathrm{N}(0.5R_i, 1)$. The unit of time is second, and m stands for the cluster size and p denotes the number of parameters to be estimated.

	Independent	Dependent
m=2,p=3	79.39	78.59
m=3,p=3	208.3	201.28
m=4,p=3	321.74	310.9
m=3,p=1	41.16	39.54
m=3,p=2	112.47	113.76
m=3,p=3	205.56	202.47

TABLE 10: Results from Huntington disease (HD) data analysis based on semiparametric estimator and normal-based maximum likelihood estimator (MLE). est: Parameter estimate, SE: standard error, 95% CI: 95% Wald-Type confidence interval, $\widehat{\beta}_{tms}$: Coefficient for total motor score, $\widehat{\beta}_{sdmt}$: Coefficient for symbol Digit Modalities Test, $\widehat{\beta}_{scolor}$: Coefficient for stroop color score, $\widehat{\beta}_{sword}$: Coefficient for stroop word score , $\widehat{\beta}_{sinter}$: Coefficient for stroop interference score. SE are multiplied by 10.

	Semiparametric Estimator			Normal-based MLE		
	Est	SE	95% CI	Est	SE	95% CI
$\widehat{\beta}_{tms}$	0.133	0.012	(0.065, 0.201)	0.266	0.004	(0.229, 0.303)
$\widehat{\beta}_{sdmt}$	0.028	0.012	(−0.040, 0.097)	−0.029	0.004	(−0.066, 0.009)
$\widehat{\beta}_{scolor}$	0.008	0.014	(−0.066, 0.081)	−0.029	0.003	(−0.063, 0.006)
$\widehat{\beta}_{sword}$	0.009	0.004	(−0.032, 0.048)	−0.014	0.002	(−0.039, 0.012)
$\widehat{\beta}_{sinter}$	0.074	0.002	(0.043, 0.104)	−0.014	0.004	(−0.053, 0.024)

Received 10 August 2017
Accepted 8 August 2018