GOODNESS-OF-FIT TEST ISSUES IN GENERALIZED LINEAR MIXED MODELS

A Dissertation

by

NAI-WEI CHEN

Submitted to the Office of Graduate Studies of
Texas A&M University
in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

December 2011

Major Subject: Statistics

GOODNESS-OF-FIT TEST ISSUES IN GENERALIZED LINEAR MIXED MODELS

A Dissertation

by

NAI-WEI CHEN

Submitted to the Office of Graduate Studies of
Texas A&M University
in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

Approved by:

Chair of Committee,	Thomas E. Wehrly
Committee Members,	Jeffrey D. Hart
	Dominique Lord
	Simon Sheather
Head of Department,	Simon Sheather

December 2011

Major Subject: Statistics

ABSTRACT

Goodness-of-Fit Test Issues in Generalized Linear Mixed Models.

(December 2011)

Nai-Wei Chen, B.B.A., Tunghai University;

M.S., National Central University

Chair of Advisory Committee: Dr. Thomas E. Wehrly

Linear mixed models and generalized linear mixed models are random-effects models widely applied to analyze clustered or hierarchical data. Generally, random effects are often assumed to be normally distributed in the context of mixed models. However, in the mixed-effects logistic model, the violation of the assumption of normally distributed random effects may result in inconsistency for estimates of some fixed effects and the variance component of random effects when the variance of the random-effects distribution is large. On the other hand, summary statistics used for assessing goodness of fit in the ordinary logistic regression models may not be directly applicable to the mixed-effects logistic models. In this dissertation, we present our investigations of two independent studies related to goodness-of-fit tests in generalized linear mixed models.

First, we consider a semi-nonparametric density representation for the random-effects distribution and provide a formal statistical test for testing normality of the random-effects distribution in the mixed-effects logistic models. We obtain estimates of parameters by using a non-likelihood-based estimation procedure. Additionally, we not only evaluate the type I error rate of the proposed test statistic through asymptotic results, but also carry out a bootstrap hypothesis testing procedure to

control the inflation of the type I error rate and to study the power performance of the proposed test statistic. Further, the methodology is illustrated by revisiting a case study in mental health.

Second, to improve assessment of the model fit in the mixed-effects logistic models, we apply the nonparametric local polynomial smoothed residuals over within-cluster continuous covariates to the unweighted sum of squares statistic for assessing the goodness-of-fit of the logistic multilevel models. We perform a simulation study to evaluate the type I error rate and the power performance for detecting a missing quadratic or interaction term of fixed effects using the kernel smoothed unweighted sum of squares statistic based on the local polynomial smoothed residuals over x-space. We also use a real data set in clinical trials to illustrate this application.

To my parents

ACKNOWLEDGMENTS

I have tremendous appreciation for my advisor Dr. Thomas E. Wehrly. Finite words cannot express how blessed I am in my life. I thank him for his patient guidance, trust, and encouragement on the pursuit of my Ph.D. degree. I not only gained confidence to explore and develop my statistical ideas through his inspiration, but also learned a lot for future development through our talks. I will bear whatever he instructed in mind forever.

I also sincerely thank Dr. Jeffrey D. Hart for his generous support and advice throughout my research. Along with Dr. Hart, I would like to express my appreciation to Dr. Simon Sheather and Dr. Dominique Lord for their support as members of my advisory committee. Additionally, I particularly thank Dr. Michael Longnecker for offering me financial support during the final term of my graduate studies.

Moreover, I offer special thanks to my colleagues, Junbum Lee, Ganggang Xu, and Paul Martin, for their friendship. Finally, I extend a sincere thanks to my spiritual family. I am grateful to my parents and my sisters for their love and encouragement. Without their firm support, I could not accomplish the goal and complete this amazing journey.

TABLE OF CONTENTS

CHAPTER Page

LIST OF TABLES

TABLE Page

TABLE Page

LIST OF FIGURES

CHAPTER I

INTRODUCTION

In most data analysis, linear models (LMs) have been widely used in cases where the observed outcome variables are continuous. When the observed outcome variables are categorical or discrete, generalized linear models (GLMs) possessing nonnormal outcome distributions and their mean functions (McCullagh and Nelder, 1989; Agresti, 2002) play an important role. Additionally, in practice, clustered or hierarchical data often occur in many fields, for instance, in the biomedical field where repeated measurements are taken over time on each of many subjects in a sample or in the field of education where we can group students by the district and measurements are taken on students within districts. On the above examples, outcome variables are usually correlated because repeated measurements are made on each subject or subjects within clusters may exhibit similar characteristics. Therefore, LMs and GLMs are not applicable in accounting for this dependence.

In the past decade, models including a vector of unobserved subject-specific effects, namely random effects, in the linear predictor component of the model are used to deal with multiple sources of variation. In random-effects models, it is often assumed that conditional on the random effects, the observed outcomes within each subject or cluster are independent and random parts can be between subjects or within subjects. Nowadays, linear mixed models (LMMs) and generalized linear mixed models (GLMMs) are random-effects models used to model normal and nonnormal observed outcomes, respectively. They have been widely applied in many different fields such as epidemiological studies of diseases, toxicology, and so on (Ver-

---

The format follows the style of *Journal of the American Statistical Association.*

beke and Molenberghs, 2000; Diggle et al., 2002; Molenberghs and Verbeke, 2005).

Undoubtedly, in the context of mixed models, estimation and inference depend that the structure of random effects is correctly specified. In general, random effects are unobserved and in most inferential procedures and computation implementation, random effects are often assumed to be normally distributed in LMMs and GLMMs. In LMMs, Verbeke and Lesaffre (1997) showed that maximum likelihood estimators (MLEs) for fixed effects and variance components of random effects obtained under the assumption of normally distributed random effects are consistent, even when the random-effects distribution is misspecified. Unlike LMMs some research has been done to discover the impact of misspecifying the random-effects distribution in the mixed-effects logistic model, a broadly discussed case with binary outcomes in GLMMs.

Neuhaus et al. (1992) carried out a set of simulation studies to find that when the distribution of random effects is misspecified and a random-intercept logistic model is fitted, the MLEs of model parameters for the fixed effects are inconsistent, but the magnitude of the bias is not large. However, estimates of the variance of the random-effects distribution exhibit large biases. Heagerty and Kurland (2001) used the Kullback-Leibler Information Criterion to evaluate the consistency of MLEs of model parameters on conditional and marginal mean models. The authors showed that for conditionally specified models, misspecification of the random-effects distribution may lead to seriously biased estimators for a cluster-level (between-subject) parameter and the intercept term when the variance of the random-effects distribution is large. Agresti et al. (2004) showed that the MLEs for fixed effect and variance component of the random-effects distribution appear inconsistent when the true random-effects distribution is a two-points mixture with a large variance in a simple one-way random-effects model. Litière et al. (2008), through simulations, found that MLEs of between-subject parameters for the mean structure may be affected by misspecification of the random-effects distribution when the variance of the true random-effects distribution is large and estimates of the variance component are severely affected by misspecification in most situations.

Moreover, Litière et al. (2007) studied the impact of the misspecification of the random-effects distribution on the type I and type II error rates related to the Wald test for the mean structure parameters. They found that misspecification of the random-effects distribution and the variance component of random effects can severely affect the power of the analysis and the type I error rate related to the tests for the intercept parameter. Huang (2009) proposed a novel two-step parametric diagnostic method that makes use of both observed data and a reconstructed data set induced from the observed data to verify misspecification of the random-effects model. In terms of one of the simulation results, the author showed that when the distribution family of random effects is misspecified, the test for the corresponding variance component of the random-effects distribution tends to be significant but the proposed test statistic loses power on testing the parameters of fixed effects. Therefore, an assessment of the goodness of fit of the random-effects distribution increasingly becomes a study issue in generalized linear mixed models.

On the other hand, since the mixed-effects logistic models have been widely used for analyzing clustered or naturally hierarchical data with binary outcomes, methods for assessment of the model fit need to be well developed. Evans and Hosmer (2004) extended summary statistics used on assessing goodness of fit in the ordinary logistic regression models to mixed-effects logistic models. The authors showed that the performance of type I error rates is not good in some situations. Additionally, Sturdivant (2005) and Sturdivant and Hosmer (2007) proposed a kernel smoothed unweighted sum of squares statistic by smoothing residuals in the y-space to assess the adequacy of the logistic multilevel models. They demonstrated satisfactory adherence to type I error rates by the proposed statistic. However, for a case with fewer subjects per cluster, the simulation results showed very limited or no power to detect the missing quadratic term of fixed effects. Therefore, finding a method for improving the existing methods of assessment of the model fit in any specified model is worth being discussed.

The objective of this dissertation includes two independent studies in mixed-effects logistic models. The first work is to provide a method for testing normality of the random-effects distribution and the second work is to apply the nonparametric local polynomial smoothed residuals to improve assessment of the model fit. In Chapter II, we present a literature review for our two studies. Chapter III of this dissertation is devoted to our first study. We consider a semi-nonparametric (SNP) density representation for the random-effects distribution and provide a formal statistical test that has a close connection to an order selection-type goodness-of-fit test for testing normality of the random-effects distribution in GLMMs. This test is non-parametric in the sense that we do not assume a parametric form for the alternative model. In addition, estimation is fundamental to any hypothesis test. Unlike LMMs the likelihood function under GLMMs may have no analytic expression and numerical approximations may be needed. As a result, likelihood-based inference is computationally challenging and non-likelihood-based estimation is an attractive approach. Zeger et al. (1988) used the generalized estimating equations (GEEs) approach to fit subject-specific and population-averaged models. Jiang (2007) and Jiang et al. (2007) proposed a procedure to solve estimating equations for parameter estimation. Throughout this study, a non-likelihood-based estimation procedure will be adopted for estimation of parameters. In a set of simulation studies, we conduct a bootstrap hypothesis testing procedure to evaluate the power performance and the type I error rate for the proposed test statistic. Further, we apply our method to revisit a case study in mental health (Alonso et al., 2004; 2008).

Chapter IV is devoted to our second study. We apply the nonparametric local polynomial smoothed residuals over within-cluster continuous covariates to the unweighted sum of squares statistic (Hosmer et al., 1997; Sturdivant and Hosmer, 2007) for assessing the goodness-of-fit of the logistic multilevel models, namely, the mixed-effects logistic models for hierarchical data with binary outcomes. We carry out a simulation study which is performed to evaluate the type I error rate of the kernel smoothed unweighted sum of squares statistic by using the local polynomial smoothed residuals and the power performance for detecting a missing quadratic or interaction term of fixed effects. Moreover, to illustrate this application, we use a real data set in clinical trials provided by Cancer Biostatistics Center, Vanderbilt University. Finally, Chapter V recapitulates all our findings and provides discussions of future research.

CHAPTER II

LITERATURE REVIEW

In this chapter, we shall introduce the basic formulation of generalized linear mixed models and review some selected articles on parameter estimation in detail. Additionally, in this dissertation, one of our studies is related to an application of smoothed residuals in the logistic multilevel model for model checking. Therefore, we also introduce the original idea through reviewing an application of the smoothed residuals in the goodness-of-fit test of the ordinary logistic regression model.

2.1 Generalized Linear Mixed Models

Suppose there are m observed subjects (or clusters). Let us denote by y_{ij} the jth response measured, for instance, the jth time point in longitudinal data, for the ith subject (or cluster), $i = 1, \ldots, m$ and $j = 1, \ldots, n_i$. Further, for subject i, conditional on random effects $\mathbf{b}_i$, all the responses y_{ij} are assumed to be independent with conditional density belonging to the exponential family (Molenberghs and Verbeke, 2005; Litière et al., 2007),

$$f(y_{ij}|\mathbf{b}_i; \boldsymbol{\beta}, \phi) = \exp\left\{ \frac{y_{ij}\theta_{ij} - \psi(\theta_{ij})}{\phi} + c(y_{ij}, \phi) \right\},$$

where $\psi(\cdot)$ is a function satisfying $E(y_{ij}|\mathbf{b}_i) = \psi'(\theta_{ij})$, $Var(y_{ij}|\mathbf{b}_i) = \phi\psi''(\theta_{ij})$, and ϕ is a dispersion parameter whose value may be known and $c(\cdot, \cdot)$ is a known function. Let $\mu_{ij}^b = E(y_{ij}|\mathbf{b}_i) = \psi'(\theta_{ij})$. A generalized linear mixed model for y_{ij} is given by

$$g(\mu_{ij}^b) = g(E[y_{ij}|\mathbf{b}_i]) = \mathbf{x}_{ij}^T\boldsymbol{\beta} + \mathbf{z}_{ij}^T\mathbf{b}_i,$$

where $g(\cdot)$ is a monotonically increasing link function depending on known $\mathbf{x}_{ij}$ and $\mathbf{z}_{ij}$ p-dimensional and q-dimensional vectors of fixed covariate values, including the intercept term; $\boldsymbol{\beta}$, a p-dimensional vector of unknown fixed regression coefficients and $\mathbf{b}_i$, a q-dimensional vector of the random effects. The subject-specific effects $\mathbf{b}_i$ are generally assumed to be normally distributed with mean zero vector and variance-covariance matrix D, denoted by $f(\mathbf{b}_i|D) \sim N(\mathbf{0}, D)$.

In clustered or hierarchical data analysis, the mixed-effects logistic models are often used to analyze binary outcome data collected in subjects (or clusters). Herein, we illustrate a special and important mixed-effects logistic model as follows.

Example 1. (*Random-Intercept Logistic Model*) Suppose the intercept terms b_i are independent and identically distributed random effects. Within the ith subject (or cluster), binary responses y_{ij} are conditionally independent Bernoulli with

$$\text{logit}(\mu_{ij}^b) = b_i + \mathbf{x}_{ij}^T \boldsymbol{\beta},$$

where $\mu_{ij}^b = p(y_{ij} = 1|b_i, \mathbf{x}_{ij})$ and the dispersion parameter ϕ is assumed to be 1. This is a common case of generalized linear mixed models where the conditional exponential family is Bernoulli, b_i is normally distributed with mean zero and variance σ^2, and the link function is the logit-link, namely, the random-intercept logistic model.

2.2 Estimation Procedure

Generally, a random-effects model can be fitted by maximizing the marginal likelihood. The likelihood function is derived as

$$\begin{aligned} L(\boldsymbol{\beta}, D, \phi) &= \prod_{i=1}^{m} \int \prod_{j=1}^{n_i} f(y_{ij}|\mathbf{b}_i, \boldsymbol{\beta}, \phi) f(\mathbf{b}_i|D) d\mathbf{b}_i \\ &= \prod_{i=1}^{m} f_i(\mathbf{y}_i|\boldsymbol{\beta}, D, \phi), \end{aligned}$$

where $\mathbf{y}_i = (y_{i1}, \ldots, y_{in_i})^T$. The likelihood function under GLMMs may have no analytic expression and it cannot be further simplified. Thus, numerical approximations may be needed. As a result, likelihood-based inference in GLMMs is computationally challenging.

So far, several approaches about the inference of GLMMs have been developed. Based on Bayesian techniques, Zeger and Karim (1991) used Gibbs sampling techniques to take repeated samples from the posterior distributions to avoid the need of numerical integration, and Booth and Hobert (1999) used Monte Carlo EM algorithm for maximum likelihood estimation. Moreover, Breslow and Clayton (1993) proposed not only approximation of the marginal quasi-likelihood using the Laplace method which leads to estimating equations based the penalized quasi-likelihood for mean parameters in the marginal model, but also a penalized quasi-likelihood for approximated inference on mean parameters and realizations of random effects in the conditional model. Lin and Breslow (1996) further developed first-order and second-order correction procedures for parameter estimation under the penalized quasi-likelihood. On the other hand, some computational methods based on the non-likelihood viewpoint are attractive. Zeger et al. (1988) used the generalized estimating equations (GEEs) approach to fit subject-specific and population-averaged models. Jiang (2007) and Jiang et al. (2007) proposed an iterative procedure to solve estimating equations for parameter estimation. In this section, we review two methods of parameter estimation for generalized linear mixed models in detail. One excerpted from Breslow and Clayton (1993) is an approximation of the quasi-likelihood function in the conditional case; the other one excerpted from Jiang (2007) is an iterative procedure for solving estimating equations in the marginal case.

2.2.1 Parameter Estimation in the Conditional Case

When the exact likelihood function is difficult to compute, Breslow and Clayton (1993) applied the Laplace method for integral approximation to a quasi-likelihood function and modified the Fisher scoring algorithm developed by Green (1987) for parameter estimation. This method has been implemented in several statistical packages.

Herein, we simplify the situation to a specified subject with n response measures. Suppose that, given a q-dimensional vector $\mathbf{b}$ of random effects, the responses $\mathbf{y} = (y_1, \cdots, y_n)^T$ are conditionally independent and the conditional mean satisfies

$$E(\mathbf{y}|\mathbf{b}) = h(\mathbf{x}\boldsymbol{\beta} + \mathbf{z}\mathbf{b}),$$

where $\mathbf{x} = (\mathbf{x}_1, \cdots, \mathbf{x}_n)^T$, $\mathbf{z} = (\mathbf{z}_1, \cdots, \mathbf{z}_n)^T$ and $h(\cdot)$ is the inverse function of a link function $g(\cdot)$. Further, suppose that a random-effects vector $\mathbf{b}$ has the multivariate normal distribution with mean $\mathbf{0}$ and covariance matrix D depending on an unknown vector of variance components $\boldsymbol{\zeta}$. A quasi-likelihood function can be expressed as

$$qL(\boldsymbol{\beta}, \boldsymbol{\zeta}) \propto |D|^{-\frac{1}{2}} \int \exp\left\{ -\frac{1}{2\phi} \sum_{j=1}^{n} d_j - \frac{1}{2}\mathbf{b}^T D^{-1} \mathbf{b} \right\} d\mathbf{b},$$

where

$$d_j = -2 \int_{y_j}^{\mu_j^b} \frac{y_j - \tau}{a_j \upsilon(\tau)} d\tau$$

is known as the deviance, $\mu_j^b = E(y_j|\mathbf{b})$, $Var(y_j|\mathbf{b}) = \phi a_j \upsilon(\mu_j^b)$, a_j is a known constant, $\upsilon(\cdot)$ is a known variance function, and ϕ is a dispersion parameter that may or may not be known. In Bernoulli cases, similar to Example 1, we can consider that the dispersion parameter ϕ is fixed at unity.

The approximation of the logarithm of the quasi-likelihood function is obtained by using the Laplace approximation (see Breslow and Clayton, 1993, page 10–11 for more details). Maximizing $qL(\boldsymbol{\beta}, \boldsymbol{\zeta})$ is equivalent to maximizing the penalized quasi-likelihood (PQL) (Green, 1987) denoted by

$$-\frac{1}{2}\sum_{j=1}^{n} d_j - \frac{1}{2}\mathbf{b}^T D^{-1}\mathbf{b}.$$

Differentiating the PQL with respect to $\boldsymbol{\beta}$ and $\mathbf{b}$ to obtain score equations for the mean parameters as follows,

$$\sum_{j=1}^{n} \frac{(y_j - \mu_j^b)\mathbf{x}_j}{a_j \upsilon(\mu_j^b) g'(\mu_j^b)} = \mathbf{0}$$

and

$$\sum_{j=1}^{n} \frac{(y_j - \mu_j^b)\mathbf{z}_j}{a_j \upsilon(\mu_j^b) g'(\mu_j^b)} = D^{-1}\mathbf{b}.$$

Next, we can define a working vector $\mathbf{y}^* = (y_1^*, \ldots, y_n^*)^T$ and $y_j^* = g(\mu_j^b) + (y_j - \mu_j^b)g'(\mu_j^b)$ where μ_j^b is computed at current estimates of $\boldsymbol{\beta}$ and $\mathbf{b}$. The solution to score equations through Fisher scoring can be expressed as the iterative solution to the system

$$\begin{bmatrix} \mathbf{x}^T\mathbf{w}\mathbf{x} & \mathbf{x}^T\mathbf{w}\mathbf{z} \\ \mathbf{z}^T\mathbf{w}\mathbf{x} & D^{-1} + \mathbf{z}^T\mathbf{w}\mathbf{z} \end{bmatrix} \begin{bmatrix} \boldsymbol{\beta} \\ \mathbf{b} \end{bmatrix} = \begin{bmatrix} \mathbf{x}^T\mathbf{w}\mathbf{y}^* \\ \mathbf{z}^T\mathbf{w}\mathbf{y}^* \end{bmatrix},$$

where $\mathbf{w}$ is a $n \times n$ diagonal matrix with diagonal terms $w_j = \left[a_j \upsilon(\mu_j^b)\left\{g'(\mu_j^b)\right\}^2\right]^{-1}$.

For fixed variance components, parameter estimation of fixed effects and estimation of random effects are obtained in the conditional model. Additionally, the variance components in D are often unknown and need to be estimated before we make any inference (see Breslow and Clayton, 1993, page 11–12 for details).

2.2.2 Parameter Estimation in the Marginal Case

In some situations, we may be interested in the marginal mean $E(y_{ij})$ obtained from averaging the conditional mean $E(y_{ij}|\mathbf{b}_i)$ over the random effects. In this situation, the marginal mean can be expressed as

$$\mu_{ij} = E(y_{ij}) = E[E(y_{ij}|\mathbf{b}_i)] = \int h(\mathbf{x}_{ij}^T\boldsymbol{\beta} + \mathbf{z}_{ij}^T\mathbf{b}_i)f(\mathbf{b}_i)d\mathbf{b}_i,$$

where $h(\cdot)$ is the inverse function of a link function $g(\cdot)$. Especially, if $g(\cdot)$ is not the identity link, it is not true that $\mu_{ij} = E(y_{ij}) = h(\mathbf{x}_{ij}^T\boldsymbol{\beta})$ (see Molenberghs and Verbeke, 2005, page 298–301 for more details).

In the case of generalized linear mixed models, the optimal estimating equations can be denoted by

$$G^* = \sum_{i=1}^{m} \dot{\boldsymbol{\mu}}_i^T V_i^{-1}(\mathbf{y}_i - \boldsymbol{\mu}_i) = \mathbf{0},$$

where $\mathbf{y}_i = (y_{i1}, \ldots, y_{in_i})^T$, $E(\mathbf{y}_i) = \boldsymbol{\mu}_i$, and $V_i = Var(\mathbf{y}_i)$. In general, the variance of the response measures, V_i, is assumed known. Jiang (2007) argued that a parametric model for V_i may increase the risk of model misspecification which affects the consistency of parameter estimators. Jiang (2007) and Jiang et al. (2007) proposed a semiparametric regression model for V_i, which can be used for either a balanced or an unbalanced data and estimated by the method of moments. In the following, we sketch this idea.

Let us start by considering a study conducted over a set of visit times $t_1, t_2, \cdots, t_b$. Suppose that response measures are collected from subject i at the visit times t_j, $j \in J_i \subset J = \{1, \ldots, b\}$. Let $\mathbf{y}_i = (y_{ij})_{j \in J_i}$ and $\mathbf{x}_i = (\mathbf{x}_{ij})_{j \in J_i}$. We assume that $(\mathbf{x}_i, \mathbf{y}_i), i = 1, \ldots, m$ are independent and the mean function is given by

$$E(y_{ij}|\mathbf{x}_i) = g_j(\mathbf{x}_i, \boldsymbol{\beta}),$$

where $\boldsymbol{\beta}$ is a $p \times 1$ vector of unknown parameters and $g_j(\cdot, \cdot)$ are fixed functions. The covariance matrix is given by

$$V_i = Var(\mathbf{y}_i|\mathbf{x}_i),$$

whose (j,k)th element is $v_{ijk} = \mathrm{cov}(y_{ij}, y_{ik}|\mathbf{x}_i) = E\{(y_{ij} - \mu_{ij})(y_{ik} - \mu_{ik})|\mathbf{x}_i\}, j,k \in J_i$ with $\mu_{ij} = E(y_{ij}|\mathbf{x}_i)$ and $\boldsymbol{\mu}_i = (\mu_{ij})_{j \in J_i} = E(\mathbf{y}_i|\mathbf{x}_i)$. In addition, let $U = \{(j,k) : j,k \in J_i$ for some $1 \leq i \leq m\}$. Suppose that L_{jk} is the number of different v_{ijk}s and $v_{ijk} = v(j,k,l), i \in I(j,k,l)$, where $I(j,k,l)$ is a subset of $\{1,\ldots,m\}, 1 \leq l \leq L_{jk}$. For any $(j,k) \in U$, $1 \leq l \leq L_{jk}$, we can define

$$\hat{v}(j,k,l) = \frac{1}{n(j,k,l)} \sum_{i \in I(j,k,l)} \{y_{ij} - g_j(\mathbf{x}_i, \boldsymbol{\beta})\}\{y_{ik} - g_k(\mathbf{x}_i, \boldsymbol{\beta})\},$$

where $n(j,k,l)$ is the cardinality of set $I(j,k,l)$, $|I(j,k,l)|$, equal to the number of elements in $I(j,k,l)$. The estimated V_i is given by

$$\hat{V}_i = (\hat{v}_{ijk})_{j,k \in J_i},$$

where $\hat{v}_{ijk} = \hat{v}(j,k,l)$, $i \in I(j,k,l)$.

Once V_is are estimated (or known), we can estimate $\boldsymbol{\beta}$ by the estimating equations. When $\boldsymbol{\beta}$ are estimated (or known), V_i can be estimated by the above proposed method of moments. This procedure is named as iterative estimating equations, or IEE. The authors further showed that the IEE estimator is asymptotically as efficient as the optimal estimator obtained by solving generalized estimating equations with the true V_is (see Jiang, 2007 and Jiang et al., 2007 for more details).

2.3 Application of Smoothed Residuals in the Ordinary Logistic Models

In practice, the ordinary logistic regression model is widely used in analyzing data with independent binary outcome variables. Several goodness-of-fit tests for the binary regression model are available, such as the Pearson residual statistic, the likelihood ratio statistic and the Hosmer-Lemeshow statistic, etc. However, le Cessie and van Houwelingen (1991) mentioned potential problems of some of these test statistics.

For instance, the asymptotic distribution of the likelihood ratio test statistic is based on letting the number of observations in each status tend to infinity and it fails when there are no replicated measurements. Therefore, they proposed a global test statistic based on nonparametric kernel methods. In this way, there is no need to partition the data into subsets, it is possible to deal properly with continuous covariates, and each observation is treated in the same manner that a weighted average of the standardized residuals in its neighborhood is calculated. The authors also argued that in this way, the individual contribution of the observations to the test statistic can be used as a diagnostic tool to detect the parts of data where the model does not fit. On the other hand, Hosmer et al. (1997) smoothed the standardized residuals by using the kernel smoothing weights for the residuals in either the x-space or the y-space. In the x-space, all covariates are used in developing the weights and in the y-space, the weights are produced using the relative distances of the model-predicted probabilities of the outcome. Next, we sketch the idea of Hosmer et al. (1997) as follows.

In the ordinary logistic regression models, assume that we observe n independent pairs $(\mathbf{x}_i, y_i)$, $i = 1, \cdots, n$, where $\mathbf{x}_i^T = (1, x_{1i}, \cdots, x_{pi})$ denotes vector of $(p+1)$ fixed covariates for the i subject and $y_i = 0, 1$ denotes an observation of outcome. Hosmer et al. (1997) suggested weight functions by using the uniform kernel in the x-space and a cubic weight in the y-space. The x-space weight defining the distance between subjects i and j is $w_{ij} = \prod_{k=1}^{p} u(x_{ik}, x_{jk})$ with

$$u(x_{ik}, x_{jk}) = \begin{cases} 1 & \text{if } \dfrac{|x_{ik} - x_{jk}|}{s_k} \leq c_u \\ 0 & \text{if } \dfrac{|x_{ik} - x_{jk}|}{s_k} > c_u, \end{cases}$$

where s_k is the sample standard deviation of x_k and the value of $c_u = \dfrac{1}{2}\left(\dfrac{4}{n^{1/2p}}\right)$.

The cubic weights defining the y-space weight can be given by

$$w_{ij} = \begin{cases} 1 - \left(\dfrac{|\hat{\pi}_i - \hat{\pi}_j|}{c_{ci}} \right)^3 & \text{if } |\hat{\pi}_i - \hat{\pi}_j| \leq c_{ci} \\ 0 & \text{if } |\hat{\pi}_i - \hat{\pi}_j| > c_{ci}, \end{cases}$$

where the constant c_{ci} depends on i and is chosen such that $\sqrt{n}$ weights are non-zero for each subject. Then, the authors smoothed standardized residuals by using $\hat{r}_{si} = \sum_{j=1}^{n} w_{ij}\hat{r}_j$ with $\hat{r}_j = \dfrac{y_j - \hat{\pi}_j}{\sqrt{\hat{\pi}_j(1-\hat{\pi}_j)}}$ and defined the test statistic of goodness-of-fit by $\hat{T}_r = \sum_{i=1}^{n} \dfrac{\hat{r}_{si}^2}{\widehat{Var}(\hat{r}_{si}^2)}$.

Additionally, le Cessie and van Houwelingen (1991) and Hosmer et al. (1997) suggested that when the models are complicated, the smoothed-residuals-based tests, compared with some other tests, have greater power for checking goodness-of-fit of some specified models. Further, Sturdivant (2005) and Sturdivant and Hosmer (2007) applied the smoothed residuals over y-space in the logistic multilevel model and showed limited or no power for model checking in some situations. Later, in Chapter IV, we shall review this application and investigate whether using the local polynomial smoothed residuals over within-cluster continuous covariates can improve power performance for checking the adequacy of fit of the logistic multilevel model.

CHAPTER III

A TEST FOR NORMALITY OF RANDOM EFFECTS IN GENERALIZED LINEAR MIXED MODELS

3.1 Introduction

In generalized linear mixed models, estimation and inference depend the random-effects distribution being correctly specified, and have been implemented under the assumption of normally distributed random effects in many statistical packages. However, misspecification of the random-effects distribution may result in (1) bias in the estimates of the mean structure parameters associated with a large variance of random effects, (2) bias in the estimates of the variance component of the random-effects distribution (Neuhaus et al., 1992; Heagerty and Kurland, 2001; Agresti et al., 2004; Litière et al., 2008), and (3) low power performance in testing the intercept parameter or the parameters of fixed effects (Litière et al., 2007; Huang, 2009). Recently, some research has been devoted to detect the departure from the normality assumption of random effects.

Chen et al. (2002) constructed an informal test of the normality of random effects by choosing the order of a semi-nonparametric (SNP) estimator based on a Hermite expansion for GLMMs. The authors considered a Monte Carlo EM algorithm using a rejection sampling scheme to estimate parameters and applied information criteria such as AIC, BIC, or Hannan and Quinn's criterion to find that the performance of the SNP approach for detecting a departure from normality is encouraging. Waagepetersen (2006) used the adaptive rejection sampling to simulate random effects conditional on the observations and found the empirical distribution function based on conditional simulated random effects. Then, the author considered a discretized version of the Anderson-Darling statistic to assess the goodness of fit of the random-effects distribution and briefly commented that poor powers only slightly bigger than the nominal level (5%) are obtained for certain nonnormal distributions of random effects in a simple random-intercept logistic model. A key to the above methods is the use of the estimated distribution of random effects for assessing departure from normality of the random-effects distribution.

On the other hand, Tchetgen and Coull (2006) argued that misspecification of the random-effects distribution may induce asymptotic bias in the marginal MLEs of the fixed effect, whereas the conditional MLEs are robust to any misspecification of the random-effects distribution. They used this property to propose a diagnostic test based on the difference between the marginal MLEs and conditional MLEs of a subset of the fixed effects in the model to detect misspecification of the random-effects distribution. In a simulation result, they showed that for relatively large samples and moderate cluster size, their proposed test statistic is able to detect departures of the random-effects distribution in most settings. Alonso et al. (2008) proposed a set of diagnostic tests, two determinant tests and the determinant-trace test, based on the eigenvalues of the variance-covariance matrix of the maximum likelihood estimators to detect misspecification of the random-effects distribution. They found that the determinant-trace test has a reasonable type I error rate and a good power performance for all misspecification studies, and all tests perform considerably better when variance of the random-effects distribution is large. Furthermore, Alonso et al. (2010) proposed two diagnostic tests based on the information matrix associated with parameter estimators. They showed that both tests have a satisfactory power performance when variance of the random-effects distribution is large and sample size is large in some settings. A key to these methods is the use of the impact of a misspecified random-effects distribution on the maximum likelihood estimators and the inferential procedures.

Lastly, Claeskens and Hart (2009) used the SNP Hermite expansion to approximate the random-effects distribution and proposed an order-selection goodness-of-fit test depending on the likelihood function to detect normality of the random-effects distribution in linear mixed models. In terms of their simulation results, this proposed order-selection test is fairly conservative, but has good power performance under the specified alternative distributions of random effects, for instance, a mixture of normal distributions. The authors also proposed a novel viewpoint that they used information from the asymptotic distribution of their proposed test statistic to modify the penalty term of the traditional AIC criterion. It can be applied by selecting the order of SNP density representation for the random-effects distribution.

In this chapter, we start with reviewing a robust score statistic involved with generalized estimating equations for testing the parametric mean function in generalized linear models (Aerts et al., 1999) and derive its asymptotic results. We combine works of Claeskens and Hart (2009) and Aerts et al. (1999) to propose a formal non-likelihood statistical test for testing the hypothesis of normality of the random-effects distribution. Additionally, we not only evaluate the type I error rate and the power of the proposed test statistic by using the parametric bootstrap procedure and calculating the kernel smoothed bootstrap p-value in a simulation study, but also carry out a test for normality of the random-effects distribution by revisiting a case study in mental health.

3.2 Robust Score Test Statistic with Estimating Equations

Let us start by recalling an idea of testing the fit of a parametric function which was applied to test mean functions in GLMs (Aerts et al., 1999). Suppose that the observed data $(Z_1, \cdots, Z_n)$ have a joint density of the form $\kappa_n(z_1, \cdots, z_n; \gamma(\cdot))$, where κ_n is known up to $\gamma(\cdot)$. Their interest is in testing the null hypothesis,

$$H_0 : \gamma(\cdot) \in \Omega,$$

where $\Omega = \{\gamma(\cdot; \boldsymbol{\theta}_0) : \boldsymbol{\theta}_0 = (\theta_1, \cdots, \theta_p) \in \Theta\}$ and Θ is a subset of a p-dimensional Euclidean space. Under this circumstance, we can consider sequences of approximators $\{\gamma(\cdot; \theta_1, \cdots, \theta_{p+r}) : r = 1, 2, \cdots\}$ as alternative models for $\gamma(\cdot)$.

In the absence of a likelihood function, we form a set of estimating equations

$$\sum_{i=1}^{n} \boldsymbol{\psi}_r(z_i; \theta_1, \cdots, \theta_{p+r}) = \mathbf{0}_{p+r},$$

where $\boldsymbol{\psi}_r$ is a $p + r$ vector of statistics, $r = 0, 1, \cdots$. Let $\hat{\boldsymbol{\theta}}_0$ be the solution to the set of equations corresponding to $r = 0$, define $\hat{\boldsymbol{\delta}}_{r0} = (\hat{\boldsymbol{\theta}}_0, \mathbf{0}_r)$, and take $\boldsymbol{\xi}_r$ to be the length $p + r$ vector equal to $\sum_{i=1}^{n} \boldsymbol{\psi}_r(z_i; \hat{\boldsymbol{\theta}}_0, \mathbf{0}_r)$. Then, define a robust score statistic as follows, $\Re_0 = 0$ and

$$\begin{aligned}
\Re_r &= (\boldsymbol{\xi}_r)_r^T (\tilde{\mathbf{A}}_{nr}^{-1}(\hat{\boldsymbol{\delta}}_{r0}))_r \\
&\quad \times \left[(\tilde{\mathbf{A}}_{nr}^{-1}(\hat{\boldsymbol{\delta}}_{r0}) \tilde{\mathbf{B}}_{nr}(\hat{\boldsymbol{\delta}}_{r0}) \tilde{\mathbf{A}}_{nr}^{-1}(\hat{\boldsymbol{\delta}}_{r0}))_r \right]^{-1} \\
&\quad \times (\tilde{\mathbf{A}}_{nr}^{-1}(\hat{\boldsymbol{\delta}}_{r0}))_r (\boldsymbol{\xi}_r)_r
\end{aligned}$$

for $r = 1, 2, \cdots$, where $\tilde{\mathbf{B}}_{nr}(\hat{\boldsymbol{\delta}}_{r0}) = \sum_{i=1}^{n} \boldsymbol{\psi}_r(z_i; \hat{\boldsymbol{\theta}}_0, \mathbf{0}_r) \boldsymbol{\psi}_r(z_i; \hat{\boldsymbol{\theta}}_0, \mathbf{0}_r)^T$ and $\tilde{\mathbf{A}}_{nr}(\cdot)$ is a $(p + r) \times (p + r)$ matrix of partial derivatives of $\boldsymbol{\psi}_r$ with respect to θ_1, θ_2, $\cdots$, θ_{p+r}. For a $(p + r) \times 1$ vector $\boldsymbol{\xi}_r$, $(\boldsymbol{\xi}_r)_r$ denotes the subvector of the last r components; for any above $(p+r) \times (p+r)$ matrix $\boldsymbol{\Sigma}$, a $r \times r$ submatrix can be defined as $(\boldsymbol{\Sigma})_r = U^T \boldsymbol{\Sigma} U$ where $U^T = [\mathbf{0}_{r,p}, \mathbf{I}_p]$ with $\mathbf{I}_p$ the $p \times p$ identity matrix and $\mathbf{0}_{r,p}$ the zero matrix of dimension $r \times p$.

Next, in order to derive the asymptotic distribution of $\Re_r$, we assume that z_is,

$i = 1, \ldots, n$, are independent observations with expectations μ_i and variances $\zeta(\mu_i)$, where $\zeta(\cdot)$ is some known function. It is obvious that μ_i is some known function of a set of parameters. Wedderburn (1974) defined the quasi-likelihood equations (or generalized estimating equations) as

$$\sum_{i=1}^{n} \frac{z_i - \mu_i}{\zeta(\mu_i)} \cdot \frac{\partial \mu_i}{\partial \boldsymbol{\delta}_r} = \sum_{i=1}^{n} \boldsymbol{\psi}_r(z_i; \boldsymbol{\delta}_r) = \mathbf{0}_{p+r}, \tag{3.1}$$

where $\boldsymbol{\delta}_r = (\boldsymbol{\theta}_0, \theta_{p+1}, \cdots, \theta_{p+r})$.

In general, the parameter vector $\boldsymbol{\delta}_r$ is defined as the solution to

$$\sum_{i=1}^{n} E\left[\boldsymbol{\psi}_r(z_i; \boldsymbol{\delta}_r)\right] = \mathbf{0}_{p+r},$$

where all expectations are with respect to the true (or unknown) p.d.f., $\kappa(z_i)$. The idea here is that solving a set of score equations in likelihood models is generalized to the construction of quasi-likelihood equations. Therefore, as we solve the system of equations (3.1), it leads to the estimator $\hat{\boldsymbol{\delta}}_r$ for $\boldsymbol{\delta}_r$. Moreover, White (1982) proved that $\hat{\boldsymbol{\delta}}_r$ is generally a strongly consistent estimator for $\boldsymbol{\delta}_r^*$, the parameter vector which minimizes the Kullback-Leibler Information Criterion (KLIC) and observed that when the true distribution is unknown, the maximum likelihood estimator is a natural estimator for the parameters which minimize the KLIC.

Under this viewpoint, suppose that the partial derivatives and appropriate inverses exist, we can use the following matrices $\tilde{\mathbf{A}}_{nr}(\boldsymbol{\delta}_r)$ and $\tilde{\mathbf{B}}_{nr}(\boldsymbol{\delta}_r)$ to construct the variance-covariance matrix of the estimator $\hat{\boldsymbol{\delta}}_{r0}$,

$$\tilde{\mathbf{A}}_{nr}(\boldsymbol{\delta}_r) = \sum_{i=1}^{n} \frac{\partial}{\partial \boldsymbol{\delta}_r} \boldsymbol{\psi}_r(z_i; \boldsymbol{\delta}_r),$$

$$\tilde{\mathbf{B}}_{nr}(\boldsymbol{\delta}_r) = \sum_{i=1}^{n} \boldsymbol{\psi}_r(z_i; \boldsymbol{\delta}_r) \boldsymbol{\psi}_r(z_i; \boldsymbol{\delta}_r)^T.$$

The variance-covariance matrix of the estimator $\hat{\boldsymbol{\delta}}_{r0}$ can be given by

$$\tilde{\mathbf{A}}_{nr}^{-1}(\hat{\boldsymbol{\delta}}_{r0})\tilde{\mathbf{B}}_{nr}(\hat{\boldsymbol{\delta}}_{r0})\tilde{\mathbf{A}}_{nr}^{-1}(\hat{\boldsymbol{\delta}}_{r0}).$$

Since it includes the information matrix $\tilde{\mathbf{A}}_{nr}(\boldsymbol{\delta}_r)$ and a correction term $\tilde{\mathbf{B}}_{nr}(\boldsymbol{\delta}_r)$, it is called the sandwich variance-covariance matrix (Hardin and Hilbe, 2003). Then, we have the following crucial asymptotic result by applying Theorem 3.5 of White (1982).

Result 1. *When we have a set of estimating equations* (3.1) *for unknown parameters,* $\boldsymbol{\delta}_r = (\boldsymbol{\theta}_0, \theta_{p+1}, \cdots, \theta_{p+r})$ *and under some appropriate assumptions shown in White (1982, page 2–5), the robust score test statistic* $\Re_r$ *with* $(\boldsymbol{\xi}_r)_r$*,* $(\tilde{\mathbf{A}}_{nr}^{-1}(\hat{\boldsymbol{\delta}}_{r0}))_r$*, and* $(\tilde{\mathbf{A}}_{nr}^{-1}(\hat{\boldsymbol{\delta}}_{r0})\tilde{\mathbf{B}}_{nr}(\hat{\boldsymbol{\delta}}_{r0})\tilde{\mathbf{A}}_{nr}^{-1}(\hat{\boldsymbol{\delta}}_{r0}))_r$ *has an asymptotic* χ_r^2 *distribution.*

Further, we define a robust score test statistic analogous to the order selection test by

$$\tilde{T}_n = \max_{1 \le r \le R_n} \frac{\Re_r}{r}. \tag{3.2}$$

This test is equivalent to one that H_0 is rejected whenever $\tilde{T}_n \geqslant C_n$ and C_n is a critical value of the statistic $\tilde{T}_n$. Aerts et al. (1999) also presented that under H_0 and appropriate regularity conditions, $\tilde{T}_n$ has the same limiting distribution as that of $T_n = \max_{1\le r \le R_n} \dfrac{2(L_r - L_0)}{r}$ where L_r denotes the maximized log-likelihood under the alternative and L_0 denotes the maximized log-likelihood under the null if the likelihood is specified. The limiting distribution of T_n has been shown in Theorem 1 of Aerts et al. (2000). As a result, we can have the limiting distribution of $\tilde{T}_n$ as the following.

Result 2. $\tilde{T}_n \xrightarrow{d} \tilde{T}$*, as* $n \to \infty$*, with*

$$\tilde{T} = \max_{r \geqslant 1} \frac{Q_r}{r},$$

where Q_r is equal to χ^2_r, $\chi^2_r = s_1^2 + \cdots + s_r^2$ for all r and s_1, s_2, $\ldots$,s_r are identically independent distributed standard normal random variables.

3.3 A Test of Normality of Random Effects in GLMMs

In this section, we shall introduce the semi-nonparametric (SNP) density representation of the random-effects distribution, use the marginal approach to parameters estimation, develop a robust score statistic involved with generalized estimating equations for testing normality of the random-effects distribution in GLMMs, and demonstrate how to obtain the smoothed bootstrap p-value under a bootstrap hypothesis test procedure.

3.3.1 SNP Density Representation of the Random-Effects Distribution

Gallant and Nychka (1987) suggested that densities satisfying certain smoothness restrictions could be approximated by a truncated version of an infinite Hermite series expansion. This idea has been applied to approximate the distribution of random effects in some articles (Zhang and Davidian, 2001; Chen et al., 2002; Claeskens and Hart, 2009).

First, let us recall the generalized linear mixed models in Section 2.1. Assume that random effects are mutually independent across subject i and denoted by

$$\mathbf{b}_i = \mathbb{J}\mathbf{u}_i,$$

where $\mathbb{J}$ is a $q \times q$ upper triangular matrix and $\mathbf{u}_i$ is a $q \times 1$ random vector. Suppose that $\mathbf{u}_i$ is a q-variate random vector with density proportional to a truncated Hermite expansion around the standard normal density ϕ,

$$f_M(\mathbf{u}_i) \propto P_M^2(\mathbf{u}_i)\phi_q(\mathbf{u}_i) = \left\{ \sum_{|\boldsymbol{\alpha}|=0}^{M} a_{\boldsymbol{\alpha}} \mathbf{u}_i^{\boldsymbol{\alpha}} \right\}^2 \phi_q(\mathbf{u}_i),$$

where $\phi_q(\cdot)$ is the density function of $N_q(\mathbf{0}, I_q)$, $\boldsymbol{\alpha} = (\alpha_1, \cdots, \alpha_q)^T$, $|\boldsymbol{\alpha}| = \sum_{k=1}^{q} \alpha_q$ and $\mathbf{u}_i^{\boldsymbol{\alpha}} = \prod_{k=1}^{q} u_{ik}^{\alpha_k}$. For instance, as $\mathbf{u}_i = (u_{i1}, u_{i2})^T$ $(q = 2)$ and $M = 2$,

$$P_M(\mathbf{u}_i) = a_{00} + a_{10}u_{i1} + a_{01}u_{i2} + a_{20}u_{i1}^2 + a_{11}u_{i1}u_{i2} + a_{02}u_{i2}^2,$$

where the number of terms in the expansion is $N_{q,M} = \binom{M+q}{q} = 6$.

Then, the density of $\mathbf{b}_i$ is represented as

$$f_M(\mathbf{b}_i) \propto P_M^2(\mathbf{u}_i) N_q\left(\mathbf{b}_i; \mathbf{0}, \Sigma\right),$$

where $\mathbf{u}_i = \mathbb{J}^{-1}\mathbf{b}_i$ and $\Sigma = \mathbb{J}\mathbb{J}^T$. When $M = 0$, it reduces to a standard q-variate normal density, $N_q(\mathbf{0}, \mathbb{J}\mathbb{J}^T)$.

Therefore, when the random-intercept logistic model from Example 1 in Section 2.1 is adopted, the SNP representation of the random-effects distribution can be simplified as follows,

$$f_M(b_i) \propto P_M^2(u_i) N\left(b_i; 0, \sigma^2\right) = \left\{ \sum_{\alpha=0}^{M} a_\alpha u_i^\alpha \right\}^2 N\left(b_i; 0, \sigma^2\right),$$

where $u_i = \dfrac{b_i}{\sigma}$ is a random variable and M is the value of the order representing the degree of the tuning. Specifically, when $M = 0$, it reduces to a normal density, $b_i \sim N(0, \sigma^2)$.

3.3.2 Implementation of the Test for Normality of Random Effects

In this part, we consider a random-intercept logistic model which is considered by some authors (Litière et al., 2007; Alonso et al., 2008; Litière et al., 2008). Let

y_{ij} be the response for subject i, $i = 1, 2, \cdots, m$, collected at time point t_{ij} and r_i be the treatment group to which subject i is allocated. We assume that given a subject-specific random effect b_i, binary responses y_{ij}, $j = 1, \cdots, n$ are conditionally independent with conditional probability $\mu_{ij}^b = p(y_{ij} = 1|b_i)$, which satisfies

$$\text{logit}(\mu_{ij}^b) = b_i + \beta_0 + \beta_1 r_i + \beta_2 t_{ij},$$

where $b_i \sim f_M(\cdot)$ and M is the value of the order representing the degree of the tuning. Additionally, Zhang and Davidian (2001) showed that order M in the SNP density representation need be no larger than one or two to approximate complicated shapes, including multimodality and skewness, via a simulation experiment. Hence, in this research, our interest lies in testing hypotheses denoted by

$$H_0 : b_i \sim N(0, \sigma^2),$$

and the alternative, H_a, for instance,

$$M = 1,\ f_1(b_i) \propto P_1^2(u_i; a_0, a_1) N\left(b_i; \sigma^2\right) = (a_0 + a_1 u_i)^2 \frac{1}{\sqrt{2\pi}\sigma} \exp\left\{-\frac{b_i^2}{2\sigma^2}\right\};$$
$$M = 2,\ f_2(b_i) \propto P_2^2(u_i; a_0^*, a_1^*, a_2^*) N\left(b_i; \sigma^2\right) = \left(a_0^* + a_1^* u_i + a_2^* u_i^2\right)^2 \frac{1}{\sqrt{2\pi}\sigma} \exp\left\{-\frac{b_i^2}{2\sigma^2}\right\}.$$

Furthermore, for $u_i \sim N(0, 1)$, a reparametrization of $P_M(\cdot)$ using the polar coordinate transformation is useful since it ensures that the integral of $f_M(b_i)$ is equal to one (Chen et al., 2002). After we do this reparametrization, we can obtain

$$\begin{cases} a_0 = \cos(\psi_1), a_1 = \sin(\psi_1); \\ a_0^* = \cos(\psi_1) - \frac{1}{\sqrt{2}} \sin(\psi_1)\sin(\psi_2), a_1^* = \sin(\psi_1)\cos(\psi_2), a_2^* = \frac{1}{\sqrt{2}} \sin(\psi_1)\sin(\psi_2), \end{cases}$$

where $\psi_1, \psi_2 \in \left(-\frac{\pi}{2}, \frac{\pi}{2}\right]$.

Specifically, under the null hypothesis, H_0, we have the marginal mean function

$$\begin{aligned} E_0(y_{ij}) &= E\left\{E(y_{ij}|b_i)\right\} \\ &= E\left\{h(\beta_0+\beta_1 r_i+\beta_2 t_{ij}+b_i)\right\} \\ &= \int h(\beta_0+\beta_1 r_i+\beta_2 t_{ij}+\sigma u_i) f(u_i) du_i \\ &\equiv \mu_0(r_i, t_{ij}; \varphi_0), \end{aligned}$$

where $h(x) = \dfrac{e^x}{1+e^x}$, $\varphi_0 = (\beta_0, \beta_1, \beta_2, \sigma)^T$, and $u_i \sim N(0,1)$. Define $\mu_{0ij} = \mu_0(r_i, t_{ij}; \varphi_0)$ and $\boldsymbol{\mu}_{0i} = (\mu_{0ij})_{1\le j\le n}$ for subject i. The first derivatives of μ_{0ij} are as follows,

$$\begin{aligned} \frac{\partial \mu_{0ij}}{\partial \beta_0} &= \int h'(\beta_0+\beta_1 r_i+\beta_2 t_{ij}+\sigma u_i) f(u_i) du_i, \\ \frac{\partial \mu_{0ij}}{\partial \beta_1} &= r_i \int h'(\beta_0+\beta_1 r_i+\beta_2 t_{ij}+\sigma u_i) f(u_i) du_i, \\ \frac{\partial \mu_{0ij}}{\partial \beta_2} &= t_{ij} \int h'(\beta_0+\beta_1 r_i+\beta_2 t_{ij}+\sigma u_i) f(u_i) du_i, \\ \frac{\partial \mu_{0ij}}{\partial \sigma} &= \int h'(\beta_0+\beta_1 r_i+\beta_2 t_{ij}+\sigma u_i) u_i f(u_i) du_i. \end{aligned}$$

The marginal mean function μ_{0ij} can be approximated by a simple Monte Carlo method (Jiang, 2007),

$$\mu_{0ij} \approx \frac{1}{L}\sum_{l=1}^{L} h(\beta_0+\beta_1 r_i+\beta_2 t_{ij}+\sigma\lambda_{il}) = \frac{1}{L}\sum_{l=1}^{L} \frac{\exp(\beta_0+\beta_1 r_i+\beta_2 t_{ij}+\sigma\lambda_{il})}{1+\exp(\beta_0+\beta_1 r_i+\beta_2 t_{ij}+\sigma\lambda_{il})},$$

where $\lambda_{il}, l = 1, \cdots, L$ are independent $N(0,1)$. Again, similar approximations can be obtained for the first derivatives,

$$\begin{aligned} \frac{\partial \mu_{0ij}}{\partial \beta_0} &\approx \frac{1}{L}\sum_{l=1}^{L} \frac{\exp(\beta_0+\beta_1 r_i+\beta_2 t_{ij}+\sigma\lambda_{il})}{\{1+\exp(\beta_0+\beta_1 r_i+\beta_2 t_{ij}+\sigma\lambda_{il})\}^2}, \\ \frac{\partial \mu_{0ij}}{\partial \beta_1} &\approx \frac{1}{L}\sum_{l=1}^{L} r_i \frac{\exp(\beta_0+\beta_1 r_i+\beta_2 t_{ij}+\sigma\lambda_{il})}{\{1+\exp(\beta_0+\beta_1 r_i+\beta_2 t_{ij}+\sigma\lambda_{il})\}^2}, \end{aligned}$$

$$\frac{\partial \mu_{0ij}}{\partial \beta_2} \approx \frac{1}{L}\sum_{l=1}^{L} t_{ij} \frac{\exp(\beta_0+\beta_1 r_i+\beta_2 t_{ij}+\sigma\lambda_{il})}{\{1+\exp(\beta_0+\beta_1 r_i+\beta_2 t_{ij}+\sigma\lambda_{il})\}^2},$$

$$\frac{\partial \mu_{0ij}}{\partial \sigma} \approx \frac{1}{L}\sum_{l=1}^{L} \lambda_{il} \frac{\exp(\beta_0+\beta_1 r_i+\beta_2 t_{ij}+\sigma\lambda_{il})}{\{1+\exp(\beta_0+\beta_1 r_i+\beta_2 t_{ij}+\sigma\lambda_{il})\}^2}.$$

Then, the generalized estimating equations for estimating φ_0 can be given by

$$G_0^* = \sum_{i=1}^{m} \dot{\boldsymbol{\mu}}_{0i}^T V_{0i}^{-1}(\mathbf{y}_i - \boldsymbol{\mu}_{0i}) = \mathbf{0}, \tag{3.3}$$

where $V_{0i} = Var(\mathbf{y}_i)$ is an unspecified (unknown) $n\times n$ covariance matrix, $1 \le i \le m$;

$$\dot{\boldsymbol{\mu}}_{0i} = \left[\left(\frac{\partial \mu_{0ij}}{\partial \beta_0}\right)_{n\times 1} \left(\frac{\partial \mu_{0ij}}{\partial \beta_1}\right)_{n\times 1} \left(\frac{\partial \mu_{0ij}}{\partial \beta_2}\right)_{n\times 1} \left(\frac{\partial \mu_{0ij}}{\partial \sigma}\right)_{n\times 1} \right].$$

In order to obtain the optimal estimators, we have to know the true covariance V_{0i}. However, in practice, the true V_{0i}s are unknown. We can adopt a method of moments introduced in Section 2.2.2 to estimate V_{0i}s. It can be applied to either a balanced or an unbalanced data set. For instance, when we consider a case with balanced data set, if φ_0 is known, V_{0i} can be estimated by

$$\hat{V}_{0i} = \frac{1}{m}\sum_{i=1}^{m}(\mathbf{y}_i - \boldsymbol{\mu}_{0i})(\mathbf{y}_i - \boldsymbol{\mu}_{0i})^T. \tag{3.4}$$

Further, Jiang (2007) and Jiang et al., (2007) suggested a procedure, namely, iterative estimating equations (IEE) procedure to obtain the optimal estimators, $\hat{\varphi}_0 = (\hat{\beta}_0, \hat{\beta}_1, \hat{\beta}_2, \hat{\sigma})^T$, by iterating between (3.3) and (3.4).

Similarly, under $M = 1$ and $M = 2$, we can define marginal mean functions $\boldsymbol{\mu}_{1i} = (\mu_{1ij})_{1\le j\le n}$ and $\boldsymbol{\mu}_{2i} = (\mu_{2ij})_{1\le j\le n}$ in each subject i with $\mu_{1ij} = \mu_1(r_i, t_{ij}; \varphi_1)$ and $\mu_{2ij} = \mu_2(r_i, t_{ij}; \varphi_2)$ where $\varphi_1 = (\beta_0, \beta_1, \beta_2, \sigma, \psi_1)^T$ and $\varphi_2 = (\beta_0, \beta_1, \beta_2, \sigma, \psi_1, \psi_2)^T$, respectively. Then, we form a set of estimating equations with respect to φ_1 and φ_2 as follows,

$$G_1^*(\varphi_1) = \sum_{i=1}^{m} \dot{\boldsymbol{\mu}}_{1i}^T V_{1i}^{-1}(\mathbf{y}_i - \boldsymbol{\mu}_{1i}) \equiv \sum_{i=1}^{m} \phi_{1i},$$

$$G_2^*(\varphi_2) = \sum_{i=1}^{m} \dot{\boldsymbol{\mu}}_{2i}^T V_{2i}^{-1}(\mathbf{y}_i - \boldsymbol{\mu}_{2i}) \equiv \sum_{i=1}^{m} \phi_{2i}.$$

In addition, we also can derive that

$$\tilde{\mathbf{A}}_{m1} = \sum_{i=1}^{m} \frac{\partial \phi_{1i}}{\partial \varphi_1^T}, \quad \tilde{\mathbf{B}}_{m1} = \sum_{i=1}^{m} \phi_{1i}\phi_{1i}^T,$$

$$\tilde{\mathbf{A}}_{m2} = \sum_{i=1}^{m} \frac{\partial \phi_{2i}}{\partial \varphi_2^T}, \quad \tilde{\mathbf{B}}_{m2} = \sum_{i=1}^{m} \phi_{2i}\phi_{2i}^T,$$

which are shown more details in Appendix A. Let $\hat{\varphi}_0$ be the solution to the set of equations corresponding to $M = 0$, $\hat{\varphi}_{10} = (\hat{\varphi}_0, 0)$, $\hat{\varphi}_{20} = (\hat{\varphi}_0, \mathbf{0}_2)$, $\boldsymbol{\xi}_1$ be the length 4+1 vector equal to $G_1^*(\hat{\varphi}_0, 0)$ and $\boldsymbol{\xi}_2$ to be the length 4+2 vector equal to $G_2^*(\hat{\varphi}_0, \mathbf{0}_2)$. A robust score test statistic is defined by

$$\begin{aligned} \Re_M &= (\boldsymbol{\xi}_M)_M^T (\tilde{\mathbf{A}}_{mM}^{-1}(\hat{\varphi}_{M0}))_M \\ &\quad \times \left[(\tilde{\mathbf{A}}_{mM}^{-1}(\hat{\varphi}_{M0}) \tilde{\mathbf{B}}_{mM}(\hat{\varphi}_{M0}) \tilde{\mathbf{A}}_{mM}^{-1}(\hat{\varphi}_{M0}))_M \right]^{-1} \\ &\quad \times (\tilde{\mathbf{A}}_{mM}^{-1}(\hat{\varphi}_{M0}))_M (\boldsymbol{\xi}_M)_M, \quad M = 1, 2. \end{aligned}$$

Finally, we construct the proposed robust score statistic analogous to the order selection test as follows,

$$T_{RS,m} = \max_{1 \le M \le 2} \frac{\Re_M}{M}$$

for testing normality of random effects in GLMMs. Under the same construction and using results shown in Section 3.2, as $m \to \infty$, $T_{RS,m} \xrightarrow{d} T_{RS}$ with

$$T_{RS} = \max_{1 \le M \le 2} \frac{Q_M}{M},$$

where Q_M is equal to χ_M^2, $\chi_M^2 = s_1^2 + \cdots + s_M^2$ for all M and s_1, s_2, $\ldots$,s_M are identically distributed independent standard normal random variables.

3.3.3 Bootstrap Hypothesis Testing

In this research, we also consider a bootstrap approach to evaluate the performance of the proposed test statistic. Let $\hat{T}_{RS,m}$ denote a test statistic computed from a sample of size m and $\hat{T}^{*(l)}_{RS,m}$ denote a bootstrap statistic computed from the lth bootstrap sample, which is generated under the null hypothesis where $l = 1, 2, \cdots, B$. We assume that the limiting distribution of $\hat{T}^{*(l)}_{RS,m}$ is the same as the limiting distribution of $\hat{T}_{RS,m}$ under the null hypothesis.

In our case, when the distribution of random effects is the normal distribution, the null distribution of the test statistic is approximated by using the conventional parametric bootstrap approach to obtain the bootstrap p-value for assessing the type I error rate. It is detailed as follows:

step 1. Generate a random sample b_i^* from $N(0, \hat{\sigma}^2)$, $i = 1, \cdots, m$.

step 2. Generate a random sample u_{ij} from U(0,1), $i = 1, \cdots, m$; $j = 1, \cdots, n$.

step 3. Construct the bootstrap data set (y^*_{ij}, r_i, t_{ij}) with

$$y^*_{ij} = \begin{cases} 1 \ \text{ if } \ \dfrac{\exp(b_i^* + \hat{\beta}_0 + \hat{\beta}_1 r_i + \hat{\beta}_2 t_{ij})}{\left\{1 + \exp(b_i^* + \hat{\beta}_0 + \hat{\beta}_1 r_i + \hat{\beta}_2 t_{ij})\right\}} > u_{ij} \\ 0 \ \text{ if } \ \dfrac{\exp(b_i^* + \hat{\beta}_0 + \hat{\beta}_1 r_i + \hat{\beta}_2 t_{ij})}{\left\{1 + \exp(b_i^* + \hat{\beta}_0 + \hat{\beta}_1 r_i + \hat{\beta}_2 t_{ij})\right\}} < u_{ij}. \end{cases}$$

step 4. Compute the test statistic $\hat{T}^*_{RS,m}$ from the bootstrap data using all the steps used in computing $\hat{T}_{RS,m}$ from the original data.

step 5. Repeat step 1 to step 4, B times, and reject H_0 at level of significance α if $\hat{T}_{RS,m}$ exceeds the $(1 - \alpha)$ percentile of all bootstrap statistics.

step 6. Repeat step 5 for N simulated data sets to assess the type I error rate.

Formally, for a test that rejects in the upper tail as in step 5, we can compute the bootstrap p-value by

$$P_B^* = 1 - \hat{F}\left(\hat{T}_{RS,m}\right) = 1 - \frac{1}{B}\sum_{l=1}^{B} I\left(\hat{T}_{RS,m}^{*(l)} \leqslant \hat{T}_{RS,m}\right) = \frac{1}{B}\sum_{j=1}^{B} I\left(\hat{T}_{RS,m}^{*(l)} > \hat{T}_{RS,m}\right),$$

where $\hat{F}(\cdot)$ is the empirical distribution function of the bootstrap statistics. When $\alpha(B+1)$ is an integer, step 5 yields exactly the same test as rejecting when P_B^* is less than α.

However, Racine and MacKinnon (2007) argued that when calculating $\hat{T}_{RS,m}$ and $\hat{T}_{RS,m}^{*(l)}$ is computationally burdensome, and if B is chosen poorly, size distortion and loss in power may happen. Therefore, the authors provided a tractable way to perform a classical hypothesis test based on a kernel estimate of the cumulative distribution function of the bootstrap statistics. Their proposed method is to replace P_B^* by the smoothed bootstrap p-value,

$$P_B^h = 1 - \hat{F}_h\left(\hat{T}_{RS,m}\right) = 1 - \frac{1}{B}\sum_{l=1}^{B} K\left(\hat{T}_{RS,m}^{*(l)}, \hat{T}_{RS,m}, h\right),$$

where $K(\cdot,\cdot,\cdot)$ is a cumulative kernel and h is the bandwidth. Moreover, they claimed that the greatest advantage of this proposed method is that it uses the information in the bootstrap statistics more efficiently than the conventional approach and yields a reasonable type I error rate and delivers solid improvements in power when the bootstrap sample size B is small.

Due to the burden of computing our proposed test statistic for testing normality of the random-effects distribution, we shall implement this smoothed bootstrap test procedure to evaluate the type I error rate and the power performance of the proposed test statistic in the simulation study. Furthermore, we also apply it to the analysis of a case study.

3.4 Simulation Study

Some work with the generalized linear mixed model has been on the study of misspecification. It indicates that misspecification of the random-effects distribution associated with a large variance component of random effects has influence on parameter estimation and testing parameters of fixed effects. In the following, we carry out a simulation study with 500 simulated data sets to evaluate the performance of $T_{RS,m}$ that we propose to detect a misspecified random-effects distribution in Section 3.3.2 when a normal random intercept is assumed. Data are generated from the random-intercept logistic model (Litière et al., 2007; Alonso et al., 2008; Litière et al., 2008) given by

$$\text{logit}(\mu_{ij}^{b}) = b_i + \beta_0 + \beta_1 r_i + \beta_2 t_{ij}, \tag{3.5}$$

where y_{ij}s are responses for subject i, collected at time point t_{ij} and r_i is the treatment group to which subject i is allocated with $i = 1, \ldots, m$, $m = 50, 100, 200$; $j = 1, \ldots, n$; $r_i = I(i \leq \frac{m}{2})$; $\beta_0 = -8$, $\beta_1 = 2$ and $\beta_2 = 1$ which is the same setting as in Alonso et al. (2008). For each situation, we determine the proportion of cases in which a significant result is detected at a nominal 5% significance level. When the random effects are generated from a normal distribution, this proportion corresponds to the type I error rate; otherwise, it represents the power of the test. Additionally, in our simulation study, we shall concentrate on scenarios in which variances of the random-effects distribution (σ^2) are large or extremely large, such as 16, 32, 48 or 64.

3.4.1 Type I Error Rate via Asymptotic Results

In this part, we evaluate the influence of cluster size and magnitude of the variance component of random effects on the quality of asymptotic results for the proposed test statistic. We consider three cluster sizes (n), namely, number of repeated measurements per cluster (or subject) with different time points t_{ij} as follows:

$$(1)\ n = 5, \quad t_{ij} = (-0.1, 0, 0.1, 0.2, 0.3),$$

$$(2)\ n = 10, \quad t_{ij} = (-0.4, -0.3, -0.2, -0.1, 0, 0.1, 0.2, 0.3, 0.4, 0.5),$$

$$(3)\ n = 15, \quad t_{ij} = (-0.7, -0.6, -0.5, -0.4, -0.3, -0.2, -0.1, 0, 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7).$$

The results shown in Table 1 indicate that the test exhibits a reasonable type I error rate in most sample sizes and cluster sizes with an extremely large variance of the random-effects distribution ($\sigma^2 = 64$). However, in the scenario of less large or large variance ($\sigma^2 = 16$ or 32), it shows a considerable inflation on the type I error rate even if we try to enhance the amount of information available by increasing the cluster size or sample size.

Table 1. Results of the type I error rate using the asymptotic distribution of $T_{RS,m}$ for three cluster sizes when σ^2=16, 32, 48 and 64.

		$\sigma^2 = 16$			$\sigma^2 = 32$		
Distribution	n	$m = 50$	100	200	$m = 50$	100	200
	5	0.132	0.196	0.234	0.058	0.072	0.120
Normal	10	0.106	0.200	0.184	0.074	0.112	0.200
	15	0.118	0.180	0.210	0.092	0.114	0.192
		$\sigma^2 = 48$			$\sigma^2 = 64$		
Distribution	n	$m = 50$	100	200	$m = 50$	100	200
	5	0.018	0.046	0.072	0.016	0.048	0.060
Normal	10	0.032	0.072	0.104	0.036	0.058	0.096
	15	0.046	0.072	0.130	0.036	0.062	0.082

In other words, when the variance component of random effects becomes extremely large and either cluster size or sample size is moderate, the proposed test $T_{RS,m}$ with the ordinary p-value will tend to be less liberal. In this situation, the proposed test is appropriate for testing normality of random effects when the true distribution of random effects is unknown.

3.4.2 Type I Error Rate via Smoothed Bootstrap Test

In order to improve the inflation of the type I error rate shown in Section 3.4.1, the smoothed bootstrap test procedure introduced in Section 3.3.3 is adopted to evaluate the type I error rate of the proposed test $T_{RS,m}$. Additionally, Racine and MacKinnon (2007), in terms of their simulation experiment, observed that a smoothed bootstrap test will overreject when h is sufficiently small and underreject when h is sufficiently large and the smoothed p-value becomes less sensitive to h when the bootstrap sample size (B) increases. Herein, to avoid overrejection when h is too small and underrejection when h is too large, we try selected bandwidths not far from $h = 1.575B^{-4/9}$ which Racine and MacKinnon (2007) used when the test statistic was asymptotically the standard normal, and the cumulative standard Gaussian kernel is adopted. For each simulated data set, the smoothed bootstrap p-value is calculated by

$$K(w) = \int_{-\infty}^{w} k(W)dW,$$

where $k(W)$ is the standard normal density and $w = (\hat{T}_{RS,m} - \hat{T}^{*(l)}_{RS,m})/h$.

First, we use the bootstrap sample size $B = 60$ to conduct a small experiment for the cluster size $n = 5$ and $t_{ij} = (-0.1, 0, 0.1, 0.2, 0.3)$. Table 2 reveals that there is no severe overrejection or underrejection and the inflation of type I error rate shown in Table 1 has been improved, even for the less large variance scenario (σ^2 =16).

Table 2. Results of the type I error rate using the smoothed bootstrap test procedure with bootstrap sample size B=60 for cluster size n=5 and t_{ij}=$(-0.1, 0, 0.1, 0.2, 0.3)$ when σ^2=16, 32 and 64.

		$\sigma^2 = 16$			$\sigma^2 = 32$		
Distribution	h	$m = 50$	100	200	$m = 50$	100	200
Normal	0.10	0.068	0.064	0.062	0.052	0.056	0.048
	0.25	0.068	0.064	0.062	0.054	0.056	0.048
	0.50	0.068	0.064	0.064	0.052	0.056	0.048

		$\sigma^2 = 64$		
Distribution	h	$m = 50$	100	200
Normal	0.10	0.046	0.040	0.038
	0.25	0.046	0.042	0.036
	0.50	0.036	0.042	0.036

Second, we consider another situation with a data set that has cluster size $n = 6$ with $t_{ij} = (0, 1, 2, 4, 6, 8)$ (Alonso et al., 2008). Table 3 also indicates that the type I error rate has been well-controlled compared with the prespecified 5% significance level for all sample sizes. Moreover, in Table 4, we observe that performance of the proposed test statistic under the smoothed bootstrap test procedure seems to become much better when the bootstrap sample size increases. Overall speaking, we believe that the proposed test $T_{RS,m}$ with the smoothed bootstrap p-value is reliable for testing normality of the random-effects distribution on this specified model (3.5).

Table 3. Results of the type I error rate using the smoothed bootstrap test procedure with bootstrap sample size B=60 for cluster size n=6 and t_{ij}=$(0, 1, 2, 4, 6, 8)$ when σ^2=16 and 32.

		$\sigma^2 = 16$			$\sigma^2 = 32$		
Distribution	h	$m = 50$	100	200	$m = 50$	100	200
	0.10	0.038	0.038	0.038	0.040	0.062	0.054
Normal	0.25	0.036	0.040	0.038	0.042	0.062	0.054
	0.50	0.038	0.040	0.042	0.040	0.062	0.056

Table 4. Results of the type I error rate using the smoothed bootstrap test procedure with bootstrap sample size B=100 for cluster size n=6 and t_{ij}=$(0, 1, 2, 4, 6, 8)$ when σ^2=16 and 32.

		$\sigma^2 = 16$		$\sigma^2 = 32$	
Distribution	h	$m = 50$	100	$m = 50$	100
	0.10	0.044	0.048	0.054	0.056
	0.25	0.042	0.048	0.054	0.060
Normal	0.50	0.042	0.048	0.054	0.056
	0.75	0.040	0.046	0.056	0.056
	1.00	0.040	0.046	0.056	0.054

3.4.3 Power Analysis via Smoothed Bootstrap Test

In this part, we evaluate the power performance of the proposed test statistic $T_{RS,m}$. Herein, six alternative distributions of random effects, including mixture, skewed or heavy-tailed distributions, are generated and adjusted with variance 16 or 32. These distributions are listed as follows:

(1) Mixture normal distribution of $0.3N(-1, 1^2) + 0.7N(6, (21.41/0.7)^2)$,

(2) Skewed mixture distribution of $0.25N(14, 10^2) + 0.75\chi^2(4)$,

(3) Discrete distribution with $p(b_i = 1) = 1/6$, $p(b_i = 2) = 1/2$ and $p(b_i = 3) = 1/3$,

(4) Gamma distribution with shape 1/2 and scale 8,

(5) t distribution with degrees of freedom 4,

(6) Lognormal distribution.

In addition, all simulation results are based on the bootstrap smoothed test procedure with bootstrap sample size $B = 60$ and generated data with cluster size $n = 6$ and $t_{ij} = (0, 1, 2, 4, 6, 8)$. Table 5 exhibits a large power for bimodal or multimodal alternatives, especially, when the random-effects distribution possibly follows a skewed mixture distribution or a discrete distribution with 3 support points.

Table 5. Results of the power performance of bimodal and multimodal alternatives using the smoothed bootstrap test procedure with bootstrap sample size B=60 for cluster size n=6 and t_{ij}=$(0, 1, 2, 4, 6, 8)$ when σ^2=16 and 32.

		$\sigma^2 = 16$			$\sigma^2 = 32$		
Distribution	h	$m = 50$	100	200	$m = 50$	100	200
Mixture Normal	0.10	0.366	0.568	0.622	0.480	0.632	0.744
	0.25	0.366	0.576	0.626	0.496	0.642	0.742
	0.50	0.364	0.576	0.618	0.500	0.648	0.734
Skewed Mixture	0.10	0.750	0.774	0.802	0.790	0.840	0.902
	0.25	0.756	0.782	0.802	0.800	0.846	0.900
	0.50	0.760	0.786	0.806	0.808	0.852	0.900
Discrete	0.10	0.814	0.842	0.874	0.906	0.936	0.954
	0.25	0.814	0.844	0.874	0.908	0.936	0.952
	0.50	0.824	0.844	0.874	0.908	0.936	0.952

However, in Table 6, it shows that when the random-effects distribution possibly follows a skewed or heavy-tailed unimodal alternative, the proposed test $T_{RS,m}$ has poor power for testing normality of random effects for this specified model (3.5), especially in a scenario of the less large variance (σ^2=16) t distribution.

Table 6. Results of the power performance of skewed and heavy-tailed alternatives using the smoothed bootstrap test procedure with bootstrap sample size B=60 for cluster size n=6 and t_{ij}=$(0, 1, 2, 4, 6, 8)$ when σ^2=16 and 32.

		$\sigma^2 = 16$			$\sigma^2 = 32$		
Distribution	h	$m = 50$	100	200	$m = 50$	100	200
Gamma	0.10	0.076	0.082	0.086	0.078	0.082	0.096
	0.25	0.078	0.082	0.088	0.078	0.084	0.096
	0.50	0.076	0.080	0.086	0.080	0.084	0.096
t	0.10	0.058	0.062	0.074	0.070	0.072	0.094
	0.25	0.058	0.060	0.074	0.068	0.076	0.098
	0.50	0.058	0.060	0.070	0.068	0.076	0.102
Lognormal	0.10	0.078	0.082	0.088	0.094	0.094	0.100
	0.25	0.080	0.082	0.088	0.092	0.098	0.104
	0.50	0.080	0.082	0.088	0.092	0.098	0.110

Nevertheless, Litière et al.(2008) showed that an asymmetric mixture of two normals random-effects distribution has more severe influence than some non-normal unimodal random-effects distributions on maximum likelihood estimators of the between-subject effect under assumption of normality of random effects. As a result, it is necessary that a proposed test has power for detecting this misspecified random-effects distribution. Fortunately, our proposed test is very powerful for detecting this type of misspecification of the random-effects distribution.

3.5 Application

In this section, we shall apply our proposed test to revisit a data set obtained from a case study in mental health (Alonso et al., 2004). Alonso et al. (2008) had applied their proposed tests on testing normality of the random-effects distribution to this data set. The data collection is briefly sketched as follows. In this case study, the authors studied the effect of risperidone compared to an active control for the treatment of chronic schizophrenia on 128 patients. During the period of the trial, half of patients were assigned to the treatment group; the others were assigned to the control group. Each patient was measured at the following time points: 0, 1, 2, 4, 6 and 8 weeks. For each measurement, the patient's mental condition was classified as normal to mildly ill ($y = 1$) or moderately to severely ill ($y = 0$) where y is a binary response variable. In our analysis, we consider two data structures as follows:

(1) The balanced data is the part of original data set (66 patients) without any missing measured responses at each time point.

(2) The unbalanced data set is the original data set (128 patients) where some patients had missing observed responses at some time points.

Earlier, someone proposed a straightforward diagnostic tool, using empirical Bayes (EB) estimates of random effects to detect departures from normality. We review this basic idea as described by Molenberghs and Verbeke (2005, page 268) as follows. Let us recall the conditional density of response variables shown in Section 2.1. The estimation of random effects is based on their posterior distribution with the following density distribution,

$$f_i(\mathbf{b}_i|\mathbf{y}_i, \boldsymbol{\beta}, D, \phi) = \frac{f_i(\mathbf{y}_i|\mathbf{b}_i, \boldsymbol{\beta}, \phi)f(\mathbf{b}_i|D)}{\int f_i(\mathbf{y}_i|\mathbf{b}_i, \boldsymbol{\beta}, \phi)f(\mathbf{b}_i|D)d\mathbf{b}_i}, \quad i = 1, 2, \cdots, m.$$

The estimator of random effects, $\hat{\mathbf{b}}_i$, is the value of $\mathbf{b}_i$ that maximizes $f_i(\mathbf{b}_i|\mathbf{y}_i, \boldsymbol{\beta}, D, \phi)$ where the unknown parameters, $\boldsymbol{\beta}$, D and ϕ, have been replaced by their maximum likelihood estimators based on the likelihood function shown in Section 2.2.

Alonso et al. (2008) provided a histogram plot of the EB estimates of random effects based on the specified model (3.5) for the unbalanced (original) data set shown on the right of Figure 1. Herein, we also provide a histogram plot of the EB estimates for the balanced data set shown on the left of Figure 1. Obviously, it seems to reveal that normality of random effects is questionable on these two data sets. However, it can be shown that in generalized linear mixed model, the empirical Bayes estimates no longer follow a normal distribution, even while the random-effects distribution is correctly specified as normal (Litière et al., 2007; Alonso et al., 2008).

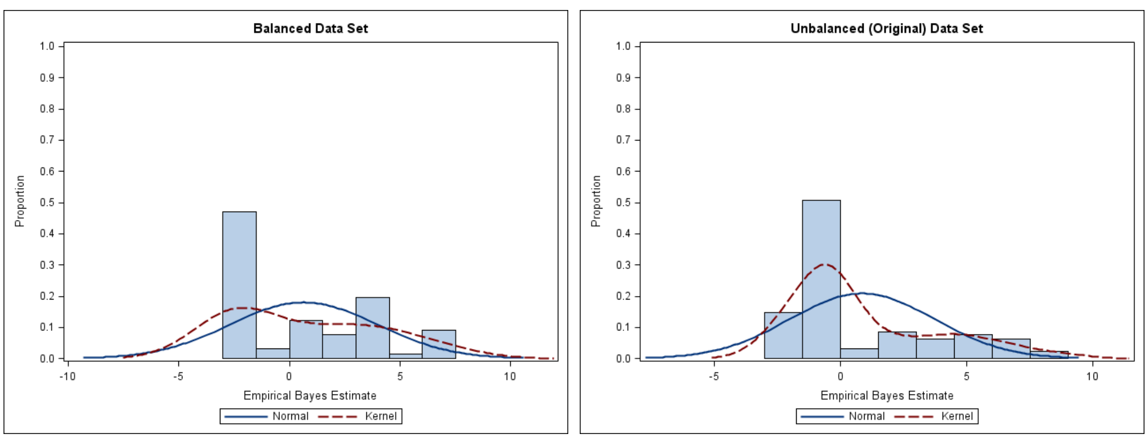

Figure 1. The distribution of empirical Bayes estimates of random effects based on the specified model (3.5) for balanced and unbalanced data sets in mental health study.

Therefore, in order to avoid making an error, we should use a formal procedure to test normality of the random-effects distribution. In the following, we apply our proposed test statistic and adopt the smoothed bootstrap test procedure to test normality of random-effects by approximating the random-effects distribution by the SNP Hermite expansion in the random-intercept logistic model (3.5) shown in Section 3.4. Furthermore, in practice, for the purpose of obtaining a more reliable smoothed bootstrap p-value, we can apply a bandwidth selection method that Racine and MacKinnon (2007) suggested, maximum likelihood cross-validation, to search an appropriate bandwidth for smoothing the cumulative distribution function of the bootstrap test statistics. It can be implemented by using a package, npudist(np), in R and the smoothed bootstrap p-value is calculated based on the selected bandwidth.

Parameter estimates shown on Table 7 are obtained from fitting the specified model (3.5) under the assumption of normally distributed random effects (H_0) that we showed in Section 3.3.2. The proposed test statistic leads to $\hat{T}_{RS,m} = 2.2814$ for the balanced data set and $\hat{T}_{RS,m} = 13.9173$ for the unbalanced data set.

Table 7. Estimates of parameters under IEE procedure and the proposed test statistic value, $\hat{T}_{RS,m}$, on testing normality of random effects for balanced and unbalanced data sets in mental health study.

	Intercept	Treatment Effect	Time Effect	Standard Deviation (b_i)	Test Statistic
	$\hat{\beta}_0$	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\sigma}$	$\hat{T}_{RS,m}$
Balanced Data Set	-5.6109	1.3887	0.6245	4.3026	2.2814
Unbalanced Data Set	-8.2952	2.1769	0.4051	5.0805	13.9173

Then, we carry out the smoothed bootstrap test procedure to test normality of the random-effects distribution. On this step, we not only calculate the smoothed bootstrap p-value but also inspect the stability of the bootstrap parameter estimates.

Results of the bootstrap parameter estimates are shown on Table 8 and Table 9 for balanced and unbalanced data sets, respectively. Overall speaking, when the bootstrap size increases, the bootstrap parameter estimates become consistent for both types of data sets; however, the variability of the bootstrap parameter estimates are not very stable for the treatment effect, namely, the between-subject effect for the unbalanced data set.

Table 8. Mean and standard deviation (S.D.) of bootstrap parameter estimates with bootstrap sample size B=50, 100, 200, 300, 400 and 500 for the balanced data set in mental health study.

Bootstrap Size B	Intercept $\hat{\beta}_0^{boot}$	Treatment Effect $\hat{\beta}_1^{boot}$	Time Effect $\hat{\beta}_2^{boot}$	Standard Deviation (b_i) $\hat{\sigma}^{boot}$
50	-5.4827	1.1668	0.6292	3.8264
	(0.8536)	(0.7039)	(0.1004)	(0.6152)
100	-5.5574	1.2039	0.6292	3.8633
	(0.8436)	(0.5952)	(0.1027)	(0.6181)
200	-5.4758	1.2338	0.6255	3.8600
	(0.8236)	(0.5142)	(0.1138)	(0.6209)
300	-5.4227	1.2320	0.6183	3.8449
	(0.8411)	(0.4991)	(0.1138)	(0.6328)
400	-5.3873	1.2094	0.6155	3.8310
	(0.8204)	(0.5143)	(0.1098)	(0.6125)
500	-5.3965	1.2200	0.6164	3.8466
	(0.8290)	(0.4950)	(0.1114)	(0.6208)

*S.D. of bootstrap parameter estimates are within parentheses.

Table 9. Mean and standard deviation (S.D.) of bootstrap parameter estimates with bootstrap sample size B=50, 100, 200, 300, 400 and 500 for the unbalanced data set in mental health study.

Bootstrap Size B	Intercept $\hat{\beta}_0^{boot}$	Treatment Effect $\hat{\beta}_1^{boot}$	Time Effect $\hat{\beta}_2^{boot}$	Standard Deviation (b_i) $\hat{\sigma}^{boot}$
50	-7.6770	2.1603	0.3622	4.9766
	(1.4157)	(0.0137)	(0.1293)	(1.0344)
100	-7.8241	2.1619	0.3772	5.0622
	(1.3785)	(0.0135)	(0.1296)	(1.0463)
200	-7.8916	2.1612	0.3692	5.1815
	(1.4206)	(0.0146)	(0.1449)	(1.0852)
300	-7.8428	2.1581	0.3652	5.1712
	(1.4219)	(0.0508)	(0.1424)	(1.0784)
400	-7.7877	2.1581	0.3577	5.1430
	(1.4253)	(0.0446)	(0.1436)	(1.0774)
500	-7.7700	2.1547	0.3558	5.1412
	(1.4439)	(0.0899)	(0.1468)	(1.1011)

*S.D. of bootstrap parameter estimates are within parentheses.

Additionally, Figure 2 and Figure 3 show empirical and smooth kernel estimates of a distribution function for the bootstrap test statistics for balanced and unbalanced data sets, respectively. For the unbalanced data set, it is clear that the smooth kernel estimate is much better than the estimate based on the empirical distribution function, especially, when the bootstrap sample size is small. For instance, when $B = 50$, for the test statistic, 8, the empirical p-value is 0.000, while the smoothed p-value is 0.042; for the test statistic, 13.917, the empirical p-value is 0.000, while the smoothed p-value is 0.018.

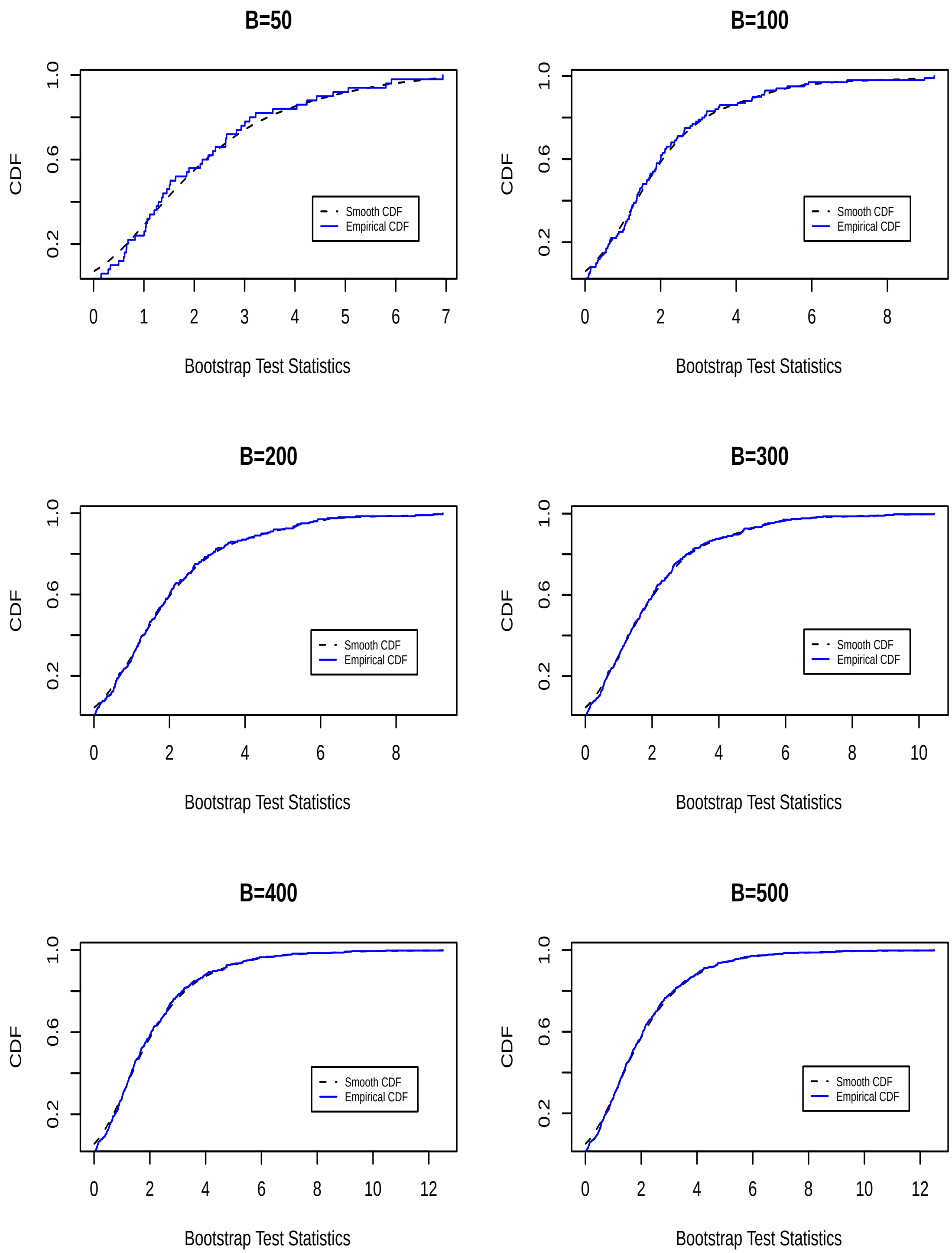

Figure 2. Empirical and kernel estimates of a distribution function for bootstrap test statistics with bootstrap sample size $B = 50$, 100, 200, 300, 400 and 500 in the analysis of the balanced data set in mental health study.

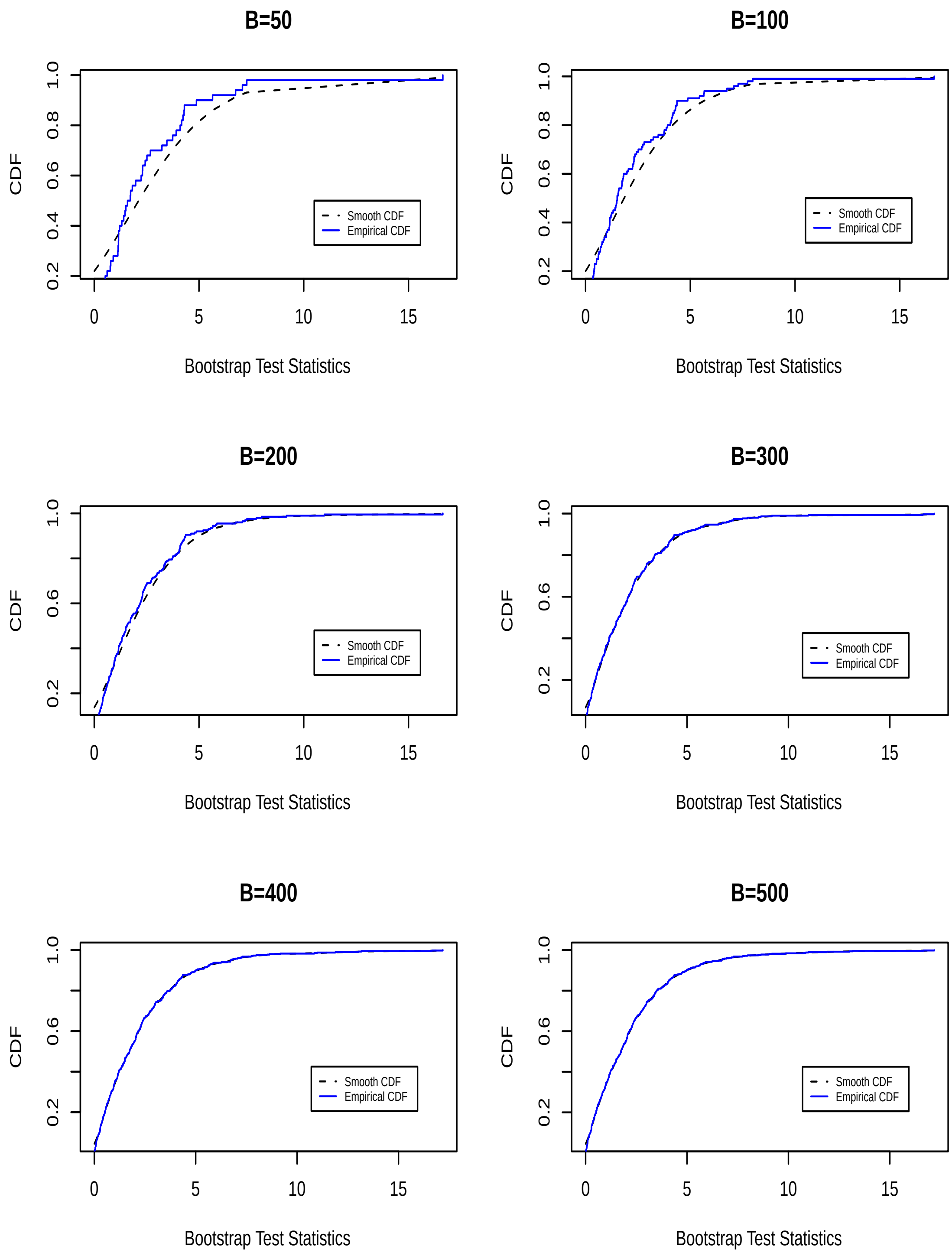

Figure 3. Empirical and kernel estimates of a distribution function for bootstrap test statistics with bootstrap sample size $B = 50$, 100, 200, 300, 400 and 500 in the analysis of the unbalanced data set in mental health study.

Further, from the analyzed results exhibited on Table 10, we observe that the smoothed bootstrap p-values become much more stable when the bootstrap sample size increases. No matter which bootstrap sample size, it provides the same decision with only a small difference in the magnitude. The results suggest that the balanced data set provides insufficient evidence that the random-effects distribution departs from normality. On the other hand, the unbalanced data set provides sufficient evidence at 5% significance level that normality of the random-effects distribution does not hold. Interestingly, the conclusion based on our proposed test for the unbalanced (original) data set is different from that (p-value > 0.75) of Alonso et al. (2008) who assumed that the missing data generating mechanism is missing at random (MAR) making their likelihood approach a valid option. Since the sample size of this case study is not large and there are missing values, the results based on our proposed test statistic and bootstrap test procedure may be more convincing.

Table 10. Results of normality test of the distribution of random effects under the smoothed bootstrap procedure in mental health study.

	Bootstrap Sample Size (B)					
	50	100	200	300	400	500
	Balanced Data Set					
Smoothed p-value	0.3886	0.3480	0.3443	0.3457	0.3651	0.3624
Selected bandwidth	0.6493	0.4781	0.3699	0.3868	0.5001	0.4720
	Unbalanced Data Set					
Smoothed p-value	0.0179	0.0095	0.0049	0.0067	0.0050	0.0040
Selected bandwidth	2.1799	1.6152	1.0017	0.4024	0.2675	0.2859

3.6 Discussion

In this chapter, we investigate a diagnostic test for the random-effects distribution without assuming any parametric form. We propose a robust score test analogous to the order selection test that is not likelihood-based for testing the distributional assumptions on the random-effects distribution in generalized linear mixed models. The proposed test statistic involved generalized estimating equations and an estimating procedure (IEE) that can be applied to analyze either a balanced or an unbalanced data set. However, the IEE procedure (Jiang, 2007; Jiang et al., 2007) is limited only to the situation where the outcomes are independently clustered.

Through a simulation study, we discover that the smoothed bootstrap test (Racine and MacKinnon, 2007) combined with a parametric bootstrap procedure can indeed reduce or eliminate the inflation of the type I error rate under non-optimal bandwidths. Even so, the issue of finding the optimal bandwidth may need to be explored. In addition, results of preliminary studies of power demonstrate satisfactory power to detect misspecification of the random-effects distribution when unobserved random effects have a distribution with obvious multiple modes; on the other hand, it does not hold for a unimodal distribution of random effects which is highly skewed or heavy-tailed in a specified random-intercept logistic model. As a result, it also would be useful to study whether similar results hold for more complex models when a complex experiment is designed for clinical research.

Moreover, since we just focus on the balanced data structure in our simulation study, for an unbalanced data set where there are missing outcomes, it is not clear how reliable the proposed test is, especially with a high percentage of missing outcomes. Nevertheless, in this situation, we believe that our proposed non-likelihood test may be more reliable than a likelihood-based test. Overall, except for possible restrictions described above, no matter whether the sample size is small or moderate, our proposed test statistic using the smoothed bootstrap test procedure performs well for testing misspecification of the random-effects distribution, especially when the true distribution of random effects is a mixture.

CHAPTER IV

LOCAL POLYNOMIAL SMOOTHED RESIDUALS APPLIED TO ASSESS THE LOGISTIC MULTILEVEL MODEL

4.1 Introduction

In recent years, mixed-effects logistic models have been widely used for analyzing clustered binary data or naturally hierarchy data. Again, estimation and inference depend on the model being correctly specified. Therefore, methods for assessment of model fit need to be well developed. Evans and Hosmer (2004) extended summary statistics used to assess goodness-of-fit in the ordinary logistic regression, such as unweighted sums of squares and Pearson statistics, to the mixed-effects logistic model. Their simulation results indicated that the performance of the type I error rate is not good in some situations. Pan and Lin (2005) developed model-checking techniques for generalized linear mixed models based on the cumulative sums of residuals over covariates or predicted values of the response variable. They indicated that a faulty functional form of a fixed covariate may cause a plot of the cumulative residuals against the predicted values to exhibit a systematic tendency.

Additionally, Sturdivant (2005) and Sturdivant and Hosmer (2007) proposed a kernel smoothed unweighted sum of squares statistic by smoothing residuals in the y-space to assess the adequacy of the logistic multilevel (hierarchical) regression model, namely, a mixed-effects logistic model for hierarchical data. Their simulation results demonstrated satisfactory adherence of the type I error rate of their proposed statistic. However, especially, for settings with fewer subjects per cluster, the simulation results showed very limited or no power to detect a missing quadratic term of fixed effects. Moreover, Lin et al. (2008) adopted a local nonparametric smoothing method for assessing population-averaged models with longitudinal binary data. Their proposed test statistic is based on smoothing the standardized residuals and has satisfactory power property for models with both categorical and continuous covariates. One of its advantages is that it can avoid the dependence on the kernel weights obtained by the partition of the predicted probabilities from the fitted model or the covariate space (Hosmer et al., 1997).

In this chapter, we extend the nonparametric local polynomial smoothed residuals over continuous covariates to the unweighted sum of squares statistic for assessing the goodness-of-fit in the logistic multilevel regression model. We investigate whether it can improve and enhance the power performance for detecting some specified alternative models. The remainder of this chapter is organized as follows. In Section 4.2, we review a kernel smoothed unweighted sum of squares statistic by smoothing residuals over y-space in the logistic multilevel regression models and briefly introduce how to implement the multivariate local polynomial smoothing technique on smoothing residuals over continuous and within-cluster covariates. In Section 4.3, a simulation study is performed to evaluate the type I error rate of the kernel smoothed unweighted sum of squares statistic by using the local polynomial smoothed residuals and the power analysis for detecting a missing quadratic or interaction term of the fixed effects. Finally, we carry out an application to a real data set.

4.2 Goodness-of-fit Tests for the Logistic Multilevel Models

4.2.1 Multilevel Models for Binary Data

We first consider a two-level model for binary outcomes with a single covariate. For instance, in the field of education, suppose we have data consisting of students (level one) indexed by $j = 1, \cdots, n_i$ in different districts (level two) indexed by $i = 1, \cdots, m$.

We observe y_{ij}, a binary response for student j in district i and x_{ij}, a covariate at the student level. The combined two-level logistic model accounting for the clustering structure by adding a random intercept across level two, namely cluster, can be written as

$$\text{logit}(\mu_{ij}^b) = \beta_{0i} + \beta_1 x_{ij}, \tag{4.1}$$

where $\beta_{0i} = \beta_0 + b_{0i}$ with $b_{0i} \sim N(0, \sigma_0^2)$. When adding both a random intercept and a random slope across level two, it can be written as

$$\text{logit}(\mu_{ij}^b) = \beta_{0i} + \beta_{1i} x_{ij}, \tag{4.2}$$

where $\beta_{0i} = \beta_0 + b_{0i}$ with $b_{0i} \sim N(0, \sigma_0^2)$ and $\beta_{1i} = \beta_1 + b_{1i}$ with $b_{1i} \sim N(0, \sigma_1^2)$. Additionally, for instance, model (4.2) can be rewritten as follows,

$$\text{logit}(\mu_{ij}^b) = (\beta_0 + \beta_1 x_{ij}) + b_{0i} + b_{1i} x_{ij}$$

or

$$\text{logit}(\boldsymbol{\mu}^b) = \mathbf{x}\boldsymbol{\beta} + \mathbf{z}\mathbf{b},$$

where $\mathbf{z}$ is a $N \times 2m$ design matrix for the random effects with $N = n_1 + n_2 + \cdots + n_m$, $\mathbf{b}$ is a $2m \times 1$ vector of random effects and $\mathbf{b} \sim N(\mathbf{0}, \boldsymbol{\Omega})$ with a block diagonal covariance matrix.

The standard assumption is that conditional on the random effects, at level one, $y_{ij} \sim \text{Bernoulli}(\mu_{ij}^b)$ and y_{ij}s are independent. More specifically, one can consider the decomposition (Molenberghs and Verbeke, 2005),

$$y_{ij} = \mu_{ij}^b + \epsilon_{ij} \Leftrightarrow \mathbf{y} = \boldsymbol{\mu}^b + \boldsymbol{\varepsilon},$$

where $\mathbf{y}$ is a $N \times 1$ vector of the binary responses with $N = n_1 + n_2 + \cdots + n_m$, $\boldsymbol{\mu}^b$ is the vector of the related probabilities, and $\boldsymbol{\varepsilon}$ has mean zero and variance given by the diagonal matrix of binomial variances conditional on the random effects.

For parameter estimation, we can consider the use of Bayesian techniques, algorithms involving quasi-likelihood, or numerical integration, such as Gaussian quadrature, to optimize the likelihood function. In practice, procedures for estimating parameters of the logistic multilevel model are available in many statistical software packages, such as R, SAS, Stata, and so on. One of the popular methods involves the penalized quasi-likelihood (PQL) estimation in conditional models (Breslow and Clayton, 1993) which was reviewed in Section 2.2. Although this procedure may suffer from some bias in parameter estimates, there are some suggested methods to reduce the bias (Lin and Breslow, 1996; Goldstein and Rasbash, 1996). Nevertheless, an advantage of this method is that it is easily implemented and involves less computational effort. Through a simulation experiment, Austin (2010) showed that except for some special situations, parameter estimation under this method is still reliable.

Evans and Hosmer (2004) and Sturdivant and Hosmer (2007) used the SAS GLIMMIX macro to carry out their study. In this research, we shall use glmmPQL in statistical package R to implement a version of PQL estimation. It also can be used to fit the logistic multilevel model with multivariate normal random effects.

4.2.2 Goodness-of-Fit Test Statistic

There are various goodness-of-fit test statistics available in the ordinary logistic regression model (Hosmer et al., 1997). Sturdivant (2005) extended some of these statistics by using smoothed residuals in the logistic multilevel model. One of them is the Pearson kernel smoothed statistic and the other one is based on the unweighted sum of squares statistic given by

$$S = \sum_{i=1}^{m} \sum_{j=1}^{n_i} (y_{ij} - \hat{\mu}_{ij}^{b})^2. \tag{4.3}$$

Sturdivant (2005) and Sturdivant and Hosmer (2007) not only modified the unweighted sum of squares statistic (4.3) by using the smoothed rather than raw residuals but also produced expressions to approximate the moments of the statistic by using the approximation of the residuals in terms of the level one errors. Then, the authors used these moments to form a standardized statistic, which should have an asymptotic standard normal distribution when the model is correctly specified. Herein, we shall review and sketch some of their derived results.

Firstly, under model (4.2), an approximation of the estimated residual in terms of the residual using the penalized quasi-likelihood (PQL) parameter estimation is given by

$$\begin{aligned} \hat{\boldsymbol{e}} = \mathbf{y} - \hat{\boldsymbol{\mu}}^b &\approx \mathbf{y} - \left[\boldsymbol{\mu}^b - \mathbf{g} + \mathbf{M}(\mathbf{y} - \boldsymbol{\mu}^b)\right] \\ &= (\mathbf{I} - \mathbf{M})(\mathbf{y} - \boldsymbol{\mu}^b) + \mathbf{g}, \end{aligned}$$

where $\mathbf{M} = \mathbf{AQ}[\mathbf{Q}^T\mathbf{AQ} + \boldsymbol{\Sigma}]^{-1}\mathbf{Q}^T$ with $\mathbf{Q} = [\mathbf{x}\ \mathbf{z}]$ and $\mathbf{A} = \mathrm{diag}[\mu_{ij}^b(1 - \mu_{ij}^b)]$, $\mathbf{g} = \mathbf{AQ}[\mathbf{Q}^T\mathbf{AQ} + \boldsymbol{\Sigma}]^{-1}\boldsymbol{\Sigma}\boldsymbol{\delta}$ with $\boldsymbol{\delta} = [\boldsymbol{\beta}\ \mathbf{b}]^T$, and

$$\boldsymbol{\Sigma} = \begin{bmatrix} \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \boldsymbol{\Omega}^{-1} \end{bmatrix}.$$

Then, the vector of kernel smoothed residuals in the y-space can be given by

$$\hat{\boldsymbol{e}}_m = \boldsymbol{\Gamma}\hat{\boldsymbol{e}},$$

where $\boldsymbol{\Gamma}$ is the matrix of smoothing weights as follows,

$$\begin{bmatrix} \lambda_{11} & \lambda_{12} & \cdots & \lambda_{1N} \\ \vdots & \vdots & \vdots & \vdots \\ \lambda_{N1} & \lambda_{N2} & \cdots & \lambda_{NN} \end{bmatrix},$$

with $N = n_1 + n_2 + \cdots + n_m$ and λ_{ij} is produced by using the kernel density,

$$\lambda_{ij} = \frac{K\left(\dfrac{|\hat{\mu}_i^b - \hat{\mu}_j^b|}{h}\right)}{\sum_j K\left(\dfrac{|\hat{\mu}_i^b - \hat{\mu}_j^b|}{h}\right)},$$

where $K(\cdot)$ is the kernel density function and h is the bandwidth weighting the residuals.

Secondly, the kernel smoothed unweighted sum of squares statistic is defined as

$$\begin{aligned} S_m &= \hat{\boldsymbol{e}}_m^T \hat{\boldsymbol{e}}_m \\ &\approx \left[(\mathbf{I} - \hat{\mathbf{M}})(\mathbf{y} - \boldsymbol{\mu}^b) + \hat{\mathbf{g}}\right]^T \boldsymbol{\Gamma}^T \boldsymbol{\Gamma} \left[(\mathbf{I} - \hat{\mathbf{M}})(\mathbf{y} - \boldsymbol{\mu}^b) + \hat{\mathbf{g}}\right], \end{aligned}$$

with the approximate mean and variance as the following,

$$\begin{aligned} E(S_m) &= tr\left\{(\mathbf{I} - \hat{\mathbf{M}})^T \boldsymbol{\Gamma}^T \boldsymbol{\Gamma} (\mathbf{I} - \hat{\mathbf{M}}) \hat{\mathbf{A}}\right\} + \hat{\mathbf{g}}^T \boldsymbol{\Gamma}^T \boldsymbol{\Gamma} \hat{\mathbf{g}}, \\ Var(S_m) &= Var(\boldsymbol{e}^T \boldsymbol{\psi} \boldsymbol{e}) + Var(\boldsymbol{q}^T \boldsymbol{e}) + 2Cov(\boldsymbol{e}^T \boldsymbol{\psi} \boldsymbol{e}, \boldsymbol{q}^T \boldsymbol{e}) \\ &= \sum_{i=1}^{N} \left\{\psi_{ii}^2 \hat{a}_i (1 - 6\hat{a}_i)\right\} + 2tr\left(\boldsymbol{\psi} \hat{\mathbf{A}} \boldsymbol{\psi} \hat{\mathbf{A}}\right) + \boldsymbol{q}^T \hat{\mathbf{A}} \boldsymbol{q} \\ &\quad + 2\sum_{i=1}^{N} \psi_{ii} q_i \hat{\mu}_i^b (1 - \hat{\mu}_i^b)(1 - 2\hat{\mu}_i^b), \end{aligned}$$

where $\boldsymbol{e} = \mathbf{y} - \boldsymbol{\mu}^b$, $\boldsymbol{\psi} = (\mathbf{I} - \hat{\mathbf{M}})^T \boldsymbol{\Gamma}^T \boldsymbol{\Gamma} (\mathbf{I} - \hat{\mathbf{M}})$, $\boldsymbol{q}^T = 2\hat{\mathbf{g}}^T \boldsymbol{\Gamma}^T \boldsymbol{\Gamma} (\mathbf{I} - \hat{\mathbf{M}})^T$, ψ_{ii}, and $\hat{a}_i$ are the ith diagonal elements of matrices $\boldsymbol{\psi}$ and $\hat{\mathbf{A}}$, and q_i and $\hat{\mu}_i^b$ are the ith elements of vectors $\boldsymbol{q}$ and $\hat{\boldsymbol{\mu}}^b$.

Under the null hypothesis that the model was correctly specified, the authors assumed that the kernel smoothed unweighted sum of squares test statistic

$$Z_{S_m} = \frac{S_m - E(S_m)}{\sqrt{Var(S_m)}}$$

has approximately a standard normal distribution. However, Sturdivant (2005) had showed that the assumption of the asymptotic standard normal distribution was problematic in many of their simulation settings. Especially, the behavior in the tail departed from the standard normal distribution.

Since the kernel smoothed unweighted sum of squares statistic, S_m, has a quadratic form that would have an asymptotic chi-squared distribution and the scaled chi-squared distribution is a commonly used approximation for distributions of non-negative random variables (Cox and Hinkley, 1974), we suggest that the asymptotic distribution of S_m can be approached by a scaled chi-squared distribution with the same moments,

$$cS_m \sim \chi^2(\nu),$$

where $c = \dfrac{2E(S_m)}{Var(S_m)}$ and $\nu = \dfrac{2\left\{E(S_m)\right\}^2}{Var(S_m)}$. Moreover, we also provide another way to transform the possible scaled chi-squared statistic to a standard normal statistic. We can use the fact that

$$cS_m^{tran} = \left[\left(\frac{cS_m}{\nu}\right)^{1/3} + \frac{2}{9\nu} - 1\right]\left(\frac{9\nu}{2}\right)^{1/2}$$

is approximately a standard normal variate (Stuart and Ord, 1987).

Later, in a part of Section 4.3, we carry out a short simulation experiment for determining which asymptotic distribution of the kernel smoothed unweighted sum of squares statistic, S_m, is approached in the logistic multilevel model.

4.2.3 Local Polynomial Smoothed Residuals

Ruppert and Wand (1994) developed the general theory for multivariate local polynomial regression in the usual situation that the covariate has p-dimensional compact support in $\mathbb{R}^p$. Generally, when we use the nonparametric smoothing method in the multidimensional situations, boundary modifications in higher dimensions are a very difficult task. Fan and Gijbels (1996) argued that an advantage of local polynomial fitting is that with local polynomial fitting no boundary modifications are required, and this is an important merit, especially when dealing with multidimensional situations.

In this part, we introduce how to obtain a smoothing function of $\hat{\boldsymbol{e}}$ with respect to continuous and within-cluster covariates by local polynomial estimation. Let us start by assuming that we have m clusters (or subjects). For each cluster, $1 \leq i \leq m$, there are n_i binary outcome values $\mathbf{y}_i = (y_{i1}, \cdots, y_{in_i})^T$ and a $n_i \times p$ covariate matrix $\mathbf{x}_i^* = (\mathbf{x}_{i1}, \cdots, \mathbf{x}_{in_i})^T$ with p-dimensional covariate vector $\mathbf{x}_{in_i}$. For simplicity, in model (4.1), suppose that $n_i = n$ for all i and total observations $N = mn$. Conditional on the random effects, let $\boldsymbol{\mu}_i^b = E(\mathbf{y}_i|\mathbf{x}_i, b_{0i}) = (\mu_{i1}^b, \mu_{i2}^b, \cdots, \mu_{in}^b)^T$, and $\boldsymbol{\beta} = (\beta_0^b, \beta_1, \cdots, \beta_p)^T$. The logistic multilevel model specifies that $\text{logit}(\boldsymbol{\mu}_i^b) = \mathbf{x}_i\boldsymbol{\beta}$ with $\mathbf{x}_i = (\mathbf{1}, \mathbf{x}_i^*)$.

Local linear regression estimates the population regression function by $\hat{\beta}_0^b$ where $(\hat{\beta}_0^b, \hat{\beta}_1, \cdots, \hat{\beta}_p)$ minimize

$$\sum_{i=1}^{m}\sum_{j=1}^{n}\left\{y_{ij} - \beta_0^b - \boldsymbol{\beta}^{*T}(\mathbf{x}_{ij} - \mathbf{x}_0)\right\}^2 K_{\mathbf{h}}(\mathbf{x}_{ij} - \mathbf{x}_0),$$

where $\boldsymbol{\beta}^* = (\beta_1, \cdots, \beta_p)$, and $K_{\mathbf{h}}(\cdot)$ is a p-variate kernel function. Then, the multivariate nonparametric local polynomial estimator of $f(\mathbf{x}_0) = E(\mathbf{y}|\mathbf{x}_0, b_{0i})$ with degree 1, $\hat{\beta}_0^b$, is given by

$$\hat{f}(\mathbf{x}_0) = \mathbf{a}_1^T(\mathbf{t}_{\mathbf{x}_0}^T\mathbf{w}_{\mathbf{x}_0}\mathbf{t}_{\mathbf{x}_0})^{-1}\mathbf{t}_{\mathbf{x}_0}\mathbf{w}_{\mathbf{x}_0}\mathbf{y} = \mathbf{s}_{\mathbf{x}_0}\mathbf{y},$$

where $\mathbf{y} = (\mathbf{y}_1^T, \cdots, \mathbf{y}_m^T)^T$, $\mathbf{s}_{\mathbf{x}_0} = \mathbf{a}_1^T(\mathbf{t}_{\mathbf{x}_0}^T\mathbf{w}_{\mathbf{x}_0}\mathbf{t}_{\mathbf{x}_0})^{-1}\mathbf{t}_{\mathbf{x}_0}\mathbf{w}_{\mathbf{x}_0}$, $\mathbf{a}_1$ is a $(p + 1) \times 1$ vector having 1 in the first entry and zero elsewhere, $\mathbf{t}_{\mathbf{x}_0}$ is a $N \times (p+1)$ matrix given by

$$\mathbf{t}_{\mathbf{x}_0} = \begin{bmatrix} 1 & (\mathbf{x}_{11} - \mathbf{x}_0)^T \\ 1 & (\mathbf{x}_{12} - \mathbf{x}_0)^T \\ \vdots & \vdots \\ 1 & (\mathbf{x}_{mn} - \mathbf{x}_0)^T \end{bmatrix},$$

and $\mathbf{w}_{\mathbf{x}_0} = \mathrm{diag}\left[K_{\mathbf{h}}(\mathbf{x}_{11} - \mathbf{x}_0), \cdots, K_{\mathbf{h}}(\mathbf{x}_{mn} - \mathbf{x}_0)\right]$ with $\mathbf{h} = \mathrm{diag}(h_1^2, \cdots, h_p^2)$. Here $\mathbf{x}_0$ is an arbitrary but fixed interior point of the domain of $\mathbf{x}^*$ and $K_{\mathbf{h}}(\cdot)$ is a multivariate kernel function. Furthermore, we can define the local polynomial smoothed residuals (Lin et al., 2008) as follows,

$$\hat{\boldsymbol{e}}_m = \mathbf{s}\hat{\boldsymbol{e}},$$

where $\mathbf{s} = [\mathbf{s}_{\mathbf{x}_{11}}^T, \mathbf{s}_{\mathbf{x}_{12}}^T, \cdots, \mathbf{s}_{\mathbf{x}_{mn}}^T]^T$.

4.3 Simulation Study

In order to study whether the local polynomial smoothed residuals improve upon the unweighted sum of squares statistic for assessing the goodness-of-fit of the logistic multilevel model, we also carry out a simulation study of the control of the type I error rate and power analysis by using the smoothed residuals over y-space. When we smooth residuals in the y-space, the bandwidth, h, $\frac{1}{2}\sqrt{N}$ or $\frac{1}{4}\sqrt{N}$, controlling the number of observations with non-zero weights, is adopted according to the suggestion of Sturdivant and Hosmer (2007). Additionally, we consider the cubic kernel function (Fowlkes, 1987; Hosmer et al., 1997; Sturdivant, 2005; Sturdivant and Hosmer, 2007) given by

$$K(u) = \begin{cases} 1 - |u|^3 & \text{if } |u| < 1 \\ 0 & \text{if } |u| \geq 1. \end{cases}$$

Moreover, Austin (2010) implemented a simulation study to show that when glmmPQL is used to analyze data with five observations per cluster in the R package, there appears to be severe bias in estimation of parameters. However, parameter estimation becomes much stable when the number of observations per cluster are greater than ten. Due to this limitation, we adopt dimension of each simulated data set as follows:

Case 1: 10 clusters with 10 observations per cluster and total observations $N = 100$,

Case 2: 15 clusters with 10 observations per cluster and total observations $N = 150$,

Case 3: 25 clusters with 8 observations per cluster and total observations $N = 200$,

Case 4: 20 clusters with 20 observations per cluster and total observations $N = 400$.

Further, the logistic multilevel models including either random intercept or random intercept and slope are considered in the simulation study. To evaluate the type I error rate, we determine the proportion of 500 simulated data sets where significant results, at 1%, 5% and 10% significance levels are detected; the power performance is evaluated at a 5% significance level.

4.3.1 Type I Error Rate for the Random Intercept Model

In this part, we investigate the performance of cS_m, Z_{S_m} and cS_m^{tran}, when we smooth the residuals in the y-space to determine an appropriate asymptotic distribution of S_m in the logistic multilevel model. Then, we assess the type I error rate of the kernel smoothed unweighted sum of squares statistic based on the nonparametric local polynomial smoothed residuals over continuous and within-cluster covariates.

We consider a logistic multilevel model (under the null hypothesis) with random intercept across level two, namely cluster,

$$\text{logit}(\mu_{ij}^b) = \beta_0 + \beta_1 x_{ij} + \beta_2 \xi_{ij} + \beta_3 c_i + b_{0i},$$

where c_i is a cluster-level covariate, x_{ij} and ξ_{ij} are covariates within the cluster-level, and b_{0i} is the random part over the cluster-level. The cluster-level covariate is generated such that the values taken on are randomly perturbed z-scores about the standard normal cumulative probability $i/(m+1)$ where m is the number of clusters and $i = 1, 2, \cdots, m$ (Evans and Hosmer, 2004). Two covariates within the cluster-level, x_{ij} and ξ_{ij} follow a uniform distribution (-1,1) and the standard normal distribution, respectively. The random intercept follows normal distribution with mean zero and variance 0.49. The vector of parameters is set as $(\beta_0, \beta_1, \beta_2, \beta_3)$ equal to $(0, 0.9, 0.5, 0.7)$. Additionally, bandwidths (h_1, h_2) from $(0.5, 0.5)$ to $(1.5, 1.5)$ by $(0.25, 0.25)$ are adopted when we use the local polynomial technique to smooth residuals over continuous and within-cluster covariates .

First, through a short simulation experiment shown in Table 11, we find that the performance of cS_m for controlling the type I error rate is relatively better than cS_m^{tran} and Z_{S_m}. On the other hand, although the performance of cS_m^{tran} and Z_{S_m} do not differ significantly in our simulation settings, the distribution of cS_m^{tran} is much closer to the standard normal distribution, especially the behavior in the tail, than that of Z_{S_m} presented in Figure 4 and Appendix B. Moreover, in Table 12, it clearly shows that the assumption of the standard normal distribution of Z_{S_m} (Sturdivant, 2005; Sturdivant and Hosmer, 2007) may be questionable and the transformed scaled chi-squared variables can replace it.

Overall, we believe that the scaled chi-squared distribution for the asymptotic distribution of S_m is the most appropriate one in our simulation settings. Therefore, we shall use cS_m to evaluate performance of the kernel smoothed unweighted sum of squares statistic by using the local polynomial smoothed residuals on the goodness-of-fit test in the logistic multilevel models.

Table 11. Comparisons of the type I error rate for three types of asymptotic distributions of the kernel smoothed unweighted sum of squares statistic by smoothing residuals in the y-space.

		Case 1 $(N=100)(10,10)$				Case 2 $(N=150)(15,10)$		
	α	0.1	0.05	0.01	α	0.1	0.05	0.01
$\frac{1}{4}\sqrt{N}$	cS_m	0.090	0.044	0.006	cS_m	0.092	0.044	0.012
	Z_{S_m}	0.078	0.036	0.006	Z_{S_m}	0.074	0.036	0.012
	cS_m^{tran}	0.074	0.034	0.004	cS_m^{tran}	0.074	0.036	0.008
$\frac{1}{2}\sqrt{N}$	cS_m	0.094	0.038	0.008	cS_m	0.104	0.050	0.014
	Z_{S_m}	0.068	0.034	0.010	Z_{S_m}	0.080	0.048	0.020
	cS_m^{tran}	0.070	0.034	0.010	cS_m^{tran}	0.072	0.042	0.012
		Case 3 $(N=200)(25,8)$				Case 4 $(N=400)(20,20)$		
	α	0.1	0.05	0.01	α	0.1	0.05	0.01
$\frac{1}{4}\sqrt{N}$	cS_m	0.100	0.052	0.006	cS_m	0.098	0.044	0.010
	Z_{S_m}	0.082	0.046	0.006	Z_{S_m}	0.076	0.032	0.010
	cS_m^{tran}	0.086	0.036	0.006	cS_m^{tran}	0.080	0.036	0.006
$\frac{1}{2}\sqrt{N}$	cS_m	0.110	0.060	0.012	cS_m	0.098	0.048	0.012
	Z_{S_m}	0.078	0.056	0.018	Z_{S_m}	0.088	0.042	0.014
	cS_m^{tran}	0.080	0.040	0.008	cS_m^{tran}	0.092	0.042	0.012

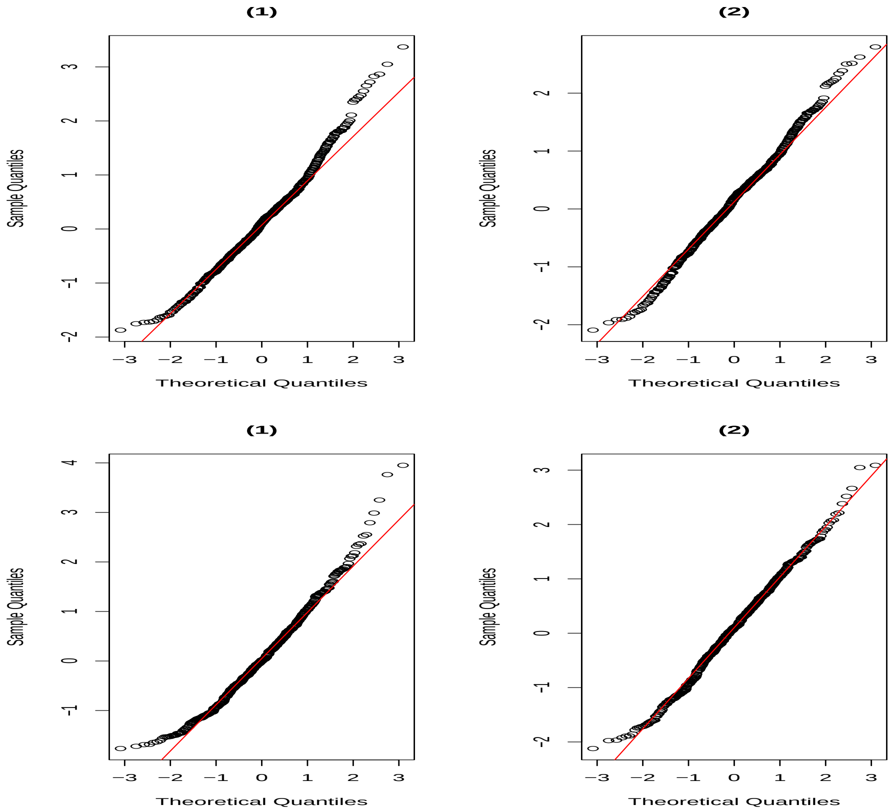

Figure 4. Normal QQ plots of test statistic values of Case 1 under $(1)Z_{S_m}$ and $(2)cS_m^{tran}$: the smoothed residuals over y-space based on $\frac{1}{4}\sqrt{N}$ and $\frac{1}{2}\sqrt{N}$ are shown on the top and bottom panels, respectively.

Table 12. P-values of normality checking of Case 1 for $(1)Z_{S_m}$ and $(2)cS_m^{tran}$.

		Anderson-Darling	Cramer-von Mises
$\frac{1}{4}\sqrt{N}$	(1)	0.0015	0.0068
	(2)	0.3035	0.3257
$\frac{1}{2}\sqrt{N}$	(1)	0.0008	0.0058
	(2)	0.3988	0.6544

Second, Table 13 reveals that when using the local polynomial smoothed residuals over continuous and within-cluster covariates, for all selected bandwidths the type I error rates of S_m compared with a scaled chi-squared distribution are well controlled at significance levels of 1%, 5% and 10% in the logistic multilevel model with only random intercept. On average, the type I error rates do not appear either too liberal or conservative in most situations.

Table 13. Results of the type I error rate of S_m by using local polynomial smoothed residuals are computed based on the scaled chi-squared distribution cS_m.

	Case 1 $(N=100)(10,10)$				Case 2 $(N=150)(15,10)$		
α	0.1	0.05	0.01	α	0.1	0.05	0.01
(h_1 , h_2)				(h_1 , h_2)			
(0.50 , 0.50)	0.106	0.044	0.006	(0.50 , 0.50)	0.094	0.040	0.008
(0.75 , 0.75)	0.088	0.036	0.008	(0.75 , 0.75)	0.108	0.062	0.016
(1.00 , 1.00)	0.090	0.048	0.010	(1.00 , 1.00)	0.104	0.054	0.012
(1.25 , 1.25)	0.094	0.048	0.008	(1.25 , 1.25)	0.092	0.054	0.008
(1.50 , 1.50)	0.098	0.058	0.012	(1.50 , 1.50)	0.096	0.052	0.012
	Case 3 $(N=200)(25,8)$				Case 4 $(N=400)(20,20)$		
α	0.1	0.05	0.01	α	0.1	0.05	0.01
(h_1 , h_2)				(h_1 , h_2)			
(0.50 , 0.50)	0.106	0.044	0.008	(0.50 , 0.50)	0.090	0.054	0.010
(0.75 , 0.75)	0.086	0.058	0.010	(0.75 , 0.75)	0.096	0.040	0.008
(1.00 , 1.00)	0.088	0.042	0.008	(1.00 , 1.00)	0.094	0.050	0.010
(1.25 , 1.25)	0.096	0.040	0.008	(1.25 , 1.25)	0.100	0.060	0.014
(1.50 , 1.50)	0.098	0.056	0.008	(1.50 , 1.50)	0.102	0.058	0.014

4.3.2 Power Analysis for the Random Intercept Model

Through a set of simulations, Hosmer et al. (1997) showed that the unweighted sum of squares statistic has a reasonable power for detecting a missing quadratic effect in the ordinary logistic regression models. However, it is far clear that the smoothed kernel unweighted sum of squares statistic by smoothing residuals in the y-space would detect such a departure in the fixed effects portion of the model with the random effects (Sturdivant, 2005; Sturdivant and Hosmer, 2007). In Section 4.3.1, we have known that the kernel smoothed unweighted sum of squares statistic by using local polynomial smoothed residuals over continuous and within-cluster covariates rejects the null model appropriately under the null hypothesis. Next, we study its power performance in detecting a missing within-cluster quadratic term in the logistic multilevel model. The alternative model is assumed to be

$$\text{logit}(\mu_{ij}^{b}) = \beta_0 + \beta_1 x_{ij} + \beta_{1q} x_{ij}^2 + (\beta_2 \xi_{ij} + \beta_3 c_i + b_{0i}). \tag{4.4}$$

Herein, we determine β_0, β_1, and β_{1q} by using $\text{logit}(\mu(x)) = \beta_0 + \beta_1 x + \beta_{1q} x^2$ to choose a strong or moderate quadratic term (Hosmer et al., 1997). We use $\mu(-0.5) = 0.05$, $\mu(1) = 0.95$, and $\mu(-1) = 0.40$ to choose $(\beta_0, \beta_1, \beta_{1q})$ equal to $(-3.23, 1.67, 4.5)$ for a strong quadratic term. On the other hand, $\mu(-0.5) = 0.05$, $\mu(1) = 0.95$, and $\mu(-1) = 0.1$ are used to choose $(\beta_0, \beta_1, \beta_{1q})$ equal to $(-2.34, 2.57, 2.71)$ for a moderate quadratic term.

The power performance for detecting a missing within-cluster quadratic term in each case is shown in Table 14 for strong and moderate quadratic terms. Based on the reported results, the local polynomial smoothed residuals enhance the power of S_m in detecting a missing quadratic term. In the situation of a strong quadratic term, even when the sample size is small, using the local polynomial smoothed residuals is still better than using the smoothed residuals over y-space. As for the situation of a moderate quadratic term, when the sample size increases to a large one, using the local polynomial smoothed residuals presents an obvious increase on power. On average, the performance of the kernel smoothed unweighted sum of squares statistic for detecting a missing quadratic term can be significantly improved when we apply the local polynomial technique to smooth residuals over within-cluster covariates.

Additionally, Sturdivant (2005) also showed that the smoothed kernel unweighted sum of squares statistic by smoothing residuals in the y-space has no real power for detecting the omission of an interaction term in the fixed effects portion of the model with the random effects. Thus, in the following part, we shall investigate whether the power performance of S_m by using the local polynomial smoothed residuals can make an improvement on detecting a missing within-cluster interaction term.

In the first part of this discussion, we consider that the null model is $\text{logit}(\mu_{ij}^b) = \beta_0 + \beta_1 x_{ij} + \beta_d d_{ij} + \beta_2 \xi_{ij} + \beta_3 c_i + b_{0i}$. The alternative model is assumed to be

$$\text{logit}(\mu_{ij}^b) = \beta_0 + \beta_1 x_{ij} + \beta_d d_{ij} + \beta_{1d} x_{ij} d_{ij} + (\beta_2 \xi_{ij} + \beta_3 c_i + b_{0i}), \tag{4.5}$$

where d_{ij} follow a Bernoulli distribution with probability 0.5. Similarly, we determine β_0, β_1, β_d and β_{1d} by using $\text{logit}(\mu(x,d)) = \beta_0 + \beta_1 x + \beta_d d + \beta_{1d} xd$ to choose a strong interaction term. In terms of $\mu(-1,0) = 0.1$, $\mu(-1,1) = 0.1$, $\mu(1,0) = 0.2$ and $\mu(1,1) = 0.9$, we choose $(\beta_0, \beta_1, \beta_d, \beta_{1d})$ equal to $(-1.792, 0.406, 1.792, 1.792)$.

Unfortunately, the results shown in Table 15 reveal that there is no improvement in the power performance of S_m by using local polynomial smoothed residuals for detecting a missing within-cluster interaction term of fixed effects between Bernoulli and continuous covariates.

Table 14. Results of the power performance of detecting a missing strong or moderate within-cluster quadratic term of fixed effects when the alternative model (4.4) is assumed.

	(1) Strong Quadratic Term			
	Case 1 ($N = 100$) $(10, 10)$	Case 2 ($N = 150$) $(15, 10)$	Case 3 ($N = 200$) $(25, 8)$	Case 4 ($N = 400$) $(20, 20)$
	Smooth over y-space			
$\frac{1}{4}\sqrt{N}$	0.102	0.128	0.140	0.232
$\frac{1}{2}\sqrt{N}$	0.180	0.222	0.288	0.452
	Local Polynomial Smooth over x-space			
$(h_1\ ,\ h_2)$				
(0.50 , 0.50)	0.546	0.820	0.952	1.000
(0.75 , 0.75)	0.658	0.876	0.966	1.000
(1.00 , 1.00)	0.592	0.810	0.938	0.998
(1.25 , 1.25)	0.524	0.778	0.924	1.000
(1.50 , 1.50)	0.418	0.642	0.820	0.998
	(2) Moderate Quadratic Term			
	Case 1 ($N = 100$) $(10, 10)$	Case 2 ($N = 150$) $(15, 10)$	Case 3 ($N = 200$) $(25, 8)$	Case 4 ($N = 400$) $(20, 20)$
	Smooth over y-space			
$\frac{1}{4}\sqrt{N}$	0.018	0.016	0.020	0.020
$\frac{1}{2}\sqrt{N}$	0.022	0.042	0.048	0.078
	Local Polynomial Smooth over x-space			
$(h_1\ ,\ h_2)$				
(0.50 , 0.50)	0.096	0.204	0.346	0.684
(0.75 , 0.75)	0.154	0.268	0.402	0.738
(1.00 , 1.00)	0.148	0.238	0.336	0.694
(1.25 , 1.25)	0.130	0.222	0.296	0.592
(1.50 , 1.50)	0.114	0.174	0.260	0.494

Table 15. Results of the power performance of detecting a missing strong within-cluster interaction term of fixed effects between Bernoulli and continuous covariates when the alternative model (4.5) is assumed.

	Case 1 ($N = 100$) (10, 10)	Case 2 ($N = 150$) (15, 10)	Case 3 ($N = 200$) (25, 8)	Case 4 ($N = 400$) (20, 20)
		Smooth over y-space		
$\frac{1}{4}\sqrt{N}$	0.018	0.022	0.024	0.022
$\frac{1}{2}\sqrt{N}$	0.030	0.040	0.044	0.062
		Local Polynomial Smooth over x-space		
(h_1 , h_2)				
(0.50 , 0.50)	0.020	0.032	0.036	0.068
(0.75 , 0.75)	0.032	0.046	0.050	0.084
(1.00 , 1.00)	0.032	0.050	0.046	0.064
(1.25 , 1.25)	0.042	0.046	0.058	0.078
(1.50 , 1.50)	0.036	0.038	0.048	0.070

On the other hand, we also consider that the null model is $\text{logit}(\mu_{ij}^b) = \beta_0 + \beta_1 x_{ij} + \beta_\delta \delta_{ij} + \beta_3 c_i + b_{0i}$. The alternative model is assumed to be

$$\text{logit}(\mu_{ij}^b) = \beta_0 + \beta_1 x_{ij} + \beta_\delta \delta_{ij} + \beta_{1\delta} x_{ij} \delta_{ij} + (\beta_3 c_i + b_{0i}), \tag{4.6}$$

where δ_{ij} follow a uniform distribution $(-3, 3)$. Again, we determine β_0, β_1, β_δ and $\beta_{1\delta}$ by using $\text{logit}(\mu(x, \delta)) = \beta_0 + \beta_1 x + \beta_\delta \delta + \beta_{1\delta} x \delta$ to choose a strong interaction term. Based on $\mu(-1, -3) = 0.1$, $\mu(-1, 3) = 0.1$, $\mu(1, -3) = 0.2$ and $\mu(1, 3) = 0.9$, we choose $(\beta_0, \beta_1, \beta_\delta, \beta_{1\delta})$ equal to $(-0.896, 1.301, 0.299, 0.299)$.

Interestingly, Table 16 indicates that when the the sample size is large with a moderate or large cluster size, for the wider selected bandwidths, the power performance of S_m by using local polynomial smoothed residuals is improved for detecting a missing strong within-cluster interaction term of fixed effects between two continuous covariates.

Therefore, on average, we believe that the smoothed kernel unweighted sum of squares statistic by using the local polynomial smoothed residuals still has an advantage in detecting a missing strong within-cluster interaction term in some special situations.

Table 16. Results of the power performance of detecting a missing strong within-cluster interaction term of fixed effects between two continuous covariates when the alternative model (4.6) is assumed.

	Case 1 ($N = 100$) $(10, 10)$	Case 2 ($N = 150$) $(15, 10)$	Case 3 ($N = 200$) $(25, 8)$	Case 4 ($N = 400$) $(20, 20)$
	Smooth over y-space			
$\frac{1}{4}\sqrt{N}$	0.024	0.028	0.032	0.032
$\frac{1}{2}\sqrt{N}$	0.028	0.046	0.070	0.078
	Local Polynomial Smooth over x-space			
$(h_1\ ,\ h_2)$				
(0.50 , 0.50)	0.016	0.044	0.052	0.138
(0.75 , 0.75)	0.046	0.074	0.102	0.224
(1.00 , 1.00)	0.078	0.104	0.156	0.282
(1.25 , 1.25)	0.090	0.126	0.136	0.320
(1.50 , 1.50)	0.094	0.150	0.188	0.370

4.3.3 Power Analysis for the Random Intercept and Slope Model

In this part, we carry out a short discussion in the logistic multilevel model including random intercept and slope across level two, namely cluster. First, we assume a model (under the null hypothesis) is given by

$$\text{logit}(\mu_{ij}^b) = \beta_0 + \beta_1 x_{ij} + \beta_2 \xi_{ij} + \beta_3 c_i + b_{0i} + b_{1i}\xi_{ij}.$$

All parameter and generated data settings are the same as we adopted in Section 4.3.1 except that b_{1i} follows normal distribution with mean zero and variance 0.1225. Moreover, we assume that b_{0i} and b_{1i} are uncorrelated. The results shown in Table 17 show that the type I error rates compared with those of the model without random slope seem to become conservative at all significance levels when the kernel smoothed unweighted sum of squares statistic is adopted by using the local polynomial smoothed residuals over covariates. For larger sample sizes and wider selected bandwidths, the type I error rate for a significance level 5% test becomes much more reasonable.

Second, we study the power performance of S_m by using the local polynomial smoothed residuals for detecting a missing within-cluster quadratic term in the logistic multilevel model with random intercept and slope. The alternative model is assumed to be

$$\text{logit}(\mu_{ij}^b) = \beta_0 + \beta_1 x_{ij} + \beta_{1q} x_{ij}^2 + (\beta_2 \xi_{ij} + \beta_3 c_i + b_{0i} + b_{1i}\xi_{ij}). \tag{4.7}$$

All parameter settings of model (4.7) are the same as those used in model (4.4). The power performance for detecting a missing within-cluster quadratic term shown in Table 18 does not differ much from results of the model without random slope presented in Table 14.

Overall, we conclude that applying the local polynomial smoothed residuals to the kernel smoothed unweighted sum of squares statistic can be a good choice for detecting a missing quadratic term in the logistic multilevel model including either random intercept or random intercept and slope.

Table 17. Results of controlling type I error rate of S_m by using local polynomial smoothed residuals are computed based on cS_m when the model includes random intercept and slope.

	Case 1 $(N=100)(10,10)$				Case 2 $(N=150)(15,10)$		
α	0.1	0.05	0.01	α	0.1	0.05	0.01
			Smooth over y-space				
$\frac{1}{4}\sqrt{N}$	0.084	0.036	0.012	$\frac{1}{4}\sqrt{N}$	0.086	0.048	0.008
$\frac{1}{2}\sqrt{N}$	0.108	0.056	0.008	$\frac{1}{2}\sqrt{N}$	0.096	0.046	0.018
			Local Polynomial Smooth over x-space				
$(h_1\ ,\ h_2)$				$(h_1\ ,\ h_2)$			
(0.50 , 0.50)	0.064	0.026	0.004	(0.50 , 0.50)	0.064	0.038	0.002
(0.75 , 0.75)	0.078	0.048	0.006	(0.75 , 0.75)	0.072	0.042	0.010
(1.00 , 1.00)	0.068	0.036	0.010	(1.00 , 1.00)	0.088	0.040	0.004
(1.25 , 1.25)	0.084	0.032	0.008	(1.25 , 1.25)	0.078	0.036	0.008
(1.50 , 1.50)	0.074	0.042	0.012	(1.50 , 1.50)	0.074	0.032	0.004
	Case 3 $(N=200)(25,8)$				Case 4 $(N=400)(20,20)$		
α	0.1	0.05	0.01	α	0.1	0.05	0.01
			Smooth over y-space				
$\frac{1}{4}\sqrt{N}$	0.118	0.062	0.010	$\frac{1}{4}\sqrt{N}$	0.084	0.050	0.006
$\frac{1}{2}\sqrt{N}$	0.152	0.072	0.012	$\frac{1}{2}\sqrt{N}$	0.094	0.048	0.010
			Local Polynomial Smooth over x-space				
$(h_1\ ,\ h_2)$				$(h_1\ ,\ h_2)$			
(0.50 , 0.50)	0.068	0.034	0.006	(0.50 , 0.50)	0.076	0.046	0.006
(0.75 , 0.75)	0.080	0.040	0.006	(0.75 , 0.75)	0.076	0.044	0.004
(1.00 , 1.00)	0.068	0.032	0.006	(1.00 , 1.00)	0.086	0.048	0.014
(1.25 , 1.25)	0.080	0.042	0.016	(1.25 , 1.25)	0.078	0.044	0.006
(1.50 , 1.50)	0.078	0.048	0.008	(1.50 , 1.50)	0.082	0.048	0.016

Table 18. Results of the power performance of detecting a missing strong or moderate within-cluster quadratic term of fixed effects when the alternative model (4.7) with random intercept and slope is assumed.

	(1) Strong Quadratic Term			
	Case 1 ($N = 100$) $(10, 10)$	Case 2 ($N = 150$) $(15, 10)$	Case 3 ($N = 200$) $(25, 8)$	Case 4 ($N = 400$) $(20, 20)$
	Smooth over y-space			
$\frac{1}{4}\sqrt{N}$	0.030	0.052	0.066	0.104
$\frac{1}{2}\sqrt{N}$	0.060	0.112	0.218	0.308
	Local Polynomial Smooth over x-space			
$(h_1\ ,\ h_2)$				
(0.50 , 0.50)	0.490	0.774	0.930	1.000
(0.75 , 0.75)	0.584	0.832	0.936	0.998
(1.00 , 1.00)	0.538	0.766	0.898	0.996
(1.25 , 1.25)	0.436	0.716	0.888	0.992
(1.50 , 1.50)	0.378	0.590	0.784	0.974
	(2) Moderate Quadratic Term			
	Case 1 ($N = 100$) $(10, 10)$	Case 2 ($N = 150$) $(15, 10)$	Case 3 ($N = 200$) $(25, 8)$	Case 4 ($N = 400$) $(20, 20)$
	Smooth over y-space			
$\frac{1}{4}\sqrt{N}$	0.012	0.020	0.024	0.018
$\frac{1}{2}\sqrt{N}$	0.034	0.050	0.088	0.098
	Local Polynomial Smooth over x-space			
$(h_1\ ,\ h_2)$				
(0.50 , 0.50)	0.060	0.152	0.256	0.578
(0.75 , 0.75)	0.114	0.186	0.290	0.616
(1.00 , 1.00)	0.098	0.172	0.254	0.590
(1.25 , 1.25)	0.086	0.140	0.218	0.514
(1.50 , 1.50)	0.074	0.126	0.156	0.362

4.4 Application

In this section, we illustrate the application of the nonparametric local polynomial smoothing residuals over continuous and within-cluster covariates on the unweighted sum of squares statistic for assessing the goodness of fit in the logistic multilevel models. We use part of real data set from a clinical trial, called "support", from the Cancer Biostatistics Center, Vanderbilt University. The objective of the support (Study to Understand Prognoses Preferences Outcomes and Risks of Treatment) was to improve decision-making in order to address the growing national concern over the loss of control that patients have near the end of life and to reduce the frequency of a mechanical, painful, and prolonged process of dying (see http://www.icpsr.umich.edu/icpsrweb/ICPSR/studies/02957/ for details). In our analysis, there are 392 patients taken from the support data set. We consider the hierarchical structure of patients (level one) in 8 clusters (level two; health status of patients), such as cirrhosis, coma, colon cancer, lung cancer, etc. The outcome of interest is whether or not patients die in the hospital during the period of the trial. Part of covariates which can be used to assess the physiological status of patients are listed as follows,

(1) Wblc: White blood cell count Day 3,

(2) Crea: Serum creatinine Day 3,

(3) Resp: Respiration rate Day 3,

(4) Temp: Temperature – average of all patients' temperature (celsius) Day 3.

On the other hand, in order to select appropriate bandwidths, we re-define a leave-one-out cross-validation method (Hart, 1997) as the follow by using $\hat{\mu}_{ij}^{b(i)}$ the nonparametric smoother of $E(y_{ij}|\mathbf{x}_{ij}, \mathbf{b}_i)$ for all data except the ith (level two) cluster,

$$CV(h) = \frac{1}{N}\sum_{i=1}^{m}\sum_{j=1}^{n_i}\left(y_{ij} - \hat{\mu}_{ij}^{b(i)}\right)^2,$$

where h is the smoothing parameter for the local polynomial estimate $\hat{\mu}_{ij}^{b(i)}$ shown in Section 4.2.3 and $N = n_1 + n_2 + \cdots + n_m$. The cross-validation smoothing parameter is the value of h by minimizing $CV(h)$. Additionally, we subjectively focus on six logistic multilevel models listed in Table 19 for demonstrating the application. For the model including the higher-order term, we can try to add a higher-degree component to the related covariate in computing the nonparametric local polynomial estimator. Furthermore, we not only implement the model checking by using the kernel smoothed weighted sum of squares statistic where parameter estimates of the model are based on the PQL estimation from glmmPQL in R, but also use glmer which is based on the adaptive Gaussian-Hermite approximation to the likelihood in R to fit models and obtain the corresponding AIC, BIC and deviance.

Table 19. List of six logistic multilevel models with the random intercept for demonstrating the application.

Model 1:	$\text{logit}(\mu_{ij}^b) = \beta_0 + \beta_w \text{Wblc}_{ij} + \beta_c \text{Crea}_{ij} + b_{0i}$
Model 2:	$\text{logit}(\mu_{ij}^b) = \beta_0 + \beta_w \text{Wblc}_{ij} + \beta_c \text{Crea}_{ij} + \beta_{w2} \text{Wblc}_{ij}^2 + b_{0i}$
Model 3:	$\text{logit}(\mu_{ij}^b) = \beta_0 + \beta_w \text{Wblc}_{ij} + \beta_c \text{Crea}_{ij} + \beta_{c2} \text{Crea}_{ij}^2 + b_{0i}$
Model 4:	$\text{logit}(\mu_{ij}^b) = \beta_0 + \beta_r \text{Resp}_{ij} + \beta_t \text{Temp}_{ij} + b_{0i}$
Model 5:	$\text{logit}(\mu_{ij}^b) = \beta_0 + \beta_r \text{Resp}_{ij} + \beta_t \text{Temp}_{ij} + \beta_{r2} \text{Resp}_{ij}^2 + b_{0i}$
Model 6:	$\text{logit}(\mu_{ij}^b) = \beta_0 + \beta_r \text{Resp}_{ij} + \beta_t \text{Temp}_{ij} + \beta_{t2} \text{Temp}_{ij}^2 + b_{0i}$

First, we carry out a cross-validation procedure to select the appropriate bandwidths based on 30×30 grid points in the lower terms of each model. The selected values of the smoothing parameter are shown in Table 20.

Table 20. Results of the selected values of smoothing parameter based on the cross–validation method for each model.

	$(h_{\text{Wblc}}, h_{\text{Crea}})$		$(h_{\text{Resp}}, h_{\text{Temp}})$
Model 1	(6.6460,4.5408)	Model 4	(9.3448,3.9082)
Model 2	(8.3398,4.5408)	Model 5	(11.5172,3.6153)
Model 3	(10.0336,1.8446)	Model 6	(10.7931,4.2012)

Second, for each model, the kernel smoothed unweighted sum of squares statistic and the corresponding p-value compared with the scaled chi-squared distribution are presented in Tables 21 and 22. The results are based on the selected values of the smoothing parameter in each model. In terms of reported p-values, we find that model 1, model 2, and model 3 provide an adequate fit; however, model 4, model 5, and model 6 do not fit adequately.

In analysis of this example, we believe that the kernel smoothed unweighted sum of squares statistic by using the local polynomial smoothed residuals is reliable for the model checking. For instance, in our analysis, parameter estimates based on PQL estimation do not differ from those converged estimates based on the adaptive Gaussian-Hermite approximation to the likelihood. Therefore, all test results based on the stable and reasonable PQL estimation are acceptable. Additionally, in the cases shown in Table 22, we observe that when adding a higher-order effect into a non-adequate model, the model checking results still reveal no improvement on the lack of fit even if the higher-order effect is significant. Overall, the model 1 can be selected as the most appropriate model among six fitted models by using AICs and results of the model checking.

Table 21. Results of the model fit based on glmmPQL and glmer and model checking based on S_m by using local polynomial smoothing residuals over within continuous cluster-level covariates for Model 1, Model 2 and Model 3.

	Model 1		
Covariates	Estimates	S.E.	p-value
Intercept	-2.0409 (-2.0873)	0.4482 (0.4606)	0.0000 (0.0000)
Wblc	0.0385 (0.0394)	0.0198 (0.0205)	0.0525 (0.0545)
Crea	0.5129 (0.5525)	0.1190 (0.1238)	0.0000 (0.0000)
$c\hat{S}_m$	$E(\hat{S}_m)$	Var$(\hat{S}_m)$	p-value of χ^2_ν
9.9499	0.9550	0.2264	0.2732
AIC	BIC	Deviance	
423.9	439.8	415.9	
	Model 2		
Covariates	Estimates	S.E.	p-value
Intercept	-1.7191 (-1.7578)	0.5323 (0.5474)	0.0013 (0.0013)
Wblc	-0.0217 (-0.0223)	0.0597 (0.0617)	0.7160 (0.7183)
Crea	0.5044 (0.5137)	0.1190 (0.1237)	0.0000 (0.0000)
Wblc^2	0.0022 (0.0023)	0.0021 (0.0022)	0.2906 (0.2957)
$c\hat{S}_m$	E$(\hat{S}_m)$	Var$(\hat{S}_m)$	p-value of χ^2_ν
8.7523	0.8516	0.1996	0.2948
AIC	BIC	Deviance	
424.8	444.7	414.8	
	Model 3		
Covariates	Estimates	S.E.	p-value
Intercept	-2.3343 (-2.3870)	0.5178 (0.5327)	0.0000 (0.0000)
Wblc	0.0371 (0.0379)	0.0199 (0.0206)	0.0639 (0.0661)
Crea	0.8311 (0.8470)	0.2891 (0.2997)	0.0043 (0.0047)
Crea^2	-0.0495 (-0.0504)	0.0389 (0.0405)	0.2048 (0.2133)
$c\hat{S}_m$	E$(\hat{S}_m)$	Var$(\hat{S}_m)$	p-value of χ^2_ν
15.9619	2.0132	0.5931	0.2936
AIC	BIC	Deviance	
424.6	444.5	414.6	

*Estimation and inference based on glmer are within parentheses.

Table 22. Results of the model fit based on glmmPQL and glmer and model checking based on S_m by using local polynomial smoothing residuals over within continuous cluster-level covariates for Model 4, Model 5 and Model 6.

	Model 4		
Covariates	Estimates	S.E.	p-value
Intercept	-1.2895 (-1.3227)	0.4981 (0.5103)	0.0100 (0.0095)
Resp	0.0189 (0.0192)	0.0129 (0.0133)	0.1457 (0.1471)
Temp	-0.0622 (-0.0626)	0.0975 (0.0999)	0.5237 (0.5312)
$c\hat{S}_m$	E($\hat{S}_m$)	Var($\hat{S}_m$)	p-value of χ^2_ν
37.5774	1.5011	0.2768	0.0019
AIC	BIC	Deviance	
451.9	467.8	443.9	
	Model 5		
Covariates	Estimates	S.E.	p-value
Intercept	-0.2422 (-0.2519)	0.7872 (0.8037)	0.7585 (0.7540)
Resp	-0.0740 (-0.0757)	0.0571 (0.0582)	0.1953 (0.1936)
Temp	-0.0764 (-0.0773)	0.0989 (0.1009)	0.4404 (0.4436)
Resp^2	0.0018 (0.0019)	0.0011 (0.0011)	0.0967 (0.0956)
$c\hat{S}_m$	E($\hat{S}_m$)	Var($\hat{S}_m$)	p-value of χ^2_ν
35.7047	1.3438	0.2922	0.0005
AIC	BIC	Deviance	
451.2	471.0	441.2	
	Model 6		
Covariates	Estimates	S.E.	p-value
Intercept	-1.5114 (-1.5482)	0.4958 (0.5087)	0.0025 (0.0023)
Resp	0.0129 (0.0132)	0.0132 (0.0137)	0.3294 (0.3339)
Temp	-0.1951 (-0.1973)	0.1131 (0.1169)	0.0854 (0.0915)
Temp^2	0.2373 (0.2418)	0.0702 (0.0728)	0.0008 (0.0009)
$c\hat{S}_m$	E($\hat{S}_m$)	Var($\hat{S}_m$)	p-value of χ^2_ν
21.0582	1.0955	0.2488	0.0174
AIC	BIC	Deviance	
440.0	459.0	430.0	

*Estimation and inference based on glmer are within parentheses.

Finally, from the fitted results of model 1, we conclude that within each cluster (health status of patients), white blood cell count and serum creatinine are positively associated with the probability of dying for patients near the end of life; further, when patients in the status of coma have the largest estimated random intercept (1.268), it reveals that the probability of dying in the status of coma is relatively higher than other statuses of patients.

4.5 Discussion

In this chapter, we apply the local polynomial smoothed residuals which had been discussed in the population-averaged model with longitudinal binary data to the logistic multilevel model which contains the random-intercept part or both random intercept and slope on the model checking. Through a simulation study, in our simulation settings, we discover that the power performance of the kernel smoothed unweighted sum of squares statistic can be significantly improved by the local polynomial smoothed residuals over within-cluster continuous covariates compared with the smoothed residuals over y-space on checking the adequacy of some specific models.

However, there are some limitations in this approach. For instance, we ignore the effect of within-cluster (level one) categorical variables and between-cluster (level two) variables on smoothing residuals in the logistic multilevel model. This may be a reason why there is no power for detecting a missing strong within-cluster interaction term of fixed effects between Bernoulli and continuous covariates. Therefore, in this situation, we may have to assume that the lack of fit resulted from the incorrect modelling of within-cluster continuous covariates, so there is no mutual interaction between the categorical and continuous variables. On the other hand, a test statistic based on stratification by within-cluster categorical covariate or between-cluster covariate can be further discussed in the logistic multilevel model.

Moreover, when there are many different within-cluster covariates included in the model, the implementation of this approach is time consuming in the analysis of the case study. Although we can consider smoothing residuals according to some variables, choosing a suitable subset of covariates is difficult and subjective. Therefore, combining this approach with the mechanism of variable selection may be needed. Nevertheless, in most situations, the kernel smoothed test statistic by using the local polynomial smoothed residuals seems to provide a global and reliable measure of goodness-of-fit test on the model checking in the logistic multilevel model.

CHAPTER V

CONCLUSION

We investigate the methodology to detect misspecification of the random-effects distribution in generalized linear mixed models and the application of the local polynomial smoothed residuals on the goodness-of-fit test of the logistic multilevel models throughout the research presented in this dissertation. We summarize our findings and discuss avenues for future research as follows.

When we account for the correlated binary outcomes within subjects (or clusters) by modeling random-effects models, estimation and inference depend that the structure of random effects is correctly specified. In most situations, we often assume that random effects are normally distributed. However, recent research exhibits that inferences for the linear predictor parameters and estimation for the variance component of random effects are seriously affected by the violation of normality of the random-effects distributions, when the true random-effects distribution is not a normal distribution with a large variance component. In addition, since random effects are unobserved, it is hard to use a straightforward diagnostic tool, for instance, the histogram plot of the empirical Bayes estimates of random effects, to check departures from normality in either linear mixed models (Verbeke and Molenberghs, 2000) or generalized linear mixed models (Litière et al., 2007; Alonso et al., 2008).

As a result, one of our studies concentrates on developing a formal test for testing the normality of the random-effects distribution in generalized linear mixed models. Our proposed robust score test analogous to the order selection test for testing the distributional assumptions of random effects without any parametric form is not likelihood-based. It is motivated by constructing a test statistic involved with generalized estimating equations and approximating the distribution of random effects based on the semi-nonparametric density representation. Throughout our study, in order to control the type I error rate, we adopt the smoothed bootstrap test with a parametric bootstrap procedure to replace the test formed by using the asymptotic results of the test statistic in a random-intercept logistic model.

Our investigations reveal that the proposed test statistic has large power to detect the violation of normality of random effects when the true unobserved random effects follow a distribution with multiple modes. Unfortunately, it has no power when the true distribution of the random effects follows a unimodal highly skewed or heavy-tailed distribution. Since we approximate the random-effects distribution with a large variance component by the semi-nonparametric Hermite expansion of the standard normal density, it is desired to further explore whether a unimodal highly skewed or heavy-tailed distribution of random effects with the large variance can be distinguished from a normal distribution in this way. Additionally, we also can try to construct the test statistic under a likelihood-based Wald test and use the Monte Carlo EM algorithm for parameter estimation to explore whether there is any improvement on the power performance in future research.

Although the smoothed bootstrap test provides a good control of the type I error rate under non-optimal bandwidth in our study and we can use npudist(np) in R to search for an appropriate bandwidth in practice, the issue of finding the optimal bandwidth is still worthy of further exploration to enhance the reliability of the smoothed bootstrap test. As for the bootstrap approach, we could discuss other resampling approaches since resampling data may blindly lead to the test with very low power and performance of the test under parametric bootstrap might not be superior when the model is not correctly assumed (Lee, 1994; Shao and Tu, 1995; Aerts and Claeskens, 2001).

Moreover, we can extend our proposed test statistic to detect another type of misspecification of random effects in a future study, for instance, incorrectly ignoring a random slope, random intercept variance depending on a binary covariate or autocorrelated random effects occurring between repeated measurements within a subject (or cluster). Since the generalized linear mixed models are widely used to analyze nonnormal outcome variables in many fields, the development of a suitable diagnostic tool for detecting misspecification of random effects including the violation of normality of the random-effects distribution would be useful and encouraged.

On the other hand, the objective of the second study is to apply the nonparametric local polynomial smoothed residuals over within-cluster continuous covariates to the unweighted sum of squares statistic for checking adequacy of a logistic multilevel model, namely, a mixed-effects logistic model for hierarchical data with binary outcomes. As asserted in the discussion part of Chapter IV, the kernel smoothed test statistic formed by using the local polynomial smoothed residuals provides a global and reliable measure for the model checking. It performs better than the kernel smoothed test statistic by using the smoothed residuals over y-space. Especially, it has significant improvement of power for detecting a missing strong within-cluster quadratic term in the logistic multilevel model containing either the random-intercept part or both random intercept and slope. Furthermore, our investigations also indicate that when the cluster size or sample size increases, the power for detecting a missing moderate within-cluster quadratic term or strong interaction term between two within-cluster continuous covariates gradually increases.

Finally, as discussed in Section 4.5, without ignoring the effect of within-cluster (level one) categorical variables and between-cluster (level two) variables on smoothing residuals, a kernel smoothed test statistic based on stratification by within-cluster categorical or between-cluster covariates in the logistic multilevel model can be further considered. In addition, the bandwidth selection mechanism over y-space in the logistic multilevel model still deserve to be formally explored (Sturdivant and Hosmer, 2007) and extending the kernel smoothed test statistic by using the local polynomial smoothed residuals to detect misspecification of random effects is also worthy of attempt in future study.

REFERENCES

Aerts, M., and Claeskens, G. (2001), "Bootstrap Tests for Misspecified Models, with Application to Clustered Binary Data," *Computational Statistics and Data Analysis*, 36, 383–401.

Aerts, M., Claeskens, G., and Hart, J. D. (1999), "Testing the Fit of a Parametric Function," *Journal of the American Statistical Association*, 94, 869–879.

—— (2000)," Testing Lack of Fit in Multiple Regression," *Biometrika*, 87, 405–424.

Agresti, A. (2002), *Categorical Data Analysis* (2nd ed.), New York: Wiley.

Agresti, A., Caffo, B., and Ohman-Strickland, P. (2004), "Examples in which Misspecification of a Random Effects Distribution Reduces Efficiency, and Possible Remedies," *Computational Statistics and Data Analysis*, 47, 639–653.

Alonso, A., Geys, H., Molenberghs, G., Kenward, M. G., and Vangeneugden, T. (2004), "Validation of Surrogate Markers in Multiple Randomized Clinical Trials with Repeated Measurements: Canonical Correlation Approach," *Biometrics*, 60, 845–853.

Alonso, A., Litière, S., and Molenberghs, G. (2008), "A Family of Tests to Detect Misspecifications in the Random-Effects Structure of Generalized Linear Mixed Models," *Computational Statistics and Data Analysis*, 52, 4474–4486.

—— (2010), "Testing for Misspecification in Generalized Linear Mixed Models," *Biostatistics*, 11, 771–786.

Austin, P. C. (2010), “Estimating Multilevel Logistic Regression Models When the Number of Clusters is Low: A Comparison of Different Statistical Software Procedures,” *International Journal of Biostatistics*, 6, 1–18.

Booth, J. G., and Hobert, J. P. (1999), “Maximizing Generalized Linear Mixed Models Likelihoods with an Automated Monte Carlo EM Algorithm,” *Journal of the Royal Statistical Society*, Ser. B, 61, 265–285.

Breslow, N. E., and Clayton, D. G. (1993), “Approximate Inference in Generalized Mixed Models,” *Journal of the American Statistical Association*, 88, 9-25.

Chen, J., Zhang, D., and Davidian, M. (2002), “A Monte Carlo EM Algorithm for Generalized Linear Mixed Models with Flexible Random Effects Distribution,” *Biostatistics*, 3, 347–360.

Claeskens, G., and Hart, J. D. (2009),“ Goodness-of-Fit Tests in Mixed Models,” *Test*, 18, 213–239.

Cox, D. R., and Hinkley, D. V. (1974), *Theoretical Statistics*, London: Chapman and Hall.

Diggle, P. J., Heagerty, P. J., Liang, K. Y., and Zeger, S. L. (2002), *Analysis of Longitudinal Data*, New York: Oxford University Press Inc..

Evans, S. R., and Hosmer, D. W. (2004), “Goodness of Fit Tests in Mixed Effects Logistic Models Characterized by Clustering,” *Communications in Statistics, Theory and Methods*, 33, 1139–1155.

Fan, J., and Gijbeles, I. (1996), *Local Polynomial Modelling and its Applications*, London: Chapman and Hall.

Fowlkes, E. (1987), "Some Diagnostics for Binary Logistic Regression via Smoothing Methods," *Biometrika*, 74, 503–515.

Gallant, A. R., and Nychka, D. A. (1987), "Seminonparametric Maximum Likelihood Estimation," *Econometrica*, 55, 363–390.

Goldstein, H., and Rasbash, J. (1996), "Improved Approximations for Multilevel Models with Binary Responses," *Journal of the Royal Statistical Society*, Ser. A, 159, 505–513.

Green, P. J. (1987), "Penalized Likelihood for General Semi-parametric Regression Models," *International Statistical Review*, 55, 245–259.

Hardin, J. W., and Hilbe, J. M. (2003), *Generalized Estimating Equations*, London: Chapman and Hall.

Hart, J. D. (1997), *Nonparametric Smoothing and Lack-of-Fit Tests*, New York: Springer-Verlag.

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

Hosmer, D. W., Hosmer, T., le Cessie, S., and Lemeshow, S. (1997), "A Comparison of Goodness-of-Fit Tests for the Logistic Regression Model," *Statistics in Medicine*, 16, 965–980.

Huang, X. (2009), "Diagnosis of Random-Effect Model Misspecification in Generalized Linear Mixed Models for Binary Response," *Biometrics*, 65, 361–368.

Jiang, J. (2007), *Linear and Generalized Linear Mixed Models and Their Applications*, New York: Springer.

Jiang, J., Luan, Y., and Y-G, Wang. (2007), "Iterative Estimating Equations: Linear Convergence and Asymptotic Properties," *Annals of Statistics*, 35, 2233–2260.

le Cessie, S., and van Houwelingen, J. C. (1991), "A Goodness-of-Fit Test for Binary Regression Models Based on Smoothing Methods," *Biometrics*, 47, 1267–1282.

Lee, S. (1994), "Optimal Choice Between Parametric and Nonparametric Bootstrap Estimates," *Mathematical Proceedings of the Cambridge Philosophical Society*, 115, 335–363.

Lin, K., Chen, Y., and Shyr, Y. (2008), "A Nonparametric Smoothing Method for Assessing GEE Models with Logitudinal Binary Data," *Statistics in Medicine*, 27, 4428–4439.

Lin, X., and Breslow, N. E. (1996), "Bias Correction in Generalized Linear Mixed Models with Multiple Components of Dispersion," *Journal of the American Statistical Association*, 91, 1007-1016.

Litière, S., Alonso, A., and Molenberghs, G. (2007), "Type I and Type II Errors under Random-Effects Misspecification in Generalized Linear Mixed Models," *Biometrics*, 63, 1038–1044.

—— (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.

McCullagh, P., and Nelder, J. A. (1989), *Generalized Linear Models* (2nd ed.), London: Chapman and Hall.

Molenberghs, G., and Verbeke, G. (2005), *Models for Discrete Longitudinal Data*, New York: Springer.

Neuhaus, J. M., and Hauck, W. W. (1992), "The Effects of Mixture Distribution Misspecification When Fitting Mixed-Effects Logistic Models," *Biometrika*, 79, 755–762.

Pan, Z., and Lin D. Y. (2005), "Goodness-of-Fit Methods for Generalized Linear Mixed Models," *Biometrics*, 57, 120–125.

Racine, J. S., and MacKinnon, J. G. (2007), "Inference via Kernel Smoothing of Bootstrap P values," *Computational Statistics and Data Analysis*, 51, 5949–5957.

Ruppert, D., and Wand, M. P. (1994), "Multivariate Locally Weighted Least Squares Regression," *Annals of Statistics*, 22, 1346–1370.

Shao, J., and Tu, D. (1995), *The Jackknife and Bootstrap*, New York: Springer.

Stuart, A., and Ord, J. K. (1987), *Kendall's Advanced Theory of Statistics* (5th ed.), Vol. 1, New York: Oxford University Press Inc..

Sturdivant, R. X. (2005), "Goodness-of-Fit in Hierarchical Logistic Regression Models," Ph.D. Dissertation, University of Massachusetts Amherst, University Microfilms International, Ann Arbor.

Sturdivant, R. X., and Hosmer, D. W. (2007), "A Smoothed Residual Based Goodness-of-Fit Statistic for Logistic Hierarchical Regression Models," *Computational Statistics and Data Analysis*, 51, 3898–3912.

Tchetgen, E. J., and Coull, B. A. (2006), "A Diagnostic Test for the Mixing Distribution in Generalised Linear Mixed Model," *Biometrika*, 93, 1003–1010.

Verbeke, G., and Lesaffre, E. (1997), "The Effect of Misspecifying the Random-Effects Distribution in Linear Mixed Models for Longitudinal Data," *Computational Statistics and Data Analysis*, 23, 541–556.

Verbeke, G., and Molenberghs, G. (2000), *Linear Mixed Models for Longitudinal Data*, New York: Springer.

Waagepetersen, R. (2006), "A Simulation-based Goodness-of-Fit Test for Random Effects in Generalized Linear Mixed Models," *Scandinavian Journal of Statistics*, 33, 721–731.

Wedderburn, R. W. M. (1974), "Quasi-likelihood Functions, Generalized Linear Models, and the Gauss-Newton Method," *Biometrika*, 61, 439–447.

White, H. (1982), "Maximum Likelihood Estimation of Misspecified Models," *Econometrica*, 50, 1–26.

Zeger, S. L., and Karim, M. R. (1991), "Generalized Linear Models with Random Effects; A Gibbs Sampling Approach," *Journal of the American Statistical Association*, 86, 79–86.

Zeger, S. L., Liang, K. Y., and Albert, P. S. (1988), "Models for Longitudinal Data: A Generalized Estimating Equation Approach," *Biometrics*, 44, 1049–1060.

Zhang, D., and Davidian, M. (2001), "Linear Mixed Models with Flexible Distributions of Random Effects for Longitudinal Data," *Biometrics*, 57, 795–802.

APPENDIX A

DERIVATIONS OF SOME COMPONENTS OF $\Re_M$ IN CHAPTER III

As M=1, the marginal mean function is defined as the follow,

$$\begin{aligned} E_1(y_{ij}) &= \int h(\beta_0 + \beta_1 r_i + \beta_2 t_{ij} + \sigma u_i)\left(a_0 + a_1 u_i\right)^2 f(u_i) du_i \\ &\equiv \mu_1(r_i, t_{ij}; \varphi_1), \end{aligned}$$

where $h(x) = \dfrac{e^x}{1+e^x}$, $u_i \sim N(0,1)$, $\varphi_1 = (\beta_0, \beta_1, \beta_2, \sigma, \psi_1)^T$, $a_0 = \cos(\psi_1)$, and $a_1 = \sin(\psi_1)$.

Let

$$\begin{aligned} Q_{1i}^{(\beta_0)} &= \left[\left(\frac{\partial^2 \mu_{1ij}}{\partial\beta_0\partial\beta_0}\right) \left(\frac{\partial^2 \mu_{1ij}}{\partial\beta_1\partial\beta_0}\right) \left(\frac{\partial^2 \mu_{1ij}}{\partial\beta_2\partial\beta_0}\right) \left(\frac{\partial^2 \mu_{1ij}}{\partial\sigma\partial\beta_0}\right) \left(\frac{\partial^2 \mu_{1ij}}{\partial\psi_1\partial\beta_0}\right)\right]^T, \\ Q_{1i}^{(\beta_1)} &= \left[\left(\frac{\partial^2 \mu_{1ij}}{\partial\beta_0\partial\beta_1}\right) \left(\frac{\partial^2 \mu_{1ij}}{\partial\beta_1\partial\beta_1}\right) \left(\frac{\partial^2 \mu_{1ij}}{\partial\beta_2\partial\beta_1}\right) \left(\frac{\partial^2 \mu_{1ij}}{\partial\sigma\partial\beta_1}\right) \left(\frac{\partial^2 \mu_{1ij}}{\partial\psi_1\partial\beta_1}\right)\right]^T, \\ Q_{1i}^{(\beta_2)} &= \left[\left(\frac{\partial^2 \mu_{1ij}}{\partial\beta_0\partial\beta_2}\right) \left(\frac{\partial^2 \mu_{1ij}}{\partial\beta_1\partial\beta_2}\right) \left(\frac{\partial^2 \mu_{1ij}}{\partial\beta_2\partial\beta_2}\right) \left(\frac{\partial^2 \mu_{1ij}}{\partial\sigma\partial\beta_2}\right) \left(\frac{\partial^2 \mu_{1ij}}{\partial\psi_1\partial\beta_2}\right)\right]^T, \\ Q_{1i}^{(\sigma)} &= \left[\left(\frac{\partial^2 \mu_{1ij}}{\partial\beta_0\partial\sigma}\right) \left(\frac{\partial^2 \mu_{1ij}}{\partial\beta_1\partial\sigma}\right) \left(\frac{\partial^2 \mu_{1ij}}{\partial\beta_2\partial\sigma}\right) \left(\frac{\partial^2 \mu_{1ij}}{\partial\sigma\partial\sigma}\right) \left(\frac{\partial^2 \mu_{1ij}}{\partial\psi_1\partial\sigma}\right)\right]^T, \\ Q_{1i}^{(\psi_1)} &= \left[\left(\frac{\partial^2 \mu_{1ij}}{\partial\beta_0\partial\psi_1}\right) \left(\frac{\partial^2 \mu_{1ij}}{\partial\beta_1\partial\psi_1}\right) \left(\frac{\partial^2 \mu_{1ij}}{\partial\beta_2\partial\psi_1}\right) \left(\frac{\partial^2 \mu_{1ij}}{\partial\sigma\partial\psi_1}\right) \left(\frac{\partial^2 \mu_{1ij}}{\partial\psi_1\partial\psi_1}\right)\right]^T. \end{aligned}$$

When we have a case with balanced data structure, V_{1i} can be defined by

$$V_{1i} = \frac{1}{m}\sum_{i=1}^{m}(\mathbf{y}_i - \boldsymbol{\mu}_{1i})(\mathbf{y}_i - \boldsymbol{\mu}_{1i})^T.$$

We obtain the derivatives of ϕ_{1i} with respect to $\varphi_1 = (\beta_0, \beta_1, \beta_2, \sigma, \psi_1)^T$,

$$\frac{\partial \phi_{1i}}{\partial \beta_0} = Q_{1i}^{(\beta_0)} V_{1i}^{-1}(\mathbf{y}_i - \boldsymbol{\mu}_{1i}) - \dot{\boldsymbol{\mu}}_{1i}^T V_{1i}^{-1} \left(\frac{\partial V_{1i}}{\partial \beta_0} \right) V_{1i}^{-1}(\mathbf{y}_i - \boldsymbol{\mu}_{1i}) - \dot{\boldsymbol{\mu}}_{1i}^T V_{1i}^{-1} \left(\frac{\partial \boldsymbol{\mu}_{1i}}{\partial \beta_0} \right),$$
$$\frac{\partial \phi_{1i}}{\partial \beta_1} = Q_{1i}^{(\beta_1)} V_{1i}^{-1}(\mathbf{y}_i - \boldsymbol{\mu}_{1i}) - \dot{\boldsymbol{\mu}}_{1i}^T V_{1i}^{-1} \left(\frac{\partial V_{1i}}{\partial \beta_1} \right) V_{1i}^{-1}(\mathbf{y}_i - \boldsymbol{\mu}_{1i}) - \dot{\boldsymbol{\mu}}_{1i}^T V_{1i}^{-1} \left(\frac{\partial \boldsymbol{\mu}_{1i}}{\partial \beta_1} \right),$$
$$\frac{\partial \phi_{1i}}{\partial \beta_2} = Q_{1i}^{(\beta_2)} V_{1i}^{-1}(\mathbf{y}_i - \boldsymbol{\mu}_{1i}) - \dot{\boldsymbol{\mu}}_{1i}^T V_{1i}^{-1} \left(\frac{\partial V_{1i}}{\partial \beta_2} \right) V_{1i}^{-1}(\mathbf{y}_i - \boldsymbol{\mu}_{1i}) - \dot{\boldsymbol{\mu}}_{1i}^T V_{1i}^{-1} \left(\frac{\partial \boldsymbol{\mu}_{1i}}{\partial \beta_2} \right),$$
$$\frac{\partial \phi_{1i}}{\partial \sigma} = Q_{1i}^{(\sigma)} V_{1i}^{-1}(\mathbf{y}_i - \boldsymbol{\mu}_{1i}) - \dot{\boldsymbol{\mu}}_{1i}^T V_{1i}^{-1} \left(\frac{\partial V_{1i}}{\partial \sigma} \right) V_{1i}^{-1}(\mathbf{y}_i - \boldsymbol{\mu}_{1i}) - \dot{\boldsymbol{\mu}}_{1i}^T V_{1i}^{-1} \left(\frac{\partial \boldsymbol{\mu}_{1i}}{\partial \sigma} \right),$$
$$\frac{\partial \phi_{1i}}{\partial \psi_1} = Q_{1i}^{(\psi_1)} V_{1i}^{-1}(\mathbf{y}_i - \boldsymbol{\mu}_{1i}) - \dot{\boldsymbol{\mu}}_{1i}^T V_{1i}^{-1} \left(\frac{\partial V_{1i}}{\partial \psi_1} \right) V_{1i}^{-1}(\mathbf{y}_i - \boldsymbol{\mu}_{1i}) - \dot{\boldsymbol{\mu}}_{1i}^T V_{1i}^{-1} \left(\frac{\partial \boldsymbol{\mu}_{1i}}{\partial \psi_1} \right).$$

Then, $\tilde{\mathbf{A}}_{m1}$ and $\tilde{\mathbf{B}}_{m1}$ can be derived.

Similarly, as $M = 2$, the marginal mean function is defined as the follow,

$$\begin{aligned} E_2(y_{ij}) &= \int h(\beta_0 + \beta_1 r_i + \beta_2 t_{ij} + \sigma u_i) \left(a_0^* + a_1^* u_i + a_2^* u_i^2\right)^2 f(u_i) du_i \\ &\equiv \mu_2(r_i, t_{ij}; \varphi_2), \end{aligned}$$

where $h(x) = \dfrac{e^x}{1+e^x}$, $u_i \sim N(0,1)$, $\varphi_2 = (\beta_0, \beta_1, \beta_2, \sigma, \psi_1, \psi_2)^T$, $a_0^* = \cos(\psi_1) - \dfrac{1}{\sqrt{2}} \sin(\psi_1)\sin(\psi_2)$, $a_1^* = \sin(\psi_1)\cos(\psi_2)$, and $a_2^* = \dfrac{1}{\sqrt{2}}\sin(\psi_1)\sin(\psi_2)$. Again, we follow the same procedure as $M = 1$, we also can obtain the derivatives of ϕ_{2i} with respect to $\varphi_2 = (\beta_0, \beta_1, \beta_2, \sigma, \psi_1, \psi_2)^T$ and construct $\tilde{\mathbf{A}}_{m2}$ and $\tilde{\mathbf{B}}_{m2}$.

APPENDIX B

ADDITIONAL RESULTS OF NORMALITY CHECKING FOR TWO ASYMPTOTIC DISTRIBUTIONS IN CHAPTER IV

Table 23. P-values of normality checking of Case 2 to Case 4 for $(1)Z_{S_m}$ and $(2)cS_m^{tran}$.

			Anderson-Darling	Cramer-von Mises
Case 2	$\frac{1}{4}\sqrt{N}$	(1)	0.0056	0.0068
		(2)	0.8889	0.8334
	$\frac{1}{2}\sqrt{N}$	(1)	<0.0001	<0.0001
		(2)	0.2440	0.1908
			Anderson-Darling	Cramer-von Mises
Case 3	$\frac{1}{4}\sqrt{N}$	(1)	0.0006	0.0008
		(2)	0.2368	0.2262
	$\frac{1}{2}\sqrt{N}$	(1)	0.0002	0.0009
		(2)	0.1403	0.1762
			Anderson-Darling	Cramer-von Mises
Case 4	$\frac{1}{4}\sqrt{N}$	(1)	0.1012	0.1902
		(2)	0.8806	0.8834
	$\frac{1}{2}\sqrt{N}$	(1)	0.0024	0.0056
		(2)	0.5956	0.5955

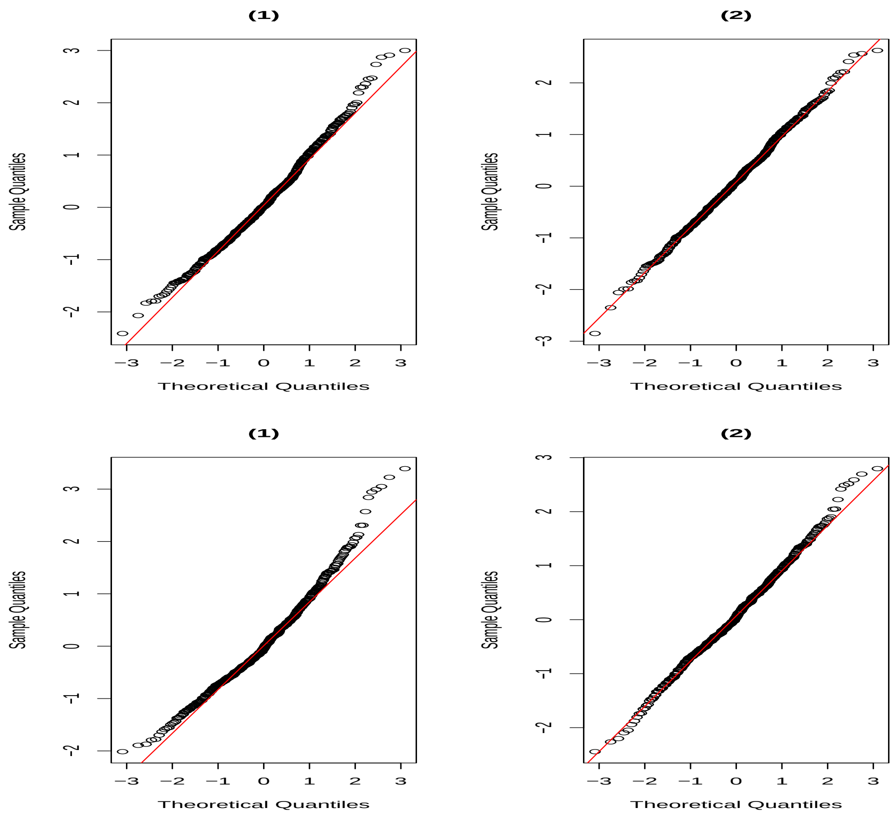

Figure 5. Normal QQ plots of test statistic values of Case 2 under (1)Z_{S_m} and (2)cS_m^{tran}: the smoothing residuals over y-space based on $\frac{1}{4}\sqrt{N}$ and $\frac{1}{2}\sqrt{N}$ are demonstrated on the top and bottom panels, respectively.

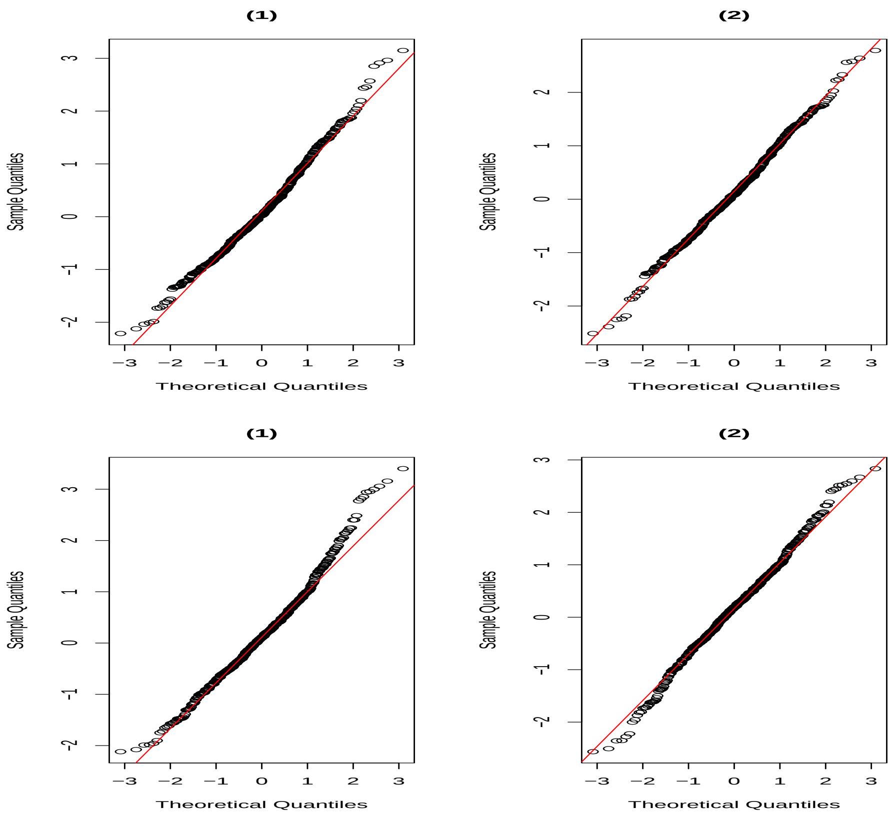

Figure 6. Normal QQ plots of test statistic values of Case 3 under (1)Z_{S_m} and (2)cS_m^{tran}: the smoothing residuals over y-space based on $\frac{1}{4}\sqrt{N}$ and $\frac{1}{2}\sqrt{N}$ are demonstrated on the top and bottom panels, respectively.

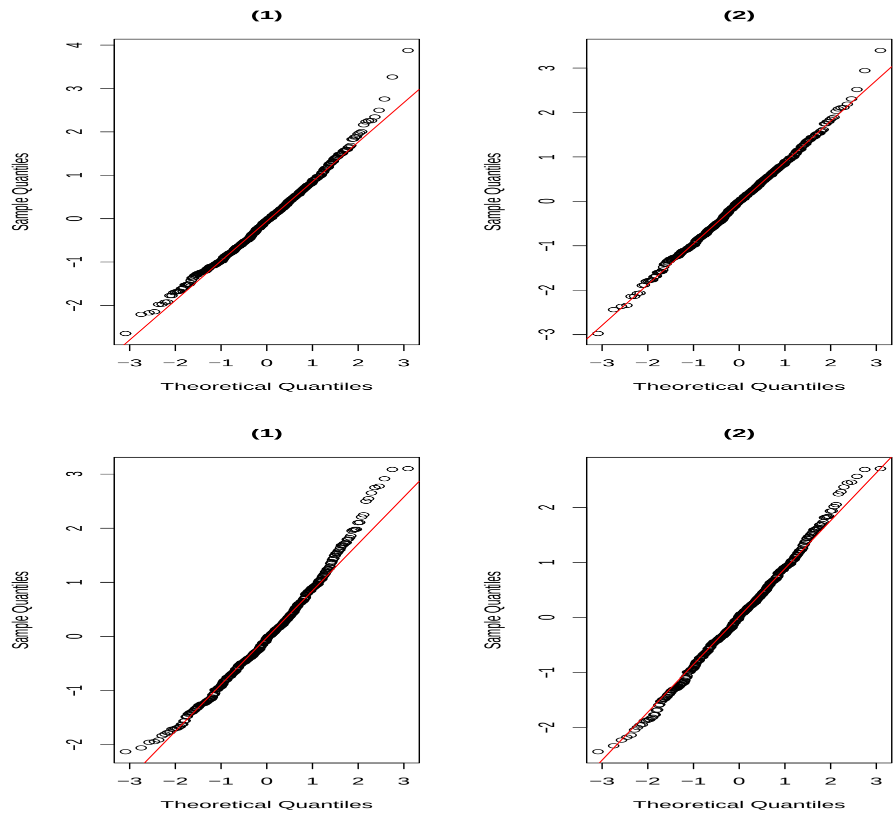

Figure 7. Normal QQ plots of test statistic values of Case 4 under (1)Z_{S_m} and (2)cS_m^{tran}: the smoothing residuals over y-space based on $\frac{1}{4}\sqrt{N}$ and $\frac{1}{2}\sqrt{N}$ are demonstrated on the top and bottom panels, respectively.

VITA

Nai-Wei Chen was born in Kaohsiung, Taiwan. He received a Bachelor of Business Administration degree in statistics at Tunghai University in Taichung, Taiwan in June 2001 and a Master of Science degree in statistics at National Central University in Taoyuan, Taiwan in June 2003. He fulfilled his obligation to perform military service at the Combined Logistics Command, Taiwan in July 2005. In August of 2006, he entered the graduate program in statistics at Texas A&M University in College Station, Texas and earned a Doctor of Philosophy degree in statistics under the advisement of Dr. Thomas E. Wehrly in December 2011. His address is:

Department of Statistics

Texas A&M University

3143 TAMU

College Station, TX 77843–3143