

# Random-intercept misspecification in generalized linear mixed models for binary responses

Shun Yu<sup>1</sup> · Xianzheng Huang<sup>2</sup>

Accepted: 5 February 2017 / Published online: 16 February 2017  
© Springer-Verlag Berlin Heidelberg 2017

**Abstract** We study properties of maximum likelihood estimators of parameters in generalized linear mixed models for a binary response in the presence of random-intercept model misspecification. Further exploiting the test proposed in an existing work initially designed for detecting general random-effects misspecification, we are able to reveal how the true random-intercept distribution deviates from the assumed. Besides this advance compared to the existing methods, we also provide theoretical insights on when and why the proposed test has low power to identify certain forms of misspecification. Large-sample numerical study and finite-sample simulation experiments are carried out to illustrate the theoretical findings.

**Keywords** Bridge distribution · Cluster data · Grouped data · Skew normal

## 1 Introduction

Cluster data arise naturally in a host of applications. For instance, in longitudinal studies, each subject is followed at multiple time points, producing repeated measures for each subject; or, at a fixed time point, measures of multiple traits of each subject are collected; in animal experiments, data on animals in each litter are recorded, with a cluster referring to a litter of animals. It is important in statistical analyses of such

---

**Electronic supplementary material** The online version of this article (doi:[10.1007/s10260-017-0376-0](https://doi.org/10.1007/s10260-017-0376-0)) contains supplementary material, which is available to authorized users.

---

✉ Xianzheng Huang  
huang@stat.sc.edu

<sup>1</sup> Wells Fargo & Company, Charlotte, NC 28282, USA

<sup>2</sup> Department of Statistics, University of South Carolina, Columbia, SC 29208, USA

data to account for the correlation among observations within a cluster. When the response of interest is discrete, generalized linear mixed models (GLMM) provide a practically useful and mathematically flexible platform for analyzing cluster data. Random effects in GLMM are hypothetical, unobservable quantities that play the key role in characterizing cluster-specific features and introducing correlation among observations within a cluster. [Molenberghs and Verbeke \(2005\)](#), [Jiang \(2007\)](#), and [McCulloch et al. \(2008\)](#) provide comprehensive reviews on theories of GLMM and abundant examples of its application.

Most off-the-shelf statistical software carry out analyses of cluster data modeled by GLMM under the assumption that random effects follow normal distributions. There is a sizable collection of literature concerning statistical inference drawn based on this normality assumption when the assumption is violated. Some researchers reported empirical evidence along with theoretical explanations suggesting that likelihood-based inference can be compromised in the presence of random-effects model misspecification ([Ten Have et al. 1999](#); [Heagerty and Kurland 2001](#); [Agresti et al. 2004](#); [Litière et al. 2007, 2008](#)). On the other hand, some studies showed that the bias in covariate (fixed) effects estimation due to this type of misspecification is usually small ([Butler and Louis 1992](#); [Neuhaus et al. 1992](#); [McCulloch and Neuhaus 2011a, b](#); [Neuhaus et al. 2013](#)), although loss of efficiency in these estimators and an inflated Type I error when testing a covariate effect are observed. The consent among these existing works is that random-effects misspecification has non-negligible adverse effects on the estimators of variance components and the fixed intercept. Misleading inference on variance components and the intercept are unattractive in many fields of research. For example, in genetics, reliable inference for variance components are important to scientists ([Scurrah et al. 2000](#)); for many models in the item response theory, the intercept represents the item difficulty, a quantity very much of interest to psychometricians ([Woods 2008](#)). More importantly, random-effects distributions themselves are the focal point of studies in many applications. These applications include surrogate marker evaluation and psychometric properties evaluation, as well as pharmacokinetics (PK) and pharmacodynamics (PD), where mixed effects models are often used to characterize a PK/PD process with practically meaningful random effects that scientists wish to understand.

Relaxing model assumptions on random effects in GLMM is one way to avoid some of the aforementioned adverse effects. This motivates semi/non-parametric modeling for random effects in GLMM, such as those proposed by [Follmann and Lambert \(1989\)](#), [Butler and Louis \(1992\)](#), [Zackin et al. \(1996\)](#), [Magder and Zeger \(1996\)](#), [Kleinman and Ibrahim \(1998\)](#), [Chen et al. \(2002\)](#), [Caffo et al. \(2007\)](#), [Lee and Thompson \(2008\)](#), [Komàrek and Lesaffre \(2008\)](#), [Wang \(2010\)](#), [Papageorgiou and Hinde \(2012\)](#), and [Lesperance et al. \(2014\)](#). In the study presented in this article, we aim at exploiting informative diagnostic methods to assess the adequacy of an assumed random-effect distribution. Many existing methods sharing the same goal as ours are designed to estimate the random effects or their distributions ([Ritz 2004](#); [Pan and Lin 2005](#); [Waagepetersen 2006](#)). There exists a collection of Hausman-type tests constructed based on contrasting a statistic that is robust to such misspecification with another one that is sensitive to it ([White 1982](#); [Tchetgen and Coull 2006](#); [Alonso et al. 2008](#); [Bartolucci et al. 2015](#)), following the general idea of the specification tests pioneered

in Hausman (1978). A very similar idea of comparing two statistics is employed in Bartolucci et al. (2015) to detect time-variant unobserved heterogeneity in generalized linear models (GLM) for panel data, with one feature distinguishing it from the aforementioned Hausman-type tests, which is that both statistics can be sensitive to the time-invariant assumption. Claeskens and Hart (2009) considered order selection-type goodness-of-fit (GOF) tests in the framework of linear mixed models (LMM), which entails nesting the normal distribution for random effects within a bigger, thus more flexible, class of distributions, and then testing for normality against a higher-order, i.e., more complicated, model in the class. In theory, this idea is applicable to GLMM, but the lack of closed-form likelihood functions in general for GLMM discourages its use. Also, most GOF tests are not informative enough to shed light on how the true distribution differs from normal. Verbeke and Molenberghs (2013) proposed an exploratory diagnostic tool based on the gradient function, which can give certain indication of how the true random-effects distribution deviates from normal. Efendi et al. (2014) further constructed a test statistic based on this idea and developed a bootstrap procedure to estimate the  $p$  value associated with the test. Drikvandi et al. (2016) proposed yet another test based on the gradient function that outperforms the test in Efendi et al. (2014), who also derived the asymptotic null distribution of the proposed test statistic. A different aspect of informativeness is achieved by the diagnostic method proposed by Huang (2009), which can distinguish model misspecification on a random intercept from that on a random slope (among possibly multiple random slopes), but it can not tell how the true distribution deviates from the assumed. In this article, we further exploit her test statistic within the subclass of GLMMs where the random intercept is the only random effect, allowing one to gain more information on the direction of deviation of the true random-intercept distribution from an assumed distribution.

Standing in stark contrast to the Hausman-type diagnostic methods, which require some component of robust (to random-effects model assumptions) inference, our method does not require robust estimators, which free us from seeking for consistent inferential methods in the presence of model misspecification. The key ingredient that frees us from robust inference is the grouped data induced from the observed cluster data, which is described in Sect. 2 following the model formulated for the observed data. In Sect. 3, effects of random-intercept misspecification on maximum likelihood estimators (MLE) based on the raw data and the MLEs based on the grouped data are investigated. We revisit the test statistics in Huang (2009) in Sect. 4 to facilitate assessing an assumed random-intercept distribution. Following this review, we seek theoretical explanations for when and why the test has low power to detect certain types of misspecification. Implementation and performance of the proposed diagnostic method are illustrated via simulation in Sect. 5, where we compare our method with that in Tchetgen and Coull (2006) and two tests proposed in Alonso et al. (2008). In Sect. 6, we apply the proposed method to two real-life applications. We conclude the article with discussions on the contributions of our study and future research directions in Sect. 7.

## 2 Data and models

Denote by  $\mathbf{Y}_i = (Y_{i1}, \dots, Y_{in_i})^t$  the vector of observed binary responses from cluster  $i$ , and by  $\mathbf{X}_i$  the  $n_i \times p$  matrix of the associated covariates, with the  $j$ th row being  $\mathbf{X}_{ij}$ , for  $i = 1, \dots, m$ ,  $j = 1, \dots, n_i$ . Suppose that the conditional mean of  $Y_{ij}$  given the covariates and the random effect is

$$E(Y_{ij} | \mathbf{X}_{ij}, b_{i0}; \boldsymbol{\beta}) = h(\mathbf{X}_{ij}\boldsymbol{\beta} + b_{i0}), \text{ for } i = 1, \dots, m, \quad j = 1, \dots, n_i, \quad (1)$$

where  $\boldsymbol{\beta}$  is the  $p \times 1$  vector of fixed effects,  $b_{i0}$  is the cluster-specific random intercept, and  $h(\cdot)$  is a known inverse link function. Under the assumed model of  $b_{i0}$ , which is specified by the probability density function (pdf) indexed by parameter(s)  $\tau$ ,  $f_b(b_{i0}; \tau)$ , the contribution of cluster  $i$  to the observed-data likelihood is  $f_{\mathbf{Y}}(\mathbf{Y}_i | \mathbf{X}_i; \boldsymbol{\beta}, \tau) = \int f_b(b_{i0}; \tau) \prod_{j=1}^{n_i} h(\mathbf{X}_{ij}\boldsymbol{\beta} + b_{i0})^{Y_{ij}} \{1 - h(\mathbf{X}_{ij}\boldsymbol{\beta} + b_{i0})\}^{1-Y_{ij}} db_{i0}$ , for  $i = 1, \dots, m$ . Note that the cluster-specific nature of  $b_{i0}$  (in the true and assumed GLMMs) rules out scenarios with endogeneity, where a random effect depends on (or is correlated with) within-cluster covariates or where there is a time-varying unobserved individual effect as considered in [Bartolucci et al. \(2015\)](#).

To implement the proposed diagnostic method, another data set induced from the above observed data is needed. To create this induced data set, one partitions cluster  $i$  into  $G_i$  groups and define grouped responses as  $Y_{ig}^* = \max_{j \in I_{ig}} Y_{ij}$ , where  $I_{ig}$  is the index set such that  $j \in I_{ig}$  indicates that observation  $j$  is in group  $g$  within cluster  $i$ , for  $i = 1, \dots, m$ ,  $g = 1, \dots, G_i$ . We defer discussions on some practical and theoretical consideration of the actual implementation of such partition to Sect. 7. Using the grouped data, letting  $\mathbf{Y}_i^* = (Y_{i1}^*, \dots, Y_{iG_i}^*)^t$ , one has the contribution of cluster  $i$  to the grouped-data likelihood as  $f_{\mathbf{Y}^*}(\mathbf{Y}_i^* | \mathbf{X}_i; \boldsymbol{\beta}, \tau) = \int f_b(b_{i0}; \tau) \prod_{g=1}^{G_i} [1 - \prod_{j \in I_{ig}} \{1 - h(\mathbf{X}_{ij}\boldsymbol{\beta} + b_{i0})\}]^{Y_{ig}^*} [\prod_{j \in I_{ig}} \{1 - h(\mathbf{X}_{ij}\boldsymbol{\beta} + b_{i0})\}]^{1-Y_{ig}^*} db_{i0}$ , for  $i = 1, \dots, m$ .

Due to its appeal to practitioners, we consider for the majority of the article a normal assumed model for  $b_{i0}$ , where  $\tau$  denotes the standard deviation of  $b_{i0}$ , and  $f_b(b_{i0}; \tau) = \tau^{-1} \phi(b_{i0}/\tau)$ , with  $\phi(\cdot)$  being the standard normal pdf. In addition, we also consider another assumed model for  $b_{i0}$  that has aroused a great deal of interest among statisticians since its proposal in [Wang and Louis \(2003\)](#), namely the bridge distribution. The pdf of a mean-zero bridge distribution (for  $b_{i0}$ ) is given by  $f_b(b_{i0}; \tau) = \sin(\tau\pi) / [2\pi \{\cosh(\tau b_{i0}) + \cos(\tau\pi)\}]$ , where  $\tau \in (0, 1)$  and the standard deviation of  $b_{i0}$  now becomes  $\pi \sqrt{(\tau^{-2} - 1)/3}$ . Like a normal distribution, the pdf of a bridge distribution is symmetric and bell-shaped but with a heavier tail. The appealing feature of a bridge distribution for  $b_{i0}$  is that, if the link function in a GLMM is the logit link and  $b_{i0}$  is the only random effect (as in our study), then the logit link is retained in the marginal model of  $Y_{ij}$  (given  $\mathbf{X}_{ij}$ ).

Define  $\boldsymbol{\Omega} = (\boldsymbol{\beta}^t, \tau)^t$  as the vector of all unknown parameters in the assumed GLMM, and denote by  $\hat{\boldsymbol{\Omega}}$  the MLE of  $\boldsymbol{\Omega}$  derived under the assumed model based on the observed data, and by  $\hat{\boldsymbol{\Omega}}^*$  as the counterpart MLE based on the grouped data. In the next section, we look into the limits of  $\hat{\boldsymbol{\Omega}}$  and  $\hat{\boldsymbol{\Omega}}^*$  as  $m \rightarrow \infty$  (with  $\max_{1 \leq i \leq m} n_i$  bounded) when the true distribution of  $b_{i0}$  deviates from the assumed model in different ways. Throughout the article, the notational conventions used to denote the (limiting) MLEs

of  $\Omega$  based on the observed/grouped data are also adopted for individual parameters in  $\Omega$ .

### 3 Limits of maximum likelihood estimators

By invoking the Kullback–Leibler (KL) divergence between the true model and an assumed model, White (1982) established that  $\hat{\Omega}$  converges almost surely to

$$\tilde{\Omega} = \arg \min_{\Omega} \lim_{m \rightarrow \infty} m^{-1} E_{\mathbf{Y}|\mathbf{X}} \left\{ \log \frac{g_{\mathbf{Y}}(\mathbf{Y}|\mathbf{X}; \Omega_0)}{f_{\mathbf{Y}}(\mathbf{Y}|\mathbf{X}; \Omega)} \right\}, \tag{2}$$

where  $g_{\mathbf{Y}}(\mathbf{Y}|\mathbf{X}; \Omega_0)$  is the likelihood function of  $\mathbf{Y} = \{\mathbf{Y}_i\}_{i=1}^m$  given  $\mathbf{X} = \{\mathbf{X}_i\}_{i=1}^m$  under the correct model evaluated at the true parameter value  $\Omega_0$ ,  $f_{\mathbf{Y}}(\mathbf{Y}|\mathbf{X}; \Omega)$  is the observed-data likelihood under the assumed model, and the expectation is taken with respect to the true model of  $\mathbf{Y}$  given  $\mathbf{X}$ . Similarly,  $\hat{\Omega}^*$  converges almost surely to

$$\tilde{\Omega}^* = \arg \min_{\Omega} \lim_{m \rightarrow \infty} m^{-1} E_{\mathbf{Y}^*|\mathbf{X}} \left\{ \log \frac{g_{\mathbf{Y}^*}(\mathbf{Y}^*|\mathbf{X}; \Omega_0)}{f_{\mathbf{Y}^*}(\mathbf{Y}^*|\mathbf{X}; \Omega)} \right\}, \tag{3}$$

where  $g_{\mathbf{Y}^*}(\mathbf{Y}^*|\mathbf{X}; \Omega_0)$  is the true-model likelihood of grouped data  $\mathbf{Y}^* = \{\mathbf{Y}_i^*\}_{i=1}^m$ ,  $f_{\mathbf{Y}^*}(\mathbf{Y}^*|\mathbf{X}; \Omega)$  is the grouped-data likelihood under the assumed model, and the expectation is taken with respect to the true model of  $\mathbf{Y}^*$  given  $\mathbf{X}$ .

In the upcoming two subsections, we first focus on a specific GLMM with the logit link, under which we are able to evaluate the KL divergence appearing in (2) and (3) nearly exactly. We then switch to a similar GLMM but with the probit link and the true distribution of  $b_{i0}$  being a skew normal, which allows for more analytic derivations to reveal some properties of the MLEs under a misspecified random-intercept model.

#### 3.1 Logistic GLMM

Consider a logistic regression model for  $Y_{ij}$  given by, for  $i = 1, \dots, m$  and  $j = 1, \dots, 8$ ,

$$E(Y_{ij}|\mathbf{X}_{ij}, b_{i0}) = \{1 + \exp(-\beta_0 - \beta_1 X_{ij,1} - \beta_2 X_{ij,2} - \beta_3 X_{ij,1} X_{ij,2} - b_{i0})\}^{-1}, \tag{4}$$

where  $X_{ij,1} = x_i$  with  $x_i$  equal to either 0 or 1,  $X_{ij,2} = (j - 1)/7$ , and the true fixed effects values are  $\beta = (\beta_0, \beta_1, \beta_2, \beta_3)^t = (-2, 1, 0.5, -0.25)^t$ . With  $n_i = 8$  for  $i = 1, \dots, m$ , the sample space of  $\mathbf{Y}_i$  consists of  $256 (= 2^8)$  distinct binary-valued  $8 \times 1$  vectors,  $\{\mathbf{y}_l\}_{l=1}^{256}$ . And, under the current covariates setting,  $\mathbf{X}_i$  is equal to either  $\mathbf{X}^{(1)} = [\mathbf{1} \mathbf{0} \mathbf{S} \mathbf{0}]$  or  $\mathbf{X}^{(2)} = [\mathbf{1} \mathbf{1} \mathbf{S} \mathbf{S}]$ , where  $\mathbf{1}$  is the  $8 \times 1$  vector of ones,  $\mathbf{0}$  is the  $8 \times 1$  vector of zeros, and  $\mathbf{S} = (0, 1, \dots, 7)^t/7$ . Now specifying  $g_{\mathbf{Y}}(\mathbf{Y}_i|\mathbf{X}_i; \Omega_0)$  boils down to finding two sets of probabilities,  $\{\pi^{(k)}(\mathbf{y}_l)\}_{l=1}^{256}$ , for  $k = 1, 2$ , where  $\pi^{(k)}(\mathbf{y}_l) = g_{\mathbf{Y}}(\mathbf{y}_l|\mathbf{X}^{(k)}; \Omega_0)$ , for  $l = 1, \dots, 256$ . Because deriving these  $512 (= 2 \times 256)$  probabilities explicitly is infeasible even when the true distribution of  $b_{i0}$

is normal, we use empirical probabilities observed from a large sample generated from the true model to approximate these true-model probabilities. More specifically, for each fixed-effect design matrix  $\mathbf{X}^{(k)}$  ( $k = 1, 2$ ), we generate a random sample  $\{\mathbf{Y}_i\}_{i=1}^{10^6}$  from the true model with  $\mathbf{X}_i = \mathbf{X}^{(k)}$  for  $i = 1, \dots, 10^6$ , and use the observed proportion of  $y_l$  to approximate  $\pi^{(k)}(y_l)$  for  $l = 1, \dots, 256$ . We further assume that  $\mathbf{X}_i$  takes  $\mathbf{X}^{(1)}$  or  $\mathbf{X}^{(2)}$  equally likely across the entire sample. Once we obtain the two sets of true-model probabilities, finding  $\tilde{\boldsymbol{\Omega}}$  defined in (2) is equivalent to maximizing

$$\sum_{l=1}^{256} \sum_{k=1}^2 \pi^{(k)}(y_l) \log f_{\mathbf{Y}}(y_l | \mathbf{X}^{(k)}; \boldsymbol{\Omega}) \quad (5)$$

over  $\boldsymbol{\Omega}$ , where the integral that defines  $f_{\mathbf{Y}}(y_l | \mathbf{X}^{(k)}; \boldsymbol{\Omega})$ , for  $k = 1, 2$  and  $l = 1, \dots, 256$ , is numerically solved using the  $M$ -point Gauss–Hermite quadrature, with a suitably large  $M$ . This recipe of finding  $\tilde{\boldsymbol{\Omega}}$  follows the idea of artificial sample proposed by Rotnitzky and Wypij (1994), which was employed by Heagerty and Kurland (2001) and Huang (2009) to numerically obtain limiting MLEs under misspecified models. The accuracy of the resulting  $\tilde{\boldsymbol{\Omega}}$  has been shown to be satisfactory in these existing works.

To fix the idea of grouped data, we set  $G_i = 2$ , for  $i = 1, \dots, m$ , and split each cluster right at the middle according to the sorted values in  $X_{ij,2}$ . We then repeat the same exercise as above to find  $\tilde{\boldsymbol{\Omega}}^*$  by maximizing the objective function,  $\sum_{l=1}^4 \sum_{k=1}^2 \pi^{(k)}(y_l^*) \log f_{\mathbf{Y}^*}(y_l^* | \mathbf{X}^{(k)}; \boldsymbol{\Omega})$ , over  $\boldsymbol{\Omega}$ , where  $\pi^{(k)}(y_l^*) = g_{\mathbf{Y}^*}(y_l^* | \mathbf{X}^{(k)}; \boldsymbol{\Omega}_0)$ , for  $k = 1, 2$  and  $l = 1, \dots, 4$ .

With the theoretical formulations needed to numerically obtain  $\tilde{\boldsymbol{\Omega}}$  and  $\tilde{\boldsymbol{\Omega}}^*$  in position, we handpick the following four mean-zero true random-intercept distributions to pair with either the normal distribution or the bridge distribution as the assumed model for  $b_{i0}$ : (I) the standard normal distribution (when the assumed model is normal) or the bridge distribution with  $\tau = 0.5$  (when the assumed model is a bridge distribution); (II) a Student's  $t$  distribution with degrees of freedom 8/3, resulting in a standard deviation of 2; (III) a shifted (via centering) and scaled (via multiplying  $-2$ ) left-skewed gamma distribution with standard deviation 2, based on a gamma distribution with shape and scale parameters both equal to 1 (before shifting and scaling); (IV) a shifted and scaled right-skewed gamma distribution that is symmetric of (III). Table 1 presents  $\tilde{\boldsymbol{\Omega}}$  and  $\tilde{\boldsymbol{\Omega}}^*$  under different true models for  $b_{i0}$  when one assumes a normal  $b_{i0}$  (upper half of Table 1) and when one assumes  $b_{i0}$  follow a bridge distribution (lower half of Table 1).

Results regarding  $\boldsymbol{\beta}$  shown in Table 1 are in line with the finding in Neuhaus et al. (1992), who concluded that the bias in the MLEs of the fixed effects is usually small, and  $\tilde{\boldsymbol{\beta}}_0$  can differ more from the truth when the true distribution of  $b_{i0}$  is skewed. We reach a similar conclusion for  $\tilde{\boldsymbol{\beta}}^*$ , which is not considered in the existing literature. Results regarding  $\tau$  shown in Table 1 are more intriguing. When the true distribution of  $b_{i0}$  is symmetric around zero but different from the assumed, a small or moderate amount of bias is present in  $\tilde{\tau}$  and  $\tilde{\tau}^*$ , but the relative change from  $\tilde{\tau}$  to  $\tilde{\tau}^*$  is minimal. In contrast, when the true random-intercept distribution is skewed, although the bias

**Table 1** Limiting MLEs based on the observed data,  $\tilde{\Omega}$ , and limiting MLEs based on the grouped data,  $\tilde{\Omega}^*$ , under different combinations of true/assumed models for  $b_{i0}$

True model	Data	$\beta_0$	$\beta_1$	$\beta_2$	$\beta_3$	$\tau$	
Assume $b_{i0} \sim N(0, \tau^2)$							
(I) $N(0, 1)$	Observed	-2.00	1.00	0.50	-0.25	1.00	
	Grouped	-2.00	1.00	0.50	-0.25	1.00	(0.00%)
(II) Student's $t$	Observed	-2.01	1.00	0.49	-0.24	1.40	
	Grouped	-2.01	1.00	0.50	-0.24	1.29	(7.86%)
(III) Left-skewed gamma	Observed	-1.67	1.00	0.51	-0.27	1.31	
	Grouped	-1.63	1.13	0.57	-0.34	1.97	(-50.38%)
(IV) Right-skewed gamma	Observed	-2.37	1.02	0.48	-0.23	2.12	
	Grouped	-2.30	0.91	0.44	-0.18	1.51	(28.77%)
Assume $b_{i0} \sim \text{bridge}(\tau)$							
(I) Bridge(0.5)	Observed	-2.00	1.00	0.50	-0.25	0.50	
	Grouped	-2.00	1.00	0.50	-0.25	0.50	(0.00%)
(II) Student's $t$	Observed	-2.00	1.00	0.50	-0.25	0.75	
	Grouped	-2.00	1.00	0.50	-0.24	0.75	(0.00%)
(III) Left-skewed gamma	Observed	-1.60	1.01	0.52	-0.28	0.74	
	Grouped	-1.62	1.10	0.56	-0.31	0.61	(17.57%)
(IV) Right-skewed gamma	Observed	-2.40	1.01	0.49	-0.23	0.62	
	Grouped	-2.30	0.92	0.45	-0.19	0.69	(-11.29%)

The last column gives values of  $100\% \times (\tilde{\tau} - \tilde{\tau}^*)/\tilde{\tau}$  in parentheses

in  $\tilde{\tau}$  and  $\tilde{\tau}^*$  may be small or moderate, the relative change from  $\tilde{\tau}$  to  $\tilde{\tau}^*$  is much more substantial. More interestingly, after translating  $\tilde{\tau}$  and  $\tilde{\tau}^*$  to the corresponding limiting MLEs of the standard deviation of  $b_{i0}$ , denoted by  $\tilde{\sigma}$  and  $\tilde{\sigma}^*$  respectively, we observe that  $\tilde{\sigma} - \tilde{\sigma}^* < 0$  when the true distribution of  $b_{i0}$  is left-skewed, whereas  $\tilde{\sigma} - \tilde{\sigma}^* > 0$  when the true model is right-skewed. This pattern is observed under both assumed models for  $b_{i0}$ . The same clean contrast between  $\tilde{\sigma}$  and  $\tilde{\sigma}^*$  is also observed when the true model is multimodal with different directions of skewness (results omitted here). Later, in Sect. 4.2, we will provide some theoretical insight on the reason for this tantalizing phenomenon. Besides the skewness of the true random-intercept distribution, [Alonso et al. \(2008\)](#) provided empirical evidence showing that how much the MLEs are affected by random-intercept misspecification also depends on the variance of the true random-intercept distribution. Bearing this finding in mind, we purposefully set the variance of distributions (II), (III), and (IV) above to be the same across these three cases, which all correspond to scenarios where the assumed random-intercept model (either normal or bridge) deviates from the true model. With this common variance changed to other values (results not included in Table 1), the relative change from  $\tilde{\sigma}$  to  $\tilde{\sigma}^*$  remains substantial when the true model is skewed, with the sign of  $\tilde{\sigma} - \tilde{\sigma}^*$  reflecting the direction of skewness.

Intrigued by the pattern of the limiting MLEs of  $\tau$ , we propose to use  $\hat{\tau} - \hat{\tau}^*$  as the basis of an informative diagnostic test to reveal the direction of skewness of the

true random-intercept distribution. Before we present a formal test statistic for this purpose, we wish to look more closely into the dependence of  $\tilde{\boldsymbol{\Omega}} - \tilde{\boldsymbol{\Omega}}^*$  (and individual limiting MLEs) on the skewness of the true random-intercept distribution without being distracted by other features of the true distribution. This close inspection is made possible by considering a probit GLMM with the true model of  $b_{i0}$  specified by a skew normal pdf, as elaborated in the next subsection.

### 3.2 Probit GLMM

Consider a probit regression model for  $Y_{ij}$ ,  $E(Y_{ij}|\mathbf{X}_{ij}, b_{i0}) = \Phi(\mathbf{X}_{ij}\boldsymbol{\beta} + b_{i0})$ , for  $i = 1, \dots, m$  and  $j = 1, \dots, n_i$ , where  $\Phi(\cdot)$  is the standard normal cumulative distribution function (cdf). Suppose that one assumes normal  $b_{i0}$ , but the true distribution of  $b_{i0}$  is a skew normal (SN, [Azzalini 1985](#)) with mean zero and variance  $\sigma^2$ , of which the pdf is  $g_b(b_{i0}; \sigma, \alpha) = (2/\omega)\phi\{(b_{i0} + \alpha d\sigma)/\omega\}\Phi\{(\alpha/\omega)(b_{i0} + \alpha d\sigma)\}$ , where  $\alpha$  is the shape parameter,  $d = \sqrt{2/\{\pi(1 + \alpha^2) - 2\alpha^2\}}$ , and  $\omega = \sigma\sqrt{\pi(1 + \alpha^2)/\{\pi(1 + \alpha^2) - 2\alpha^2\}}$ . Note that  $\alpha < 0$  results in a left-skewed distribution,  $\alpha > 0$  leads to a right-skewed distribution, and  $\alpha = 0$  yields a normal distribution.

With the probit regression model in conjunction with the SN random intercept, integrals used to define the true/assumed observed-data likelihood can be derived analytically up to the point where one only needs to evaluate the cdf of a multivariate normal. As shown in Appendix S1 of the supplementary material, the true observed-data likelihood for cluster  $i$  is, for  $i = 1, \dots, m$ ,

$$g_{\mathbf{Y}}(\mathbf{Y}_i|\mathbf{X}_i; \boldsymbol{\Omega}) = 2P \left\{ V_{i0} \leq 0, V_{ij} \leq (-1)^{Y_{ij}} (-\mathbf{X}_{ij}\boldsymbol{\beta} + \alpha d\sigma), j = 1, \dots, n_i \right\}, \tag{6}$$

where  $\mathbf{V}_i = (V_{i0}, V_{i1}, \dots, V_{in_i})^t \sim N_{n+1}(\mathbf{0}, \mathbf{S}_{\mathbf{V}_i})$ , in which

$$\mathbf{S}_{\mathbf{V}_i} = \begin{bmatrix} 1 + \alpha^2 & (-1)^{Y_{i1}+1}\alpha\omega & (-1)^{Y_{i2}+1}\alpha\omega & \dots & (-1)^{Y_{in_i}+1}\alpha\omega \\ (-1)^{Y_{i1}+1}\alpha\omega & 1 + \omega^2 & (-1)^{Y_{i1}+Y_{i2}}\omega^2 & \dots & (-1)^{Y_{i1}+Y_{in_i}}\omega^2 \\ (-1)^{Y_{i2}+1}\alpha\omega & (-1)^{Y_{i1}+Y_{i2}}\omega^2 & 1 + \omega^2 & \dots & (-1)^{Y_{i2}+Y_{in_i}}\omega^2 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ (-1)^{Y_{in_i}+1}\alpha\omega & (-1)^{Y_{i1}+Y_{in_i}}\omega^2 & (-1)^{Y_{i2}+Y_{in_i}}\omega^2 & \dots & 1 + \omega^2 \end{bmatrix}. \tag{7}$$

The observed-data likelihood function under the assumed model can be obtained by setting  $\alpha = 0$  in (6) and (7), resulting in the following likelihood evaluated at  $\tilde{\boldsymbol{\Omega}}$ , for  $i = 1, \dots, m$ ,

$$f_{\mathbf{Y}}(\mathbf{Y}_i|\mathbf{X}_i; \tilde{\boldsymbol{\Omega}}) = 2P \left\{ W_{i0} \leq 0, W_{ij} \leq (-1)^{Y_{ij}+1}\mathbf{X}_{ij}\tilde{\boldsymbol{\beta}}, j = 1, \dots, n_i \right\}, \tag{8}$$



where  $\mathbf{W}_i = (W_{i0}, W_{i1}, \dots, W_{in})^t \sim N_{n+1}(\mathbf{0}, \mathbf{S}\mathbf{W}_i)$  with

$$\mathbf{S}\mathbf{W}_i = \begin{bmatrix} 1 & 0 & 0 & \dots & 0 \\ 0 & 1 + \tilde{\sigma}^2 & (-1)^{Y_{i1}+Y_{i2}} \tilde{\sigma}^2 & \dots & (-1)^{Y_{i1}+Y_{in}} \tilde{\sigma}^2 \\ 0 & (-1)^{Y_{i1}+Y_{i2}} \tilde{\sigma}^2 & 1 + \tilde{\sigma}^2 & \dots & (-1)^{Y_{i2}+Y_{in}} \tilde{\sigma}^2 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & (-1)^{Y_{i1}+Y_{in}} \tilde{\sigma}^2 & (-1)^{Y_{i2}+Y_{in}} \tilde{\sigma}^2 & \dots & 1 + \tilde{\sigma}^2 \end{bmatrix}. \tag{9}$$

Similarly, integrals that define the true/assumed grouped-data likelihood can also be greatly simplified, as elaborated in Appendix S1 of the supplementary material.

Having the aforementioned four (nearly) explicit likelihood functions not only improves the transparency of the numerical search for  $\tilde{\boldsymbol{\Omega}}$  and  $\tilde{\boldsymbol{\Omega}}^*$  as maximizers of the corresponding objective functions (such as (5)), but also allows analytic exploration of the dependence of these limiting MLEs on  $\alpha$ . In particular, because  $\tilde{\boldsymbol{\Omega}}$  is the point in the  $r$ -dimensional parameter space that makes the wrong-model likelihood “look like” the truth-model likelihood evaluated at the true parameter values, one may compare (6)–(7) with (8)–(9) to get a sense of how  $\tilde{\boldsymbol{\Omega}}$  differs from  $\boldsymbol{\Omega}_0$ . For example, to simplify the comparison, we approximate (8) by setting  $(\tilde{\beta}_1, \tilde{\beta}_2, \tilde{\beta}_3)$  at the true parameter values,  $(\beta_1, \beta_2, \beta_3)$ . This approximation is not far-fetched because, as shown in Neuhaus et al. (1992),  $(\hat{\beta}_1, \hat{\beta}_2, \hat{\beta}_3)^t$  are very close to being consistent in the presence of most random-intercept model misspecification. With this approximation, one can heuristically compare  $\tilde{\beta}_0$  with the truth,  $\beta_0$ , by matching the upper bounds involved in the cdf in (6) with those in (8), that is, setting  $(-1)^{Y_{ij}}(-\mathbf{X}_{ij}\boldsymbol{\beta} + \alpha d\sigma) \approx (-1)^{Y_{ij}+1}\mathbf{X}_{ij}\tilde{\boldsymbol{\beta}}$ , resulting in  $\tilde{\beta}_0 \approx \beta_0 - \alpha d\sigma$ . Because  $d, \sigma > 0$ , this approximation for  $\tilde{\beta}_0$  indicates positive bias in  $\tilde{\beta}_0$  when  $\alpha < 0$ , negative bias when  $\alpha > 0$ , and zero bias when  $\alpha = 0$  (as expected).

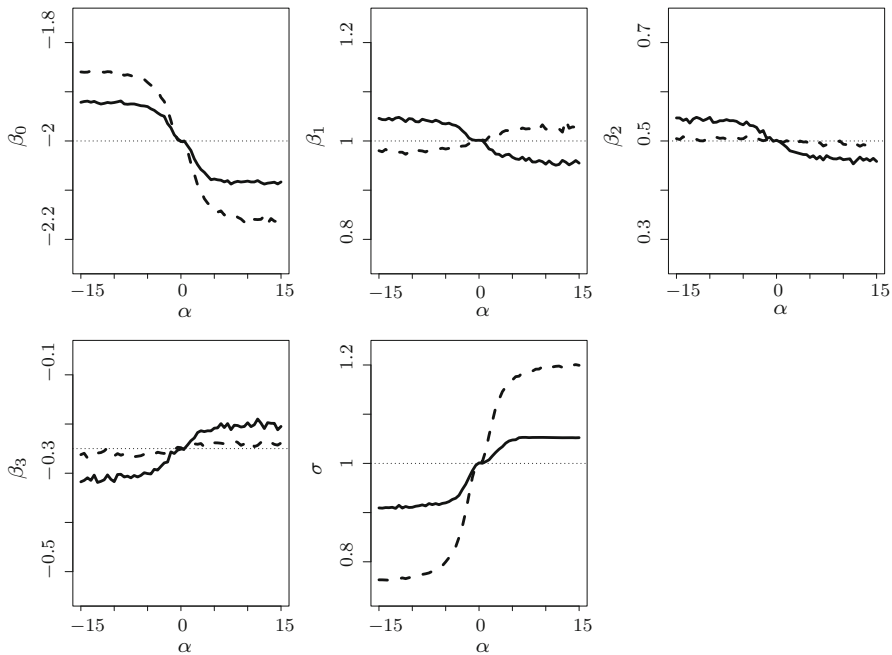
Using the same true values of  $\boldsymbol{\beta}$  and covariates setting as those in Sect. 3.1, fixing  $\sigma$  at 1 in the true SN pdf and varying  $\alpha$  from  $-15$  to  $15$  at increments of  $0.5$ , we obtain  $\tilde{\boldsymbol{\Omega}}$  and  $\tilde{\boldsymbol{\Omega}}^*$  as functions of  $\alpha$ . Figure 1 shows these functions and Fig. 2 depicts  $\tilde{\boldsymbol{\Omega}} - \tilde{\boldsymbol{\Omega}}^*$ . Note that the direction of bias in  $\tilde{\beta}_0$  demonstrated in Fig. 1 reconciles with the above heuristic arguments about  $\tilde{\beta}_0$ . Figure 2 highlights the interesting phenomenon, partly underscored in Sect. 3.1, that the sign of  $\tilde{\theta} - \tilde{\theta}^*$  solely depends on the sign of  $\alpha$ , where  $\theta$  denotes generically a parameter in  $\boldsymbol{\Omega}$ ; and the absolute difference,  $|\tilde{\sigma} - \tilde{\sigma}^*|$ , is much larger than that associated with any other parameter when  $\alpha \neq 0$ .

## 4 Model diagnostics

### 4.1 Revisit Huang’s test

Properties of limiting MLEs revealed in Sect. 3 suggest the following test statistic to facilitate diagnosing random-intercept model misspecification,

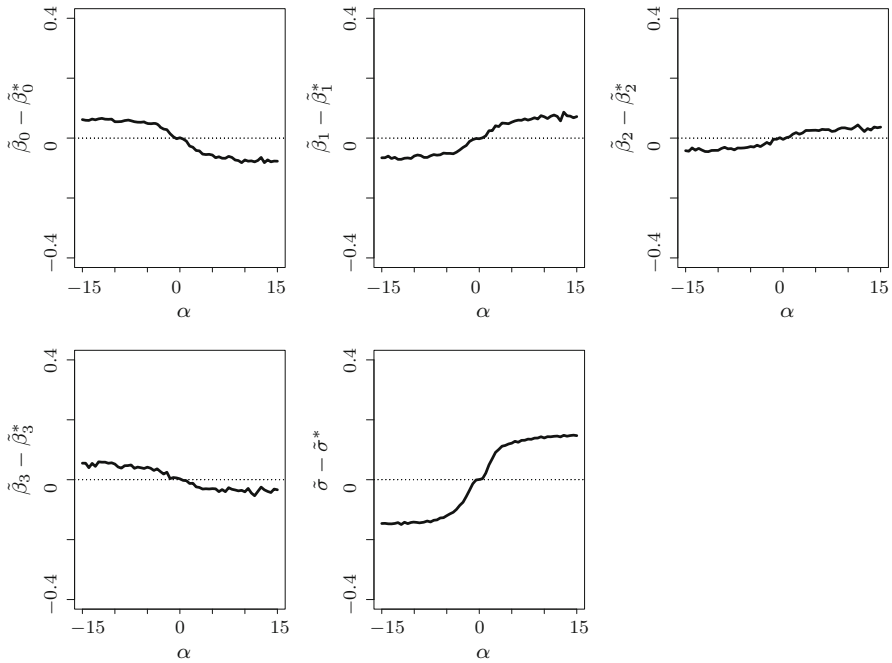
$$t_\tau = (\hat{\tau} - \hat{\tau}^*) / \hat{v}_\tau, \tag{10}$$



**Fig. 1** Limiting MIEs of  $\Omega$  based on the observed (dashed lines)/grouped (solid lines) data versus  $\alpha$  for probit GLMM with skew normal  $b_{i0}$ . The horizontal dotted reference line in each panel refers to the true value of the corresponding parameter

where  $\hat{v}_\tau$  is an estimator of the standard error of  $\hat{\tau} - \tau^*$ . Recall that, if one assumes a normal random intercept,  $\tau$  is the standard deviation of  $b_{i0}$ , i.e.,  $\sigma$ ; if one assumes a bridge distribution for  $b_{i0}$ ,  $\tau$  is the only parameter in the bridge distribution, of which  $\sigma$  is a decreasing function. As implemented in Huang (2009), where more general GLMMs that include random slopes are considered, one may certainly construct test statistics like (10) associated with other parameters estimated along with  $\tau$ . That is, denoting by  $\theta$  a generic element in  $\Omega$  and by  $\hat{U}$  an estimator of the variance–covariance matrix of  $\hat{\Omega} - \hat{\Omega}^*$ , one may consider  $t_\theta = (\hat{\theta} - \hat{\theta}^*)/\hat{v}_\theta$ , where  $\hat{v}_\theta$  is the square root of the diagonal element of  $\hat{U}$  corresponding to  $\theta$ . One may also combine  $t_\theta$ 's for all parameters in  $\Omega$  in the following test statistic,  $T^2 = (m - r)\{r(m - 1)\}^{-1}(\hat{\Omega} - \hat{\Omega}^*)' \hat{U}^{-1}(\hat{\Omega} - \hat{\Omega}^*)$ , where  $r$  is the number of parameters in  $\Omega$ . The estimator  $\hat{U}$  is derived from the influence-function approximation of  $\hat{\Omega} - \hat{\Omega}^*$ , as elaborated in Huang (2009). It is also shown there that, under the null hypothesis (to be stated momentarily),  $T^2 \sim F(r, m - r)$  and  $t_\theta \sim$  Student's  $t$  with  $m - r$  degrees of freedom asymptotically, and significant deviation from zero of any one of these test statistics signals some form of model misspecification.

Strictly speaking, these test statistics do not directly test  $H_0 : b_{i0} \sim N(0, \tau^2)$  or  $H_0 : b_{i0} \sim \text{bridge}(\tau)$  versus an alternative. Instead, take  $t_\tau$  as an example, it merely tests  $H_0^* : \tilde{\tau} - \tilde{\tau}^* = 0$  versus an alternative such as  $H_a^* : \tilde{\tau} - \tilde{\tau}^* \neq 0$ . As observed in the large-sample numerical studies in Sect. 3,  $\tilde{\tau} - \tilde{\tau}^*$  can be very close to zero even if the



**Fig. 2** Plot of  $\tilde{\Omega} - \tilde{\Omega}^*$  versus  $\alpha$  for probit GLMM with skew normal  $b_{i0}$ . Horizontal dotted reference lines signify the value zero

true random-intercept distribution is neither normal nor a bridge distribution yet is also symmetric. Hence,  $t_\tau$  (or even  $T^2$ ) typically lacks power to distinguish two symmetric distributions for  $b_{i0}$ . This is the price we pay for letting go of robust inference, i.e., not requiring consistent estimators in the presence of model misspecification, in return for a more flexible and informative diagnostic test. More importantly, this is not surprising because of the intrinsic arbitrariness in random-effects modeling according to [Verbeke and Molenberghs \(2010\)](#), who concluded that “. . .to any given model an entire class of (random-effects) models can be assigned, with all of its members producing the same fit to the observed data. . .”. To look deeper into this “gray area” of the test, in the next subsection, we search for theoretical explanations for when and why  $\tilde{\tau} - \tilde{\tau}^* \approx 0$ . A byproduct of this search is an explanation for the relationship between the sign of  $\tilde{\sigma} - \tilde{\sigma}^*$  and the skewness direction of the true random-intercept distribution observed in Sect. 3.

### 4.2 Local effects of skewness

We begin with formulating the true random-intercept distribution by contaminating the assumed distribution as  $G_b = (1 - \epsilon)F_{b, \tau_0} + \epsilon G_b^c$ , where  $G_b$  is the cdf of the true random-intercept distribution,  $F_{b, \tau_0}$  is the cdf of the assumed distribution evaluated at the true parameter value(s)  $\tau_0$ , and  $G_b^c$  can be viewed as the contaminating distribution that, along with the weight  $\epsilon \in [0, 1]$ , contributes to the discrepancy between the truth

and the assumed model. This is also the formulation used in Gustafson (1996) when studying the so-called local effect of misspecification. Similar to the assumptions made in Gustafson (1996), to retain the interpretability of  $\tau$  even in the wrong-model inference so that, for example, it is always meaningful to estimate the variance of  $b_{i0}$ , it is assumed that  $F_b$  is parameterized in terms of moments of  $b_{i0}$ , and  $G_b^c$  belongs to a class of distributions that have the same first two moments as those of  $F_{b,\tau_0}$ . This is a large enough class of contaminating distributions to keep the investigation of model misspecification interesting and practically well motivated, and also a sensible class to consider now that we wish to detect random-intercept misspecification mainly based on the difference between two estimators of the second moment of  $b_{i0}$ ,  $\hat{\sigma} - \hat{\sigma}^*$ . To adapt to the context of Sect. 3,  $F_{b,\tau_0}$  is assumed to correspond to a mean-zero symmetric distribution. Consequently, the direction of skewness of  $G_b$  is fully determined by that of  $G_b^c$  unless  $\epsilon = 0$ .

Because  $\epsilon$  partly controls the severity of distributional misspecification on  $b_{i0}$ , it is of interest to study  $\tilde{\Omega}$  and  $\tilde{\Omega}^*$  both as functions of  $\epsilon$ . Now, considering a first-order Taylor expansion of  $\tilde{\Omega}(\epsilon)$  and  $\tilde{\Omega}^*(\epsilon)$  around  $\epsilon = 0$ , while noting that  $\tilde{\Omega}(0) = \tilde{\Omega}^*(0) = \Omega_0$ , one can see that  $\epsilon \tilde{\Omega}'(0)$  and  $\epsilon \tilde{\Omega}^{*'}(0)$  are the first-order asymptotic bias of  $\tilde{\Omega}(\epsilon)$  and  $\tilde{\Omega}^*(\epsilon)$ , respectively. This motivates us to derive the first derivative,  $\tilde{\Omega}'(0)$  and  $\tilde{\Omega}^{*'}(0)$ . Like in Gustafson’s derivations, all derivatives evaluated at zero are right derivatives. Unlike Gustafson’s goal of study, we are less interested in each first-order bias term and more interested in the difference between these two bias terms for the purpose of model diagnostics. With details provided in Appendix S2 of the supplementary material, we first derive  $\tilde{\Omega}'(0) - \tilde{\Omega}^{*'}(0)$ , of which the element associated with  $\sigma$  is of most interest. Based on this difference and assuming that both random-intercept distributions specified by  $F_{b,\tau_0}$  and  $G_b^c$  have pdf’s tail off as  $b_{i0}$  (or  $b$  for short) deviates from zero, so that, under both  $F_{b,\tau_0}$  and  $G_b^c$ , the majority of probability mass is assigned at a neighborhood of  $b = 0$ , we show that

$$\tilde{\sigma} - \tilde{\sigma}^* = (\epsilon/6)\Delta_\sigma^{(3)}(0, \Omega_0)E_{g^c}(b^3) + O(b^4), \tag{11}$$

where  $E_{g^c}(b^3)$  is the third moment of  $G_b^c$ , which determines the skewness of  $G_b^c$ ;  $\Delta_\sigma^{(3)}(0, \Omega_0)$  is the third-order partial derivative of  $\Delta_\sigma(b, \Omega_0)$  with respect to  $b$  evaluated at  $b = 0$ , and  $\Delta_\sigma(b, \Omega_0)$  denotes the element in the following  $r \times 1$  vector corresponding to  $\sigma$ ,

$$E_{\mathbf{X}} \left\{ \mathbf{I}_{f^*}^{-1} \sum_{l=1}^4 \mathbf{s}^*(y_l^*; \mathbf{X}, \Omega_0) f_{\mathbf{Y}^*}(y_l^* | \mathbf{X}, b; \Omega_0) - \mathbf{I}_f^{-1} \sum_{l=1}^{2^n} \mathbf{s}(y_l; \mathbf{X}, \Omega_0) f_{\mathbf{Y}}(y_l | \mathbf{X}, b; \Omega_0) \right\}, \tag{12}$$

in which  $\mathbf{s}(y_l; \mathbf{X}, \Omega_0)$  is the observed-data score function associated with the assumed model evaluated at  $\Omega_0$  and  $y_l$ ,  $\mathbf{s}^*(y_l^*; \mathbf{X}, \Omega_0)$  is the counterpart grouped-data score function evaluated at  $y_l^*$ ,  $\mathbf{I}_f$  and  $\mathbf{I}_{f^*}$  are the negative Fisher information matrices associated with these two scores, respectively. Looking at (11), one may conclude that the sign of  $\tilde{\sigma} - \tilde{\sigma}^*$  is dominated by the sign of  $\Delta_\sigma^{(3)}(0, \Omega_0)E_{g^c}(b^3)$ . Hence, it

is expected that, if  $G_b^c$  is symmetric and thus  $E_{g^c}(b^3) = 0$ ,  $\tilde{\sigma} - \tilde{\sigma}^* \approx 0$ , resulting in the aforementioned gray area of our test based on  $t_\tau$ . Given (12), it is numerically straightforward to compute  $\Delta_\sigma^{(3)}(0, \Omega_0)$  given any true parameter configuration (i.e., the value of  $\Omega_0$ ) and design matrix configuration (i.e., the distribution of  $\mathbf{X}$ ). And we observe  $\Delta_\sigma^{(3)}(0, \Omega_0) \geq 0$  for a broad range of these configurations (see Appendix S2 for a pictorial demonstration). This explains the connection between the sign of  $\tilde{\sigma} - \tilde{\sigma}^*$  and the skewness direction of  $G_b$ , which is fully determined by the sign of  $E_{g^c}(b^3)$ .

The phenomenon that  $t_\tau$  has low power when  $G_b$  is symmetric should not cause much concern as, after all, both  $\hat{\beta}$  and  $\hat{\beta}^*$  are practically unaffected by this type of misspecification. The informativeness of  $t_\tau$  when  $G_b$  is skew deserves more attention (and celebration) as this is the case where wrong-model inference suffers more. The information revealed by  $t_\tau$  can be further exploited in choosing a more adequate model for the random intercept that is flexible yet constrained to reflect the direction of skewness. Imposing this constraint can improve the efficiency of inference compared to when one assumes a bigger class of flexible distribution family without the constraint imposed. Admittedly, such informativeness of  $t_\tau$  can be partly due to the fact that we focus on the subclass of GLMMs where  $b_{i0}$  is the only random effect, which is undoubtedly an important subclass that has been the topic of great interest among many researchers (Neuhaus et al. 1992; Wang and Louis 2003, 2004; Tchetgen and Coull 2006; Caffo et al. 2007, etc.). The operating characteristics of  $t_\tau$  in the presence of random slope(s), whose distribution(s) may be misspecified, are among the topics of our follow-up investigation.

The last step of the derivations outlined above (and elaborated in Appendix S2) leading to (11) requires the assumption that the pdf associated with  $G_b$  tails off outside a neighborhood of zero. We discuss in Appendix S3 of the supplementary material implication and consequences of violation of this assumption. The take-home message there is that  $t_\tau$  can have some power to detect model misspecification when the pdf associated with  $G_b$ , although symmetric, does not tail off as  $b$  deviates from zero.

## 5 Empirical evidence

We realize that the test statistics defined in Sect. 4.1 have a similar structure as the test statistic proposed by Tchetgen and Coull (2006). Their method is applicable for GLMMs with at least one within-cluster covariate. Denote by  $\beta^W$  the  $q \times 1$  subvector of  $\beta$  corresponding to the within-cluster covariate(s). Then an estimator of  $\beta^W$  can be obtained based on the conditional distribution of a sufficient statistic for  $\beta^W$  given a sufficient statistic for the random effects derived in Sartori and Severini (2004). Denote this estimator as  $\hat{\beta}_C^W$ , where the subscript “C” signifies that this estimator originates from a conditional likelihood rather than the marginal likelihood used to obtain the MLE of  $\beta^W$ ,  $\hat{\beta}^W$ . Note that the term “conditional distribution/likelihood” in Tchetgen and Coull (2006) should not be confused with the conditional model of the response given random effect(s) as in, say, (1). By construction,  $\hat{\beta}_C^W$  is robust to model assumptions on random effects. Making use of this feature, Tchetgen and Coull (2006) proposed to diagnose random-effects misspecification via the following test

statistic that follows a  $\chi^2_q$  asymptotically in the absence of model misspecification,

$$D = m \left( \hat{\beta}_C^W - \hat{\beta}^W \right)^t \hat{\Sigma}^{-1} \left( \hat{\beta}_C^W - \hat{\beta}^W \right), \tag{13}$$

where  $\hat{\Sigma}$  is the estimator for the variance-covariance matrix of  $\hat{\beta}_C^W - \hat{\beta}^W$  derived based on influence functions corresponding to the two estimators,  $\hat{\beta}_C^W$  and  $\hat{\beta}^W$ . We include their test as a competing method in the simulation study presented next.

In addition, we also implement two determinant tests developed by [Alonso et al. \(2008\)](#). These determinant tests exploit the equivalence between the inverse Fisher’s information matrix and the sandwich variance-covariance matrix in the absence of model misspecification ([White 1982](#)). The two test statistics are  $\delta_{d1} = \log |\mathbf{B}(\hat{\Omega})\{-\mathbf{A}^{-1}(\hat{\Omega})\}|$  and  $\delta_{d2} = |\mathbf{B}(\hat{\Omega})| - |\mathbf{A}^{-1}(\hat{\Omega})|$ , where  $\mathbf{A}(\Omega) = m^{-1} \sum_{i=1}^m \partial^2 \log f_Y(\mathbf{Y}_i|\mathbf{X}_i; \Omega)/(\partial \Omega \partial \Omega^t)$ , and  $\mathbf{B}(\Omega) = m^{-1} \sum_{i=1}^m (\partial \log f_Y(\mathbf{Y}_i|\mathbf{X}_i; \Omega)/\partial \Omega) (\partial \log f_Y(\mathbf{Y}_i|\mathbf{X}_i; \Omega)/\partial \Omega)^t$ . The authors showed that, with  $\hat{\Omega}$  replaced by  $\Omega_0$  in these two test statistics, under regularity conditions (including that the score function is normally distributed), the asymptotic null distributions of  $m\delta_{d1}^2/(2r)$  and  $m(\delta_{d2}-1)^2/(2r)$  are both  $\chi^2_r$ . In what follows, we consider two simulation settings, with the first one closely mimicking that in [Tchetgen and Coull \(2006\)](#). Both R code and SAS IML code that implement the proposed method are available from the correspondence author upon request. Computing  $D$ , as well as the test proposed by [Bartolucci et al. \(2015\)](#), involves the conditional MLE of within-cluster fixed effects. Computing our test statistics only involves the marginal MLEs, which are less readily available in standard software when it comes to grouped data than conditional MLEs.

[Setting I]: In this experiment, given a realization of  $\{b_{i0}\}_{i=1}^m$ , binary responses  $\{Y_{ij}, j = 1, \dots, n\}_{i=1}^m$  are generated according to the model considered in the accompanying technical report of [Tchetgen and Coull \(2006\)](#) at <http://www.bepress.com/harvardbiostat/>. In particular, this model is the same as that given in (4) but with the linear predictor being  $\beta_0 + \beta_1 X_{ij,2} + \beta_2 X_{ij,1} X_{ij,2} + b_{i0}$ , where  $\beta = (\beta_0, \beta_1, \beta_2)^t = (-0.5, 0.2, 0.5)^t$ , the between-cluster covariate is  $X_{ij,1} = x_i$ , with  $\{x_i\}_{i=1}^m$  simulated from Bernoulli(0.5), and the within-cluster covariate is  $X_{ij,2} = 5(j - 1)/(n - 1)$ , for  $i = 1, \dots, m, j = 1, \dots, n$ . We set  $m = 150, 300, 600$ , and, at each level of  $m$ , four levels of  $n$  (3, 5, 8, 10) are considered. When creating grouped data, we set  $G_i = 2$  for  $i = 1, \dots, m$ , and partition each cluster into two subgroups according to the sorted values of  $X_{ij,2}$ . Fixing the assumed model of  $b_{i0}$  at  $N(0, \tau^2)$ , we consider the following five true models for  $b_{i0}$  when generating  $\{b_{i0}\}_{i=1}^m$ : (I)  $N(0, 1)$ ; (II) a  $t$  distribution with standard deviation 2; (III) a shifted and scaled right-skewed gamma with standard deviation 2; (IV) a mixture normal specified by  $\lambda N(2.0, 0.5) + (1 - \lambda)N(-0.86, 0.5)$ , where  $\lambda \sim \text{Bernoulli}(0.3)$ ; (V)  $b_{i0}|X_{ij,1} \sim N\{0, (1 + 2X_{ij,1})^2\}$ . For each simulation configuration, we generate 300 Monte Carlo (MC) replicates (to be consistent with the scale of simulations in [Tchetgen and Coull \(2006\)](#)), resulting in 300 sets of  $t_\theta$  from our proposed method, 300 realizations of  $D$  defined in (13), and 300 realizations of  $(\delta_{d1}, \delta_{d2})$ . To compute Tchetgen and Coull’s  $D$ , we employ the SAS code provided in their technical report, implemented using PROC LogXact 7.0 for SAS 9.2 ([Cytel Software, Inc. 2012](#)). To compute our test statistics, we only need to go through two

rounds of relatively more straightforward marginal, as opposed to conditional, maximum likelihood estimation. And quantities needed to compute the two determinant test statistics are readily available as byproducts from computing  $\hat{v}_\theta$  involved in our test statistics.

Table 2 presents the proportions of tests among 300 MC replicates from each method that reject  $H_0^*$ , thus also reject  $H_0 : b_{i0} \sim N(0, \tau^2)$ , at significance level 0.05. Considering the “directional” nature of our test statistic, and now that we do know the sign of  $\tilde{\tau} - \tilde{\tau}^*$  based on the skewness of the true distribution of  $b_{i0}$ , we also report rejection rates of our test assuming a one-sided alternative under cases (III) and (IV), besides the rejection rates associated with a two-sided alternative  $H_a^* : \tilde{\tau} - \tilde{\tau}^* \neq 0$ . Empirical evidence here suggest that our test confers the right size when the assumed random-intercept model coincides with the true model; and when the true model is skew, our test has comparable or higher power than Tchetgen and Coull’s test, especially when  $n > 3$ . When the true model is symmetric as in case (II), where  $b_{i0}$  follows a  $t$  distribution, the determinant tests have moderate power to reject the normality assumption when  $m$  and  $n$  are large. For case (V) with  $b_{i0}|X_{ij,1} \sim N\{0, (1 + 2X_{ij,1})^2\}$ , our  $t_\tau$  performs like that under cases (I) and (II), which is not surprising because the true marginal distribution of  $b_{i0}$  is symmetric, like those under cases (I) and (II). However, as seen in Table 2 under case (V), the test statistic  $t_{\beta_2}$  sends strong signals to indicate model misspecification. This phenomenon of  $t_\theta$ , where  $\theta$  is one of the fixed effects, is explored in Huang (2009) and Huang (2011), where more general GLMMs and more different types of random-effects misspecification are considered. This observation motivates inspection of  $t_\theta$  for all parameters in  $\Omega$  if one wishes to discover existence of model misspecification besides the type we focus on in this article. When multiple  $t_\theta$ ’s are used simultaneously for model diagnosis, one should address the issue of multiple testing, say, by using the Bonferroni correction. [Setting II]: Note that both fixed effects in the model considered in Setting I are associated with within-cluster covariates, that is,  $\beta^W = (\beta_1, \beta_2)^t$ , and there is no fixed effect associated with the between-cluster covariate. We now consider another GLMM that includes a between-cluster covariate by itself so that  $\beta$  contains a fixed effect ( $\beta_1$ ) associated with the between-cluster covariate. This GLMM is specified by (4) with  $\beta = (\beta_0, \beta_1, \beta_2, \beta_3)^t = (-2, 1, 0.5, -0.25)^t$ . Except for that  $X_{ij,2} = (j - 1)/(n - 1)$ , the configurations of  $X_{ij,1}$ , the sizes of  $m$  and  $n$ , as well as the way to create grouped data, are the same as those in Setting I. Interestingly, with a between-cluster covariate by itself in the GLMM besides two within-cluster covariates, Tchetgen and Coull’s test (implemented using the authors’ SAS code applied to the model in Setting II) gives nearly no power in all considered configurations to detect random-intercept misspecification. Table 3 provides the rejection rates of our tests associated with a two-sided  $H_a^*$ , Tchetgen and Coull’s test, and the determinant tests when assuming a normal  $b_{i0}$  across 900 MC replicates. Except when the true random-intercept distribution is a Student’s  $t$ , the proposed test mostly outperforms the competing tests. Table 4 shows the empirical power of the aforementioned tests when one assumes  $b_{i0}$  follows a bridge distribution. To pair up with the assumed bridge distribution, we use the same five true distributions as in Setting I except that we replace the  $t$  distribution in (II) with the bridge distribution in order to monitor the size of the tests. The size of the determinant tests are severely inflated in this scenario (and thus their rejection rates in the presence

**Table 2** Rejection rates (in percentage) of our proposed test ( $t_\theta$ ), the test proposed by Tchetgen and Coull (2006) ( $D$ ), and the two determinant tests ( $\delta_{d1}$ ,  $\delta_{d2}$ ) in Alonso et al. (2008) across 300 MC replicates generated from the true model under Setting I in Sect. 5

True Model	$m = 150$			$m = 300$			$m = 600$					
	$t_\theta$	$D$	$\delta_{d1}$	$\delta_{d2}$	$t_\theta$	$D$	$\delta_{d1}$	$\delta_{d2}$	$t_\theta$	$D$	$\delta_{d1}$	$\delta_{d2}$
$n = 3$												
(I)	4.3	6.7	3.7	2.0	3.3	7.7	3.0	1.7	4.3	9.0	1.7	1.7
(II)	3.3	6.0	3.0	2.3	2.0	7.7	2.0	2.0	6.3	9.3	1.7	2.3
(III)	{6.0, 9.7}	13.7	9.0	3.0	{7.7, 12.3}	18.7	11.0	7.0	{10.7, 19.0}	30.7	11.3	6.7
(IV)	{8.7, 11.7}	7.0	5.7	1.0	{7.3, 10.0}	14.3	7.7	2.3	{8.3, 14.7}	26.0	8.3	4.7
(V)	73.3	13.7	1.6	0.3	93.3	23.3	2.0	3.0	99.7	73.3	5.7	2.7
$n = 5$												
(I)	6.3	6.7	4.7	2.3	2.0	7.3	3.7	3.7	4.0	6.0	4.0	3.0
(II)	1.7	7.7	2.0	4.3	3.0	11.0	4.3	10.3	4.7	7.0	12.3	16.3
(III)	{19.3, 26.7}	21.0	11.7	5.5	{32.0, 43.3}	34.7	10.3	5.3	{50.0, 61.0}	51.3	27.3	21.3
(IV)	{15.3, 23.7}	16.4	13.3	3.7	{24.0, 31.0}	25.3	13.0	7.3	{45.7, 56.3}	46.0	26.0	18.0
(V)	80.7	12.0	4.3	2.3	97.0	53.0	4.7	3.0	100.0	83.0	4.0	2.3
$n = 8$												
(I)	7.7	8.0	10.3	4.3	1.3	9.0	5.0	4.0	5.3	7.0	6.0	6.0
(II)	1.0	8.0	6.0	9.0	1.7	9.3	15.0	23.0	18.3	8.7	28.0	33.3
(III)	{36.0, 46.0}	30.0	22.3	8.7	{56.0, 64.3}	53.3	27.3	16.3	{79.0, 86.0}	83.7	44.7	37.7
(IV)	{21.7, 32.7}	23.0	22.0	6.7	{37.3, 50.3}	37.0	28.7	18.0	{64.0, 74.7}	61.0	47.3	34.7
(V)	90.0	18.7	5.0	3.0	99.7	68.0	4.7	6.3	100.0	87.0	5.3	4.3



Table 2 continued

True Model	m = 150			m = 300			m = 600					
	$t_\theta$	D	$\delta_{d1}$	$\delta_{d2}$	$t_\theta$	D	$\delta_{d1}$	$\delta_{d2}$	$t_\theta$	D	$\delta_{d1}$	$\delta_{d2}$
n = 10												
(I)	7.7	10.0	5.0	3.7	1.7	8.3	6.0	5.0	4.0	7.0	9.3	8.0
(II)	0.7	15.0	9.0	14.3	0.3	11.3	17.3	24.7	15.0	10.7	42.3	53.3
(III)	{37.3, 52.3}	33.3	27.7	9.3	{58.0, 67.7}	58.0	31.3	18.0	{88.7, 94.0}	90.3	51.0	38.7
(IV)	{20.0, 32.7}	23.3	27.7	11.7	{38.3, 50.0}	42.0	32.3	17.7	{69.7, 78.0}	72.3	51.3	41.3
(V)	92.3	25.3	8.0	5.7	100.0	72.7	6.3	8.0	100.0	83.0	9.0	11.3

The assumed model for  $b_{j0}$  is normal. The true model is (I)  $N(0, 1)$ , (II)  $t$ , (III) gamma, (IV) mixture normal, (V)  $b_{j0}|X_{j,1} \sim N\left\{0, (1 + 2X_{j,1})^2\right\}$ . Rejection rates of our test for cases (III) and (IV) are inside curly brackets, first the rejection rate assuming a two-sided test, then the rejection rate assuming a one-sided test. All rejection rates associated with our test are based on  $t_r$  except for case (V), where rejection rates associated with  $t_{\beta_2}$  are presented

**Table 3** Rejection rates (in percentage) of our proposed test ( $t_\theta$ ), the test proposed by Tchetgen and Coull (2006) ( $D$ ), and two determinant tests ( $\delta_{d1}$ ,  $\delta_{d2}$ ) in Alonso et al. (2008) when  $b_{i0}$  is assumed to follow a normal distribution across 900 MC replicates generated from the true model in Setting II in Sect. 5

True model	$m = 150$			$m = 300$			$m = 600$					
	$t_\theta$	$D$	$\delta_{d1}$	$\delta_{d2}$	$t_\theta$	$D$	$\delta_{d1}$	$\delta_{d2}$	$t_\theta$	$D$	$\delta_{d1}$	$\delta_{d2}$
$n = 3$												
(I) $N(0, 1)$	4.7	0.0	4.4	2.4	3.8	0.0	2.7	1.4	3.1	0.0	2.4	2.8
(II) Student's $t$	5.6	0.0	2.2	1.4	7.4	0.0	1.2	1.2	8.6	0.0	1.7	1.9
(III) Gamma	13.6	0.1	3.0	1.9	16.5	0.0	1.4	2.1	25.0	0.0	4.2	5.3
(IV) Mixture normal	4.7	0.0	2.8	0.7	5.8	0.0	2.1	1.1	5.2	0.0	2.9	1.9
(V) HV	4.4	0.7	17.6	5.0	3.1	0.6	22.4	10.3	4.1	0.8	33.0	24.0
$n = 5$												
(I) $N(0, 1)$	3.3	0.0	8.8	4.4	5.1	0.0	6.3	4.2	4.4	0.0	4.2	2.8
(II) Student's $t$	6.7	0.0	4.0	5.7	11.0	0.0	3.8	6.2	15.1	0.0	8.1	11.8
(III) Gamma	31.0	0.1	4.7	5.8	49.4	0.0	5.1	7.1	77.6	0.1	8.2	9.8
(IV) Mixture normal	8.5	0.0	8.4	3.6	10.0	0.0	6.4	3.2	17.6	0.0	8.3	5.7
(V) HV	13.2	0.7	23.4	8.4	23.9	0.0	31.6	19.8	33.8	0.6	47.4	35.8
$n = 8$												
(I) $N(0, 1)$	2.2	0.0	11.0	5.2	3.3	0.0	8.9	6.1	3.6	0.0	7.2	5.8
(II) Student's $t$	7.1	0.0	6.4	9.2	10.8	0.0	10.1	15.0	13.6	0.0	22.0	27.3
(III) Gamma	50.7	0.3	7.1	8.7	81.0	0.2	8.6	11.9	97.6	0.2	13.7	16.4
(IV) Mixture normal	19.5	0.1	19.1	5.7	34.9	0.0	16.8	9.3	58.2	0.1	23.1	15.8
(V) HV	39.1	0.6	26.9	8.9	64.6	0.7	32.1	20.2	89.6	0.7	45.1	33.4

Table 3 continued

True model	m = 150			m = 300			m = 600			
	$t_\theta$	D	$\delta_{d2}$	$t_\theta$	D	$\delta_{d1}$	$t_\theta$	D	$\delta_{d2}$	
$n = 10$										
(I) $N(0, 1)$	2.2	0.0	5.7	3.6	0.0	9.0	4.0	0.0	5.0	6.3
(II) Student's $t$	6.4	0.0	10.9	8.6	0.0	12.8	11.1	0.0	20.1	36.9
(III) Gamma	58.8	0.2	7.9	87.3	0.2	9.3	99.7	0.0	12.4	18.9
(IV) Mixture normal	29.5	0.0	8.4	51.9	0.0	27.7	76.8	0.0	15.2	25.9
(V) HV	45.6	0.8	11.3	70.8	0.6	33.7	95.2	0.8	19.3	36.0

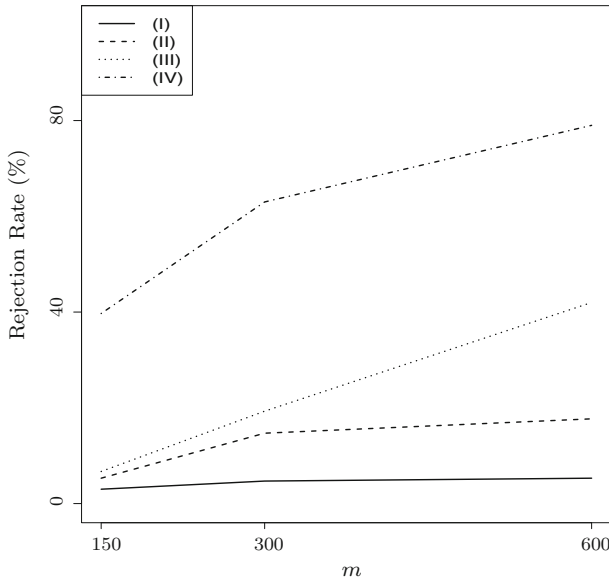
All rejection rates associated with our test are based on  $t_\tau$  except for the row with row title "(V) HV", where rejection rates associated with  $t_{\beta_2}$  are presented. "HV" refers to the case with  $b_{i0} | X_{ij,1} \sim N \left\{ 0, (1 + 2X_{ij,1}^2) \right\}$

**Table 4** Empirical size (in percentage) of our proposed test ( $t_\theta$ ), the test proposed by Tchetgen and Coull (2006) ( $D$ ), and two determinant tests ( $\delta_{d1}, \delta_{d2}$ ) proposed by Alonso et al. (2008) in the absence of model misspecification, and rejection rates associated with  $t_\theta$  and  $D$  in the presence of model misspecification when  $b_{i0}$  is assumed to follow a bridge distribution across 900 MC replicates generated from the true model in Setting II in Sect. 5

$m$	$n = 3$			$n = 5$			$n = 8$			$n = 10$		
	150	300	600	150	300	600	150	300	600	150	300	600
(I) $b_{i0} \sim \text{Bridge}$												
$t_\tau$	5.8	6.7	4.6	6.5	4.5	6.1	6.4	5.7	4.0	4.4	5.8	4.8
$D$	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
$\delta_{d1}$	50.7	71.0	92.3	28.7	34.6	54.8	21.3	26.3	32.8	17.8	20.8	23.8
$\delta_{d2}$	24.9	53.0	86.2	11.3	21.8	43.4	9.1	14.7	24.8	5.8	10.8	16.6
(II) $b_{i0} \sim N(0, 1)$												
$t_\tau$	5.1	5.8	5.0	6.3	9.2	9.0	5.1	4.5	6.2	2.8	5.1	4.8
$D$	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
(III) $b_{i0} \sim \text{Gamma}$												
$t_\tau$	7.4	8.5	9.8	12.9	21.2	39.0	27.1	48.0	78.3	34.9	60.2	89.1
$D$	0.0	0.1	0.1	0.0	0.0	0.2	0.1	0.2	0.2	0.1	0.3	0.3
(IV) $b_{i0} \sim \text{Mixture normal}$												
$t_\tau$	4.0	3.5	4.0	6.2	4.2	4.4	12.1	20.6	41.6	23.1	37.0	62.3
$D$	0.0	0.0	0.0	0.0	0.1	0.0	0.1	0.0	0.0	0.2	0.0	0.1
(V) $b_{i0} \sim N\left\{0, (1 + 2X_{ij,1})^2\right\}$												
$t_{\beta_2}$	2.9	4.7	12.5	6.1	3.8	7.3	17.1	34.6	54.6	21.0	35.4	66.5
$D$	0.4	0.0	0.9	0.0	0.6	0.9	0.4	0.7	0.3	0.7	0.8	0.6

of model misspecification in Table 4 are omitted). We again observe promising power of our test except when the true distribution is symmetric, while retaining the right size in the absence of model misspecification.

Up to this point, the assumed random-intercept distribution has always been symmetric in our theoretical analyses and simulation studies, a feature playing a key role especially in Sect. 4.2 where we establish analytically the connection between the skewness of the true random-intercept distribution and the sign of  $\tilde{\sigma} - \tilde{\sigma}^*$ . Keeping the model of  $Y_{ij}$  given  $\mathbf{X}_{ij}$  and  $b_{i0}$  under Setting II, assuming that  $b_{i0}$  follows a SN distribution specified by two parameters,  $\alpha$  and  $\sigma$ , as introduced in Sect. 3.2, we compute  $t_\sigma$  as we vary the true  $b_{i0}$ -distribution. Recall that the family of SN includes normal distributions as well as (unimodal) skewed distributions. To violate the model assumption, we consider generating  $b_{i0}$  from three bimodal mixture normals given by a symmetric mixture normal,  $0.5N(-1, 0.25) + 0.5N(1, 0.25)$ , a left-skewed mixture normal,  $0.2N(-2, 0.16) + 0.8N(0.5, 0.16)$ , and a right-skewed mixture normal,  $0.2N(2, 0.16) + 0.8N(-0.5, 0.16)$ . To monitor the size of the test, we also simulate  $b_{i0}$  from  $N(0, 1)$ . Figure 3 presents the rejection rates across 300 MC replicates when the cluster size is 10. The comparison between  $\tilde{\sigma}$  and  $\tilde{\sigma}^*$  when the assumed  $b_{i0}$ -distribution is skewed no longer follows the arguments in Sect. 4.2. But we still



**Fig. 3** Plot of rejection rates (in percentage) of  $t_\sigma$  when  $b_{j0}$  is assumed to follow a skew normal distribution across 300MC replicates generated from the true model in Setting II in Sect. 5. The true distribution is (I)  $N(0, 1)$ , (II) a symmetric mixture normal, (III) a left-skewed mixture normal, and (IV) a right-skewed mixture normal, respectively

observe some empirical power from this experiment, suggesting that, this type of discrepancy between the assumed and the truth (unimodality versus bimodality) can lead to significant disagreement between the MLE of the variance component based on the raw data and the MLE from the grouped data. In this particular case, the estimated shape parameter of the assumed SN model can provide a clue on the direction of the skewness of the underlying truth.

### 6 Real-data examples

In what follows, we entertain two real-life examples where cluster data are collected and modeled by GLMMs. For illustration purposes, we focus on using the proposed test to assess the adequacy of a GLMM with a random intercept, of which the distribution is assumed to be normal or a bridge distribution. Two-sided tests are considered in both examples when computing  $p$ -values associated with the tabulated test statistics.

*Example 1 (Respiratory infection data):* In this study 275 preschool children were examined every three months for 18 months for the presence of respiratory infection (Lin and Carroll 2001). Besides the binary response as an indicator of the presence/absence of a child’s infection status, each child’s age at the beginning of the study and the season when the examination took place were also recorded. We start from a relatively simple GLMM given by, for  $i = 1, \dots, 275, j = 1, \dots, n_i$ , with  $n_i$

ranging from 4 to 6,

$$E(Y_{ij} | \mathbf{X}_{ij}, b_{i0}; \boldsymbol{\beta}) = \{1 + \exp(-\beta_0 - \beta_1 X_{ij,1} - \beta_2 X_{ij,2} - b_{i0})\}^{-1}, \quad (14)$$

where  $X_{ij,1}$  is defined as  $[(\text{baseline age (in months)} - 36)/12]^3$ , and  $X_{ij,2}$  takes values 1, 2, 3, 4, representing the season of the  $j$ th examination for child  $i$ . The inclusion of a cubic term of baseline age in (14) is supported by Lin and Carroll (2001), who found that this term is highly statistically significant. To create grouped data, we leave the clusters with  $n_i < 3$  as they are, and for clusters with  $n_i \geq 3$ , we create two subgroups within each cluster according to the sorted values of  $X_{ij,2}$ . Using the raw observed data and the grouped data, we carry out maximum likelihood estimation to obtain  $\hat{\boldsymbol{\Omega}}$  and  $\hat{\boldsymbol{\Omega}}^*$  twice, first assuming  $b_{i0} \sim N(0, \tau^2)$ , and second assuming  $b_{i0} \sim \text{bridge}(\tau)$ . For completeness, under each random-intercept assumed model, we compute  $t_\theta$  for all parameters as well as  $T^2$ . The MLEs and the corresponding test results are tabulated in Table 5 (upper half of the table).

According to Table 5, the test based on  $t_\tau$  under neither assumed random-intercept model suggests significant discrepancy between  $\hat{\tau}$  and  $\hat{\tau}^*$ , although the tests based on  $t_{\beta_0}$  and  $t_{\beta_2}$  show border-line significant evidence of model misspecification. Referring back to our observations in the simulation study, one may conclude that, the true distribution of  $b_{i0}$  may have a variance component that depends on  $X_{ij,2}$  and the marginal distribution of  $b_{i0}$  (after “marginalizing out”  $X_{ij,2}$ ) may be approximately symmetric. Additionally, we notice that the values of  $t_{\beta_0}$ ,  $t_{\beta_2}$ , and  $T^2$  under the normal assumed model for  $b_{i0}$  are slightly more significant than those under the bridge-distribution assumption. This could be evidence that the true (marginal) distribution of  $b_{i0}$  has a heavier tail, like that of a bridge distribution, than a normal distribution. Refitting the GLMM assuming other heavy-tailed distributions for  $b_{i0}$  can be a follow-up model building exercise, which we do not further pursue in this example.

*Example 2 (Toenail infection data):* This data set is from a multi-center study, in which two oral treatments for toenail infection are compared (Fitzmaurice et al. 2004). In the study, each patient was scheduled to be evaluated for the degree of onycholysis seven times, although in the end only 224 out of a total of 294 patients provided the complete set of seven follow-up measures. The onycholysis outcome variable is the binary response of interest, with 0 representing none or mild infection and 1 indicating moderate or severe infection. Besides the response, recorded information of each patient include the treatment indicator,  $X_{ij,1}$ , and the time point (in months) when a measure was taken,  $X_{ij,2}$ , for  $i = 1, \dots, 294$  and  $j = 1, \dots, n_i$ , where  $n_i$  varies from 1 to 7. We adopt the GLMM considered in Molenberghs and Verbeke (2005, Chapter 14) given by  $E(Y_{ij} | \mathbf{X}_{ij}, b_{i0}; \boldsymbol{\beta}) = \{1 + \exp(-\beta_0 - \beta_1 X_{ij,1} - \beta_2 X_{ij,2} - \beta_3 X_{ij,1} X_{ij,2} - b_{i0})\}^{-1}$ , where, as in Example 1, we first assume  $b_{i0} \sim N(0, \tau^2)$  and then assume  $b_{i0} \sim \text{bridge}(\tau)$ . To implement the proposed test, we partition each cluster of size greater than 2 into two groups according to the sorted values of  $X_{ij,2}$ , and leave the other smaller clusters as they are. Using the raw observed data and the induced grouped data, we obtain  $\hat{\boldsymbol{\Omega}}$  and  $\hat{\boldsymbol{\Omega}}^*$ ,

**Table 5** Maximum likelihood estimates of  $\Omega$  and values of the corresponding test statistics computed in Example 1 (upper half of the table) and Example 2 (lower half of the table) in Sect. 6

	Observed	Grouped	$t_\theta$	$T^2$
<i>Example 1: Respiratory infection data</i>				
Assume $b_{i0} \sim N(0, \tau^2)$				
$\beta_0$	-2.66 (0.31)	-1.76 (0.46)	-2.15 (0.03)	1.98 (0.10)
$\beta_1$	-0.05 (0.01)	-0.05 (0.02)	0.95 (0.34)	
$\beta_2$	-0.06 (0.10)	-0.48 (0.20)	2.13 (0.03)	
$\tau$	0.97 (0.23)	1.10 (0.28)	-0.69 (0.49)	
Assume $b_{i0} \sim \text{bridge}(\tau)$				
$\beta_0$	-2.53 (0.28)	-1.67 (0.45)	-2.06 (0.04)	1.61 (0.17)
$\beta_1$	-0.04 (0.01)	-0.05 (0.02)	1.17 (0.24)	
$\beta_2$	-0.06 (0.10)	-0.48 (0.20)	2.13 (0.03)	
$\tau$	0.90 (0.03)	0.85 (0.06)	0.99 (0.32)	
<i>Example 2: Toenail infection data</i>				
Assume $b_{i0} \sim N(0, \tau^2)$				
$\beta_0$	-1.63 (0.45)	-1.77 (0.39)	0.72 (0.472)	3.57 (0.004)
$\beta_1$	-0.15 (0.64)	-0.27 (0.55)	0.43 (0.669)	
$\beta_2$	-0.39 (0.07)	-0.27 (0.09)	-1.73 (0.083)	
$\beta_3$	-0.14 (0.13)	-0.05 (0.12)	-0.79 (0.428)	
$\tau$	4.02 (0.51)	3.00 (0.49)	2.79 (0.005)	
Assume $b_{i0} \sim \text{bridge}(\tau)$				
$\beta_0$	-1.42 (0.41)	-1.72 (0.36)	1.90 (0.057)	4.92 (0.000)
$\beta_1$	-0.11 (0.61)	-0.25 (0.51)	0.53 (0.596)	
$\beta_2$	-0.39 (0.07)	-0.26 (0.08)	-2.01 (0.045)	
$\beta_3$	-0.14 (0.13)	-0.05 (0.12)	-0.77 (0.441)	
$\tau$	0.41 (0.03)	0.49 (0.04)	-2.95 (0.003)	

Numbers in parentheses following the MLEs are the sandwich estimated standard errors of the MLEs. Number in parentheses following the test statistics are the corresponding  $p$  values

along with the corresponding test statistics, under each assumed distribution for  $b_{i0}$ . These results are presented in Table 5 (lower half of the table).

Under both assumed models for  $b_{i0}$ , besides the overall test based on  $T^2$ ,  $t_\tau$  stands out as highly significant in Table 5, which is a strong signal suggesting that both assumed models for  $b_{i0}$  are questionable. Moreover, because the value of  $t_\tau$  is positive (negative) when a normal (bridge) random intercept is assumed, this test provides sufficient evidence that the true random-intercept distribution is right-skewed. We notice that, in Verbeke and Molenberghs (2013) where this data set is analyzed using the same GLMM, they estimated the distribution of  $b_{i0}$  using a mixture normal, and the estimated density (Verbeke and Molenberghs 2013, Fig. 1(c)) is indeed right-skewed, matching up with our conclusion based on  $t_\tau$ . To follow up, we repeat our tests when  $b_{i0}$  is assumed to follow a two-component mixture normal, with a common standard

deviation  $\tau$  in the two components. Now we obtain a  $p$ -value of 0.465 associated with  $t_\tau$ , suggesting that a mixture normal is likely a more adequate model for  $b_{i0}$  than a normal distribution. Although in principle the proposed test can be used to check the adequacy of any assumed random-intercept distribution, we discourage its use on a complicated assumed distribution that is flexible at the price of involving many parameters. This is because the proposed test requires one to infer all unknown parameters using the less informative grouped data, which can be inefficient and in turn compromise the power of the test.

## 7 Discussion

In this study we are able to reveal more information of the true model and gain further insight of the diagnostic test in Huang (2009) when applying to GLMM with a random intercept  $b_{i0}$  as the only random effect. Focusing on this class of GLMM, we discover an interesting connection between the discrepancy of the two limiting MLEs,  $\tilde{\sigma} - \tilde{\sigma}^*$ , and the skewness of the true random-intercept distribution, where  $\sigma$  is the standard deviation of  $b_{i0}$ . We provide theoretical explanations for this connection, as well as other observed properties of MLEs. We find reasons for the low power of the proposed test in the presence of certain misspecification. With closed-form likelihood functions rarely available for GLMM, such theoretical development are important contributions that can advance the understanding of likelihood-based inference in the framework of GLMM in the presence of random-effects model misspecification.

Besides being able to detect random-intercept misspecification of the type that can substantially compromise inference, the test based on  $t_\tau$  can reveal in what direction the true random-intercept distribution deviates from the assumed. Such valuable information is attained without estimating  $b_{i0}$  or its pdf (as being attempted by many researchers). This advantage of the proposed method is achieved by drawing wrong-model inference twice, starting from the same assumed GLMM (with the same assumption on  $b_{i0}$ ), first using the raw observed data and then using an induced grouped data set. It may be surprising, albeit expected after some in-dept theoretical exploration, that repeated wrong-model analyses can be more fruitful than exploitation of some robust/consistent inference in developing diagnostic methods. And the key to an informative diagnostic method based on wrong-model analyses is to understand how such analyses based on different types of data compare with each other, rather than how each round of wrong-model analysis is affected by model misspecification. This is a fresh view we have not seen in the existing literature on the topic of model diagnosis. To recap, when it comes to assessing model assumptions, making use of wrong-model analyses repeatedly can be a more fruitful direction than many traditional routes.

We have not addressed some practical issues related to creating grouped data. In all examples in this article, we set  $G_i = 2$  as long as  $n_i \geq 2$ , and we partition each cluster according to the sorted values of the within-cluster covariate. The motivation of partitioning according to the sorted values of a within-cluster covariate is to maximize the between-group variation of such covariate, which yields more informative groups within a cluster compared to groups randomly formed. Setting  $G_i = 2$  is mainly for simplicity, although one may legitimately concern about the information loss with such



a small  $G_i$  (compared to a bigger  $G_i$ ). Noticing that the raw observed data is a special case of the grouped data with  $G_i = n_i$ , one may wish to distinguish  $G_i$  from  $n_i$  as much as possible (i.e., making  $G_i$  as small as possible) so that the grouped data differ from the observed data more, which typically leads to more discrepancy between  $\hat{\Omega}$  and  $\hat{\Omega}^*$  (at least in limit) in the presence of model misspecification. However, with a smaller  $G_i$ , the variability of  $\hat{\Omega}^*$  is higher since more information are lost with fewer groups per cluster. Hence, there is a trade-off between  $\hat{\Omega} - \hat{\Omega}^*$  and  $\hat{U}$ . And the interplay of “discrepancy between limiting MLEs” and “finite-sample variability” is reflected in  $t_\theta$  and  $T^2$  constructed in Sect. 4.1. In other words, these test statistics are designed to balance these two factors by “standardizing” the discrepancy on the numerator, which weakens the dependence of the test statistics on  $G_i$ . Hence, in practice, one does not need to be too concerned about the choice of  $G_i$ . After all, the power of the proposed tests mainly depends on, first, the type of model misspecification, and second, the observed data structure, neither of which one can manipulate at the testing stage.

The above discussion leads to the awareness that forming grouped data is like a one-trick pony in the sense that one does not have many options to create grouped data that can influence the amount or the type of information one can obtain from the test statistics. This trick is useful enough for the subclass of GLMM considered in this article. It will be interesting to look into how the operating characteristics of the proposed tests are affected by co-existence of other model misspecification in a bigger class of GLMM. When there are multiple sources of model misspecification, we conjecture that more flexible ways of creating induced data from the observed data are needed to disentangle different inappropriate model assumptions. Rather than this one-trick pony considered in the current study, we have started looking into creating missing data within each cluster as a mechanism of constructing induced data, based on which wrong-model analyses are carried out (repeatedly). This direction has been shown to be promising in detecting model misspecification in LMM (Huang 2013). A referee brought up the possibility of applying the proposed idea to detect time-varying latent individual effects in GLM for panel data as those considered in Bartolucci et al. (2015). Although this context is beyond the scope of the current study, where exogeneity is imposed, we conjecture that some modification of the proposed test with a more careful design of the grouped data can yield tests for detecting time-varying unobserved heterogeneity. Lastly, we would like to point out that the presented study is in the frequentist framework, and the consideration of random-effects specification in the Bayesian framework is fundamentally different, as discussed in Grilli and Rampichini (2015). There, a complete specification of a random-effect distribution involves a conditional distribution of the random effect given  $\tau$ , and a prior for  $\tau$ .

**Acknowledgements** The authors are grateful to the referees, the Associate Editor, and the Editor for their helpful suggestions that lead to a substantially improved manuscript. This work was supported by National Science Foundation (US) Grant DMS-1006222.

## References

Agresti A, Caffo B, Ohman-Strickland P (2004) Examples in which misspecification of a random effects distribution reduces efficiency, and possible remedies. *Comput Statist Data Anal* 47:639–653

- Alonso A, Litière S, Molenberghs G (2008) A family of tests to detect misspecifications in random-effects structure of generalized linear mixed models. *Comput Stat Data Anal* 52:4474–4486
- Azzalini A (1985) A class of distributions which includes the normal ones. *Scand J Stat* 12:171–178
- Bartolucci F, Belotti F, Peracchi F (2015) Testing for time-invariant unobserved heterogeneity in generalized linear models for panel data. *J Econom* 184:111–123
- Bartolucci F, Bacci S, Pigini C (2015) A misspecification test for finite-mixture logistic models for clustered binary and ordered responses. MPRA Paper 64220, University Library of Munich
- Butler SM, Louis TA (1992) Random effects models with nonparametric priors. *Stat Med* 11:1981–2000
- Caffo B, Ming-Wen A, Rohde C (2007) Flexible random intercept models for binary outcomes using mixtures of normals. *Comput Stat Data Anal* 51:5220–5235
- Chen J, Zhang D, Davidian M (2002) A Monte Carlo EM algorithm for generalized linear models with flexible random effects distribution. *Biostatistics* 3:347–360
- Claeskens G, Hart JD (2009) Goodness-of-fit tests in mixed models. *Test* 18:213–239
- Cytel Software, Inc. (2012) Proc-LogXact 7.0 For SAS users. Cytel Software, Inc, Cambridge
- Drikvandi R, Verbeke G, Molenberghs G (2016) Diagnosing misspecification of the random-effects distribution in mixed models. *Biometrics*. doi:[10.1111/biom.12551](https://doi.org/10.1111/biom.12551)
- Efendi A, Drikvandi R, Verbeke G, Molenberghs G (2014) A goodness-of-fit test for the random-effects distribution in mixed models. *Stat Methods Med Res*. doi:[10.1177/0962280214564721](https://doi.org/10.1177/0962280214564721)
- Fitzmaurice GM, Laird NM, Ware JH (2004) Applied longitudinal analysis. Wiley, New York
- Follmann DA, Lambert D (1989) Generalized logistic regression by nonparametric mixing. *J Am Stat Assoc* 84:295–300
- Grilli L, Rampichini C (2015) Specification of random effects in multilevel models: a review. *Qual Quant* 49:967–976
- Gustafson P (1996) The effect of mixing-distribution misspecification in conjugate mixture models. *Can J Stat* 24:307–318
- Hausman JA (1978) Specification tests in econometrics. *Econometrica* 46:1251–1271
- Heagerty PJ, Kurland BF (2001) Misspecified maximum likelihood estimates and generalised linear mixed models. *Biometrika* 88:973–985
- Huang X (2009) Diagnosis of random-effect model misspecification in generalized linear mixed models for binary response. *Biometrics* 65:361–368
- Huang X (2011) Detecting random-effects model misspecification via coarsened data. *Comput Stat Data Anal* 55:703–714
- Huang X (2013) Tests for random effects in linear mixed models using missing data. *Stat Sin* 23:1043–1070
- Jiang J (2007) Linear and generalized linear mixed models and their applications. Springer series in statistics. Springer, New York
- Kleinman KP, Ibrahim JG (1998) A semi-parametric Bayesian approach to generalized linear mixed models. *Stat Med* 17:2579–2596
- Komàrek A, Lesaffre E (2008) Generalized linear mixed model with a penalized Gaussian mixture as a random effects distribution. *Comput Stat Data Anal* 52:3441–3458
- Lee KJ, Thompson SG (2008) Flexible parametric models for random-effects distributions. *Stat Med* 27:418–434
- Lesperance M, Saab R, Neuhaus J (2014) Nonparametric estimation of the mixing distribution in logistic regression mixed models with random intercepts and slopes. *Comput Stat Data Anal* 71:211–219
- Lin X, Carroll RJ (2001) Semiparametric regression for clustered data. *Biometrika* 88:1179–1185
- Litière S, Alonso A, Molenberghs G (2007) Type I and type II error under random-effects misspecification in generalized linear mixed models. *Biometrics* 63:1038–1044
- Litière S, Alonso A, Molenberghs G (2008) The impact of a misspecified random-effects distribution on maximum likelihood estimation in generalized linear mixed models. *Stat Med* 27:3125–3144
- Magder LS, Zeger SL (1996) A smooth nonparametric estimate of a mixing distribution using mixtures of Gaussians. *J Am Stat Assoc* 91:1141–1151
- McCulloch CE, Searle SR, Neuhaus JM (2008) Generalized, linear, and mixed models. Wiley, Hoboken, New Jersey
- McCulloch CE, Neuhaus JM (2011) Misspecifying the shape of a random effects distribution: why getting it wrong may not matter. *Stat Sci* 26:388–402
- McCulloch CE, Neuhaus JM (2011) Prediction of random effects in linear and generalized linear models under model misspecification. *Biometrics* 67:270–279

- Molenberghs G, Verbeke G (2005) Models for discrete longitudinal data. Springer series in statistics. Springer, New York
- Neuhaus JM, Hauck WW, Kalbfleisch JD (1992) The effects of mixture distribution specification when fitting mixed-effects logistic models. *Biometrics* 79:755–762
- Neuhaus JM, McCulloch CE, Boylan R (2013) Estimation of covariate effects in generalized linear mixed models with a misspecified distribution of random intercepts and slopes. *Stat Med* 32:2419–2429
- Pan Z, Lin DY (2005) Goodness-of-fit methods for generalized linear mixed models. *Biometrics* 61:1000–1009
- Papageorgiou G, Hinde J (2012) Multivariate generalized linear mixed models with semi-nonparametric and smooth nonparametric random effects densities. *Stat Comput* 22:79–92
- Ritz C (2004) Goodness-of-fit tests for mixed models. *Scand J Stat* 31:443–458
- Rotnitzky A, Wypij D (1994) A note on the bias of estimators with missing data. *Biometrics* 50:1163–1170
- Sartori N, Severini TA (2004) Conditional likelihood inference in generalized linear mixed models. *Stat Sin* 14:349–360
- Scurrah KJ, Palmer LJ, Burton PR (2000) Variance components analysis for pedigree-based censored survival data using generalized linear mixed models (GLMMs) and Gibbs sampling in BUGS. *Genet Epidemiol* 19:127–148
- Tchetgen EJ, Coull BA (2006) A diagnostic test for the mixing distribution in a generalised linear mixed model. *Biometrika* 93:1003–1010
- Ten Have TR, Kunselman AR, Tran L (1999) A comparison of mixed effects logistic regression models for binary response data with two nested levels of clustering. *Stat Med* 18:947–960
- Verbeke G, Molenberghs G (2010) Arbitrariness of models for augmented and coarse data, with emphasis on incomplete data and random effects models. *Stat Model* 14:477–490
- Verbeke G, Molenberghs G (2013) The gradient function as an exploratory goodness-of-fit assessment of the random-effects distribution in mixed models. *Biostatistics* 14:477–490
- Waagepetersen R (2006) A simulation-based goodness-of-fit test for random effects in generalized linear mixed models. *Scand J Stat* 33:721–731
- Wang J (2010) A nonparametric approach using Dirichlet process for hierarchical generalized linear mixed models. *J Data Sci* 8:43–59
- Wang Z, Louis TA (2003) Matching conditional and marginal shapes in binary mixed-effects models using a bridge distribution function. *Biometrika* 90:765–775
- Wang Z, Louis TA (2004) Marginalized binary mixed-effects with covariate-dependent random effects and likelihood inference. *Biometrics* 60:884–891
- White H (1982) Maximum likelihood estimation of misspecified models. *Econometrica* 50:1–25
- Woods CM (2008) Likelihood-ratio DIF testing: effects of nonnormality. *Appl Psychol Meas* 32:511–526
- Zackin R, De Gruttola VG, Laird N (1996) Nonparametric mixed-effects models for repeated binary data arising in serial dilution assays: an application to estimating viral burden in AIDS. *J Am Stat Assoc* 91:52–61