

PQL Estimation Biases in Generalized Linear Mixed Models

Woncheol Jang* Johan Lim†

March 18, 2006

Abstract

The penalized quasi-likelihood (PQL) approach is the most common estimation procedure for the generalized linear mixed model (GLMM). However, it has been noticed that the PQL tends to underestimate variance components as well as regression coefficients in the previous literature. In this paper, we numerically show that the biases of variance component estimates by PQL are systematically related to the biases of regression coefficient estimates by PQL, and also show that the biases of variance component estimates by PQL increase as random effects become more heterogeneous.

Keywords and Phrases: Generalized linear mixed models; heterogeneity; penalized quasi-likelihood estimator; variance components.

1 Introduction

Repeatedly measured outcomes of each subject are the data scheme which we commonly encounter in medical studies. In the situation that there is only one measurement for each subject, the generalized linear model (GLM) is a common assumption for a broad variety of continuous and discrete data. A major issue in extending the GLM for single observation per subject to models of repeatedly measured data is intra-subject correlation. In the previous

*Institute of Statistics & Decision Sciences, Duke University, wjang@stat.duke.edu

†Department of Statistics, Texas A & M University, johanlim@stat.tamu.edu

literature, two most common approaches to take into account the correlation within a subject are the marginal model approach (for example, generalized estimation equation) by Zeger and Liang (1986) and the model with random effects to explain individual heterogeneity such as the generalized linear mixed effects model (GLMM), where the variance components are the parameters related to the random effects.

In many of repeatedly measured data examples, the primary research interest is given to the change of response outcomes over time or covariates (equivalently estimating the regression coefficients) rather than the variance components. However, knowing the variance components is often as much of interest as knowing the regression coefficients. For example, in animal breeding where the GLMM has been widely used, the random effects explain the subject specific variation and the heritability of the genetic correlation can be represented as a function of variance components. In addition, although the role of variance components are not clearly specified in the marginal model approach, the optimal weight for the well known generalized estimation equation (GEE) method is a function of the variance components and their estimation is important in obtaining efficient regression coefficient estimates.

When estimating the parameters in the GLMM, the exact likelihood function involves an intractable high-dimensional integration and is hard to compute. Accordingly, several approximations to the likelihood function and approximate maximum likelihood estimators (MLE) have been proposed in the previous literature (Schall, 1991; Breslow and Clayton, 1993; Wolfinger, 1993). Among them, the penalized quasi-likelihood (PQL) by Breslow and Clayton (1993) is the most popular for the GLMM. It approximates the high-dimensional integration using the well-known Laplace approximation and the approximated likelihood function has that of a Gaussian distribution. Subsequently, it suggests to apply linear mixed model restricted maximum likelihood (REML) estimation to the normal theory problem introduced in Harville (1977).

Even though the penalized likelihood approach is widely used in many different applications, it has known that estimating the variance components is quite challenging due to their non-observability. Accordingly, several different

types of likelihood functions of the variance components have been suggested including maximum adjusted profile h-likelihood estimator (MAPHLE) by Lee and Nelder (1996). However, most such variance component estimators have not received as much attention as regression coefficient estimators have. Furthermore they are not understood well when the observations are from non-Gaussian distributions.

A main theme of this paper is to study the performance of the PQL variance component estimators, in particular, when the random effects are heterogeneous in the sense that the distributions of the random effects vary over subjects. The heterogeneity in random effects may arise in many different situations. The two common cases are: (1) the variability of individual effects can be affected by covariates such as sex, race, origin of each subject as in the (mean variance) joint-model of the generalized linear model (McCullagh and Nelder, 1989); (2) the random effects also become heterogeneous when there exist several sources of variations (Li and Zhong, 2002).

In this paper, we address two specific questions for the performance of the PQL estimators when the random effects are heterogeneous.

First, a simulation study in Section 3.1 shows that the biases of regression coefficient estimates by PQL are closely related to those of the variance component estimates by PQL, and, also shows that the PQL regression coefficient estimators are biased even we accurately estimate the variance components.

Second, Section 3.2 shows that the PQL underestimates the variance components with the heterogeneity in random effects and the biases increase as the random effects become more heterogeneous; accordingly, the biases of PQL random coefficient estimators also increase.

This paper is organized as follows. Section 2 summarizes the GLMM, REML and the PQL. Section 3 implements two simulation studies on the performance of the PQL estimators. Section 4 concludes the paper with discussions of other issues on the heterogeneous random effects not covered in this paper.

2 GLMM, REML and PQL

Let us consider the odds ratio inference problem for a series of 2×2 tables from eight clinical centers reported by Beitler and Landis (1985). Each table has counts of “success” and “failures” among 293 patients distributed in treatment and control groups at each clinical center. The experimenters were interested whether if there was a clinic specific treatment effect. Let y_{ij} be a binary indicator which is 1 for success and 0 for failure for j th subject in the i th clinic. Let the covariate x_{ij} be $-\frac{1}{2}$ for control and $\frac{1}{2}$ for treatment. A simple model for analysis is a simple random intercept model:

$$\text{logit}E(y_{ij}|b_i) = \alpha_0 + \alpha_1 x_{ij} + b_i,$$

where $\text{logit}(x) = \log\left(\frac{x}{1-x}\right)$ and $b = (b_1, \dots, b_8)^T \sim N_8(0, D)$.

The above model assumes the constant odds ratio over the clinic centers and individual clinical center effect can be explained by the random effect b .

We often assume D is the identity matrix, but it may not be true for some cases. For example, the variations of odd ratios can be different by environmental factors. Suppose that deviations of odds ratios of 4 clinical centers from certain areas are bigger than those of the others. To take it account, one may assume $D = \text{diag}(\theta_1, \theta_2) \oplus I_4$.

One can explain the above model as a general framework of the generalized linear mixed model. Let y_1, \dots, y_N are independent vectors and a sample of units with n_i measurements of response of on each unit. Suppose that $y_i = (y_{i1}, y_{i2}, \dots, y_{in_i})$ are observed along with the covariates $\{x_{i1}, x_{i2}, \dots, x_{in_i}\}$ for the fixed effects $\alpha^T = (\alpha_1, \alpha_2, \dots, \alpha_p)$ and $\{z_{i1}, z_{i2}, \dots, z_{in_i}\}$ for the random effects $b^T = (b_1, b_2, \dots, b_q)$. Here, x_{ij} are $p \times 1$ vectors and z_{ij} are $q \times 1$ vectors.

Then, it is assumed that, given b , y_{ij} are independent of each other with means and variances specified as:

$$E(y_{ij}|b) = \mu_{ij}^b = h(x_{ij}^T \alpha + z_{ij}^T b) \quad \text{and} \quad \text{Var}(y_{ij} | b) = \frac{\phi}{a_{ij}} V(\mu_{ij}^b),$$

where $g = h^{-1}$ is the link function; ϕ is a dispersion parameter; a_{ij} is a prior weight and $V(\cdot)$ is a variance function, subjectively specified.

Here, α is a $q \times 1$ vector of fixed effects and the random effects b follows a multivariate normal distribution with mean $\mathbf{0}$ and covariance matrix $D = D(\theta)$ and θ is a $c \times 1$ unknown vector of variance components. It is usually assumed that the random effects are independent of each other which means $D = \text{diag}(\theta_s \oplus I_{q_s})$ for $s = 1, \dots, c$ and $\sum_{s=1}^c q_s$.

The link function can be expressed as using matrix notation as:

$$g(\mu_i^b) = X_i^T \alpha + Z_i^T b.$$

where $\mu_i^b = (\mu_{i1}^b, \dots, \mu_{in_i}^b)^T$ and the design matrix X_i and Z_i have rows x_{ij}^T and z_{ij}^T .

To estimate the parameters, one can consider the integrated quasi-likelihood $L(\alpha, \theta)$ which is given by

$$L = \frac{1}{\sqrt{(2\pi)^q |D(\theta)|}} \int \exp \left[-\frac{1}{2\phi} \sum_{i=1}^N \sum_{j=1}^{n_i} d_{ij}(y_{ij}, \mu_{ij}^b) - \frac{1}{2} b^T D^{-1}(\theta) b \right] db,$$

where $d_{ij}(y, \mu) = -2a_{ij} \int_y^\mu \frac{y-u}{v(u)} du$.

Since the integral cannot be evaluated as a closed form, an alternative is to use a Laplace approximation to

$$\mathbf{PQL}(\alpha, b) = -\frac{1}{2\phi} \sum_{i=1}^N \sum_{j=1}^{n_i} d_{ij}(y_{ij}, \mu_{ij}^b) - \frac{1}{2} b^T D^{-1}(\theta) b,$$

which is the *penalized quasi-likelihood*.

Breslow and Clayton (1993) replaced PQL with its quadratic expansion at $\hat{b} = \arg \min \mathbf{PQL}(\alpha, b)$ for fixed θ and α and defined $\hat{\alpha} = \arg \min \mathbf{PQL}(\alpha, \hat{b})$ for fixed θ . After further approximation, they derived the standard REML estimating equation for θ (Harville, 1977) with the *working vector* $Y_i^* = (y_{i1}^*, \dots, y_{in_i}^*)^T$ which is defined as follows:

$$y_{ij}^* = g(y_{ij}) = g(\hat{\mu}_{ij}^b) + (y_{ij} - \hat{\mu}_{ij}^b) g'(\hat{\mu}_{ij}^b),$$

where $\hat{\mu}_{ij}^b = h(x_{ij}^T \hat{\alpha} + z_{ij}^T \hat{b})$.

Then, based on the original model, one can describe the distribution of y_{ij}^* as a linear model with structure,

$$Y_i^* = X_i \hat{\alpha} + Z_i \hat{b} + \epsilon_i,$$

where $\epsilon_i \sim N(\mathbf{0}, W_i^{-1})$ and W_i is the diagonal matrix of

$$w_{ij} = \{V(\hat{\mu}_{ij}^b)(g'(\hat{\mu}_{ij}^b))^2\}^{-1}, \text{ for } j = 1, \dots, n_i.$$

We briefly summarize the procedures as follows:

Step 1 Given θ and b , we can estimate fixed effect α by solving the normal equation

$$\sum_{i=1}^N X_i^T V_i^{-1} X_i \alpha = \sum_{i=1}^N X_i^T V_i^{-1} Y_i^*,$$

where $V_i = W_i^{-1} + Z_i D Z_i^T$.

Step 2 The random effect b can be estimated as

$$\hat{b} = \sum_{i=1}^N D Z_i^T V_i^{-1} (Y_i^* - X_i \hat{\alpha}).$$

Step 3 Subsequently, the REML estimator for θ is

$$\hat{\theta}_s = \frac{\sum_{n \in Q_s} \hat{b}_n^2}{\sum_{n \in Q_s} (1 - t_{nn})}, \text{ for } s = 1, \dots, c,$$

where

$$Q_s = \left\{ n : \sum_{i=1}^{s-1} q_i < n \leq \sum_{i=1}^s q_i \right\}, \quad S = W - W X (X^T W X)^{-1} X^T W,$$

$$X^T = (X_1^T, \dots, X_N^T), \quad Z^T = (Z_1^T, \dots, Z_N^T), \quad W = \text{diag}(W_1, W_2, \dots, W_N),$$

and t_{nn} is the n th diagonal element of $T = (I + Z^T S Z D)^{-1}$.

Step 4 One then updates Y_i^* at the end of each iteration. The PQL estimators are defined upon convergence.

Finally, the covariance matrix of the estimators can be computed at the value $\alpha = \hat{\alpha}$ and $b = \hat{b}$ by

$$\text{Cov}(\hat{\alpha}) = \left\{ \sum_{i=1}^N X_i^T V_i^{-1} X_i \right\}^{-1}, \quad \text{Cov}(\hat{\theta}) = H^{-1}. \quad (1)$$

Here H has components

$$h_{st} = \frac{1}{2} \sum_{i \in Q_s} \sum_{j \in Q_t} \left(Z_{(i)}^T P Z_{(j)} \right)^2,$$

where $Z_{(i)}$ is the i th row vector of Z , $V^{-1} = \text{diag}(V_1^{-1}, \dots, V_N^{-1})$ and

$$P = V^{-1} - V^{-1} X \text{Cov}(\hat{\alpha}) X^T V^{-1}.$$

For more detail on this, refer to Harville (1977) and Breslow and Clayton (1993).

3 Simulation Studies

In this section, we implemented two simulation studies to investigate the performance of the PQL estimators in the GLMM. First, we studied how the PQL regression coefficient estimators vary according to the magnitude of the biases of the variance component estimators. Second, we evaluated the performance of the estimates by PQL for heterogeneous random effects at various levels of the heterogeneity.

In both simulations, simple logistic, probit and Poisson regressions with random intercepts were used. Each simulated data set had 50 subjects with 4 repetitions in each subject. For logistic and probit regressions, Bernoulli random variables, y_{ij} , were generated at each subject with conditional mean μ_{ij}^b given by

$$g(\mu_{ij}^b) = b_i + \alpha_0 + \alpha_1 x_{ij1} + \alpha_2 x_{ij2},$$

where $g(x) = \log\left(\frac{x}{1-x}\right)$ or $\Phi^{-1}(x)$ with Φ is the normal cumulative distribution function. Here, x_{ij1} and x_{ij2} were independently from $N(0, 1)$ for $i = 1, \dots, 50$ and $j = 1, \dots, 4$. The fixed effects were set to be $\alpha^T = (0.5, 2.0, 0.0)$ and the random effects b were generated from a multivariate normal distribution with mean $\mathbf{0}$ and covariance $D = \text{diag}(\theta_1, \theta_2) \oplus I_{25}$. In other words, the first 25 subjects had random intercepts of a variance θ_1 while those in the second half of the subjects had a variance θ_2 .

For the Poisson regression, random variables were generated with conditional mean μ_{ij}^b given by

$$\log(\mu_{ij}^b) = b_i + \alpha_0 + \alpha_1 x_{ij1} + \alpha_2 x_{ij2}.$$

	$\theta_1 = \theta_2$	α_0	α_1	α_2
Binary (Logit)	0.1	0.4456 (0.2166)	1.7695 (0.2730)	-0.0038 (0.1766)
	0.5	0.4438 (0.2286)	1.7959 (0.2617)	-0.0057 (0.1821)
	1.0	0.4602 (0.2238)	1.8772 (0.2728)	-0.0030 (0.1896)
	1.5	0.4657 (0.2304)	1.9430 (0.2945)	0.0076 (0.2123)
	2.0	0.4959 (0.2397)	1.9916 (0.2837)	0.0001 (0.2031)
Binary (Probit)	0.1	0.3790 (0.1591)	1.5055 (0.2249)	0.0090 (0.1191)
	0.5	0.4267 (0.1778)	1.6694 (0.2300)	0.0032 (0.1296)
	1.0	0.4533 (0.1832)	1.8450 (0.2367)	0.0098 (0.1339)
	1.5	0.4657 (0.2304)	1.9430 (0.2945)	0.0076 (0.2123)
	2.0	0.4942 (0.1987)	2.0188 (0.2493)	-0.0027 (0.1569)
Poisson	0.1	0.7441 (0.1538)	1.9570 (0.0641)	-0.0004 (0.0483)
	0.5	0.5604 (0.1427)	1.9906 (0.0492)	-0.0001 (0.0334)
	0.75	0.5498 (0.1643)	1.9930 (0.0496)	0.0015 (0.0357)
	1.0	0.5224 (0.1791)	2.0001 (0.2390)	0.0011 (0.1421)
	1.25	0.4999 (0.1554)	2.0004 (0.0471)	0.0018 (0.0352)
	True value	0.5	2.0	0.0

Table 1: Average of fixed effects estimates over 500 data sets; the numbers in parentheses are the standard errors of the estimates. $\theta_1 = \theta_2 = 1.0$ is the true value in each simulation.

3.1 Downward biases of the PQL estimators

To show the effects of the variance components estimates to regression coefficient estimates, we computed the PQL estimates with fixed values of variance components. To be specific, 500 data sets of 200 observations (50 clusters with 4 repetitions) were generated under the above settings with $\theta_1 = \theta_2 = 1.0$ (true value). The PQL estimates were computed after fixing the variance components as constants 0.1, 0.5, 1.0, 1.5, and 2.0 for logistic and probit binary regressions and 0.1, 0.5, 0.75, 1.0, and 1.25 for Poisson regression. In the estimation process, we used the following initial values $\alpha_1 = \alpha_2 = \alpha_3 = 0, b_i = 0$ for $i = 1, \dots, 50$.

Table 1 shows the regression coefficient estimates for different values of variance components. Two interesting observations comes from Table 1.

First, while the PQL estimates work fine for the Poisson regression when the true variance component values are given, it can be found that the PQL estimates of the regression coefficient for binary regressions are still underestimated even we set the variance components as their true values.

Second, it can be found that, for binary regressions, as the fixed values of the variance component decrease, the regression coefficient estimate of α_1 also decrease toward 0. As pointed out in Section 1, these observations are compatible with the well known results that the regression coefficients are downward biased to 0 when the random effects are mistakenly disregarded (Neuhaus , 1998; Henderson and Oman , 1999).

3.2 Biases of the PQL variance component estimators

To see the effect of heterogeneous random effects, five hundred data sets of 200 observations were generated with the same data structure (50 subjects with 4 replications). A simulation study implemented under the following 4 scenarios for logistic and Poisson regressions, respectively. As we observed from the simulation in Section 3.1, the logistic and the probit regressions were not different much in their results, hence, we only considered the logistic regression for binary outcomes. The four scenarios for logistic regression were:

Scenario 1: $\theta = (0.5, 0.5)$;

Scenario 2: $\theta = (0.5, 1.0)$;

Scenario 3: $\theta = (0.5, 1.5)$;

Scenario 4: $\theta = (0.5, 2.0)$.

For Poisson regression,

Scenario 1: $\theta = (0.5, 0.5)$

Scenario 2: $\theta = (0.5, 0.75)$;

Scenario 3: $\theta = (0.5, 1.0)$;

	(θ_1, θ_2)	$\hat{\alpha}_0$	$\hat{\alpha}_1$	$\hat{\alpha}_2$	$\hat{\theta}_1$	$\hat{\theta}_2$
Binary (Logit)	(0.5, 0.5)	0.4971	1.9488	-0.0037	0.4585	0.4782
	(0.5, 0.1)	0.4824	1.8972	-0.0063	0.4257	0.7881
	(0.5, 1.5)	0.4716	1.8600	-0.0049	0.4051	1.0958
	(0.5, 2.0)	0.4653	1.8311	-0.0045	0.3879	1.3863
Poisson	(0.5, 0.5)	0.5327	1.9973	-0.0005	0.4784	0.4761
	(0.5, 0.75)	0.5345	1.9939	-0.0016	0.4866	0.7036
	(0.5, 1.0)	0.5373	1.9966	0.0029	0.4891	0.9666
	(0.5, 1.25)	0.5277	1.9958	-0.0009	0.4828	1.2174
	True value	0.5	2.0	0.0	0.5	

Table 2: Average of fixed effects and variance component estimates over 500 data sets.

Scenario 4: $\theta = (0.5, 1.25)$.

Table 2 contains the average of the estimates of fixed effects and variance components. The first four rows in Table 2 correspond to these averages for the four different scenarios, while the last row contains the true parameter values for $\alpha_0, \alpha_1, \alpha_2$ and θ_1 . Since the values of θ_2 are different for each scenarios, we report them inside the parentheses after the estimated θ_2 . Using the estimated α_1 as a typical example in the estimated fixed effects, one may find that the percentage of bias in $\hat{\alpha}_1$ is 2.6% in scenario 1, 5.1% in scenario 2, 7% in scenario 3, and 8.4% in scenario 4. In other words, the larger θ_2 is, the larger the biases of fixed effect estimates are. In variance component estimates, the percentage of biases in $\hat{\theta}_2$ dramatically increases from 0.4% to 30.7% as θ_2 increases. The effect of increasing θ_2 on the direction of biases in $\hat{\theta}_1$ is almost the same as on $\hat{\theta}_2$, but the percentage of biases in $\hat{\theta}_1$ are much smaller than those in θ_2 as θ_2 increases.

Table 3 shows the ratios of the estimated and the Monte Carlo simulated standard errors. The ratios are defined as Est./Monte where “Est” and “Monte” correspond to the estimated and Monte Carlo standard errors. The Monte Carlo standard errors are sample standard deviations of the PQL estimators and the estimated standard errors are calculated from the equation (1).

	(θ_1, θ_2)	$\hat{\alpha}_0$	$\hat{\alpha}_1$	$\hat{\alpha}_2$	$\hat{\theta}_1$	$\hat{\theta}_2$
Binary	(0.5,0.5)	0.9594	0.9652	0.9969	1.2974	1.2677
	(0.5,1.0)	0.9626	0.9633	1.0211	1.3314	1.1239
	(0.5,1.5)	0.9582	0.9619	1.0275	1.3519	1.0720
	(0.5,2.0)	0.9597	0.9850	1.0312	1.3862	1.0423
Poisson	(0.5, 0.5)	0.9879	0.9779	0.9580	0.9871	1.0538
	(0.5, 0.75)	0.9447	0.9921	0.9610	1.0692	1.0659
	(0.5, 1.0)	0.9456	0.9456	1.0114	0.9876	0.8325
	(0.5, 1.25)	0.9992	0.9865	1.0406	1.0271	0.7824

Table 3: Ratios of Estimated and Monte Carlo simulated standard errors.

In fixed effects of the logit regression, the ratios are between 0.95 and 1.04. However, for variance components of the binary regression, the Monte Carlo simulated standard errors are much smaller than the estimated standard errors. One may note that the ratios of Monte Carlo and estimated standard errors in θ_2 decreases while ratios in θ_1 increase as θ_2 increases .

For variance components of the Poisson regression, one can find similar results; the ratios of Monte Carlo and estimated standard errors in θ_2 decrease as θ_2 increases, but the ratios in θ_1 remains around 1.

Figure 1 shows the distributions of $\hat{\theta}$ for logistic regression. Density estimates are computed by kernel smoothing and the true parameter values are as indicated in each plot. Long right tails are observed in all $\hat{\theta}$ which is consistent with Lin (1997). Lin also pointed out that $\hat{\theta}$ is not exactly normally distributed, unless that the number of clusters is really large and the θ are bounded away from the boundary, 0.

4 Discussion

This paper numerically studies the performance of the PQL estimates in the GLMM when the random effects are heterogeneous. The two main results for binary outcomes are: (1) the PQL regression coefficient estimates are biased even though we precisely estimate the variance components; (2) using PQL, the variance components are underestimated while the standard er-

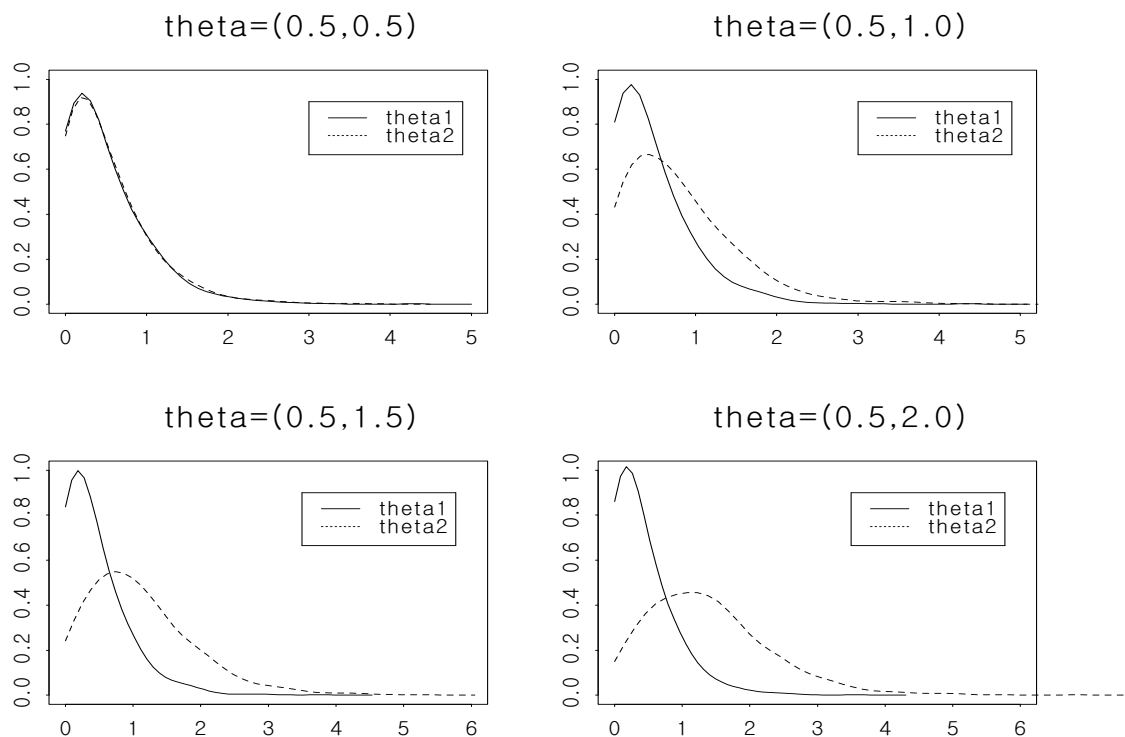


Figure 1: Distribution of $\hat{\theta}$

rors of variance components are overestimated when the random effects are heterogeneous; such phenomena is clear in all binary regression. However, as pointed out in Lin (1997) and Breslow (2005), the biases of the PQL estimators decrease as the cluster size n_i increases.

In the remainder, we finish the paper with a few discussions of further issues on heterogeneous random effects which are not covered in this paper.

Firstly, bias correction of the PQL estimates has been suggested by several authors in the previous literature. Some of them are based on higher order Laplace approximation (Breslow and Lin, 1995; Lin and Breslow, 1996; Lin, 1997). However, their adjustments do not seem to take care of the heterogeneity of the random effects in the sense that $\hat{\theta}^{ADJ} / \hat{\theta}^{PQL}$ does not change over different level of heterogeneity (see Equation (20) in Lin and

		$\hat{\alpha}_0$	$\hat{\alpha}_1$	$\hat{\alpha}_2$	$\hat{\theta}_1$	$\hat{\theta}_2$
Logit	Homo.	0.4538 (0.2248)	1.8924 (0.3017)	0.0023 (0.1784)	0.5912 (0.3384)	0.5912 (0.3384)
	Hetero.	0.4824 (0.2248)	1.8972 (0.2996)	-0.0063 (0.1897)	0.4257 (0.4662)	0.7881 (0.6681)
Poisson	Home.	0.5338 (0.1246)	1.9953 (0.0526)	0.0010 (0.0373)	0.6020 (0.1563)	0.6020 (0.1563)
	Hetero.	0.5345 (0.1302)	1.9939 (0.0511)	-0.0016 (0.0385)	0.4866 (0.1590)	0.7036 (0.2217)
True		0.5	2.0	0.0	0.5	

Table 4: The effect of mis-specified homogeneous random effects model. The numbers in the parentheses are the Monte Carlo standard errors of the corresponding estimates.

Breslow (1996) ; $\hat{\theta}^{ADJ}$ is their bias corrected PQL estimator and $\hat{\theta}^{PQL}$ is the PQL estimator. This is also found from their simulation reported in Section 7 of their paper; $\hat{\theta}^{ADJ}$ does not perform better than $\hat{\theta}^{PQL}$ when the random effects are heterogeneous.

Secondly, with the application to more complicated problems, we often need to capture several sources of variations, however, it is hard to specify all sources of variations. Upon such difficulty, it is natural to ask what if we mistakenly treat observations with heterogeneous random effects as those with homogeneous. Here, we implement a simple simulation study in extending those in Section 3; we generate 200 data sets from models with heterogeneous random effects and estimate the parameters under the assumption of homogeneous random effects which means we assume $D(\theta) = \theta \cdot I_{50}$ when the true $D(\theta) = (\theta_1, \theta_2) \oplus I_{25}$. We assume $(\theta_1, \theta_2) = (0.5, 1.0)$ for logistic and $(\theta_1, \theta_2) = (0.5, 0.75)$ for Poisson regressions. The results are reported in Table 4.

The results in Table 4 show that the estimates from the misspecified homogeneous model are smaller than those from the heterogeneous random effects model (which is correctly specified). Such phenomena is consistent with the results of Neuhaus (1998) and Henderson and Oman (1999) in

the sense that, when we mistakenly assume a shorter tail distribution than the true random effect distribution, the regression coefficient estimates are downward biased to 0. However, theoretical justification on the conjecture is demanded.

References

- Beitler, P.J. and Landis, J.R. (1985). A mixed-effects model for categorical data. *Biometrics*, **41**, 991-1000.
- Breslow, N.E. (2005). Whither PQL? In *Proceedings of the Second Seattle Symposium in Biostatistics*, Springer.
- Breslow, N.E. and Clayton, D.G. (1993). Approximate inference in generalized linear mixed models. *Journal of American Statistical Association*, **88**, 9-25.
- Breslow, N.E. and Lin, X. (1995) Bias correction in generalized linear mixed models with a single component of dispersion. *Biometrika*, **82**, 81-91.
- Harville, D.A. (1977). Maximum likelihood approaches to variance component estimation and to related problems. *Journal of American Statistical Association*, **72**, 320-340.
- Henderson, R. and Oman, P. (1999). Effect of frailty on marginal regression estimates in survival analysis. *Journal of the Royal Statistical Society: Series B*, **61**, 367-379.
- Lee, Y. and Nelder, J.A. (1996), Hierarchical generalized linear models. *Journal of the Royal Statistical Society: Series B*, **58**, 619-678.
- Li, H. and Zhong, X. (2002), Multivariate survival models induced by genetic frailties, with applications to linkage analysis. *Biostatistics*, **3**, 57-75.
- Lin, X. (1997) Variance component testing in generalised linear models with random effects. *Biometrika*, **84**, 309-326.

- Lin, X and Breslow, N.E. (1996) Bias Correction in Generalized Linear Mixed Models with Multiple Components of Dispersion. *Journal of American Statistical Association*, **91**, 1007-1016.
- McCullagh, P. and Nelder, J.A. (1989). *Generalized Linear Models*. Chapman and Hall, London.
- Neuhaus, J.M. (1998). Estimation efficiency with omitted covariates in generalized linear models. *Journal of American Statistical Association*, **93**, 1124-1129.
- Schall, R. (1991). Estimation in generalized linear models with random effects. *Biometrika*, **78**, 719-727.
- Wolfinger, R. (1993). Laplace's approximation for nonlinear mixed models. *Biometrika*, **80**, 791-795.
- Zeger, S.L. and Liang, K.Y. (1986) Longitudinal data analysis for discrete continuous outcomes. *Biometrics*, **42**, 121-130.