WORKING PAPER SERIES

WORKING PAPER NO 14, 2007

ESI

Computationally feasible estimation of the covariance structure in Generalized linear mixed models (GLMM)

By

Kenneth Carling
Dalarna University

Md Moudud Alam
Örebro University

kca@du.se

maa@du.se

http://www.oru.se/esi/wps

SE-701 82 Örebro

SWEDEN

ISSN 1403-0586

Computationally feasible estimation of the covariance structure in Generalized linear mixed models (GLMM)

Kenneth Carling*and Md Moudud Alam†

September 10, 2007

Abstract

In this paper we discuss how a regression model, with a non-continuous response variable, that allows for dependency between observations should be estimated when observations are clustered and there are repeated measurements on the subjects. The cluster sizes are assumed to be large. We find that the conventional estimation technique suggested by the literature on Generalized Linear Mixed Models (GLMM) is slow and often fails due to non-convergence and lack of memory on standard PCs. We suggest to estimate the random effects as fixed effects by GLM and derive the covariance matrix from these estimates. A simulation study shows that our proposal is feasible in terms of Mean-Square Error and computation time. We recommend that our proposal be implemented in the software of GLMM techniques so that the estimation procedure can switch between the conventional technique and our proposal depending on the size of the clusters.

Mathematics Subject Classification: Primary 62J12; Secondary 65C60

Keywords: Monte-Carlo simulations, large sample, interdependence, cluster errors.

*Dalarna University, SE 781 88 Borlange, Sweden; kca@du.se

†Corresponding author: Örebro University and Dalarna University. Mailing address- Department of Economics and Social Sciences, Dalarna University, SE 781 88 Borlange, Sweden; maa@du.se.

‡This work was partially financed by Sveriges Riksbank. The views expressed in this paper are solely the responsibility of the authors and should not interpreted as reflecting the opinion of the Executive Board of Sveriges Riksbank. The authors wish to thank the participants of the seminars at Riksbank, the Department of Statistics at Umeå University and Uppsala University for their valuable comments and suggestions. We also gratefully acknowledge the extensive comments of Kasper Roszbach and Lars Rönnegård which improved the quality of the paper.

1 Introduction

This paper introduces a way to estimate a model with a non-continuous response variable in a setting when the responses are correlated within subjects, there are repeated measures and the sample size is large. The non-linear mixed effect models are widely used to capture the dependencies of the non-continuous response variable in the model. There are several different techniques to estimate the non-linear mixed models, for example, the numerical integration technique [12], Bayesian method based on Markov chain Monte-Carlo (MCMC) [7], frequentist version of MCMC [22], Penalized Quasi Likelihood (PQL) approach [3] [26] and the hierarchical (or h-) likelihood approach [14]. Each of the techniques has some advantages and disadvantages. Nevertheless, the methods mentioned above becomes computationally heavy for large data set.

We propose a discrete response mixed model to be estimated via the fixed effect approach by deriving the covariance matrix of the random effect from the fixed effect model parameters. A successful estimation of the covariance matrix of the random effect is of central importance since it determines the uncertainty of any single prediction.

We have investigated the accuracy of the proposed fixed effect approach by the use of simulated data. The Monte-Carlo simulation shows that the proposed approach performs worse than the PQL approach, in terms of the mean sum of squared error (MSE) and some other relative measure, for small sample sizes, equally well for moderate sample sizes but it continues to work for large sample sizes when the PQL slows down.

The paper is organized in the following way. The second section presents the model and discusses the existing techniques to estimate it as well as providing a heuristic description of our proposed approach, which can be called the fixed-effect (FE) approach, for a situation with a large cluster set. The third section compares the FE-approach to the conventional PQL-approach by means of a simulation study. The fourth section is dedicated to a discussion on various practical issues including inference and the fifth section concludes the paper.

2 The mixed model and its estimation

In this section we present the model and discuss how it can be estimated. We will briefly revise the existing approaches to estimate it and conclude that they are infeasible for large clusters. For this reason we suggest an approach that can cope with this.

2.1 The model specification

The linear regression model can be formulated as

$$y_i = \mathbf{x}_i\boldsymbol{\beta} + \varepsilon_i \tag{1}$$

where $\mathbf{x}_i$ is a row-vector of p covariates for the i:th observation, $\boldsymbol{\beta}$ is a $p \times 1$ vector of parameters associated with the covariates, and ε_i is the error term. Conventionally a number of assumptions are imposed in this model. First, it is assumed that $E[\varepsilon_i] = 0$ and that $Cov[\varepsilon_i, \mathbf{x}_i] = 0$, and these assumptions will be maintained in this paper. Moreover it is assumed that $E\left[\varepsilon_i^2\right] = \sigma^2$, and that $E[\varepsilon_i \varepsilon_j] = 0$ for all $i \neq j$. The last assumption of independence between the error terms means that conditional on $\mathbf{x}$ the response is independent between two observations. In applied work there are sometimes reasons to expect that a set of observations (a cluster) are in fact dependent. Such reasons are, e.g., that there are several observations on the same individual (repeated measurements) or that the observations have been gathered by cluster sampling or that observations are spatially related. A neat way of extending the model to include (positive) dependence is to introduce a cluster specific component. Assume that there exist K clusters and let the number of observations in each cluster be denoted by n_k $(k = 1, ..., K)$ and indexed by i. The model extends to

$$y_{ki} = \mathbf{x}_{ki}\boldsymbol{\beta} + u_k + \varepsilon_{ki} \tag{2}$$

where u_k is the cluster specific component and it is by construction assumed to be independent of $\mathbf{x}$ and ε. A conceptual issue discussed elsewhere (see for instance Greene [11]) is whether u_k should be considered a fixed or a random effect. However, to introduce dependence it is useful to view u_k as a random variable from the $N(0, d_k)$ distribution. In this case, if ε also is assumed to follow the normal distribution, the correlation between two observations on y is $\frac{d_k}{d_k+\sigma^2}$, if they are in the same cluster k and zero otherwise. The $K \times K$ and symmetric covariance matrix, denoted $\mathbf{D}$, for the random effect vector $\mathbf{u} = (u_1, ..., u_k)'$ has the following form under the model in (2)

$$\mathbf{D} = \begin{pmatrix} d_1 & 0 & \cdots & 0 \\ 0 & d_2 & & \vdots \\ \vdots & & \ddots & 0 \\ 0 & \cdots & 0 & d_K \end{pmatrix}$$

and the distribution for $\mathbf{u}$ is taken to be $N_K(\mathbf{0}, \mathbf{D})$. It might also be that observations are dependent even if they are from different clusters, and the model in (2) can be extended to allow for such between cluster dependence as

$$y_{ki} = \mathbf{x}_{ki}\boldsymbol{\beta} + u_k + \sum_{k' \neq k} \delta_{k'} u_{k'} + \varepsilon_{ki} \tag{3}$$

The second and the third terms in the right hand hand side of the equation (3) construct a new random effect term, u_k, which is, indeed, a linear combination of K independent normal variates. Hence, model (3) takes the same functional form as model (2) with an extended

covariance matrix given by

$$\mathbf{D} = \begin{pmatrix} d_{11} & d_{12} & \cdots & d_{1K} \\ d_{21} & d_{22} & & \vdots \\ \vdots & & \ddots & d_{(K-1)K} \\ d_{K1} & \cdots & d_{K(K-1)} & d_{KK} \end{pmatrix}.$$

The covariance matrix has $K(K+1)/2$ free parameters due to the symmetry, for instance $d_{12} = d_{21}$.

Estimation of the model in (3) along with the covariance matrix can be done by iterative Generalized Least Squares (GLS) as long as y is a continuous variable. If, however, y is a non-continuous response variable, estimation is more challenging.

Consider the threshold model as a convenient way of expressing a regression model for a binary response variable which resembles the models that have been discussed above. Let y be an indicator variable taking on either the value unity or zero. In the threshold model y is related to an underlying continuous variable y^* in the following way,

$$y = \begin{cases} 1, \text{ if } y^* \geq 0 \\ 0, \text{ if } y^* < 0 \end{cases} \tag{4}$$

where y^* is modeled as above. Absent the random effect components the regression model for y can be either the logistic or the probit model. The threshold model turns the specification into a non-linear one that makes standard econometric procedures for estimating linear random effect models inadequate.

Much progress has been made, however, in the development of statistical procedures for the estimation of these types of models. Probit models with random effects have been widely used in econometrics. See Heckman and Willis [13] for an early example and Guilkey and Murphy [12] for details. This line of research has mainly considered panel data with a small number of repeated measurements on the same subject, often less than 10. The parameters in such models with small cluster sizes *i.e.* the number of repeated observations on the subject have typically been estimated by maximizing the likelihood of the multivariate normal distribution using numerical integration (see Butler and Moffitt [4] and Greene [9]).

The dimension of the numerical integration, in order to evaluate the likelihood function, equals the cluster size, *i.e.* n_k. High-dimensional numerical integration is slow and inaccurate [27] and therefore the approach is badly suited for problems where the size of the clusters is large. These problems of numerical integration could however be overcome in a Bayesian approach by use of importance sampling or Gibbs sampling techniques, and Bayesian methods thus present an alternative (see *e.g.* Clayton [7]) by use of Markov Chain Monte-Carlo algorithms. The basic principle of the MCMC approaches is to generate a sequence of Markov chain values on the conditional distribution of the random effect given the data and use them in the Monte-Carlo EM or Newton-Rapson algorithm to bypass the high-dimentional integration through

the Monte-Carlo expectation. McCulloch [20] provides a frequentist version of the MCMC approach using the Gibbs sampler technique for the probit model, Mcculloch [21] presented a more generalized version of the MCMC Newton-Rapson (NR) technique with the Metropolis-Hasting algorithm, Booth and Hobert [2] provide an automated MCMC-EM algorithm with the importance sampling, Quintana *et.al.* [23] provide an MCMC-EM algorithm with the importance reweighting, Chen [6] proposed a computationally feasible MCMC-EM algorithm while Leviene and Fan [16] presented an automated EM algorithm. McCulloch and Searle [22] contains a general introduction to the MCMC approach. However, it should be noted here that all these references mentioned above dealt with the random effects with a diagonal covariance matrix. Alam [1] studied the MCMC-NR technique with the Metropolice-Hasting algorithm along with an unstructured covariance matrix, *i.e. a* $\mathbf{D}$ matrix with the off diagonal terms, of the random effect and found it useful for small clusters, but computationally too expensive for large clusters.

The third approach to estimate the model in equation (2) has been developed in the work on Generalized Linear Models (GLM). McCullagh and Nelder [19] contains an introduction. A generalized linear model with a random effect component is called Generalized Linear Mixed Model (GLMM) and is described by the following five assumptions: i) y_{ki} is observed independently at a given value of the covariate (x_{ki}) and a given realization of the random effect u_k, ii) x_{ki} and u_k influence the distribution of y_{ki} through a linear function $\eta_{ki} = \mathbf{x}_{ki}\boldsymbol{\beta} + u_k$ which is called a linear predictor, iii) conditional on u_k, $\mu_{ki} = E(y_{ki}|u_k)$ satisfies $g(\mu) = \eta$ for some function g which is called a link function, iv) conditional on $\mathbf{u}$, the distribution of y_{ki} belongs to the exponential family of distributions and v) $\mathbf{u}$ follows a marginal distribution, $h(\mathbf{u})$[1].

The binary regression model with random effect components is a special case of the Generalized Linear Mixed Model (GLMM) where $y_{ki}|u_k \backsim binomial(n_{ki}, \pi_{ki})$. For the logistic mixed model the link function, $g(.)$, is given as $g(\mu_{ki}) = \log\left(\frac{\pi_{ki}}{1-\pi_{ki}}\right)$[2].The above link function, g, is called the canonical link for binomial y.

The general idea in GLMM is to linearize the non-linear models by the first (or sometimes the second) order Taylor expansion of the link function and then regress the linearized version of the response variable on $\mathbf{x}_i$ with a procedure similar to (iterative) Generalized Least Squares. Details of the procedure, which often is referred to as PQL, are given in Zhou, Perkins and Hui [28].[3] The iterative scheme of the PQL-approach makes the properties of the estimator inherently dependent on the actual computer code in use. It is also well-known that PQL suffers from small-sample bias [8] due to the first (or second) order Taylor approximation. An extensive

[1] Here $h(\mathbf{u}) \sim N_K(0, \mathbf{D})$

[2] For the probit model, $g(\mu_{ki}) = \Phi^{-1}(\pi_{ki})$ [Φ stands for the standard normal CDF] while for the complementary log-log model it is $g(\mu_{ki}) = \log(-\log(1 - \pi_{ki}))$

[3] One implementation provided by SAS is the GLIMMIX macro (Wolfinger and O'Connell [26]). In this macro, one can implement different link functions including the logistic, probit and complementary log-log ones. See for example Littell *et.al.* [17].

simulation study of the estimator in various program packages, which would contribute greatly to the understanding of the estimator's properties, is still lacking.

The fourth approach to estimate the mixed effect models has been developed with the hierarchical (or h-) likelihood [14] but they are widely criticized in the case of the binary response data [8]. Recently, Nelder and Lee [15] have developed the double hierarchical generalized linear models (DHGLM) technique. They claim that it works foor the situation of correlated random effects but still we have to wait until the implementation of the algorithm becomes available in commonly used softwares[4].

While we were re-analyzing the data used in Carling, *et.al.* [5], we found that the PQL-approach is the best in terms of computational speed of the first three approaches discussed above. It can handle small and mid-sized clusters, but we have encountered great difficulties in applications with large clusters ($n_k > 1000$).

Since none of the methods discussed above is computationally efficient for large data set, we will investigate the working of an approach based on the estimation of the fixed cluster effects, from which the $\mathbf{D}$ matrix can be derived. Often the interest is on the estimation of $\boldsymbol{\beta}$ and the $\mathbf{D}$ matrix is considered to eliminate the risk of biased estimates of $\boldsymbol{\beta}$. However, a single value prediction can be done without any reference to the covariance matrix, $\mathbf{D}$, but any kind of statement about the prediction uncertainty must rely on an understanding of the $\mathbf{D}$ matrix. For this reason, it is important to get good estimates of the covariance matrix.

2.2 Estimation of D via fixed cluster effects

Let us consider the model again,

$$y_{ki} = \mathbf{x}_{ki}\boldsymbol{\beta} + u_k + \varepsilon_{ki}$$

with the covariance matrix :

$$\mathbf{D} = \begin{pmatrix} d_{11} & d_{12} & \cdots & d_{1K} \\ d_{21} & d_{22} & & \vdots \\ \vdots & & \ddots & d_{(K-1)K} \\ d_{K1} & \cdots & d_{K(K-1)} & d_{KK} \end{pmatrix}.$$

Assume that $\mathbf{u}$ is $N_K(\mathbf{0}, \mathbf{D})$ where $\mathbf{u} = (u_1, ..., u_K)'$. Assume further that there exist T realizations on $\mathbf{u}$ where $t = (1, ..., T)$. Let the $T \times K$ matrix $\mathbf{U} = (\mathbf{u}_1', ..., \mathbf{u}_T')'$ collects all realizations of the random-effect components. A natural estimator for $\mathbf{D}$ is

$$\mathbf{D} = \frac{1}{T}\mathbf{U}'\mathbf{U} \tag{5}$$

since $E[u_{kt}] = 0$. Such an estimator would be consistent since,

$$\underset{T\to\infty}{P\lim}\frac{1}{T}\mathbf{U}'\mathbf{U} = \mathbf{D}$$

[4] A Genstat implementation of the algorithm is available from: j.nelder@imperial.ac.uk[15]

given that T is large enough (at least $(K+1)/2$) to allow identification of all parameters in $\mathbf{D}$. The random effect, u_{kt}, is not observable, but it can be estimated in a model like (3) by treating the component u_{kt} as a fixed effect that is estimated jointly with $\boldsymbol{\beta}$. For expositional simplicity, assume that $n_{kt} = n_{k't'}$ $\forall t, t', k$, and k'. The basic intuition here is that

$$\underset{n_{kt}\to\infty}{P\lim}\,\widehat{u}_{kt}=u_{kt}$$

such that a large cluster size would reveal the realization of the underlying, unobserved random-effect. Moreover, if $\widehat{\mathbf{U}}$ denotes the estimates of the components in the matrix $\mathbf{U}$, we would expect the following to hold,

$$\underset{T\to\infty}{P\lim}\left(\underset{n_{kt}\to\infty}{P\lim}\frac{1}{T}\widehat{\mathbf{U}}'\widehat{\mathbf{U}}\right)=\mathbf{D}.$$

Thus, it seems to be possible to estimate the covariance structure provided that the estimation procedure yields consistent estimates of the realizations of the random-effects. This might not be the case for the implemented versions of the GLM estimation method. And even if this is the case, it is unclear how this consistency argument should be interpret in a practical situation where neither n_{kt} nor T will be close to infinity.

The proposed fixed effect approach (FE) is almost surely bad compared with the PQL approach for n_{kt} small. The question is if it can match the PQL for cluster sizes where PQL tends to fail. To answer this question we conduct a simulation study where we compare the two approaches at varying values of n_{kt}, K, and T by checking the bias, for all the model parameters.

3 Monte-Carlo experiment

To check the accuracy of the parameter estimates in FE method in comparison with PQL method we have conducted some Monte-Carlo simulations. We have used a probit mixed model and a Poisson mixed model with no intercept and one covariate with a coefficient value $\beta = 0.5$ along with different covariance structures, namely diagonal, compound symmetry and unstructured, of the random effect $\mathbf{u}$. The specific construction of the different covariance matrices, $\mathbf{D}$, are shown in the left hand side of the Table 1.1-1.3. Among those matrices the diagonal and compound symmetric $\mathbf{D}$ matrices were rather judgemental. The diagonal matrix was the identity matrix and the compound symmetric matrix implied a correlation of 0.5. The unstructured $\mathbf{D}$ was constructed from the Swedish inter industry input-otput matrix after merging the companies into 7 clusters in almost the same way as done in Carling *et.al.*[5], adjusting the row sums equal to one by dividing all the elements in each row by their respective row sums and then performing a matrix product of it with its transpose. Reason behind constructing such an unstructured $\mathbf{D}$ matrix is that we assume this unstructured matrix might explain the dependencies among the companies in Sweden.

Moreover, for $K = 3$ we have considered the first 3×3 minor of $\mathbf{D}$. Starting with a reasonably small cluster size ($n_{kt} = 10$) and small number of clusters ($K = 3$) we have sequentially increased the cluster size (up to $n_{kt} = 500$) and number of clusters (up to $K = 7$) and the number of repeated measure were T=10, 20 and 40 to study the speed of convergence of the expected estimates ($E[\widehat{\mathbf{D}}]$ and $E[\widehat{\boldsymbol{\beta}}]$) to the real parameter values which gives a sense of the consistency of the estimators.

Following the structure of the real data set used in Carling *et.al.*[5], we considered a maximum of 7 cross sectional clusters (K) and 20 time points (T).

Table 1.1 A diagonal covariance matrix of the random effects and its mean estimate through the fixed effect approach (cluster size is 500)

	Original variance covariance matrix							Mean estimate of variance covariance matrix						
	1	2	3	4	5	6	7	1	2	3	4	5	6	7
1	1	0	0	0	0	0	0	1.011	0.009	-0.013	0.002	-0.008	0.007	-0.017
2	0	1	0	0	0	0	0	0.009	1.090	-0.008	-0.061	-0.001	-0.001	0.011
3	0	0	1	0	0	0	0	-0.013	-0.008	1.036	0.005	0.018	-0.003	0.005
4	0	0	0	1	0	0	0	0.002	-0.061	0.005	1.080	0.017	-0.007	0.006
5	0	0	0	0	1	0	0	-0.008	-0.001	0.018	0.017	1.068	0.002	-0.004
6	0	0	0	0	0	1	0	0.007	-0.001	-0.003	-0.007	0.002	1.079	0.024
7	0	0	0	0	0	0	1	-0.017	0.011	0.005	0.006	-0.004	0.024	0.992

Table 1.2 A compound symmetric covariance matrix of the random effect and its mean estimate through fixed effect approach (cluster size is 500)

	Original variance covariance matrix							Mean estimate of variance covariance matrix						
	1	2	3	4	5	6	7	1	2	3	4	5	6	7
1	1.0	0.5	0.5	0.5	0.5	0.5	0.5	1.005	0.471	0.473	0.485	0.474	0.477	0.477
2	0.5	1.0	0.5	0.5	0.5	0.5	0.5	0.471	1.050	0.481	0.488	0.485	0.483	0.483
3	0.5	0.5	1.0	0.5	0.5	0.5	0.5	0.473	0.481	0.951	0.481	0.476	0.473	0.456
4	0.5	0.5	0.5	1.0	0.5	0.5	0.5	0.485	0.488	0.481	1.027	0.486	0.494	0.477
5	0.5	0.5	0.5	0.5	1.0	0.5	0.5	0.474	0.481	0.476	0.486	0.988	0.476	0.480
6	0.5	0.5	0.5	0.5	0.5	1.0	0.5	0.477	0.483	0.473	0.494	0.476	1.000	0.467
7	0.5	0.5	0.5	0.5	0.5	0.5	1.0	0.477	0.483	0.456	0.477	0.480	0.467	0.982

Table 1.3 A unstructured covariance matrix of the random effects and its mean estimate through fixed effect approach (cluster size is 500)

	Original variance covariance matrix							Mean estimate of variance covariance matrix						
	1	2	3	4	5	6	7	1	2	3	4	5	6	7
1	0.192	0.143	0.151	0.171	0.123	0.213	0.144	0.198	0.146	0.158	0.177	0.129	0.214	0.148
2	0.143	0.237	0.168	0.160	0.097	0.125	0.182	0.146	0.249	0.176	0.164	0.100	0.126	0.187
3	0.151	0.168	0.281	0.208	0.059	0.131	0.151	0.158	0.176	0.294	0.217	0.063	0.136	0.158
4	0.171	0.160	0.208	0.231	0.197	0.173	0.154	0.177	0.164	0.217	0.243	0.205	0.177	0.159
5	0.123	0.097	0.059	0.197	0.522	0.159	0.090	0.129	0.100	0.063	0.205	0.537	0.164	0.097
6	0.213	0.125	0.131	0.173	0.159	0.256	0.124	0.214	0.126	0.136	0.177	0.164	0.259	0.127
7	0.144	0.182	0.151	0.154	0.090	0.124	0.188	0.148	0.187	0.158	0.159	0.097	0.127	0.197

All simulations have been performed in R 2.2.0 with a standard Pentium-IV computer (3 GHz processor and 1 GB RAM). It should be noted here that for the lack of sufficient computer memory we had to limit our experiments to a maximum cluster size of 500 observations. From the results of our Monte-Carlo experiments (see subsections 3.1 and 3.2) it seems unproblematic to extrapolate the results to $n_{kt} > 500$. Since the Monte-Carlo experiments are computatinally expensive, especially for the PQL approach, we have conducted 200 replications for each setting in the Monte-Carlo simulation.

The Monte-Carlo experiments are mainly conducted to address two issues. The first objective is to compare the speed of convergence of the $\boldsymbol{\beta}$ parameter estimates in the fixed effect approach (when cluster size, n_{kt} increases) with that in the mixed effect approach obtained through PQL[5] method. The second objective is to study the relative bias in the estimate of the variance parameters, $\mathbf{D}$, when increasing the cluster size (n_{kt}). We use two different measures of the preciseness of the parameter estimates namely, mean squared error (MSE) for β and relative absolute error (RAE) for $\mathbf{D}$ respectively. The first measures the average squared distance of the estimates from the true parameter value over the 200 replications. The second measures the ratio of the total absolute bias in all variance parameter estimates to the total of the true parameter values. If d_i is the true value of a parameter in $\mathbf{D}$ and $\widehat{d_i}$ the corresponding estimate, then the relative absolute error is given by

$$RAE = \frac{\sum_{i=1}^{\frac{K(K+1)}{2}} |\widehat{d_i} - d_i|}{\sum_{i=1}^{\frac{K(K+1)}{2}} |d_i|}. \tag{6}$$

[5]For the practical implementation we have used the glmmPQL() function of the R-library, MASS. Readers are referred to Venables and Ripley [25] for details about this function.

Table 2: MSE of the $\boldsymbol{\beta}$ parameter estimates in the fixed effect (FE) and PQL approaches for the probit model for different combination of cluster sizes

n_{KT}	Cluster size	Covariance structure of the random effect					
		Diagonal		Compound symmetry		Unstructured	
		FE	PQL	FE	PQL	FE	PQL
10	K=3, T=10	20e-3	10e-3	21.3e-3	10.4e-3	15.8e-3	7.5e-3
	K=7, T=10	10.7e-3	4.2e-3	10.7e-3	4.2e-3	8.2e-3	3.3e-3
	K=3, T=20	9.9e-3	5.1e-3	11.0e-3	4.1e-3	9.3e-3	4.0e-3
	K=7, T=20	6.3e-3	2.0e-3	7.4e-3	2.2e-3	5.8e-3	1.7e-3
50	K=3, T=10	18.8e-4	16.1e-4	19.3e-4	15.9e-4	17.5e-4	15.9e-4
	K=7, T=10	8.4e-4	6.8e-4	11.0e-4	9.2e-4	7.9e-4	5.9e-4
	K=3, T=20	11.5e-4	9.4e-3	9.3e-4	8.0e-4	8.9e-4	6.6e-4
	K=7, T=20	5.4e-4	4.7e-4	5.4e-4	3.9e-4	4.1e-4	3.2e-4
	K=3, T=40	7.2e-4	-	7.7e-4	-	4.0e-4	-
	K=7, T=40	3.2e-4	-	3.0e-4	-	4.0e-4	-
200	K=3, T=10	45.4e-5	44.4e-5	49.5e-5	48.4e-5	36.3e-5	35.3e-5
	K=7, T=10	21.6e-5	20.0e-5	22.3e-5	21.0e-5	17.9e-5	16.3e-5
	K=3, T=20	24.6e-5	24.1e-5	20.9e-5	20.6e-5	17.4e-5	16.5e-5
	K=7, T=20	8.8e-5	9.0e-5	9.5e-5	8.6e-5	8.2e-5	7.7e-5
500	K=3, T=10	18.6e-5		16.3e-5		15.2e-5	
	K=7, T=10	7.47e-5		8.5e-5		6.5e-5	
	K=3, T=20	9.69e-5		8.1e-5		9.1e-5	
	K=7, T=20	3.8e-5		3.9e-5		3.3e-5	

3.1 The probit mixed model

For the probit mixed model our data generating process was as follows: $\eta = x\beta + \mathbf{u}$ and $y = \mathbf{1}[(\boldsymbol{\eta} + \boldsymbol{\varepsilon}) > 0]$ where, β and $\mathbf{u}$ are as discussed above, $x \backsim N(0,1)$ and $\varepsilon \backsim N(0,1)$.

The results from the Monte-Carlo experiments in Table 2 reveal that both the fixed and PQL approaches estimate the $\boldsymbol{\beta}$ parameter with a reasonable precision. Although the fixed effect approach is less efficient with a small cluster size, this difference becomes negligible as the sample size increases. Table 2 also shows that irrespective of the covariance structure, both methods can estimate the $\boldsymbol{\beta}$ parameter consistently. Maddala [18] mentions (without any empirical evidence) that the pooled probit model gives a consistent (but inefficient) estimator of $\boldsymbol{\beta}$ even if there is serial correlation in the errors due to random effects. So, these results need not be surprising. Since the computation of the mixed effect model is time consuming when the cluster size increases, we did not compare the two approaches at cluster size, $n_{kt} = 500$. However, the decreasing trend of the MSE in the fixed effect approach indicates better performance of the fixed effect approach as the cluster size (n_{kt}) increases. It seems the estimates of the β parameters in the two approaches become qualitatively equal somewhere between the cluster size (n_{kt}) 50 and 200 observations.

Table 3 shows a comparison of the relative absolute errors in the covariance parameter, $\mathbf{D}$, estimates between the two methods. For the FE approach the bias in the estimates of the covariance parameters is remarkably large until the cluster size equals 200, whereas the PQL

Table 3: RAE of the covariance parameter estimates (**D**) in the fixed effect (FE) and PQL approaches for the probit model for different combination of cluster sizes

n_{KT}	Cluster size	Covariance structure of the random effect					
		Diagonal		Compound symmetry		Unstructured	
		FE	PQL	FE	PQL	FE	PQL
10	K=3, T=10	6.68	0.57	4.85	0.47	3.90	0.59
	K=7, T=10	9.10	0.93	9.10	0.93	3.02	0.63
	K=3, T=20	6.30	0.43	4.78	0.35	3.66	0.54
	K=7, T=20	8.26	0.69	4.01	0.40	2.75	0.57
50	K=3, T=10	1.34	0.55	1.06	0.40	0.55	0.48
	K=7, T=10	1.91	0.98	0.96	0.45	0.57	0.45
	K=3, T=20	1.19	0.40	0.86	0.29	0.39	0.33
	K=7, T=20	1.65	0.71	0.89	0.33	0.42	0.37
	K=3, T=40	1.03	-	0.73	-	0.42	-
	K=7, T=40	1.39	-	0.70	-	0.42	-
200	K=3, T=10	0.75	0.58	0.54	0.42	0.43	0.42
	K=7, T=10	1.27	1.05	0.55	0.45	0.49	0.47
	K=3, T=20	0.55	0.41	0.41	0.32	0.31	0.31
	K=7, T=20	0.96	0.76	0.42	0.32	0.33	0.31
500	K=3, T=10	0.63		0.46		0.41	
	K=7, T=10	1.16		0.49		0.45	
	K=3, T=20	0.52		0.32		0.29	
	K=7, T=20	0.85		0.36		0.34	

approach provides comparatively good estimate, in terms of RAE, for small cluster sizes. From equation (6) we see that RAE will approach zero as the bias goes to zero. The bias might not be a matter of great worry as long as it is reasonably small (and tends to zero with the increase of T, which means the estimates are consistent). Table 2 shows that none of the methods give us an RAE close to zero. However, the speed with RAE decreases in the fixed effect approach is impressive. Although we could not conduct any simulation for n_{kt} greater than 500, the results suggest that the variance parameter estimates will approximate the true value when the number of clusters and cluster size are both quite large (i.e. $n_{kt} \longrightarrow \infty, K \times T \longrightarrow \infty$). Note also that RAE falls when T is increased.

The true covariance matrices namely, the diagonal, the compound symmetric and the unstructured, that we used in the Monte-Carlo experiments are shown in the left hand sides of the Tables 1.1-1.3 while their corresponding Monte-Carlo mean estimates are given in the right hand sides. From those tables it is evident that, on an average, the estimate of the covariance parameters in the fixed effect approach are close to their true values.

3.2 The Poisson mixed model

The results for the probit model are promising but we would like to know if the fixed effect approach also works in other generalized linear models. In particular, it would be useful to see the performance of the fixed effect approach in at least one another setting. Therefore we will

Table 4: MSE of the **β** parameter estimates in the fixed effect (FE) and PQL approaches for the Poisson mixed model

n_{KT}	Cluster size	Covariance structure of the random effect					
		Diagonal		Compound symmetry		Unstructured	
		FE	PQL	FE	PQL	FE	PQL
10	K=,3 T=10	2.13e-3 (200)	2.07e-3 (193)	2.21e-3 (200)	2.31e-3 (196)	3.21-3 (200)	2.86e-3 (200)
	K=7, T=20	0.44e-3 (200)	0.44e-3 (178)	0.44e-3 (200)	0.42e-3 (172)	0.67e-3 (200)	0.67e-3 (200)
50	K=3, T=10	3.90e-4 (200)	3.94e-4 (197)	5.56e-4 (200)	5.68e-4 (198)	5.90e-4 (200)	5.94e-4 (200)
	K=7, T=20	7.69e-5 (200)	7.96e-5 (167)	8.26e-5 (200)	8.52e-5 (177)	11.13e-5 (200)	11.46e-5 (200)
200	K=3, T=10	9.16e-5 (200)	9.17e-5 (196)	10.33e-5 (200)	10.58e-5 (195)	12.66e-5 (200)	12.52e-5 (200)
	K=7, T=20	2.01e-5 (200)	1.99e-5 (173)	1.77e-5 (200)	1.70e-5 (178)	2.69e-5 (200)	2.73e-5 (200)
500	K=3, T=10	3.42e-5 (200)		3.82e-5 (200)		6.46e-5 (200)	
	K=7, T=20	7.19e-6 (200)		8.66e-6 (200)		10.88e-6 (200)	

Note: Values within parenthesis show the number of iteration used for the calculation of the MSE. Any deviation of these values from 200 indicates the number of times the respective iteration failed to converge during the Monte-Carlo simulation.

conduct similar experiments for a Poisson model.

For the Poisson model we have generated random data according to the following protocol $\eta = x\beta + u$ and $y \backsim Poisson(e^{\eta})$ *i.e.* y's are drawn randomly from a Poisson distribution with $E(y) = e^{\eta}$ when x, β and $\mathbf{u}$ are the same variables and parameters as the the probit model. This implies the link function, $g(\mu) = \log(\mu)$, which is the canonical link for Poisson distribution.

Since the PQL approach was not always converging for the Poisson model, we pay some attention to its convergence properties. Tables 4 and 5 show, along with other results, the number of time that the PQL estimation of the Poisson mixed model succeeded within 200 replications. In the probit model case, the default setting for the maximum number of iterations (niter= 10) in glmmPQL() function was sufficient for the procedure to converge. However, for the Poisson mixed model estimation, the procedure was not always converging within the default setting for "niter". Therefore, we increased the maximum number of iteration, niter, for PQL estimation from 10 to 20 for the Poisson mixed models. Any specific estimation that reached the maximum number of iteration was considered as not converged. And these replicates were not included in the calculation of the MSE and the RAE. Moreover, only the lowest $(K = 3, T = 10)$ and the highest $(K = 7, T = 20)$ settings have been studied. From the results for the probit models in Tables 2 and 3 we see that the MSE and the RAE of the other settings come somewhere in between of these two situations, therefore we considered it reasonable to study only these two situations in order to get an idea about the relative accuracy of the model parameter estimates of the fixed and mixed effect approaches.

Table 5: RAE of the covariance parameter estimates (**D**) in the fixed effect (FE) and the PQL approaches for the Poisson mixed model

n_{KT}	Cluster size	Covariance structure of the random effect					
		Diagonal		Compound symmetry		Unstructured	
		FE	PQL	FE	PQL	FE	PQL
10	K=3, T=10	6.59	0.56	4.19	0.41	2.34	0.50
		(200)	(193)	(200)	(196)	(200)	(200)
	K=7, T=20	7.38	0.71	3.26	0.36	1.62	0.42
		(200)	(178)	(200)	(172)	(200)	(200)
50	K=3, T=10	0.64	0.56	0.48	0.41	0.42	0.38
		(200)	(197)	(200)	(198)	(200)	(200)
	K=7, T=20	0.94	0.73	0.41	0.33	0.35	0.33
		(200)	(167)	(200)	(177)	(200)	(200)
200	K=3, T=10	0.63	0.61	0.41	0.41	0.42	0.41
		(200)	(196)	(200)	(195)	(200)	(200)
	K=7, T=20	0.80	0.76	0.33	0.32	0.31	0.31
		(200)	(173)	(200)	(178)	(200)	(200)
500	K=3, T=10	0.59		0.43		0.41	
		(200)		(200)		(200)	
	K=7, T=20	0.80		0.34		0.32	
		(200)		(200)		(200)	

Note: Values within parenthesis show the number of iteration used for the calculation of the RAE. Any deviation of these values from 200 indicates the number of times the respective iteration failed to converge during the Monte-Carlo simulation.

The results from the Monte-Carlo studies are shown in Table 4 and 5. One can see that both the fixed and the PQL estimates, ignoring the problem of non-convergence, are more efficient in terms of MSE and RAE in the Poisson mixed models than in the probit mixed model. This result is not surprising since the Poisson response variable contains more information than a binary response variable. Comparing the results in Table 2 and 4 it is interesting to notice that the degree of improvement in performance, in terms of lower MSE, in fixed effect approach for the Poisson models is much higher than in the PQL approach. Table 4 also shows that both the fixed and the PQL approaches have almost the same level of accuracy, in terms of MSE's, for the Poisson model.

Table 5 shows that the mixed effect approach can estimate the variance components more precisely than the fixed effect approach when the cluster size is small ($n_{kt} = 10$). However, the differences between the RAE's of the two approaches tends to disappear as the cluster size (n_{kt}), number of clusters (K) and the number repeated measures (T) increase. This tells us that the estimates of the two approaches coincide qualitatively somewhere between the cluster sizes 50 and 200. Furthermore, the differences in $RAE's$ between the two approaches come closer in the Poisson mixed model cases than for the probit models. In other words, as the cluster size (n_{kt}) increases, the fixed effect approach approximates the PQL approach more quickly for the Poisson model than the probit model.

To get an idea about the relative efficiency of the covariance parameter estimates between the two approaches, we have calculated the standard deviation of the RAE within 200 Monte-Carlo

Table 6: Variation in RAE (within 200 simulations) for probit model with K=3 and T=20

n_{kt}	Approach	Covariance structure of the random effect					
		Diagonal		Compound Symmetric		Unstructured	
		Mean RAE	SD	Mean RAE	SD	Mean RAE	SD
50	FE	1.26	0.85	0.87	0.67	0.50	0.24
	PQL	0.41	0.12	0.29	0.12	0.44	0.19
200	FE	0.54	0.29	0.35	0.22	0.29	0.15
	PQL	0.42	0.13	0.27	0.11	0.30	0.14

iterations for a Poisson mixed model with $n_{kt} = 50$ and 200, $K = 3$ and $T = 20$. Since we got the impression form the results in Tables 2-5 that the fixed effect and the PQL approach produce almost the same results somewhere between the cluster size 50 and 200, we have concentrated our investigation on the efficient issue within this limit. Moreover, it is evident from our results in Tables 2-5 that a large number of clusters (K) and a more time points (T) also influence the better performance of the estimator, we investigated the efficiency at reasonably small number of cluster ($K = 3$) to be able to claim that the efficiency of the estimators will not be worse than what we have showed when both K and T increases. The results presented in the Table 6 show that the fixed effect approach is relatively inefficient for small cluster size ($n_{kt} = 50$). Column 4, 6 and 8 in Table 6 also show that the difference in the standard deviation of the RAE between the two approaches decreases as the cluster size increases. However, for cluster size, $n_{kt} = 200$, the standard deviations of the RAE in the fixed effect approach, in some cases, twice those of the mixed effect approach. For example, at $n_{kt} = 200$ with the diagonal covariance matrix of the random effect, the standard deviation of the RAE for the mixed effect approach is 0.13 which is only 44.83% of that of the fixed effect approach (0.29) at the same setting.

These results lead us to make two conclusions about the accuracy of the estimates in the fixed and the PQL approaches. First, for large cluster sizes, large number of clusters and lengthy time increases both the fixed and mixed effect approaches have a similar degree of accuracy in estimation of the parameter. Both of them are more efficient in estimating the β parameter than the covariance matrix. In the case of the fixed effect approach, reduced efficiency of the model parameter estimate may be a concern as shown in Table 6. However, for sufficiently large data *i.e.* $n_{kt} > 200$, $K \geq 7$ and $T \geq 20$ we can ignore this problem.

Second, the performance of the models varies with the specific response variable. In particular, both approaches perform better in the Poisson model than in the probit model. Greene [11] studied the performance of mixed effect model in different settings[6] and found that the discrete response mixed models' behavior differs significantly from that of the continuous response models. Now our simulation results show that within the sub-group of the discrete response mixed models difference in efficiency occurs for example, probit model does not behave better than the

[6] Though those experiments did not include the poisson model

Poisson model.

Thirdly, with large data sets, either with a large number of cluster or large sample size, the PQL approach is slow and it frequently fails to converge. For example, for a probit model with $K = 7$, $T = 20$, $n_{kt} = 500$ and an unstructured $\mathbf{D}$ matrix, the fixed effect approach requires, on an average, 29 seconds to estimate the model whereas the PQL approach takes 51 seconds in the same situation. Moreover, in some particular setting the rate of failure to converge for the PQL approach can be as high as 17% cases. This favors the fixed effect approach to be preferred in the large sample cases.

We have also investigated the reasons of the non-convergence and we have found that on an average, in nine out of ten cases, the non-convergence is due to the fact that the PQL algorithm did not converge within a prefixed number of iteration (with is 20 in this case). However, we did not dig much into the reason why the PQL estimate did not converge in the remaining one out of ten cases.

4 Inference on D and related topics

In Section 2.2 we discussed a procedure to estimate the covariance matrix of the random effect. In section 3 we showed the performance for a range of specifications. For simplicity, we assumed a zero intercept. In practice, models with no intercept are of less interest. Therefore, we would like to show how the fixed effect approach can be extended to deal with the situation when there is an intercept term in the model. A second issue we discuss below is statistical hypothesis testing of the structure of the covariance matrix, $\mathbf{D}$.

4.1 Estimation of D with a fixed effect model having an intercept term

When the model in (2) does not contain any intercept term, using the fixed effect approach, the estimates of cluster and time interaction terms provides the realization of the random effects directly. These terms serve as the elements in matrix, $\widehat{\mathbf{U}}$, which we use to estimate the covariance matrix, $\mathbf{D}$.The estimation of the covariance matrix $\mathbf{D}$ is less straightforward for a model with an intercept term. This is because the time and cluster interactions effect, which we consider as the realizations of the random effect, altogether with the intercept term make the model unidentifiable. One remedy of this problem is to estimate one parameter less than the actual number of parameters in the interaction terms. However, the estimates of the intercept and the interaction parameters we get through this techniques are not directly the estimate of the true parameters but an estimable function of those parameters. To distinguish the estimates of the interaction terms in the models with intercept from the realizations of $\mathbf{u}_k = (u_{1k}, u_{2k}, ..u_{Tk})'$ we use the superscript ("*") *e.g.* $\widehat{u}^*_{kt}$ to indicates an estimate of ktth interaction term for a model with an intercept. If we omit u_{11} from the model, the estimate $\widehat{u}^*_{kt}$ can be expressed as

$\widehat{u^*_{kt}} = \widehat{(u_{kt} - u_{11})}$. By a similar reasoning the estimate of the intercept term ($\widehat{\alpha}^*$, say) represents the original intercept plus the term we left out from the model *i.e.* $\widehat{\alpha + u_{11}}$. Now adding the intercept terms with all interaction parameter estimates we have

$$\widehat{\alpha}^* + \widehat{u}^*_{kt} = \widehat{\alpha + u_{11}} + \widehat{(u_{kt} - u_{11})}$$

$$\Longrightarrow \widehat{\alpha}^* + \widehat{u}^*_{kt} = \widehat{\alpha + u_{kt}}$$

Since the interactions (u_{kt}) have a zero mean, the average of all interaction estimates ($\widehat{u}^*_{kt}$) after adding intercept term eventually gives us the estimate of the original intercept term. Then it is not difficult to get the estimate of the other parameter (u_{kt}). The calculations are as follows

$$\frac{\sum_k \sum_t (\widehat{\alpha}^* + (\widehat{\alpha}^* + \widehat{u}^*_{kt}))}{KT} = \widehat{\alpha} \tag{7}$$

$$\widehat{\alpha}^* - \widehat{\alpha} = \widehat{u}_{11} \tag{8}$$

$$\widehat{u}^*_{kt} - \widehat{\alpha} = \widehat{u}_{kt} \tag{9}$$

These estimates of the u_{kt}'s can be used as the elements of $\widehat{\mathbf{U}}$ to estimate the covariance matrix, $\mathbf{D}$.

4.2 Inference on the structure of D

Sometimes it is of interest to test whether the $\mathbf{D}$ matrix has a specific structure *e.g.* whether it is diagonal which is equivalent to say whether the random effects are independent or, in a genetic application, one might be interested to test whether a genetic relation matrix can precisely express the correlation structure of the random effects, or in credit risk modeling, whether the inter-industry input-output matrix can explain the correlation structure of the random effect. Here we offer a detailed discussion on the test of statistical significance of the $\mathbf{D}$ matrix for a certain structure.

From equation (5) we have

$$\widehat{\mathbf{D}} = \frac{1}{T}\widehat{\mathbf{U}}'\widehat{\mathbf{U}}$$

$$\Rightarrow \widehat{\mathbf{D}} = \frac{1}{T}(\widehat{\mathbf{u}}'_1, \widehat{\mathbf{u}}'_2, ..., \widehat{\mathbf{u}}'_T)(\widehat{\mathbf{u}}'_1, \widehat{\mathbf{u}}'_2, ..., \widehat{\mathbf{u}}'_T)'$$

$$\Longrightarrow \begin{pmatrix} \widehat{d}_{11} & \widehat{d}_{12} & \cdots & \widehat{d}_{1K} \\ \widehat{d}_{12} & \widehat{d}_{22} & & \vdots \\ \vdots & & \ddots & \widehat{d}_{(K-1)K} \\ \widehat{d}_{1K} & \cdots & \widehat{d}_{(K-1)K} & \widehat{d}_{KK} \end{pmatrix} = \frac{1}{T} \begin{pmatrix} \sum_{t=1}^{T} u_{1t}^2 & \sum_{t=1}^{T} \widehat{u}_{1t}\widehat{u}_{2t} & \cdots & \sum_{t=1}^{T} \widehat{u}_{1t}\widehat{u}_{Kt} \\ \sum_{t=1}^{T} \widehat{u}_{2t}\widehat{u}_{1t} & \sum_{t=2}^{T} u_{2t}^2 & & \sum_{t=1}^{T} \widehat{u}_{2t}\widehat{u}_{Kt} \\ \vdots & & \ddots & \vdots \\ \sum_{t=1}^{T} \widehat{u}_{Kt}\widehat{u}_{1t} & \cdots & \sum_{t=1}^{T} \widehat{u}_{Kt}\widehat{u}_{(K-1)t} & \sum_{t=1}^{T} \widehat{u}_{Kt}^2 \end{pmatrix}$$

Assume we want to test the following hypothesis $H_0: \ \mathbf{D} = \sigma^2\mathbf{I}$ vs. $H_1: \ \mathbf{D} \neq \sigma^2\mathbf{I}$ where, $\mathbf{I}$ is a $k \times k$ identity matrix and σ^2 is a known scalar. Then the above test can be performed by using the Wald-test[7] for testing non-linear constraints on the parameters. In our case, we have $\frac{K\times(K+1)}{2}$ non-linear restrictions on the parameters as presented in the lower (or upper) triangular elements of the symmetric matrix $\frac{1}{T}\widehat{\mathbf{U}}\widehat{\mathbf{U}}$. This test does not require any additional estimate than obtained through a general estimation procedure of the model. The detailed calculations for the Wald-test are as follows.

The Wald-test statistic for testing non-linear restriction can be written as

$$W = [c(\theta) - q]'(Asy.Var[c(\theta) - q])^{-1}[c(\theta) - q]$$

where $c(\theta)$ is the vector of non-linear restrictions (in our case the lower triangular elements in $\frac{1}{T}\widehat{\mathbf{U}}\widehat{\mathbf{U}}$) of length $\frac{K\times(K+1)}{2}$, q is a vector of constants (in our case the lower triangular elements of $\boldsymbol{\sigma}^2\mathbf{I}$) having the same length as $c(\theta)$. And, $c(\theta) - q = 0$ represents the formal equation of restrictions. In this particular case the system of equations for the non-linear restrictions on the parameters can be presented as

$$\begin{pmatrix} \sum_{t=1}^{T} u_{1t}^2 \\ \sum_{t=1}^{T} \widehat{u}_{2t}\widehat{u}_{1t} \\ \vdots \\ \sum_{t=1}^{T} \widehat{u}_{Kt}^2 \end{pmatrix} - \begin{pmatrix} \sigma^2 \\ 0 \\ \vdots \\ \sigma^2 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{pmatrix} \tag{10}$$

Under the null hypothesis, in large sample, W has a Chi-square distribution with degrees of freedom equal to the number of equations in $c(\theta) - q = 0$ (*i.e.* $\frac{K\times(K+1)}{2}$). On the other hand, the variance of $[c(\theta) - q]$ can be given as

[7] Greene [10] provides a general discussion on the computational procedure for the Wald-test for non-linear restrictions on the parameters.

$$Est.\quad Asy.\quad Var[c(\widehat{\theta})-q] \;=\; \widehat{\mathbf{C}}\quad Est.\quad Asy.\quad Var[\widehat{\theta}]\quad \widehat{\mathbf{C}}'$$

where, $\widehat{\mathbf{C}}=\left[\frac{\delta c(\widehat{\theta})}{\delta\widehat{\theta}'}\right]$ is a matrix of order $\frac{K\times(K+1)}{2}\times(K\times T)$ whose mth row is the derivatives of the mth constraint with respect to nth element in θ.

Again we do not obtain the estimate of $V(\widehat{\mathbf{U}})$ directly when there is an intercept term in the model. However, we do obtain an estimate of $V(\widehat{u}^*_{kt})$. Then the calculation for the estimation of $V(\widehat{U})$ is straight forward using equations (7), (8) and (9) and $V(\widehat{u}^*_{kt})$. We have from (7)

$$\widehat{\alpha}=\frac{\sum_k\sum_t(\widehat{\alpha}^*+(\widehat{\alpha}^*+\widehat{u}^*_{kt}))}{KT}$$

$$\Rightarrow V(\widehat{\alpha})=V\left(2\widehat{\alpha}^*+\frac{\sum_i\sum_j\widehat{u}^*_{kt}}{KT}\right)$$

$$\Rightarrow V(\widehat{\alpha})=4V(\widehat{\alpha}^*)+V\left(\frac{\sum_k\sum_t\widehat{u}^*_{kt}}{KT}\right)+2Cov\left(2\widehat{\alpha}^*,\frac{\sum_k\sum_t\widehat{u}^*_{kt}}{KT}\right)$$

$$\Rightarrow V(\widehat{\alpha})=4V(\widehat{\alpha}^*)+\frac{1}{K^2T^2}\left[\sum_k\sum_t V(\widehat{u}^*_{kt})+\sum_k\sum_t\sum_{(k'\neq k\,or\;t'\neq t)}\sum Cov(\widehat{u}^*_{kt}.\widehat{u}^*_{k't'})\right]+\frac{4}{KT}\sum_k\sum_t Cov(\widehat{\alpha}^*,\widehat{u}^*_{kt}) \tag{11}$$

From (8) we have

$$V(\widehat{u}_{11})=V(\widehat{\alpha}^*-\widehat{\alpha})$$

$$\Rightarrow V(\widehat{u}_{11})=V(\widehat{\alpha}^*)+V(\widehat{\alpha})-2Cov\left(\widehat{\alpha}^*,2\widehat{\alpha}^*+\frac{\sum_k\sum_t\widehat{u}^*_{kt}}{KT}\right)$$

$$\Rightarrow V(\widehat{u}_{11})=V(\widehat{\alpha})-V(\widehat{\alpha}^*)-\frac{2}{KT}\sum_k\sum_t Cov(\widehat{\alpha}^*,\widehat{u}^*_{kt}) \tag{12}$$

On the other hand, from (9) we have

$$V(u_{kt})=V(\widehat{u}^*_{kt}-\widehat{\alpha})$$

$$\Rightarrow V(\widehat{u}_{kt})=V(\widehat{u}^*_{kt})+V(\widehat{\alpha})-2Cov\left(\widehat{u}^*_{kt},2\widehat{\alpha}^*+\frac{\sum_k\sum_t\widehat{u}^*_{kt}}{KT}\right)$$

$$\Rightarrow V(\widehat{u}_{kt})=\frac{KT-2}{KT}V(\widehat{u}^*_{kt})+V(\widehat{\alpha})-4V(\widehat{u}^*_{kt},2\widehat{\alpha}^*)-\frac{2}{KT}\sum_{l\neq k}\sum_{m\neq t}Cov(\widehat{u}^*_{kt},\widehat{u}^*_{lm}) \tag{13}$$

Therefore, all the variance terms can be calculated through the equations (11)-(13) where everything we need for those calculations are obtained form the estimate of the variance-covariance matrix of the model parameters. Again, the calculations for the covariance terms are as follows

$$Cov(\widehat{u}_{11}, \widehat{u}_{kt}) = Cov\left[(\widehat{\alpha}^* - \widehat{\alpha}), (\widehat{u}^*_{kt} - \widehat{\alpha})\right]$$

$$\Rightarrow Cov(\widehat{u}_{11}, \widehat{u}_{kt}) = V(\widehat{\alpha}) + Cov(\widehat{\alpha}^*, \widehat{u}^*_{kt}) - Cov(\widehat{u}^*_{kt}, \widehat{\alpha}) - Cov(\widehat{\alpha}^*, \widehat{\alpha}) \tag{14}$$

Furthermore,

$$Cov(\widehat{u}_{kt}, \widehat{u}_{k't'}) = V(\widehat{\alpha}) + Cov(\widehat{u}^*_{kt}, \widehat{u}^*_{k't'}) - Cov(\widehat{u}^*_{kt}, \widehat{\alpha}) - Cov(\widehat{u}^*_{k't'}, \widehat{\alpha}) \tag{15}$$

All the terms in the equations (14) and (15) can either be obtained directly from the formal model estimation procedure or from the calculation of the above equations, (11)-(13). Therefore any further simplification of these two equations is not shown here.

So, using the same technique we can test for any other covariance structures provided that the null hypothesis is simple *i.e.* all the elements in $\mathbf{D}$ matrix must be completely specified under the null hypothesis. For example, if we want to test the hypothesis $H_0: \ \mathbf{D} = \mathbf{D}^0$ vs. $H_1: \ \mathbf{D} \neq \mathbf{D}^0$ where $\mathbf{D}^0$ is any known symmetric and positive definite matrix, we just have to replace the second term in the left hand side of the equation (10) with the upper (or lower) triangular elements of $\mathbf{D}^0$. Note, however, that the empirical evidence on the performance of the Wald-test in this particular situation is yet to be explored.

5 Conclusion

The fixed effect approach provides us with a consistence estimate of both the $\boldsymbol{\beta}$ and the covariance parameters even when there is a complex structure of correlation in the dependent variable due to the random effect. The efficiency of the proposed estimator may be a concern for the small sample cases but it matters little in the case of large sample size (greater than 200 observation per cluster). The PQL method becomes more and more infeasible comparing the estimation time with the fixed effect approach as sample size increases. In addition to the longer estimation time, in some cases, the PQL algorithm does not converge within a prefixed number of iterations when the fixed effect approach does. Moreover, we can perform hypothesis test on the structure of the covariance matrix through the fixed effect approach. Therefore, considering all these issues we favour the fixed effect model to be preferred in the large sample cases with the repeated measures on the individuals when the observations are correlated both within and between individuals.

Throughout this paper we have considered the $\mathbf{U}$ matrix consisting of T independent realizations from a K variate normal distribution. A situation might arise where the components

u_{kt} are exposed to autocorrelation over time, T. An appropriate technique to deal with such a situation when using the fixed effect approach is not yet available.

References

[1] Alam, M. (2005), Is it feasible to use MCMC to estimate GLMM with a complex correlation structure in an R-enviroment?, Master's Thesis, Dalarna University, Borlange.

[2] Booth, J. G. and Hobert, J. P. (1999), Maximizing generalized linear mixed model likelihoods with an automated Monte Carlo EM algorithm, *Journal of the Royal Statistical Society (B)* 61(1), pp. 265-285.

[3] Breslow, N. E. and Clayton, D. G. (1993), Approximate inference in generalized linear mixed models, *Journal of the American Statistical Association* 88(421), pp. 9-25.

[4] Butler, J. S. and Moffitt, R. (1982), A computational efficient quadrature procedure for the one-factor multinomial probit model, *Econometrica* 50(3), pp. 761-764.

[5] Carling, K., Rönnegård, L. and Roszbach, K. (2004), Is firm interdependence within industries important for portfolio credit risk?, Sverige Riksbank Working Paper Series No. 168.

[6] Chen, Y. H. (2006), Computationally efficient monte carlo EM algorithm for generalized linear mixed model, *Journal of Statistical Computation and Simulation* 76(9), pp. 817-828.

[7] Clayton, D. G. (1996), Generalized linear mixed models in Gilks, W. R., Richardson S. and Spiegelhalter, D. J. (eds.), *Markov Chain Monte Carlo in Practice*, Chapman and Hall, London, pp. 275–301.

[8] Engel, B. (1998), A simple illustration of the failure of the PQL, IRREML and APHL as approximate ML methods for mixed models for binary data, *Biometrical Journal* 40(2), pp. 141-154.

[9] Greene, W. H. (1991), Limdep version 6, Econometric Software, New York.

[10] Greene, W. H. (2003), *Econometric Analysis*, Pearson Education, Inc., Upper Saddle River, New Jersey.

[11] Greene, W. H. (2004), The behaviour of the maximum likelihood estimator of limited dependent variable models in the presence of fixed effects, *The Econometrics Journal* 7(1), pp. 98-127.

[12] Guilkey, D. K. and Murphy, J. L. (1993), Estimation and testing in the random effects probit model, *Journal of Econometrics* 59, pp.301-317.

[13] Heckman, J. J. and Willis, R. (1976), Estimation of a stochastic model of reproduction: An econometric approach, in: N. Terleckyj, ed. Household production and consumption, National Bureau of Economic Research, New York.

[14] Lee, Y. and Nelder, J. A. (1996) Hierarchical generalized linear models, *Journal of the Royal Statistical Society (B)* 58(4), pp. 619-678.

[15] Lee, Y. and Nelder J. A. (2006), Double hierarchical generalize linear models, *Journal of the Royal Statistical Society (C)* 55(2), pp. 1-29.

[16] Levine, R. A. and Fan, J. (2006), An automated (Markov chain) Monte Carlo EM algorithm, *Journal of Statistical Computation and Simulation* 74(5), pp. 349-360.

[17] Littell, R. C., Milliken, G. A., Stroup, W. W. and Wolfinger R. D. (1996), The SAS system for mixed models, SAS Inst. Inc., Cary, North Carolina.

[18] Maddala, G. S. (1987), Limited dependent variable models using panel data, *Journal of Human Resources* 22(3), pp. 307-338.

[19] McCullagh, P. and Nelder, J. A. (1989), *Generalized Linear Models*, London: Chapman and Hall.

[20] McCulloch, C. E. (1994), Maximum likelihood variance component estimation for binary data, *Journal of the American Statistical Association* 89(425), pp. 330-335.

[21] McCulloch, C. E. (1997), Maximum likelihood algorithms for generalized linear mixed models, *Journal of the American Statistical Association* 92(437), 162-170.

[22] McCulloch, C. E. and Searle, S. R. (2001), *Generalized Linear and Mixed Models*, John Wiley & Sons, Inc. New York.

[23] Quintana, F. A., Liu, J. S. and del Pino, G. E. (1999), Monte Carlo EM with importance reweighting and its applications in random effects models, *Computational Statistics and Data Analysis* 29 (1999), pp. 429-444.

[24] Song, J. X. and Ahn, C. W. (2003), An evaluation of methods for the stratified analysis of clustered binary data in community intervention trials, *Statistics in Medicine* 22(2), pp. 205-226.

[25] Venables, W. N., and Ripley, B. D. (2002), *Modern Applied Statistics with S*, Springer, New York.

[26] Wolfinger, R. and O'Connell, M. (1993), Generalized linear mixed models: a pseudo-likelihood approach. *Journal of Statistical Computation and Simulation* 48, pp. 233-243.

[27] Yun, S. and Lee, Y. (2004), Comparison of hierarchical and marginal likelihood estimators for binary outcomes, *Computational Statistics and Data Analysis* 45 (2004), pp. 639-650.

[28] Zhou, X.-H., Perkins, A. J. and Hui, S. L. (1999), Comparisons of software packages for generalized linear multilevel models, *The American Statistician* 53(3), pp. 282-290.