Comparisons of Estimation Procedures for Nonlinear Multilevel Models

Ali Reza Fotouhi, University of Mashhad, Mashhad, Iran.

Postal Address: P. O. Box 91775-1133, Mashhad, Iran.

Email Address: alirezafotouhi@yahoo.com

Abstract

We introduce General Multilevel Models and discuss the estimation procedures that may be used to fit multilevel models. We apply the proposed procedures to three-level binary data generated in a simulation study. We compare the procedures by two criteria, Bias and efficiency. We find that the estimates of the fixed effects and variance components are substantially and significantly biased using Longford's Approximation and Goldstein's Generalized Least Squares approaches by two software packages VARCL and ML3. These estimates are not significantly biased and are very close to real values when we use Markov Chain Monte Carlo (MCMC) using Gibbs sampling or Nonparametric Maximum Likelihood (NPML) approach. The Gaussian Quadrature (GQ) approach, even with small number of mass points results in consistent estimates but computationally problematic. We conclude that the MCMC and the NPML approaches are the recommended procedures to fit multilevel models.

Keywords: Multilevel Models, Simulation, Longford's Approximation, Goldstein's Generalized Least Squares, Nonparametric Maximum Likelihood, Gaussian Quadrature, Markov Chain Monte Carlo, Gibbs Sampling

1 Introduction

Many kinds of observational and design data are such that the observations are clustered within groups and the groups in terms are nested in upper groups and form multilevel data. Two types of multilevel data are longitudinal and multicenter clinical trial data. In longitudinal studies, we investigate changes over time in characteristics which is measured repeatedly for each individual. In medical studies, the measurements may be the number of epileptic seizures which are recorded for each patient over time (Thall and Vail, 1990). In social sciences, the measurements may be residential states which are observed for each individual over time (Clark et al., 1979). Longitudinal data are examples of two level data and the observations, in first level for each individual, in second level, are correlated. In a multicenter clinical study, subjects within a given site are prospectively studied over time. An example of this type of data is the data from the Multicenter AIDS Cohort Study or MACS (Kaslow et al., 1987). Here data exists at three levels: measurement occasion, subject, and site. There may be correlation between the repeated experiences of an individual subject as well as possible correlation between the experiences of subjects within a sit.

Generally speaking, in multilevel data, the observations within the same group are more likely to be correlated than the observations from different groups. So in each level we have some type of correlation. The correlations from all levels should be taken into account and ignoring any one of them may lead to inconsistent estimates and misleading inferences. A well known method of representing this common variation is by adding a common unobserved random effect to the linear predictor for each lower level unit in the same upper level unit. If the distribution of this random effects is conjugate to the distribution of the responses, then maximum likelihood is straightforward. Otherwise the likelihood function does not have a closed form and we need an approach to deal with the problem when we assume a specific distribution for the random effects in each level. A common

distribution for the random effects is normal distribution with mean zero (Breslow and Clayton, 1993; McGilchrist, 1994). If the distribution of the responses is not normal then we will have an integral in each level (except for the random coefficient if there exists). In this case the dimension of integrals in the likelihood function is equal to the number of the levels. Some approaches to solve the integrals are:(a) The likelihood can be integrated numerically using Gaussian Quadrature (GQ) points. (b) The log likelihood function can be approximated by a second order Taylor series expansion. (c) A fully Bayesian approach can be used with the additional structure of a prior distribution on all the model parameters. The Markov Chain Monte Carlo (MCMC) methods can be used to obtain marginal posterior distributions of the parameters.

In these three approaches we assume a specific parametric form of the mixing distribution of the unobserved random effects. Heckman and Singer (1984) and Davies (1987) have shown that the parameter estimation is sensible to the choice of the mixing distribution. This problem can be solved by Nonparametric Maximum Likelihood (NPML) estimation on mixing distribution on a finite number of mass points. This approach is used by Aitkin (1999) for fitting two-level data.

For the data that are clustered and / or longitudinal, random effects regression models have been developed to model continuous data (Laid and Ware, 1982; Bock, 1989; Jennrich and Schluchter, 1986; Bryk and Raudenbush, 1987). The same models have also been developed for dichotomous data (Stiratelli, Laird and Ware, 1984; Anderson and Aitkin, 1985; Wong and Mason, 1985; Zeger and Liang, 1986; Gibbons and Bock,1987; Qu et al. , 1972). Goldstien (1991) has discussed nonlinear multilevel models and their application to discrete response data. Qaqish and Liang (1992) have presented marginal models for correlated binary response within multiple classes and multiple levels of nesting. Hedeker and Gibbons (1994) have proposed a random effects ordinal regression model for multilevel analysis. Their model is proposed for analyzing the clustered longitu-

dinal ordinal response data. Gibbons and Hedeker (1997) have explained random effects probit and logistic regression models for three level data. Aitkin (1999) has described an EM algorithm for nonparametric maximum likelihood estimation in generalized linear models with variance components structure. Various approaches have been proposed in recent years to model survival data with two level clustering. For example, Clayton (1991), Gray (1992), Klein (1992), Nielsen et al. (1992), McGilchrist (1993), Lin (1994). Currently Yau (2001) has described and applied a method for modeling survival data with multilevel clustering and random effects.

There are many software for multilevel data analysis and a number of comparisons have been done by some researchers. Kreft et al. (1994) have compared HLM, ML3, VARCL, BMDP5-V, and GENMOD. Rodriguez and Goldman (1995) have evaluated two software packages VARCL and ML3 for fitting models to binary response data by using a Monte Carlo study. Van der Leeden et al. (1996) have compared HLM, ML3, and BMDP5-V on repeated measures data. De Leeuw and Kreft (2001) have compared software MLn, HLM, VARCL, MIXFOO, MLA, BMDP, SAS, SPSS, and MLwiN for multilevel analysis. They have indicated what the programs can do, where one can get the program, the cost, on what systems the software run, and how easy it is to use them.

Very little work has been done on using and comparing the four mentioned approaches, GQ, Taylor series, MCMC, and NPML in analyzing multi level data. The purpose of this paper is to model a multilevel data in a general form and explain, apply and compare the above approaches through simulation study. Our analysis focuses on bias and efficiency of estimates produced by the mentioned approaches. However the results will compare some software in fitting multilevel models.

The rest of this paper is organized as follows. In section 2 we introduce multilevel model. In section 3 we describe the estimation procedures. In section 4 we

present the simulation study and compare the results from different procedures. Section 5 summarizes our conclusions.

2　Multilevel Models

Before introducing multilevel models in a general form consider a simple three level model. Suppose that the data consists of N_1 communities and each community consists of N_2 families with N_3 children within each family. Here communities, families and children define level one, two and three for a three level data respectively. Suppose that we are interested in estimating the effect of x_{ijk}, x_{ij}, and x_i, the explanatory variables in levels one, two, and three respectively, on the binary response measured for each child. Moreover assume that we believe the community and the family populations are heterogeneous. To control for heterogeneity we introduce two random effects u_{ij} and u_i at the second and first level. With these specifications the linear predictor will be of the form $\beta_1 x_{ijk} + \beta_2 x_{ij} + \beta_3 x_i + u_i + u_{ij}$, $i = 1, 2, ..., N_1$, $j = 1, 2, ..., N_2$, $k = 1, 2, ..., N_3$. Assuming any link function, for example logit, one may estimate $\beta_1, \beta_2, \beta_3$ the fixed effects of x_{ijk}, x_{ij}, and x_i and the variances of u_{ij} and u_i by using appropriate estimation procedure. Following we introduce multilevel models in a general form and in the next section we will discuss about various estimation approaches.

2.1　Multilevel Linear Model

consider the situation that the data have L levels, such that in level i we have f_i fixed effects, r_i random effects, and g_i units. We first define the following notations.

$\mathbf{X}_i = diag_{g_i}[X_{(j)}]$ ($diag_{g_i}$ refers to a $g_i \times g_i$ diagonal matrix), where $X_{(j)}$ is $a_i \times f_i$ model matrix for the fixed effects for j^{th} unit of level i. $\boldsymbol{\beta}_i = [\beta'_1 \vdots \beta'_2 \vdots \cdots \vdots \beta'_{g_i}]'$, where : β_j is $f_i \times 1$ column vector of fixed effects for j^{th} unit of level i. Now if $\mathbf{X} = diag_L[\mathbf{X}_i]$ and $\boldsymbol{\beta} = [\boldsymbol{\beta}'_1 \vdots \boldsymbol{\beta}'_2 \vdots \cdots \vdots \boldsymbol{\beta}'_L]'$ then $\mathbf{X}$ is a matrix of $N \times P$ where

$N = \sum\limits_{i=1}^{L} \sum\limits_{j=1}^{g_i} a_j$ and $P = \sum\limits_{i=1}^{L} g_i f_i$ and $\boldsymbol{\beta}$ is $P \times 1$ vector of fixed effects across all levels. $\mathbf{Z}_i = diag_{g_i}[Z_{(j)}]$, where $Z_{(j)}$ is $c_j \times r_i$ model matrix for the random effects for j^{th} unit of level i. $\mathbf{u}_i = [u'_1 \vdots u'_2 \vdots \cdots \vdots u'_{g_i}]'$, where : u_j is $r_i \times 1$ column vector of random effects for j^{th} unit in level i. Now if $\mathbf{Z} = diag_L[\mathbf{Z}_i]$ and $\mathbf{u} = [\mathbf{u}'_1 \vdots \mathbf{u}'_2 \vdots \cdots \vdots \mathbf{u}'_L]'$ then $\mathbf{Z}$ is a matrix of $N \times Q$ where $N = \sum\limits_{i=1}^{L} \sum\limits_{j=1}^{g_i} c_j$ and $Q = \sum\limits_{i=1}^{L} g_i f_i$ and $\mathbf{u}$ is $Q \times 1$ vector of random effects across all levels. Using these notations a multilevel linear model is defined as,

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{u} + \boldsymbol{\varepsilon}. \tag{1}$$

where $\mathbf{Y}$ is $N \times 1$ vector of responses and $\boldsymbol{\varepsilon}$ is $N \times 1$ vector of error terms.

Assume that the random effects from different units are mutually independent with mean $\mathbf{0}$ and $Var(\mathbf{u}_i) = \boldsymbol{\Omega}_i$. We then have $Var(\mathbf{u}) = \boldsymbol{\Omega}$ and $\boldsymbol{\Omega} = diag_L[\mathbf{I}_{g_i} \otimes \boldsymbol{\Omega}_i]$. Moreover assume that $Var(\boldsymbol{\varepsilon}) = \sigma^2 \mathbf{W}$ and $Cov(\boldsymbol{\varepsilon}, \mathbf{u}) = \mathbf{0}$. With these assumptions we have $E(\mathbf{Y}) = \mathbf{X}\boldsymbol{\beta}$ and,

$$\mathbf{V} = Var(\mathbf{Y}) = \mathbf{Z}\boldsymbol{\Omega}\mathbf{Z}' + \sigma^2 \mathbf{W} = \sum\limits_{i=1}^{L} \mathbf{Z}_i(\mathbf{I}_{g_i} \otimes \boldsymbol{\Omega}_i)\mathbf{Z}'_i + \sigma^2 \mathbf{W}.$$

The model introduced by Rodriguez and Goldman (1995) is a especial case of our model in equation (1). Their model includes three levels and no random effects in the first level. There are many cases in applications that the analyst is interested in considering random coefficients for individuals in the first level especially a random effect to control omitted variables.

2.2 Multilevel Nonlinear Model

Following Goldstein (1991) a Multilevel Nonlinear Model consists of a nonlinear component and a linear component, both of which may contain fixed and random effects as,

$$\mathbf{Y} = f(\mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{u}) + \mathbf{G}\boldsymbol{\eta} + \mathbf{H}\boldsymbol{\delta} \tag{2}$$

where $\mathbf{X}, \mathbf{Z}, \boldsymbol{\beta}$, and $\mathbf{u}$ have the same definition as above and $\mathbf{G}, \mathbf{H}, \boldsymbol{\eta}$, and $\boldsymbol{\delta}$ have the similar structure to $\mathbf{X}, \mathbf{Z}, \boldsymbol{\beta}$, and $\mathbf{u}$. Model (1) is a special case of model (2) if we omit the second and the third terms and f is an identity function. An important especial case of model (2) is when the response is a vector of proportions, there is no linear component, and f is logit function. The multilevel logit model then is of the form,

$$logit(\boldsymbol{\mu}) = \boldsymbol{\eta} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{u} \tag{3}$$

where $\mu_i = P_r(Y_i = 1|\boldsymbol{\beta}, \boldsymbol{\Omega}, \mathbf{X}, \mathbf{Z})$; for $i = 1, ..., N$ and $\boldsymbol{\eta}$ is a conditional linear predictor.

3 Estimation Approaches

We assume that the elements of $\boldsymbol{\Omega}_i$ are unknown parameters and $\mathbf{W}$ is a known matrix. Further we assume that $\boldsymbol{\varepsilon}$ has multivariate normal distribution and we look for Maximum Likelihood Estimation (MLE) of the parameters. Under these assumptions the conditional likelihood function for a multilevel model is,

$$L(\boldsymbol{\beta}|\mathbf{u}) = \prod_{i=1}^{N} f(\mathbf{y}_i|\mathbf{X}, \mathbf{Z}, \boldsymbol{\beta}, \mathbf{u}) \tag{4}$$

where f is the density function of $\mathbf{y}_i$. The marginal likelihood function of the parameters is,

$$L(\boldsymbol{\beta}, \boldsymbol{\Omega}) = \int_{\mathbf{u}} L(\boldsymbol{\beta}|\mathbf{u})g(\mathbf{u})d\mathbf{u} \tag{5}$$

where g is the density function of the vector $\mathbf{u}$. If we assume that $\mathbf{u}$ is normally distributed (Breslow and Clayton, 1993) then,

$$L(\boldsymbol{\beta}, \boldsymbol{\Omega}) = \int_{\mathbf{R}^Q} L(\boldsymbol{\beta}|\mathbf{u})\boldsymbol{\Phi}(\mathbf{u})d\mathbf{u} \qquad (6)$$

where $\boldsymbol{\Phi}$ is the multivariate normal density. For the multilevel logit model (3) the conditional likelihood function is,

$$L(\boldsymbol{\beta}|\mathbf{u}) = \prod_{i=1}^{N} \mu_i^{y_i}(1 - \mu_i)^{1-y_i} \qquad (7)$$

The marginal likelihood function can be obtained by replacing from (7) in (5) or in (6). Note that for other multilevel nonlinear models such as multilevel probit model we replace μ_i in (7) by $\boldsymbol{\Phi}(\eta_i)$, where η_i is the linear predictor for individual i.

The numerical integration in (5) is intractable unless the distribution of $\mathbf{u}$ is conjugate to the distribution of $\mathbf{y}$. Expression (6) is also intractable if the distribution of $\mathbf{y}$ is not normal. However, simple two level exponential family models other than the normal with a normal random effect may be fitted with difficulty and slow by maximum likelihood. Therefore, in order to maximize the likelihood function (6) many algorithms have been used. Goldstein (1986) have proposed a generalized least squares algorithm that has been implemented in the package ML3. Longford (1987) has proposed a Fisher scoring algorithm that has been implemented in program VARCL. Raudenbush and Bryk (1986) have used an EM algorithm in the package HLM. Mason et al. (1983) have also used an EM algorithm in the program GENMOD. Rodrigues and Goldman (1995) have compared least squares algorithm and Fisher scoring algorithm through package ML3 and program VARCL. Tsutakawa (1985) has used full Bayes and empirical

Bayes approaches to analyze two-level models with a random effect at upper level and no explanatory variable at the lower level. Aitkin (1999) has compared the results of Tsutakawa (1985) with his results of applying GQ and NPML approaches using GL1M4 software. Hedeker and Gibbons (1994) have used Gauss Hermite Quadrature approach to analyze two level data set. Gibbons and Hedeker (1997) have used numerical quadrature approach to find the maximum marginal likelihood estimation in random effects probit and Logistic regression models for the three level data. They have also compared parameter estimates for a normal versus rectangular prior to determine the degree to which their estimates are robust to deviation from the assumed normality of the prior distribution of the random effects. Wong and Mason (1985) have proposed empirical Bayes approach and have applied it to estimate the parameters of a two-level model. Qaqish and Liang (1992) have used the Generalized Estimation Equations (GEE) approach of Liang and Zeger (1986). They have estimated the parameters of marginal models for correlated binary responses with multiple classes and multiple levels of nesting. A regression coefficient in this marginal model is interpreted as the change in the "population average" responses rather than the changes in any one cluster's expected response with covariate X.

Except the work of Aitkin (1999) and Rodriguez and Goldman (1995) that have compared some of these approaches, no research has been done to compare the above approaches. For example, Rodriguez and Goldman (1995) have used simulation but have only compared Longford's approximate likelihood (1988 and 1994) and Goldstein's Generalized least squares approach (1991). In this paper we use simulation and will show that other approaches perform better than approaches used in VARCL and ML3. Following we briefly explain the approaches that are tractable to deal with the integrals in (5) and (6).

3.1 Gaussian Quadrature Approach

To explain approaches we consider a three level logit model with one random effect at each of the levels two and three. If we consider one explanatory variable at each level then model (5) reduces to

$$L(\boldsymbol{\beta}, \boldsymbol{\Omega}) = \prod_{i=1}^{L} \int_{-\infty}^{+\infty} \left(\prod_{j=1}^{n_i} \int_{-\infty}^{+\infty} \left(\prod_{k=1}^{n_{ij}} \mu_{ijk} \right) g_1(u_{ij}) du_{ij} \right) \times g_2(u_i) du_i \qquad (8)$$

$$\mu_{ijk} = \frac{\exp\left[(\beta_1 x_{ijk} + \beta_2 x_{ij} + \beta_3 x_i + u_i + u_{ij}) \, y_{ijk} \right]}{1 + \exp\left[\beta_1 x_{ijk} + \beta_2 x_{ij} + \beta_3 x_i + u_i + u_{ij} \right]}$$

where x_{ijk}, x_{ij}, and x_i are the explanatory variables in levels one, two, and three respectively. u_{ij} and u_i are the random effects with means zero and standard errors σ_1, σ_2 and density g_1, g_2 related to second and third levels respectively. y_{ijk} is the response for the kth individual in the jth unit of level two and i^{th} unit of level one. $\beta_1, \beta_2, \beta_3$ are the fixed effects of x_{ijk}, x_{ij}, and x_i. Here we need to calculate one dimensional integral.

If we assume that the random effects u_{ij} and u_i are distributed normally with means zero and standard errors σ_1, σ_2 then we can approximate the integrals in (8) by a number of GQ points. The likelihood function (8) becomes,

$$L(\boldsymbol{\beta}, \boldsymbol{\Omega}) = \prod_{i=1}^{L} \sum_{l=1}^{s_2} \left(\prod_{j=1}^{n_i} \sum_{m=1}^{s_1} \left(\prod_{k=1}^{n_{ij}} \eta_{ijk} \right) p_m \right) q_l \qquad (9)$$

$$\eta_{ijk} = \frac{\exp\left[(\beta_1 x_{ijk} + \beta_2 x_{ij} + \beta_3 x_i + \sigma_1 \alpha_m + \sigma_2 \tau_l) \, y_{ijk} \right]}{1 + \exp\left[\beta_1 x_{ijk} + \beta_2 x_{ij} + \beta_3 x_i + u_1 + u_{ij} \right]}$$

where $\alpha_1, \alpha_2, ..., \alpha_{s_1}$ are s_1 Gaussian mass points with probability masses $p_1, p_2,$ ..., p_{s_1} at level two and $\tau_1, \tau_2, ..., \tau_{s_2}$ are s_2 Gaussian mass points with probability masses $q_1, q_2, ..., q_{s_2}$ at level one. We may use equal number of masses for different levels.

3.2 Nonparametric Approach

A disadvantage of any approach using a specified parametric form for the distribution of the unobserved random effects is the sensitivity of the parameter estimation to the choice of the distribution of the random effects (Heckman and Singer ,1984 and Davies, 1987). This important problem can be solved by considering the mass points and probability masses as parameters and estimate them together with the structural parameters. So the nonparametric approach is not a simplification of parametric approach. The identification of the number, location and masses of these points of support present formidable computational problems. This approach has been used by some researches. Hind (1982) and Anderson and Hind (1988) used EM algorithm for fitting the finite mixture distribution used in (9). Aitkin (1999) has used EM algorithm in GLIM4 software to estimate a two level model using GQ and NPML approaches. A general method for estimating a multilevel model using either GQ or NPML is to consider (9) as a multivariate function of parameters. Then maximize this function by using an appropriate programing software as NAG, Fortran library or Fortran power station.

3.3 Longford's Approximation Approach

Longford (1988 (b)) has proposed an approximation to the likelihood function (6). The approximation relies on a second order Taylor expansion of the logarithm of the conditional likelihood (4) about $\mathbf{u} = \mathbf{0}$. Longford (1988 (b)) has implemented this estimation strategy in the software package VARCL. This method provides the basis for a Fisher scoring procedure which can be applied alternately to $\boldsymbol{\beta}$ and $\boldsymbol{\Omega}$. Although, Longford's approximation has solved the problem of high dimensionality of the integrals in (6) for some models but care should be taken in applying this approximation. Since the true likelihood function is not maximized and the remainder of the Taylor expansion is not controlled the parameter estimate may by biased. Even if all the necessary conditions needed to write

the Taylor series of the likelihood function are attained we need to control the remainder of the estimation of the likelihood function by its finite Taylor series. For more details about the expansion of the likelihood function in equation (6) one may see Rodriguez and Goldman (1995).

3.4   Goldstein's Generalized Least Square Approach

Goldstein (1991) has proposed an alternative approach to the estimation of non-linear multilevel models, including the logistic model with random coefficients. His method needs a single integration instead of iterating to convergence. In a special case if there are no random effects the proposed procedure is equivalent to the standard algorithm, iteratively reweighted least square used by McCullagh and Nelder (1989) to fit generalized linear models. This approach has used first-order Taylor series expansion and the same problems that mentioned in 3.3 may arise here. However, this approximation has been implemented in the software package ML3 and currently in MLwiN. For more details see Rodriguez and Goldman (1995).

3.5   Markov Chain Monte Carlo Approach

As explained in sections 3.1 to 3.4 the computationally burden has limited the analysis of multilevel data in several ways. First , investigators have largely restricted their attention to random intercept models to avoid higher dimensional numerical integration . Second , specialized software is required and is typically optimized for a particular random effects distribution. For example the Gaussian is used in VARCL explained in 3.3.

MCMC techniques is being increasingly used as an approach for dealing with the problems for which there is no exact analytic solution, and for which standard approximation technique, have difficulties. The basic philosophy behind MCMC is to take a Bayesian approach and carry out the necessary numerical integrations using simulation, Gelfand and Smith (1990); Smith and Roberts(1993). Instead

of calculating exact or approximate estimates, this computer-intensive technique generates a stream of simulated values for each quantity of interest. The conditional independence assumptions , common in data analysis , mean that the full distribution of all quantities has a simple factorization in terms of conditional distribution of each node given its parents. Thus we only need to provide the parent-child distributions in order to fully specify the model. Bayesian inference Using Gibbs sampling (BUGS) program (Spiegelhalter, Thomas, Best, and Gilks, 1996) carries out Bayesian inference on statistical problems using a simulation technique known as Gibbs sampling. We can use BUGS to analyze multilevel data which uses MCMC techniques.

In multilevel models we assume specific parametric priors for the random effects and non-informative priors with extremely small precision for the structural parameters and the precisions of the random effects. For a three level binary data the logit of μ_{ijk}, the mean response for the ijk^{th} individual, is

$$logit\left(\mu_{ijk}\right) = \beta_1 x_{ijk} + \beta_2 x_{ij} + \beta_3 x_i + u_i + u_{ij} \tag{10}$$

where u_i and u_{ij} have informative priors with non-informative priors for their precisions as their parents. $\beta_1, \beta_2, \beta_3$ have no parents and so have non-informative priors. In order to use Gibbs sampling we need to successively sample from the distribution of each node given all the others in equation 10. These are known as full conditional distributions. Geman and Geman (1984) have shown that under mild conditions this process eventually provides samples from the joint posterior distribution of the unknown quantities. For model 10 we start with initial values $\beta_1^{(0)}, \beta_2^{(0)}, \beta_3^{(0)}, \sigma_1^{(0)}, \sigma_2^{(0)}$, draw $\beta_1^{(1)}$ from the distribution of $\left[\beta_1 | \beta_2^{(0)}, \beta_3^{(0)}, \sigma_1^{(0)}, \sigma_2^{(0)}\right]$, then draw $\beta_2^{(1)}$ from the distribution of $\left[\beta_2 | \beta_1^{(1)}, \beta_3^{(0)}, \sigma_1^{(0)}, \sigma_2^{(0)}\right]$ and finally complete the first iteration by drawing σ_2 from the distribution of $\left[\sigma_2 | \beta_1^{(1)}, \beta_2^{(1)}, \beta_3^{(1)}, \sigma_1^{(1)}\right]$. After a large number of iterations we consider the last sample as the initial val-

ues for the parameters and continue the sampling for another large number of sampling. The average of the observations, for each parameter, then will be the estimate of the parameter. It should be mentioned that BUGS software only provides a simple method to solve the high dimensionality problem of the integrals. So the sensitivity of the parameter estimation to the choice of random effects distribution mentioned in section 3.2 is still a serious problem. But since our comparison of methods of estimation is based on using the true distributions of the random effects that generate the data, this problem does not arise. So we can focus on the differences of the approaches from other points of view.

4 Simulation Study

In empirical study, since the true value of the parameters are not known we can never be certain if the results of empirical work are accurate and so we may have misleading comparisons of underlying approaches. In simulation studies we know the true value of the parameter and so we can compare the approaches to see which one is more accurate in estimating the parameters of the model. On the other hand in simulation studies we are not certain if the simulation results are relevant in practice. But if the simulation's structure is built related to the real data structure then we can rely on the simulation's results. For comparisons of estimation procedures we followed the simulation's structure proposed by Rodriguez and Goldman (1995). They have simulated data sets using the same hierarchial structure as one of the Guatemalan data sets analyzed by Pebley and Goldman (1992). We have used their rectangular structure design in order to compare the estimation procedures mentioned in section 3 with their results.

4.1 Simulation of Data

Consider 20 units in each level of the three level model (8). Suppose that $x_{ijk}, x_{ij},$ and x_i are dummy variables in fully balanced design, so the covariates are inde-

pendent and each of the eight combinations of values occur equally often. the fixed effects $\beta_1, \beta_2, \beta_3$ are set to be one. The random effects u_{ij} and u_i are generated from independent normal distributions with means zero and large variances (1.0) and small variances (0.16). The Fortran program that generates the data are reported in Appendix B.

4.2 Results of simulation study and comparisons of approaches

To compare the approaches two criteria, bias and efficiency are used. Tables 1 and 2 report values of the estimated fixed effects and the estimated standard errors of the random effects averaged over the 100 simulations when the variance of random effects are 1 and 0.16 respectively. The standard deviation of the estimates are reported in the parenthesis and the MSE of estimates are reported in Bold. First we discuss the case that the variances of random effects are large, i.e. $\sigma_1^2 = \sigma_2^2 = 1$.

The results from Rodriguez and Goldman (1995), reported in table 1, show large significant biases for all the parameters. When they used VARCL software, except the fixed effect at third level, all the other estimates are significantly biased. The estimates of the parameters $\beta_1, \beta_2, \beta_3, \sigma_1, \sigma_2$ are 24.4, 22.5, 9.4, 19.9, 25.1 percent downward bias respectively. Their performance in ML3 results in substantial significant biases especially for the standard error of the random effect at second level. The biases are 89.7 and 72.2 percent using linear and quadratic approximations respectively. They have not reported the standard deviations of the estimates to check if the biases are statistically significant.

To implement the GQ approach we have used the subroutine BCONF from Fortran Power Station 4.0 software to maximize the likelihood function in equation 9. The subroutine BCONF minimize a function of N variables subject to bounds on the variables using a quasi-Newton method and a finite-difference gra-

dient (See Appendix C for codes). The GQ approach, even for 6 mass points, results in less biases than the results from VARCL and ML3, except for the standard error of the random effect at level three. The estimates of the parameters $\beta_1, \beta_2, \beta_3, \sigma_2$ are $14.9, 1.7, 3.5, 99.4$ percent upward bias respectively and σ_1 is 4.3 percent downward bias. None of these biases are statistically significant. We received the same pattern of biases by using 14 mass points in limited simulations and also in simulations with larger sample size at the third level (50 and 100 instead of 20). The average time for each run was 6 and 28 minutes, using a 500 pentium, for 6 and 14 mass points respectively. During the minimization 48% of runs have fixed the value of the standard error of the random effect of the third level at lower or upper bound. We found that this approach behave poorly in estimating the standard error of the third level and is computationally prolematic.

To apply the NPML approach we have used the subroutine LCONF from Fortran Power Station 4.0 software to maximize the likelihood function in equation 9. The subroutine LCONF minimize a general objective function subject to linear equality/inequality constraints (See Appendix D for codes). Table 1 shows that the results from the performance of the NPML approach are better than the results from the GQ approach in estimating the standard errors of the random effects. With 3 mass points β_1, β_2 are 0.3 and 2.8 percent downward bias and $\beta_3, \sigma_1, \sigma_2$ are $24.4, 35.0, 24.3$ percent upward bias respectively. With 2 mass points $\beta_1, \beta_2, \beta_3, \sigma_2$ are $2.5, 11.6, 24.9, 20.7$ percent downward bias and σ_1 is 2.9 percent upward bias respectively. The increment of about 88 in deviance (from 7286.49 to 7375.01) with the reduction of 4 in degree of freedom confirm to report the results for 3 mass points in table 1. Non of these biases are significant. The average time for each run was 1 and 4 minutes, using a 500 pentium, for 2 and 3 mass points respectively.

To apply the MCMC approach using Gibbs sampling we have used BUGS

software (See Appendix E for codes). It is assumed that the prior distributions of u_i and u_{ij} to be normal with means 0 and standard errors σ_1 and σ_2 respectively. $\beta_1, \beta_2, \beta_3$ have non-informative normal prior with mean 0 and standard error 1000, σ_1 and σ_2 have non-informative gamma prior with mean 1 and variance 1000. In order to get over the influence of the initial values we have performed 500 iterations of the Gibbs sampler and then have updated another 1000 iterations to estimate the parameters. Table 1 shows that this approach performs excellent with at most 2.8% bias for the fixed effect at the second level. The standard deviation of estimates are small and none of the biases are statistically significant. Table 1 shows that the MCMC approach results in very small MSE. The average time for each run was 5 minutes by using a 500 pentium.

From the point of bias the least biases are obtained by using BUGS program that uses MCMC method using Gibbs sampling. The NPML approach performs better than VARCL and ML3. The GQ results better estimation for β_2, β_3 than VARCL, ML3, and NPML. The results from BUGS, NPML, and GQ are not significantly biased while only the estimation of β_3 produced by VARCL is not significantly biased. The values of MSE reported in Table 1 show that the efficiency of the NPML is more than VARCL in estimating β_1, β_2 but this is vice versus in estimating $\beta_3, \sigma_1, \sigma_2$. Both NPML and VARCL are more efficient than GQ. The MSE obtained from using BUGS is much less than those obtained from using other approaches and consequently this approach estimates the parameters more efficient than other approaches. Since the standard deviations of estimates from ML3 have not been reported by Rodriguez and Goldman (1995) we can not compare the efficiency of their estimates with the other approaches. All runs in applying NPML and BUGS perform without any problem but, as mentioned, in performing GQ many runs results poor estimates for the standard error of the random effect in the third level. The GQ approach is used in GLIM4 (Aitkin, 1999) and SABR (Barry, Francis, and Davies, 1989) software to evaluate a one

dimensional integral. Our simulations show that although the GQ approach may performs well in the problems containing one dimensional integral but care should be done in using this approach in fitting multilevel data.

Table 2 shows that when the variances of the random effects are small, i.e. $\sigma_1^2 = \sigma_2^2 = 0.16$, non of the estimates are significantly biased. Using GQ or NPML results in large absolute biases for the standard errors of the random effects. But since the standard deviations of the estimates are also large these biases are not statistically significant. VARCL and BUGS perform almost the same but with less biases using BUGS.

Fig. 1 reports the Q-Q plots of the estimates produced by three approaches NPML, GQ, and MCMC (See Appendix A). The plots for NPML show that, except for a few outliers, the asymptotic normality is attained for all the estimates. For the GQ approach, except the plot for the second level fixed effect, the plots show some lack of symmetry. For the MCMC approach we have a bit deviation from asymptotic normality for the fixed effect of the first level. Our conclusion is that, non of the NPML and MCMC approaches result in significant deviation from asymptotic normality.

To complete the comparisons of approaches two points should be mentioned. Firstly, our main aim is to estimate the explanatory variables effects. The estimation of the variances of the random effects is for identifying the levels of correlations explained in section 1. So comparisons of approaches with respect to the explanatory variables effects is more important than comparisons with respect to the variances of the random effects. The above comparisons show that if the analyst know the distributions of the random effects the MCMC approach is the best method. While, if there is no prior information about the distribution of the random effects the NPML approach is recommended. This approach is also recommended if the analyst believes in the results of Heckman and Singer (1984) and Davies (1987).

Secondly, Our simulations are based on a three level model, while there are many examples in application with four or more levels. For example in an investigation the data may consists of students within classes, classes within schools, and schools within school districts. The performance of the various methods may be affected by the number of levels. Our simulations show that the MCMC approach performs uniformly across the levels. There is no evidence that this approach performs differently across levels. The NPML performs better in lower levels while VARCL performs better in higher levels. The performance of the GQ approach depends on the variances of the error terms. With a low and high error variance this procedure performs better in lower and upper levels respectively. The sample size is an important factor which affect the performance of some methods. For small sample size or where there are few lower level units within each higher level unit and for binary response, the Taylor series approximation, used in VARCL, may produce downwardly biased parameter estimates. While MCMC method using Gibbs and Metropolis-Hastings sampling produces unbiased estimates (see Gilks et al. 1996). Although this investigation recommend the use of MCMC approach but still it is recommended that more than one approach is tried; if similar results are obtained then more confidence can be placed in the estimates. If the results from approaches are different then a simulation study based on the data is recommended before choosing any software. Further research is needed to clearly indicate that how the performance of the various methods is affected by the number of levels considered and which factors play a significant role. Another simulation study may be proposed to investigate the effect of sample size, variance of random effects, type of explanatory variable, number of levels, and other factors on the performance of the various methods.

5 Conclusions

In this paper we reviewed multilevel models and introduced some estimation procedures that may be applied to fit these models. Some of them have been used by other researchers but no comparisons have been made. We have compared these approaches through simulation study for binary responses. The structure of simulations are the same as rectangular structure proposes by Rodriguez and Goldman (1995). We showed that the substantial significant biases coming from VARCL and ML3, reported by Rodriguez and Goldman (1995), can be vanished by applying the MCMC method using Gibbs sampling. The efficiency of the MCMC approach is considerably high and recommended if we have prior information about the distributions of the random effects. If there is no prior information, then parameters estimates may be sensible to the choice of the mixing distributions (Heckman and Singer, 1984). In this case the NPML approach is recommended and our simulation study shows that the nonparametric approach performs better than VARCL and ML3. Although our simulation study was for binary data but both the NPML and MCMC approaches are easily applicable to other types of multilevel data.

Table 1. *Simulation results for large error variance. The figures in the parentheses are the standard deviations of estimates. Bold figures are MSE of the estimates. *Significantly biased estimates.*

Approach	$\beta_1 = 1$	$\beta_2 = 1$	$\beta_3 = 1$	$\sigma_1 = 1$	$\sigma_2 = 1$
VARCL	0.756*	0.775*	0.906	0.801*	0.749*
	(0.062)	(0.089)	(0.378)	(0.044)	(0.115)
	0.063	**0.059**	**0.152**	**0.042**	**0.076**
GQ	1.149	1.017	1.035	0.957	1.994
	(0.408)	(0.378)	(0.674)	(0.425)	(1.073)
	0.189	**0.143**	**0.456**	**0.182**	**2.139**
NPML	1.003	0.972	0.756	1.350	1.243
	(0.063)	(0.155)	(0.467)	(0.244)	(0.315)
	0.004	**0.025**	**0.278**	**0.182**	**0.158**
MCMC	0.992	0.972	1.010	1.000	0.997
	(0.115)	(0.118)	(0.350)	(0.062)	(0.199)
	0.013	**0.015**	**0.123**	**0.009**	**0.040**
ML3-Linear	0.738	0.74	0.771	0.103	0.732
ML3-Quadratic	0.854	0.860	0.910	0.278	0.764

Table 2. *Simulation results for low error variance. The figures in the parentheses are the standard deviations of estimates. Bold figures are MSE of the estimates.*

Approach	$\beta_1 = 1$	$\beta_2 = 1$	$\beta_3 = 1$	$\sigma_1 = 0.4$	$\sigma_2 = 0.4$
VARCL	0.944	0.950	1.020	0.356	0.356
	(0.060)	(0.070)	(0.180)	(0.056)	(0.063)
	0.007	**0.007**	**0.033**	**0.005**	**0.006**
GQ	1.051	1.041	1.340	0.293	0.676
	(0.307)	(0.343)	(0.488)	(0.249)	(0.749)
	0.097	**0.119**	**0.354**	**0.073**	**0.637**
NPML	1.001	1.008	0.967	0.801	0.630
	(0.054)	(0.079)	(0.293)	(0.251)	(0.229)
	0.003	**0.006**	**0.087**	**0.224**	**0.105**
MCMC	1.006	0.975	1.001	0.396	0.386
	(0.068)	(0.060)	(0.177)	(0.052)	(0.091)
	0.005	**0.004**	**0.031**	**0.003**	**0.008**

References

Aitkin, M. 1999. A General Maximum Likelihood Analysis of Variance Components in Generalized Linear Models. Biometrics **55**, 117-128.

Anderson, B. A. and Hind, J. P. 1988. Random Effects in Generalized Linear Models and the EM Algorithm. Communications in Statistics - Theory and Methods **17**, 3847-3856.

Anderson, D. A. and Aitkin, M. 1985. Variance Components Models with Binary Response: Interviewer Variability. Journal of the Royal Statistical Society, Series B **47**, 203-210.

Barry, J. T., Francis, B. J., and Davies, R. B. 1990. A Guide for Users, Centre for Applied Statistics, Lancaster University, U.K.

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

Bryk, A. S. and Raudenbush, S. W. 1987. Application of Hierarchical Linear Models to Assessing Change. Psychological Bulletin **101**, 147-158.

Bock, R. D. 1987. Measurement of Human Variation: A Two-Stage Model. In: R. D. Bock (Ed.), Multilevel analysis of Educational Data. San Diego: Academic Press, 319-342.

Clark, W. A. V., Huff, J. O. and Burt, J. E. 1979. Calibrating a Model of the Decision to Move. Environment and Planing **A 11**, 689-704.

Clayton, D. G. 1991. A Monte Carlo Method for Bayesian Inference in Frailty Models. Biometrics **47**, 467-485.

Davies, R. B. 1987. Mass Point Methods for Dealing with Nuisance Parameters in Longitudinal Studies. In: R. Crouchley (ED.) Longitudinal Data analysis (Avebury, 1987) 88-109.

De Leeuw, J. and Kreft, I. G. G. 2001. Software for Multilevel Analysis. In: Leyland A. H. & Goldstein H. (Eds.), Multilevel Modelling of Health Statistics.

Chichester: Wiley, 187-204.

Gelfand, A. E. and Smith, A. F. M. 1990. Sampling Based Approaches to Calculating Marginal Densities. Journal of the American Statistical Association, **85**, 398-409.

Geman, S. and Geman, D. 1984. Stochastic Relaxation, Gibbs Distribution and the Bayesian Restoration of Images. IEEE Transactions on Pattern Analysis and Machine Intelligence **6(6)**, 721–741.

Gibbons, R. D. and Bock, R. D. 1987. Trend in Correlated Proportions. Psychometrika **52**, 113-124.

Gibbons, R. D. and Hedeker, D. 1997. Random Effects Probit and Logistic Regression Models for Three Level Data. Biometrics **53**, 1527-1537.

Gilks, W. R., Richardson, s. and Spiegelhalter, D. J. 1996. Markov Chain Monte Carlo in Practice. Chapman and Hall, London.

Goldstien, H. 1987. Multilevel Models in Educational and Social Research. New York: Oxford University Press.

Goldstein, H. I. 1986 Multilevel Mixed Linear Model Analysis Using Iterative Generalized Least Squares. Biometrika, **73**, 43-56.

Goldstein, H. 1991. Nonlinear Multilevel Models with an Application to Discrete Response Data. Biometrika **78**, 1, 45-51.

Grey, R. J. 1992. Flexible Methods for Analyzing Survival Data Using Splines with Applications to Breast Cancer Prognosis. Journal of the American Statistical Association **87**, 942-951.

Heckman, J. J. and Singer, B. 1984. A Method for Minimizing the Impact of Distributional Assumptions in Econometric Models for Duration Data. Econometrica **52**, 271-320.

Hedeker, D. and Gibbons, R. D. 1994. A Random Effect Ordinal Regression Models for Multilevel Analysis. Biometrics **50**,933-944.

Jennrich, R. I. and Schluchter, M. D. 1986. Unbalanced Repeated Measures

Models with Structured Covariance Matrices. Biometrics **42**, 805-820.

Kaslow, R. A., Ostrow, D. G., et al., 1987. The Multicenter AIDS Cohort Study: Rationale, Organization and Selected Characteristics of the Participants. American Journal of Epidemiology, **126**, 310-318.

Klein, J. P. 1992. Semiparametric Estimation of Random Effects Using the Cox Model Based on the EM Algorithm. Biometrics 48, 795-806.

Kreft, I. G. G., de Leeuw, J. and van der Leeden, R. 1994. Review of Five Multilevel Analysis Programs: BMDP-5V, GENMOD, HLM, ML3, VARCL. American Statistician **48**, 324-1994.

Laird, N. M. and Ware, J. H. 1982. Random Effects Models for Longitudinal Data. Biometrics **38**, 963-974.

Liang, K. Y. and Zeger, S. L. 1986. Longitudinal Data Analysis Using Generalized Linear Models. Biometrika **73**, 13-22.

Lin, D. Y. 1994. Cox Regression Analysis of Multivariate Failure Time Data: The Marginal Approach. Statistics in Medicine **13**, 2233-2247.

Longford, N. T. 1987. A Fast Scoring Algorithm for Maximum Likelihood Estimation in Unbalanced Mixed Models with Nested Random Effects. Biometrika, **74**, 817-827.

—— 1988. (a) A Quasi - Likelihood Adaptation for Variance Component Analysis. In Proceedings of the Section on Statistical Computing of the Journal of the American Statistical Association.

—— 1988. (b) VARCL: Software for Variance Component Analysis of Data With Hierarchically Nested Random Effects (Maximum Likelihood). Princeton, N. J., Educational Testing Service.

—— 1994. Logistic regression with Random Coefficients. Computational Statistics and Data Analysis, **17**, 1-15.

Mason, W. M., Wong, G. W. and Entwisle, B. 1983. Contextual Analysis Through the Multi Level Linear Model. In: Samuel Leinhardt (ED.), Sociological

Methodology 1983-1984. San Francisco: Jossey - Bass, 72-103.

McCullagh, P. and Nelder, J. A. 1989. Generalized Linear Models. Chapman and Hall, New York.

McGilchrist, C. A. (1993. REML Estimation for Survival Models with Frailty. Biometrics **49**, 221-225.

McGilchrist, C. A. 1994. Estimation in Generalized Mixed Models. Journal of the Royal Statistical Society, Series B **56**, 61-69.

Nielsen, G. G., Gill, R. D., Anderson, P. K., and Sorensen, T. I. A. 1992. A Counting Process Approach to Maximum Likelihood Estimation in Frailty Models. Scandinavian Journal of Statistics **19**, 25-44.

Pebley, A. R. and Goldman, N. 1992. Family, Community, Ethnic Identity and the use of Formal Health Care Services in Guatemala. OPR Working Paper 92-102. Office of Population Research, Princeton.

Qaqish, B. F. and Liang, K. Y. 1992. Marginal Models for Correlated Binary Responses with Multiple Classes and Multiple Levels of Nesting. Biometrics **48**, 939-950.

Qu, Y., Williams, G. W., Beck, G. J., and Medendorp, S. V., 1992. Latent Variable Models for Clustered Dichotomous Data with Multiple Subclusters, Biometrics **48**, 1095-1102.

Raudenbush, S.W. and Bryk, A. S. 1986. A Hierarchical Model for Studying School Effects. Sociology of Education, **59**, 1-17.

Rodriguez, G. and Goldman, N. 1995. An Assessment of Estimation Procedures for Multilevel Models with Binary Responses. Journal of the Royal Statistical Society, Series A **158**, part 1, 73-89.

Smith, A. F. M. and Robberts, G. O. 1993. Bayesian Computation Via the Gibbs Sampler and Related Markov Chain Monte Carlo Methods (with discussion). Journal of the Royal Statistical Society, Series B **55**, 3-24.

Spiegelhalter, D., Thomas, A., Best, N., and Gilks, W. 1996. The BUGS

0.5 Manual.. MRC Biostatistics Unit, Institute of Public Health.

Stiratelli, R., Laird, N., and War, J. H. 1984. Random Effects Models for Serial Observations with Binary Responses. Biometrics **40**, 961-971.

Thall, P. F. and Vail, S. C., 1990. Some Covariance Models for Longitudinal Count Data with Overdispersion. Biometrics **46**, 657-671.

Tsutakawa, R. K. 1985. Estimation of cancer mortality rates: A Bayesian Analysis of Small Frequencies. Biometrics **41**,69-79.

Van der Leeden, R., Verijburg, K. and de Leeuw, J. 1996. A Review of Two Different Approaches for the Analysis of Growth Data Using Longitudinal Mixed Linear Models. Computational Statistics and Data Analysis 21, 583-605.

Wong, G. Y. and Mason, W. M. 1985. The Hierarchical Logistic Regression Model for Multilevel Analysis. Journal of the American Statistical Association **80,** 513-524**.**

Yau, K. K. W. 2001. Multilevel Models for Survival Analysis with Random Effects. Biometrics **57** (1), 96-102.

Zeger, S. L. and Liang, K. Y. 1986. Longitudinal Data Analysis for Discrete and Continuous Outcomes. Biometrics **42**, 121-130.

Appendix A

Fig 1: Q-Q Plots of Estimates.
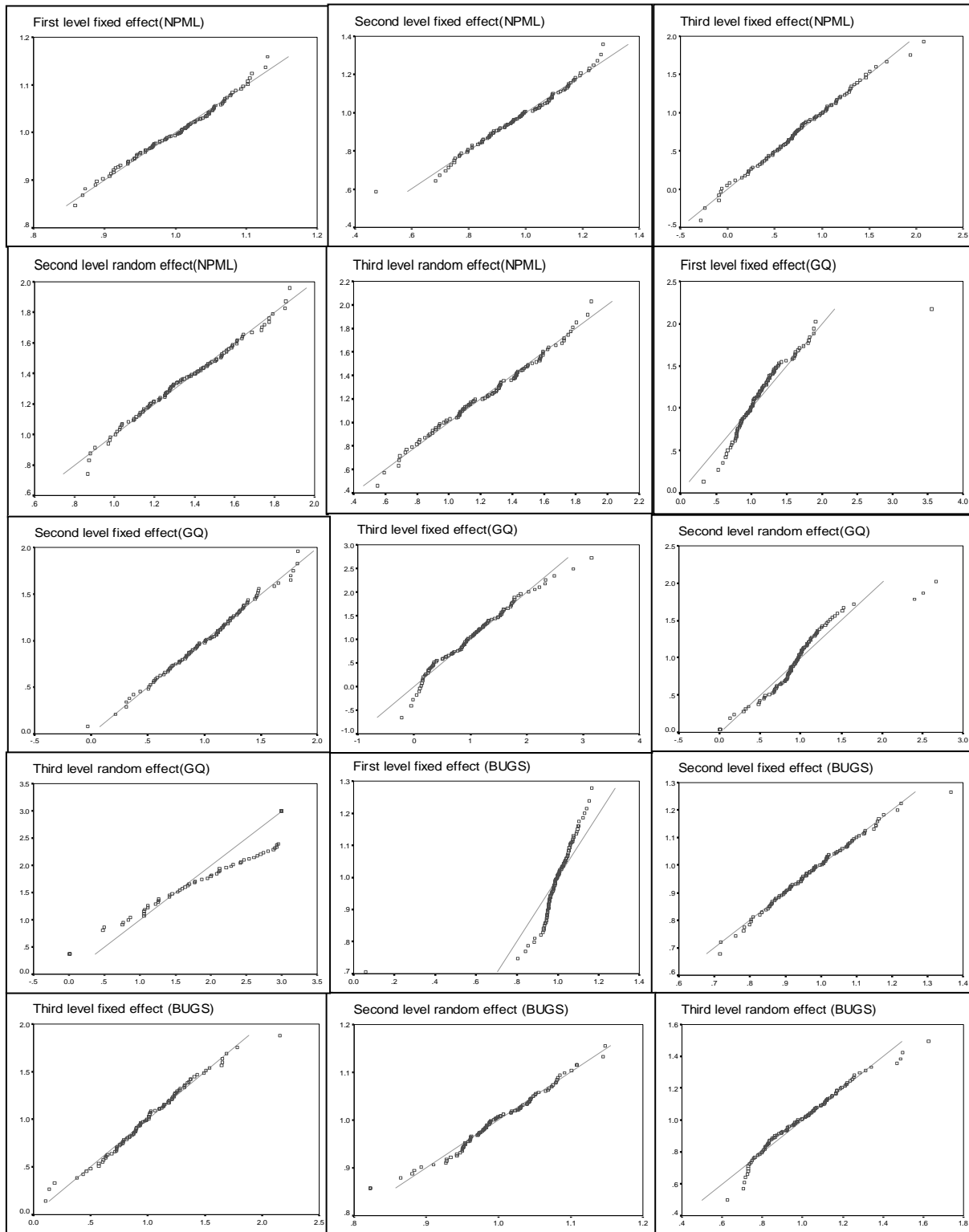
```
!     This program generates the data sets according to the simulation plan explained in the
!     paper.
!     The program is written in Microsoft Developer Studio (Fortran Power Station 4.0).
!
!     Defining of variables. The variables COMMUNITY and GAMMA with one dimenSion is for third
!     level, variables FAMILY and ETA with two dimensions are for second level, and variables
!     CHILD, RESPONSE, W, and E are for the first level.
!
      INTEGER      i,j,k, NOUT, NR,COMMUNITY(20),FAMILY(20,20),CHILD(20,20,20)
      INTEGER      IR,  RESPONSE(20,20,20)
      REAL         RNNOF,E(20,20,20),ETA(20,20),GAMMA(20), UNIFORM,UNI,W(20,20,20)
      EXTERNAL     RNNOF,  UMACH,RNBIN
!
!     Three files COMMUNITY, FAMILY, AND CHILD are used to save the sumulated data at
!     levels 3, 2, and
!     1 respectively. The file RESPONSE is for saving the response at the third level.
!     The structure of printing in the following files are set so that they are readable
!     in WinBugs31.
!
      OPEN(1,FILE='COMMUNITY.TXT')
      OPEN(2,FILE='FAMILY.TXT')
      OPEN(3,FILE='CHILD.TXT')
      OPEN(4,FILE='RESPONSE.TXT')
      WRITE(1,9991)
      WRITE(2,9992)
      WRITE(3,9993)
      WRITE(4,9994)
!
!     NR=20 is the sample size in each level.
!
      NR=20
!
!     The following three nested loops are used to generate a random effect and a fixed
!     effect at level 3, a random effect and a fixed effect at level 2, and a fixed
!     effect and a binomial response at level 1.
!
      CALL UMACH (2, NOUT)
      DO 30  I=1, NR
!
!     Generating Normal random number for third level as random effect.
!
      GAMMA(I)= RNNOF()
!
!     Generating Binomial random number in third level for using as two level factor.
!
      CALL RNBIN (1, 1, 0.5, IR)
      COMMUNITY(I)=IR
      DO 20  j=1, NR
!
!     Generating Normal random number for second level as random effect.
!
      ETA(I,J)= RNNOF()
!
!     Generating Binomial random number in second level for using as two level factor.
!
      CALL RNBIN (1, 1, 0.5, IR)
      FAMILY(I,J)=IR
      DO 10  K=1, NR
!
!     Generating Logistic random number for first level to establish logit model.
!
      CALL RNUN (1, UNIFORM)
      UNI=(UNIFORM)/(1-UNIFORM)
      E(I,J,K)=LOG(UNI)
!
!     Generating Binomial random number in first level for using as two level factor.
!
      CALL RNBIN (1, 1, 0.5, IR)
      CHILD(I,J,K)=IR
!
!     Generating Binomial random number in first level for using as response.
```

```
!      W(I,J,K) is used as a latent variable to generate the binomial response.
!
       W(I,J,K)=COMMUNITY(I)+GAMMA(I)+FAMILY(I,J)+ETA(I,J)+CHILD(I,J,K)+E(I,J,K)
       IF(W(I,J,K).GE.0.0)THEN
       RESPONSE(I,J,K)=1
       ELSE
       RESPONSE(I,J,K)=0
       END IF
   10 CONTINUE
!
!      The following commands are provided to produce Splus format output readable
!      in WinBugs31.
!      For information about the splus format one may see the splus manual or WinBugs31 manual.
!
       IF(I.EQ.NR.AND.J.EQ.NR)THEN
       WRITE(4,9986)(RESPONSE(I,J,K),K=1,NR)
       WRITE(3,9986)(CHILD(I,J,K),K=1,NR)
       ELSE
       WRITE (4,9995)(RESPONSE(I,J,K),K=1,NR)
       WRITE (3,9996)(CHILD(I,J,K),K=1,NR)
       END IF
   20 CONTINUE
       IF(I.EQ.NR)THEN
       WRITE(2,9986)(FAMILY(I,J),J=1,NR)
       ELSE
       WRITE (2,9998)(FAMILY(I,J),J=1,NR)
       END IF
   30 CONTINUE
       WRITE (1,9997)(COMMUNITY(I),I=1,NR)
       WRITE(1,9990)
       WRITE(2,9989)
       WRITE(3,9988)
       WRITE(4,9987)
 9998 FORMAT(55X,I1,19(",",I1),",")
 9997 FORMAT(55X,I1,19(",",I1))
 9996 FORMAT(55X,I1,19(",",I1),",")
 9995 FORMAT(55X,I1,19(",",I1),",")
 9994 FORMAT("list(response=structure( .Data=c(")
 9993 FORMAT("list(child=structure( .Data=c(")
 9992 FORMAT("list(family=structure( .Data=c(")
 9991 FORMAT("list(N=20,NUM=1,community=c(")
 9990 FORMAT("))")
 9989 FORMAT(".Dim=c(20,20)))")
 9988 FORMAT(".Dim=c(20,20,20)))")
 9987 FORMAT(".Dim=c(20,20,20)))")
 9986 FORMAT(55X,I1,19(",",I1),"),")
       END
```

Appendix C

```
!
!     This program generates the data sets according to the simulation plan explained in
!     paper.
!     Six Guassian Quadrature points are used to estimate the parameters in equation 9 in
!     paper.
!     The program is written in Microsoft Developer Studio (Fortran Power Station 4.0).
!
!     Defining of variables.  The variables COMMUNITY and GAMMA with one dimenSion is for third
!     level, variables FAMILY and ETA with two dimensions are for second level, and variables
!     CHILD, RESPONSE, W, and E are for the first level.
!
      INTEGER    I,J,K, NOUT,COMMUNITY(20),FAMILY(20,20),CHILD(20,20,20),RESPONSE(20,20,20)
      REAL       RNNOF,E(20,20,20),ETA(20,20),GAMMA(20),UNIFORM,UNI,W(20,20,20)
!
!     Introducing the subroutines that should be called and is external to
!     this program.
!
      EXTERNAL   RNNOF, UMACH,RNBIN
!
!     Declaration of variables FOR MAXIMIZATION. NSAMPLE is the sample size.
!     For the definition of the other parameters one should see subroutines
!     BCONF and U4INF in the manual of Fortran Power Station 4.0.
!     FCN is the subroutine for this program.
!
      INTEGER    N,NSAMPLE
      PARAMETER  (N=5, NSAMPLE=100)
      INTEGER    IPARAM(7), IBTYPE
      REAL       FVALUE, FSCALE, RPARAM(7), X(N), XGUESS(N),XLB(N), XSCALE(N), XUB(N)
!
!     Introducing the subroutines that should be called and is external to
!     this program.
!
      EXTERNAL   BCONF, FCN, U4INF
!
      DO 40 SAMPLE=1,NSAMPLE
!
!     Defining the output files for simulated data and parameter estimation.
!     Three files COMMUNITY, FAMILY, AND CHILD are used to save the sumulated data at
!     levels 3, 2, and 1 respectively.
!     The file RESPONSE is for saving the response at the third level.
!     The file NONPAOUT is for saving the parameters estimate.
!
      OPEN(1,FILE='COMMUNITY.TXT')
      OPEN(2,FILE='FAMILY.TXT')
      OPEN(3,FILE='CHILD.TXT')
      OPEN(4,FILE='RESPONSE.TXT')
      OPEN(9,FILE='NONPAOUT.TXT')
!
!     Generating the data.  NR=20 is the sample size in each level.
!
      NR=20
!
!     The following three nested loops are used to generate a random effect and a fixed
!     effect at level 3, a random effect and a fixed effect at level 2, and a fixed
!     effect and a binomial response at level 1.
!
      CALL UMACH (2, NOUT)
      DO 30  I=1, NR
!
!     Generating Normal random number for third level as random effect.
!
      GAMMA(I)= RNNOF()
!
!     Generating Binomial random number in third level for using as two level factor.
!
      CALL RNBIN (1, 1, 0.5, IR)
      COMMUNITY(I)=IR
      DO 20  j=1, NR
!
!     Generating Normal random number for second level as random effect.
!
      ETA(I,J)= RNNOF()
```

Page 1

```
!
!       Generating Binomial random number in second level for using as two level factor.
!
        CALL RNBIN (1, 1, 0.5, IR)
        FAMILY(I,J)=IR
        DO 10  K=1, NR
!
!       Generating Logistic random number for first level to establish logit model.
!
        CALL RNUN (1, UNIFORM)
            UNI=(UNIFORM)/(1-UNIFORM)
        E(I,J,K)=LOG(UNI)
!
!       Generating Binomial random number in first level for using as two level factor.
!
        CALL RNBIN (1, 1, 0.5, IR)
        CHILD(I,J,K)=IR
!
!       Generating Binomial random number in first level for using as binary response.
!       W(I,J,K) is used as a latent variable to generate the binomial response.
!
        W(I,J,K)=COMMUNITY(I)+GAMMA(I)+FAMILY(I,J)+ETA(I,J)+CHILD(I,J,K)+E(I,J,K)
        IF(W(I,J,K).GE.0.0)THEN
        RESPONSE(I,J,K)=1
        ELSE
        RESPONSE(I,J,K)=0
        END IF
   10 CONTINUE
        WRITE(4,9986)(RESPONSE(I,J,K),K=1,20)
        WRITE(3,9986)(CHILD(I,J,K),K=1,20)
   20 CONTINUE
        WRITE(2,9986)(FAMILY(I,J),J=1,20)
   30 CONTINUE
        WRITE (1,9986)(COMMUNITY(I),I=1,20)
 9986 FORMAT(20(1X,I1))
        CLOSE (1)
        CLOSE (2)
        CLOSE (3)
        CLOSE (4)
!
!       Introducing the parameters for calling subroutine BCONF.
!       XGUESS introduces the intial values, XLB introduces the lower bounds,
!       XUB introduces the upper bounds for the parameters, for XSCALE and
!       FSCALE see the manual.
!
        DATA XGUESS/0.9E0, 0.7E0, 0.7E0, 0.5E0, 0.5E0/, XSCALE/1.0E0,1.0E0,1.0E0,1.0E0,1.0E0/,
             FSCALE/1.0E0/
        DATA XLB/-10.0E0, -10.0E0, -10.0E0, 0.01E0,0.01E0/, XUB/10.0E0, 10.0E0, 10.0E0,
                 3.0E0,3.0E0/
!
!       Fitting model.
!       All bounds are provided
!
        IBTYPE = 0
!
!       Default parameters are used
!
        IPARAM(1) = 0
!
!       Minimization by calling subroutine BCONF using initial guesses.
!       FCN will be called from subroutine FCN, X is th vector of estimated
!       parameters, N is the number of the parameters and for other parameters
!       see the manual.
!
        CALL BCONF (FCN, N, XGUESS, IBTYPE, XLB, XUB, XSCALE, FSCALE, IPARAM, RPARAM, X, FVALUE)
!
!       Printing the results
!
        WRITE (9,99999) X, FVALUE
99999 FORMAT (5(1X,F10.6),5X,F16.6)
        WRITE (NOUT,*)'SOLUTION # ',SAMPLE
   40 CONTINUE
```

```
      END
!
!
!
!     The following subroutine calculates the minus two log liklehood function
!     that should be minimized by the main program. This gets N and X
!     from main program and returns F as minus two log liklehood function.
!
      SUBROUTINE FCN (N, X, F)
!
!     The following variables have the same definition as in the main program.
!
      INTEGER    COMMUNITY(20), FAMILY(20,20), CHILD(20,20,20), RESPONSE(20,20,20)
!
!     NMASS1 and NMASS2 are the number of the mass points.
!
      INTEGER    N, NMASS1, NMASS2, NR
!
!     XX is the vector of the mass points and their probabilities.
!
      REAL       X(N), F, XX(24)
      REAL       LINEAR, EXPLINEAR, LOGISTIC, LOGISTIC1, LOGISTIC2, LOGISTIC3, LOGLIK
      PARAMETER  (NMASS1=6, NMASS2=6, NR=20)
!
!     Reding data simulaled in main program.
!
      OPEN(5, FILE='COMMUNITYO.TXT')
      READ (5,9999)(COMMUNITY(I), I=1,20)
      OPEN(6, FILE='FAMILYO.TXT')
      READ (6,9999)((FAMILY(I,J), J=1,20), I=1,20)
      OPEN(7, FILE='CHILDO.TXT')
      READ (7,9999)(((CHILD(I,J,K), K=1,20), J=1,20), I=1,20)
      OPEN(8, FILE='RESPONSEO.TXT')
      READ (8,9999)(((RESPONSE(I,J,K), K=1,20), J=1,20), I=1,20)
9999  FORMAT(20(1X,I1))
      CLOSE (5)
      CLOSE (6)
      CLOSE (7)
      CLOSE (8)
!
!     Defining quadraturs and their probabilities. For each random effects
!     we need 6 quadratures and their probabilities.
!
          XX(1)=3.3243
          XX(2)=1.8892
          XX(3)=0.6167
          XX(4)=-0.6167
          XX(5)=-1.8892
          XX(6)=-3.3243
          XX(7)=3.3243
          XX(8)=1.8892
          XX(9)=0.6167
          XX(10)=-0.6167
          XX(11)=-1.8892
          XX(12)=-3.3243
          XX(13)=0.0025
          XX(14)=0.0886
          XX(15)=0.4089
          XX(16)=0.4089
          XX(17)=0.0886
          XX(18)=0.0025
          XX(19)=0.0025
          XX(20)=0.0886
          XX(21)=0.4089
          XX(22)=0.4089
          XX(23)=0.0886
          XX(24)=0.0025
!
!     Deviance evaluation using model 9 in paper.
!
      LOGLIK=0.0E0
      DO 50 I=1,NR
```

```
      LOGISTIC3=0.0E0
      DO 60 L=1,NMASS2
      LOGISTIC2=1.0E0
      DO 70 J=1,NR
      LOGISTIC1=0.0E0
      DO 80 M=1,NMASS1
!
!     We define LOGISTIC=1.0E2 insted of one to avoid mathematical error then reduce
!     1842.068.74 from LOGLIK according to the model (equation 9) explained in paper.
!
      LOGISTIC=1.0E2
      DO 90 K=1,NR
      LINEAR=X(1)*COMMUNITY(I)+X(2)*FAMILY(I,J)+X(3)*CHILD(I,J,K)+X(4)*XX(M)
            +X(5)*XX(NMASS1+L)
      EXPLINEAR=EXP(LINEAR)
      LOGISTIC=LOGISTIC*(EXP(LINEAR*RESPONSE(I,J,K))/(1+EXPLINEAR))
   90 CONTINUE
      LOGISTIC1=LOGISTIC1+LOGISTIC*XX(NMASS1+NMASS2+M)
   80 CONTINUE
      LOGISTIC2=LOGISTIC2*LOGISTIC1
   70 CONTINUE
      LOGISTIC3=LOGISTIC3+LOGISTIC2*XX(6+NMASS1+NMASS2+L)
   60 CONTINUE
!
!     Small positive number is added to LOGISTIC3 to avoid mathematical error when this
!     parameter is close to zero.
!
      LOGLIK=LOGLIK+LOG(LOGISTIC3+0.000000000000001)
   50 CONTINUE
      F=(-2)*(LOGLIK-1842.068074)
      RETURN
      END
```

```
!       This program generates the data sets according to the simulation plan explained in
!       paper.
!       Three mass points are used for Non-parametric approach to estimate the parameters in
!       equation 9 introduced in paper.
!       The program is written in Microsoft Developer Studio (Fortran Power Station 4.0).
!
!       Defining of variables. The variables COMMUNITY and GAMMA with one dimenSion is for third
!       level, variables FAMILY and ETA with two dimensions are for second level, and variables
!       CHILD, RESPONSE, W, and E are for the first level.
!
        INTEGER     i,j,k,NR,NOUT,COMMUNITY(20),FAMILY(20,20),CHILD(20,20,20),RESPONSE(20,20,20)
        REAL        RNNOF,E(20,20,20),ETA(20,20),GAMMA(20),UNIFORM,UNI,W(20,20,20)
!
!       Introducing the subroutines that should be called and is external to
!       this program.
!
        EXTERNAL    RNNOF, UMACH,RNBIN
!
!       Declaration of variables FOR MAXIMIZATION. NSAMPLE is the sample size.
!       For the definition of the other parameters one should see subroutines
!       LCONF and UMACH in the manual of Fortran Power Station 4.0.
!       FCN is the subroutine for this program.
!
        INTEGER     LDA, NCON, NEQ, NVAR, SAMPLE, NSAMPLE
        PARAMETER   (NCON=2, NEQ=2, NVAR=17, LDA=NCON, NSAMPLE=100,NR=20)
        INTEGER     IACT(100), MAXFCN, NACT
        REAL        A(NCON,NVAR), ACC, ALAMDA(NVAR), B(NCON), OBJ, SOL(NVAR), XGUESS(NVAR),
        REAL        XLB(NVAR), XUB(NVAR)
!
!       Introducing the subroutines that should be called and is external to
!       this program.
!
        EXTERNAL    FCN, LCONF
!
        DO 40 SAMPLE=1,NSAMPLE
!
!       Defining the output files for simulated data and parameter estimation.
!       Three files COMMUNITY, FAMILY, AND CHILD are used to save the sumulated data at
!       levels 3, 2, and 1 respectively.
!       The file RESPONSE is for saving the response at the third level.
!       The file NONPAOUT is for saving the parameters estimate.
!
        OPEN(1,FILE='COMMUNITY.TXT')
        OPEN(2,FILE='FAMILY.TXT')
        OPEN(3,FILE='CHILD.TXT')
        OPEN(4,FILE='RESPONSE.TXT')
        OPEN(9,FILE='NONPAOUT.TXT')
!
!       Generating the data. NR=20 is the sample size in each level.
!
!
!       The following three nested loops are used to generate a random effect and a fixed
!       effect at level 3, a random effect and a fixed effect at level 2, and a fixed
!       effect and a binomial response at level 1.
!
        CALL UMACH (2, NOUT)
        DO 30  I=1, NR
!
!       Generating Normal random number for third level as random effect.
!
        GAMMA(I)= RNNOF()
!
!       Generating Binomial random number in third level for using as two level factor.
!
        CALL RNBIN (1, 1, 0.5, IR)
        COMMUNITY(I)=IR
        DO 20  j=1, NR
!
!       Generating Normal random number for second level as random effect.
!
        ETA(I,J)= RNNOF()
```

```
!
!         Generating Binomial random number in second level for using as two level factor.
!
          CALL RNBIN (1, 1, 0.5, IR)
          FAMILY(I,J)=IR
          DO 10  K=1, NR
!
!         Generating Logistic random number for first level to establish logit model.
!
          CALL RNUN (1, UNIFORM)
          UNI=(UNIFORM)/(1-UNIFORM)
          E(I,J,K)=LOG(UNI)
!
!         Generating Binomial random number in first level for using as two level factor.
!
          CALL RNBIN (1, 1, 0.5, IR)
          CHILD(I,J,K)=IR
!
!         Generating Binomial random number in first level for using as binary response.
!         W(I,J,K) is used as a latent variable to generate the binomial response.
!
          W(I,J,K)=COMMUNITY(I)+GAMMA(I)+FAMILY(I,J)+ETA(I,J)+CHILD(I,J,K)+E(I,J,K)
          IF(W(I,J,K).GE.0.0)THEN
          RESPONSE(I,J,K)=1
          ELSE
          RESPONSE(I,J,K)=0
          END IF
   10 CONTINUE
          WRITE(4,9986)(RESPONSE(I,J,K),K=1,NR)

          WRITE(3,9986)(CHILD(I,J,K),K=1,NR)
   20 CONTINUE
          WRITE(2,9986)(FAMILY(I,J),J=1,NR)
   30 CONTINUE
          WRITE (1,9986)(COMMUNITY(I),I=1,NR)
 9986 FORMAT(20(1X,I1))
          CLOSE (1)
          CLOSE (2)
          CLOSE (3)
          CLOSE (4)
!
!         Fitting model.
!         Minimization  of -2LOGLIKELIHOOD.
!
!         Sum of the probabilities for three masses should be one i.e:
!         X(12)+X(13) + X(14)  .EQ.   1 and X(15)+X(16) + X(17)  .EQ.   1
!
!         Defining bounds for the standard errors of random effects:
!         0  .LT.  X(4)  .LE.  5 and 0  .LT.  X(5)  .LE.  5
!
          DATA A/0.0,0.0,0.0,0.0,0.0,0.0,0.0,0.0,0.0,0.0,0.0,0.0,0.0,0.0,0.0,0.0,0.0,
     &          0.0,0.0,0.0,0.0,1.0,0.0,1.0,0.0,1.0,0.0,0.0,1.0,0.0,1.0,0.0,1.0/,B/1.0,1.0/
          DATA XLB/-10,-10,-10,0.01,0.01,-100,-100,-100,-100,-100,-100,0.01,0.01,0.01,
     &            0.01,0.01,0.01/, XUB/10,10,10,5,5,100,100,100,100,100,100,0.99,0.99,
     &            0.99,0.99,0.99,0.99/, XGUESS/0.7,0.7,0.7,0.9,0.9,-1.0,0.0,1.0,-1.0,
     &            0.0,1.0,0.5,0.3,0.2,0.5,0.3,0.2/
          DATA ACC/0.0/,MAXFCN/800/
          CALL UMACH (2, NOUT)
!
!         Minimization by calling subroutine LCONF using initial guesses.
!         FCN will be called from subroutine FCN. For definition of the
!         parameters see the manual.
!
          CALL LCONF (FCN, NVAR, NCON, NEQ, A, LDA, B, XLB, XUB, XGUESS, ACC, MAXFCN, SOL, OBJ,
     &                NACT, IACT, ALAMDA)
!
!         Printing the parameters estimates.
!
          WRITE (NOUT,99998) 'Solution:',SAMPLE
          WRITE (9,99999) SOL,OBJ
99998 FORMAT (//, ' ', A, I4, I4)
99999 FORMAT (17(1X,F10.6),5X,F16.6)
```

```
!
!     Setting the maximum function evaluation to 800.
!
      MAXFCN=800
  40  CONTINUE
      END


!
!
!
!     The following subroutine calculates the minus two log liklehood function
!     that should be minimized by the main program. This gets N and X
!     from main program and returns F as minus two log liklehood function.
!
      SUBROUTINE FCN (N, X, F)
!
!     The following variables have the same definition as in the main program.
!
      INTEGER     COMMUNITY(20),FAMILY(20,20),CHILD(20,20,20),RESPONSE(20,20,20)
!
!     NMASS1 and NMASS2 are the number of the mass points.
!
      INTEGER     N,NMASS1,NMASS2,NR
      REAL        X(N), F
      DOUBLE PRECISION LINEAR,EXPLINEAR,LOGISTIC,LOGISTIC1,LOGISTIC2,LOGISTIC3,LOGLIK
      PARAMETER   (NMASS1=3,NMASS2=3,NR=20)
!
!     Reding simulated data from main program.
!
      OPEN(5,FILE='COMMUNITY.TXT')
      READ (5,9999)(COMMUNITY(I),I=1,NR)
      OPEN(6,FILE='FAMILY.TXT')
      READ (6,9999)((FAMILY(I,J),J=1,NR),I=1,NR)
      OPEN(7,FILE='CHILD.TXT')
      READ (7,9999)(((CHILD(I,J,K),K=1,NR),J=1,NR),I=1,NR)
      OPEN(8,FILE='RESPONSE.TXT')
      READ (8,9999)(((RESPONSE(I,J,K),K=1,NR),J=1,NR),I=1,NR)
9999  FORMAT(20(1X,I1))
      CLOSE (5)
      CLOSE (6)
      CLOSE (7)
      CLOSE (8)
!
!     Deviance evaluation using model 9 in paper.
!
      LOGLIK=0.0D0
      DO 50 I=1,NR
      LOGISTIC3=0.0D0
      DO 60 L=1,NMASS2
      LOGISTIC2=1.0D0
      DO 70 J=1,NR
      LOGISTIC1=0.0D0
      DO 80 M=1,NMASS1
!
!     We define LOGISTIC=1.0D+7 insted of one to avoid mathematical error then reduce
!     6447.23826 from LOGLIK according to the model (equation 9) explained in paper.
!
      LOGISTIC=1.0D+7
      DO 90 K=1,NR
      LINEAR=X(1)*COMMUNITY(I)+X(2)*FAMILY(I,J)+X(3)*CHILD(I,J,K)+X(4)*X(5+M)
     +       +X(5)*X(5+NMASS1+L)
      EXPLINEAR=EXP(LINEAR)
      LOGISTIC=LOGISTIC*(EXP(LINEAR*RESPONSE(I,J,K))/(1+EXPLINEAR))
  90  CONTINUE
      LOGISTIC1=LOGISTIC1+LOGISTIC*X(5+NMASS1+NMASS2+M)
  80  CONTINUE
      LOGISTIC2=LOGISTIC2*LOGISTIC1
  70  CONTINUE
      LOGISTIC3=LOGISTIC3+LOGISTIC2*X(8+NMASS1+NMASS2+L)
  60  CONTINUE
!
!     Small positive number is added to LOGISTIC3 to avoid mathematical error when this
```

```
!      parameter is close to zero.
!
       LOGLIK=LOGLIK+DLOG(LOGISTIC3+0.000000000000001)
 50 CONTINUE
       F=(-2)*(LOGLIK-6447.23826)
       RETURN
       END
```

Appendix E

```
# Bugs Codes (Software WinBUGS31) for MCMC approach. Data are simulated using
codes in appendix A.
model simulation
{
     for (i in 1:N){
      gamma[i]~dnorm(0.0,vargamma)
      for (j in 1:N){
      eta[i,j]~dnorm(0.0,vareta)
      for (k in 1:N){
      logit(mu[i,j,k])<-
betacom*community[i]+betafam*family[i,j]+betachi*child[i,j,k]+gamma[i]+eta[i,j]
      response[i,j,k]~dbin(mu[i,j,k],NUM)
      llike[i,j,k]<-response[i,j,k]*log(mu[i,j,k])+(1-response[i,j,k])*log(1-
mu[i,j,k])
                      }
                  }
       }
     betacom~dnorm(0.0,1.0E-6)
     betafam~dnorm(0.0,1.0E-6)
     betachi~dnorm(0.0,1.0E-6)
     vargamma~dgamma(1.0E-3,1.0E-3)
     vareta~dgamma(1.0E-3,1.0E-3)
     sigmaeta<-1.0/sqrt(vareta)
     sigmagamma<-1.0/sqrt(vargamma)
     llikelihood<-(-2)*sum(llike[,,])
}
```