

# A Diagnosis and Prognosis for Generalized Linear Mixed Models with Correlated Ordinal Responses: Power and Area under ROC Curve

Veeranun Pongsapukdee (Corresponding author) & Chayanat Phonork  
Department of Statistics, Faculty of Science, Silpakorn University, Nakhon-Pathom  
Snam-Chandra Palace Campus, Thailand  
E-mail: veeranun@su.ac.th/veeranun@hotmail.com

Received: July 4, 2011

Accepted: August 2, 2011

doi:10.5539/mas.v5n5p29

## Abstract

The Generalized Linear Mixed Models (GLMMs) are designed to account for the dependency inherent in data and permit both the fixed effects in the linear predictors and the random effects in the models. However, the GLMMs' implementations are still limited only in few applications due to the complexity of the model and its efficiency. In this article, we have investigated both the diagnosis and the prognosis of the ordinal-category logit GLMMs and the probit GLMMs for their power of the tests and the ability to predict the right categories in each condition of parameters and the sample size of the number of clusters and the cluster size. It is shown that the cumulative logit GLMM is superior to the probit GLMM. Furthermore, as the number of clusters and the cluster size are increased, the precision of parameter estimates through the power of the tests is much improved. However, the increasing of the intra-cluster correlation affects the AUC estimates, which probably mean that it gets difficulty to the prognosis of the right categories when there are many units in the same cluster, that is when the intra-cluster correlation occurs. But, such impact is only little for the power of the tests. Hence, in conclusion the GLMMs may well be recommended to use in applications since their power are very high and approaching to 1 for moderate and large number of clusters and the cluster size with the satisfactorily high AUC values..

**Keywords:** Generalized linear mixed models (GLMMs), Correlated ordinal responses, Cumulative logits, Probit models, Adaptive gaussian quadrature (AGQ), Area under ROC curves, Power of the tests

## 1. Introduction

For analyses of multilevel categorical response data, observations are often observed either nested within clusters (e.g., hospital, household, clinics) or repeatedly assessed across time (e.g., longitudinal ordinal outcomes). For analyses of such data, mixed effects models for ordinal response categories have developed (Ezzet & Whitehead, 1991, Hedeker & Gibbons, 1994). In addition, the implementation of statistical models such as Generalized Linear Models (GLMs, Nelder and Wedderburn, 1972; McCullagh and Nelder, 1983) as well as Generalized Linear Mixed Models (GLMMs) is currently interested, especially in the fields of medical diagnosis and engineer (Waegeman, et al., 2006). It is often used with the assessing of fits and the predicting ability of the model through the performance of diagnostic and prognostic technology. However, the GLMMs' implementations are still limited only in few applications or even rare in practices, due to the complexity and the interpretation of the models. Simpler alternative models are recently studied such as logit and fuzzy logic models using ROC curve analysis for predicting multilateral trade credit risks (Tang & Chi, 2005). They analyze the real data from WorldScope database and show that logit model outperforms the fuzzy logic model. In addition, mixed-effects logistic regression models are described for analysis of daily life longitudinal ordinal outcomes, where observations are observed clustered within subjects (Hedeker, et al., 2009). They illustrate the application of the mixed effects models for the data from ecological momentary assessment (EMA) studies that is interest in subject heterogeneity, both between- and within-subjects.

In this article we aim to primarily study the GLMMs from simulation studies in two aspects, firstly the diagnosis tests of the model through the statistical power and secondly the prognosis ability mainly through the area under the ROC curves.

The GLMMs are further extended from GLMs that permit only the fixed effects in the linear predictors by including the random effects in the model. Following up to Searle's classic linear models (Searle, 1997, 1971)

and variance components of Searle, Casella, and McCulloch (1992), the modern perspective on GLMMs can be useful to provide a unified and accessible treatment of the newest statistical methods for analyzing correlated and non-normally distributed data (McCulloch and Searle, 2001). Traditional diagnosis tests include different kinds of information; for example, medical tests, signs, symptoms, statistical modeling. Then, all interpretations are commonly investigated and evaluated. Meanwhile, assessment of predictive accuracy is a critical aspect of evaluating and comparing models. Thus, the accuracy of diagnosis is essential especially in medical care and that the attributes of the diagnosis tests should be further studied and measured. Sensitivity, specificity, accuracy, and the Receiver Operating Characteristic (ROC) curve or the Area Under the ROC Curve (AUC) is often used for these attributes. The AUC is used as a standard performance measure in many fields where a binary classification system is needed. A meaning of the area under an ROC curve in terms of the result such that of a signal detection experiment does correspond to the probability of correctly identifying which of the two stimuli is “noise” and which is ‘signal plus noise” (Hanley & McNeil, 1982). In medical imaging studies, the more economical rating method is generally used: images from diseased and non-diseased subjects are thoroughly mixed, then presented in this random order to a reader who is asked to rate each on a categorical or discrete ordinal scale ranging from definitely normal to definitely abnormal, that generally a five- categories scale is used. The AUC can be also viewed and obtained from a rating experiment that has the same meaning as it has when it is derived from the experiment in a more indirect way, i.e., by successively considering broader categories of abnormal (e.g., category 5 alone, categories 5 plus 4, or categories 5 plus 4, plus 3, compared with the rest of categories).

Diagnostic and prognostic models such as GLMs and GLMMs are also increasingly common in the biomedical literature. They can be composed of multiple prognostic factors related to each other and an outcome by a complex statistical model such as logistic regression, logit model, loglinear model, and another form of GLMs and GLMMs. These models are useful to physicians, clinicians and other users who feel comfortable with the complexity of the models and their ability to help them to prognosticate about a clinically important patient-related situation or event under studies (Harrell, et al., 1996; Van & Lee, 1990).

Therefore, in this article, we have empirically investigated the performing of the modeling the ordinal response variable, primarily through the mixed modeling, GLMMs. The random effects usually apply to a sample; for example, the model may treat observations from a given clinics as a cluster, and it has both a random effect in each clinic and a fixed effect in each treatment. The efficiency of GLMMs’ parameter estimations is based on a suitable and efficient estimation approach for GLMMs called Adaptive Gaussian Quadrature (AGQ, Pinheiro & Bates, 1995). All corresponding simulated data are performed and carried out to compare for the accuracy and the efficacy of GLMMs considering on the statistical power and the AUC from 1,000 datasets in each sample size (number of clusters and cluster size) and the intra-cluster correlations. All simulation works are generated and processed using the authors’ developed macro run with SAS® version 9.1.

**2. Modeling procedures of GLMs and GLMMs**

*2.1 Models for fixed effects (GLMs)*

Let N be the number of independent clusters or subjects,  $n_i$  the number of observations for cluster  $i, i = 1, \dots, N$  and let  $Y_{ij}$  denote the response  $j, j = 1, \dots, n_i$ , whose values consist of  $k$  level,  $k = 1, 2, \dots, c$ . Furthermore, let  $\mathbf{x}_{ij} = (x_{1ij}, \dots, x_{pij})'$  denote a column vector of p-covariates associated with the response  $j$ , in cluster  $i$ . We then assume a cumulative logit model in the form:

$$\text{logit}[P(Y_{ij} \leq k | \mathbf{x}_{ij})] = \eta_{ijk} = \alpha_k + \mathbf{x}'_{ij}\boldsymbol{\beta}, \quad k = 1, \dots, c - 1. \quad \dots\dots\dots (1)$$

Where  $\alpha_k$  denote the cutpoint for the response level  $k, \alpha_1 < \alpha_2 < \dots < \alpha_{c-1}$  and  $\boldsymbol{\beta}$  denote the fixed effects of the explanatory variables. Thus,

$$P(Y_{ij} \leq k) = 1 / [1 + \exp(-\eta_{ijk})],$$

and  $\pi_k(\mathbf{x}_{ij}) = P(Y_{ij} = k | \mathbf{x}_{ij}) = P(Y_{ij} \leq k | \mathbf{x}_{ij}) - P(Y_{ij} \leq k - 1 | \mathbf{x}_{ij})$ .

The maximum likelihood function for the ordinal response for the ML approach is

$$L(\boldsymbol{\beta}) = \prod_{i=1}^N \prod_{j=1}^{n_i} \prod_{k=1}^c \pi_k(\mathbf{x}_{ij})^{y_{ijk}} .$$

Where  $y_{ijk} = 1$  if  $Y_{ij} = k$  and  $y_{ijk} = 0$  if  $Y_{ij} \neq k$ . The log-likelihood function is

$$l(\boldsymbol{\beta}) = \log L(\boldsymbol{\beta}) = \sum_{i=1}^N \sum_{j=1}^{n_i} \sum_{k=1}^c y_{ijk} \log \pi_k(\mathbf{x}_{ij}).$$

The estimates of  $\boldsymbol{\beta}$  (including  $\alpha_k$ ) that are the results from Fisher scoring (Agresti, 2002) have the form

$$u(\boldsymbol{\beta}) = \frac{\partial l(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}} = \sum_{i=1}^N \mathbf{D}'_i \mathbf{V}_i^{-1} (\mathbf{Y}_i - \boldsymbol{\pi}_i) = \mathbf{0}.$$

Where,  $D_i$  denotes the matrix of derivatives for the element  $i$ ,  $\frac{\partial \boldsymbol{\pi}_i}{\partial \boldsymbol{\beta}} = \frac{\partial \boldsymbol{\pi}_i}{\partial \boldsymbol{\eta}_i} \frac{\partial \boldsymbol{\eta}_i}{\partial \boldsymbol{\beta}} = \frac{\partial \boldsymbol{\pi}_i}{\partial \boldsymbol{\eta}_i} \mathbf{x}_i$

and  $\mathbf{V}_i = [\text{diag}\{\text{var}(Y_{ij})\}]$ ,  $i = 1, \dots, N$ .

In some situations the probit link function for GLMS are probably able to give the best power of the tests for every goodness-of-fit test statistic (Pongsapakdee & Sukgumphaphan, 2008). In classical general linear model the response variable values are expected to follow the normal distribution, and the link function is a simple identity function. For the developed GLMs the response variable follows the exponential family distribution models, and the most often used link functions include logit, probit, complementary-log-log, and also the log links.

### 2.2 Models for mixed and random effects (GLMMs)

Let  $\mathbf{u}_i$  denote a column vector of random effect values for cluster  $i$ . This is common to all observations in the cluster. Let  $\mathbf{z}_{ij}$  denote a column vector of their explanatory variables. The linear predictor for a cumulative logit GLMM has the form (2),

$$\text{logit}[P(Y_{ij} \leq k | \mathbf{x}_{ij}, \mathbf{u}_i)] = \alpha_k + \mathbf{x}'_{ij} \boldsymbol{\beta} + \mathbf{z}'_{ij} \mathbf{u}_i, \quad k = 1, \dots, c - 1. \tag{2}$$

The random effect  $\mathbf{u}_i$  is assumed to have a multivariate normal distribution  $N(\mathbf{0}, \boldsymbol{\Sigma}_u)$ . The covariance matrix  $\boldsymbol{\Sigma}_u$  depends on unknown variance components and possibly also the correlation parameters. The special case with univariate random effect and  $\mathbf{z}'_{ij} = 1$  are straight forward. Therefore, with  $\mathbf{u}_i$  replaced by  $u_i$ , where  $\{u_i\}$  are  $N(0, \sigma_u^2)$ , from which, (1) and (2) it follow that the cumulative logit GLMM and the cumulative probit GLMM have respectively the forms:

$$\text{logit}[P(Y_{ij} \leq k | \mathbf{x}_{ij}, u_i)] = \alpha_k + \mathbf{x}'_{ij} \boldsymbol{\beta} + u_i, \quad k = 1, \dots, c - 1; \tag{3}$$

and

$$\text{probit}[P(Y_{ij} \leq k | \mathbf{x}_{ij}, u_i)] = \alpha_k + \mathbf{x}'_{ij} \boldsymbol{\beta} + u_i, \quad k = 1, \dots, c - 1. \tag{4}$$

This has the similarly form of ordinary GLM with unobserved values  $\{u_i\}$  of a particular covariate. The random effects also provide a mechanism for explanatory over dispersion in basic models not having those effects (Breslow and Clayton, 1993). Conditional on  $\mathbf{u}_i$ , the conditional likelihood function for cluster  $i$  have the form

$$f(Y_i | \mathbf{u}_i; \boldsymbol{\beta}) = \prod_{j=1}^{n_i} \prod_{k=1}^c \pi_k^u(\mathbf{x}_{ij})^{y_{ijk}}.$$

Where,  $\pi_k^u(\mathbf{x}_{ij}) = P(Y_{ij} = k | \mathbf{x}_{ij}, \mathbf{u}_i) = P(Y_{ij} \leq k | \mathbf{x}_{ij}, \mathbf{u}_i) - P(Y_{ij} \leq k - 1 | \mathbf{x}_{ij}, \mathbf{u}_i)$ .

Without conditional on  $\mathbf{u}_i$ , the conditional likelihood function for cluster  $i$  (marginal) have the form

$$L_i(\boldsymbol{\beta}, \boldsymbol{\Sigma}_u) = f(Y_i; \boldsymbol{\beta}, \boldsymbol{\Sigma}_u) = \int f(Y_i | \mathbf{u}_i; \boldsymbol{\beta}) \times f(\mathbf{u}_i; \boldsymbol{\Sigma}_u) d\mathbf{u}_i.$$

Then its log likelihood function is  $l(\boldsymbol{\beta}, \boldsymbol{\Sigma}_u) = \log L = \sum_{i=1}^N \log L_i$ . The log likelihood function is evaluated numerically and maximized as a function of  $\boldsymbol{\beta}$  and  $\boldsymbol{\Sigma}_u$  or Monte Carlo in combination with

Newton-Rhaphson or Fisher scoring may be used iteratively, (Agresti, 2002). Instead of using this MLE, we use the Adaptive Gaussian Quadrature (AGQ) approach which was proposed by Pinheiro & Bates (1995) and can be used efficiently and conveniently in SAS package. The approximation is a finite weighted sum that evaluates the function at certain points in the form

$$\int_{-\infty}^{\infty} f(\mathbf{u}) \exp(-\mathbf{u}^2) d\mathbf{u} \approx \sum_{i=1}^q w_i f(d_i).$$

Where,  $w_i$  denote weights,  $d_i$  denote quadrature points that are tabulated. The approximation improves as  $t$ , the number of quadrature points ( $t$ ) increases.

Besides this, there are several GLMMs' parameter-estimation procedures as followings: Penalized quasi-likelihood (PQL) which was proposed by Green (1987). His integral has closed form. This type of integral approximation is called Laplace approximation that evaluates the function at certain points around  $\bar{\mathbf{u}} \approx E(\mathbf{u} | \mathbf{Y})$  in the form of  $q(\boldsymbol{\beta}, \mathbf{Y}) - (1/2) \bar{\mathbf{u}}' \boldsymbol{\Sigma}_u^{-1} \bar{\mathbf{u}}$ , where,  $q(\boldsymbol{\beta}, \mathbf{Y})$  is similar to the quasi-log-likelihood of GLMs. Then, Monte Carlo method in combination with Newton-Rhaphson or Fisher scoring may be used iteratively. However, this method still need further study as for example in case when the response is binary, PQL tends to produce estimates with negative bias (Breslow and Lin, 1995). Generalized Estimating Equation (GEE) is an alternative approach that was originally proposed by Liang and Zeger (1986) for modeling univariate marginal distribution, such that of the binomial and Poisson. GEE is an alternative to ML estimation but does not completely specify the joint distribution, then it does not have a likelihood function. The quasi-likelihood parameter estimates are the solutions of quasi-score equations,  $u(\boldsymbol{\beta}) = \frac{\partial l(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}} = \sum_{i=1}^N \mathbf{D}'_i \mathbf{V}_i^{-1} (\mathbf{Y}_i - \boldsymbol{\pi}_i) = \mathbf{0}$ ,

where,  $\boldsymbol{\pi}_i$  denote the probability associated with  $\mathbf{Y}_i$ ,  $\mathbf{D}_i$  denotes the matrix of derivatives for the element  $i$ ,  $\frac{\partial \boldsymbol{\pi}_i}{\partial \boldsymbol{\beta}}$ , and  $\mathbf{V}_i = \mathbf{A}_i^{1/2} \mathbf{R}_i(\rho) \mathbf{A}_i^{1/2}$  is the working covariance matrix for  $\mathbf{Y}_i$ , whereas  $\mathbf{A}_i$  denotes working correlation matrix. However, its consistency also follows from general results for unbiased estimating functions (Liang and Zeger, 1995). For more details Lipsitz et al. (1994) also outlined a GEE approach for cumulative logit models with repeated ordinal responses. Another last procedure that may be used for GLMMs is the Marginal Quasi-Likelihood (MQL) method which was proposed by Goldstein (1991). The model is developed for only the marginal distribution that has the form  $E(\mathbf{y}_i) = \boldsymbol{\mu}_i = \mathbf{g}^{-1}(\mathbf{x}'_i \boldsymbol{\beta}^*)$  or  $\mathbf{y}_i = \boldsymbol{\mu}_i^u + \boldsymbol{\varepsilon}_i$ , where,  $\mathbf{u} \sim N(\mathbf{0}, \boldsymbol{\Sigma}_u)$  and  $\text{var}(\boldsymbol{\varepsilon}_i) = a_i \nu(\boldsymbol{\mu}_i^u)$ ,  $a_i$  is a constant and  $\nu(\boldsymbol{\mu}_i^u)$  denotes the function of the conditional mean. The likelihood parameter estimates are the solutions for GLMs which follow the equation  $u(\boldsymbol{\beta}) = \mathbf{D}'[\text{var}(\mathbf{y})]^{-1}(\mathbf{y} - \boldsymbol{\mu})$  in the form,  $\mathbf{x}'(\mathbf{D}\mathbf{V}_0\mathbf{D} + \mathbf{z}\boldsymbol{\Sigma}_u\mathbf{z}')^{-1}\mathbf{D}(\mathbf{y} - \boldsymbol{\mu}) = \mathbf{0}$ , where,  $\text{var}(\mathbf{y}) = \mathbf{V}_0 + \mathbf{D}^{-1}\mathbf{z}\boldsymbol{\Sigma}_u\mathbf{z}'\mathbf{D}^{-1}$ ,  $\mathbf{D}$  and  $\mathbf{V}$  denote the diagonal matrices whose element are  $\frac{\partial \boldsymbol{\mu}_i}{\partial \boldsymbol{\beta}}$  and  $a_i \nu(\boldsymbol{\mu}_i^u)$ , respectively.

### 3. Simulation and Statistical Analyses

From the models (3) & (4) in Section 2.2, the simulations have been conducted for the ordinal response ( $\mathbf{Y}$ ) with three categories or  $K=3$ , and for two explanatory variables,  $X_1 \sim \text{Bernoulli}(0.5)$  and  $X_2 \sim \text{Bernoulli}(0.5)$  under the fixed effects ( $\mathbf{X}$ 's) and the random effects ( $\mathbf{Y}$ ,  $\mathbf{U}$ ), where,  $\mathbf{Y}$  is from the true model,  $\mathbf{U}$  is from  $\mathbf{U} \sim N(0, \sigma_u^2)$ . The cumulative logit GLMM (3) and the cumulative probit GLMM (4) are fitted for the data under each condition of the number of clusters ( $N = 50, 100, 250$ ), the cluster size ( $n_i = 5, 7, 10$ ), and the intra-cluster correlation  $ICC = \sigma_u^2 / (\sigma_u^2 + \sigma_\varepsilon^2)$  from which it follows that ( $ICC = 0.05, 0.20, 0.40$ ) which are corresponding to the variance of the random effects ( $\sigma_u^2 = 0.17, 0.82, 2.19$ ), where  $\sigma_\varepsilon^2 = \pi^2 / 3$ . The models' parameter values  $\alpha_1 = -1$ ,  $\alpha_2 = 1$ ,  $\beta_1 = 0.5$ , and  $\beta_2 = 0.5$ , are chosen similarly to those of Spiess & Hamerle, 2000). Each condition is processed for each model fitted and carried out under the AGQ parameter estimation procedure for at least 1,000 repeated simulations using the developed macro program run with the SAS version 9.1. Statistical analyses are assessing of the goodness of fits of models using the likelihood ratio statistics for the models (3) & (4). Then, the Area under ROC curve (AUC), corresponding to the plots of the sensitivity =  $P(\hat{Y} \leq 1 | Y \leq 1)$  vs. the specificity =  $P(\hat{Y} > 1 | Y > 1)$  (that is implemented in some advanced statistical packages such the practical SAS<sup>®</sup>

implementations) and the power of the tests (power = 1 - type II error), corresponding to the proportion of the rejection of  $H_0$  when  $H_0$  is false in 1,000 simulations, are computed and analyzed for all conditions.

#### 4. Results

The results in Table 1 show that the area under the ROC curve (AUC) and the power of the test are obtained for the parameters intra-cluster correlation (ICC = 0.05, 0.20, 0.40); the number of clusters ( $N = 50, 100, 250$ ), and the cluster size ( $n_i = 5, 7, 10$ ), under the logit GLMM and the probit GLMM models, respectively.

For the AUC, it is found that the AUC increases as the ICC and N increase but the AUC does decrease when the cluster size, n increases. These AUC's outcomes are similar results for both the logit GLMM and the probit GLMM. However, all the AUC's results are most satisfied because they are approaching 1 as N and ICC are large. Moreover, the logit GLMM is a bit superior to the probit GLMM due to the bigger AUC values (Table 1 and Figures 1-3).

When considering on the power of the tests, it is shown that for moderate and big N and n, the power is very high and approximately approaches to its maximum value, 1, for almost all ICC, N and n, except only for small N and n. However, the power increases as the ICC decreases. In conclusion, all results show that the power increases as the N and n increases. Furthermore, using the logit GLMM are probably more suitable than using the probit GLMM due to the bigger AUC values and the higher power of the tests (Table 1 and Figures 4-6).

#### 5. Discussion

We illustrate that for ordinal correlated responses with categorical explanatory variables, the logit GLMM is superior to the probit GLMM for analyses and the assessing of model fits, as the same way as the GLMs do. That is in the case when the explanatory variables are all categorical data, the logit model is preferred to both the probit model and the logistic regression models (Agresti, 2002). In addition, in the case when the explanatory variables are mixed categorical and continuous variables under a real set of data, the logit model is preferred to the fuzzy models (Tang & Chi, 2005). Moreover, in certain cases when the explanatory variables are all continuous data or are mixed continuous and categorical data, the probit model outperforms the logit models (Pongsapakdee & Sukgumpaphan, 2008). Furthermore, these GLMMs' results reveal that as the number of clusters and the cluster size are increased, the precision of parameter estimates through the power of the tests is much improved. Although, for increasing of the intra-cluster correlation, it does affect the AUC estimates, this probably means that when the intra-cluster correlation occurs it might get a bit more difficult to classify units for the prognosis of categories due to many correlated units in the same cluster. But, it is found that such impact is only in small amount for the power of the tests which is normally a well known criterion for assessing the efficiency of models. Hence, the logit GLMM model trends to be suitable model and it may well be recommended to use for the correlated ordinal response categories because not only the power of the tests are very high for both moderate and big sample sizes of the number of clusters and that of the cluster size (Table 1 and Figures 4-6), but also its AUC values are quite high (Table 1 and Figures 1-3).

#### 6. Conclusion

The results reveal that as the number of clusters and the cluster size are increased, the precision of parameter estimates through the power of the tests is much improved. Both the power estimates and the magnitude of the AUC values are getting on satisfactorily high. Although, for increasing of the intra-cluster correlation, it does affect the AUC estimates of models parameters; however, such impact is only little for the power of the tests, which is normally a well known criterion for assessing the efficiency of models. Hence, considering the bigger AUC values and that the power of the tests are still very high for moderate and big sample sizes of both the number of clusters and the cluster size, then the logit GLMM model will be preferred to the probit GLMM for the correlated ordinal response categories.

#### References

- Agresti, A. (2002). *Categorical Data Analysis*. 2nd Ed. New York: John Wiley & Sons.
- Breslow N. E. & Clayton, D. G. (1993). Approximate inference in generalized linear mixed model. *Journal of the American Statistical Association*. 88(421), 9-25, doi:10.2307/2290687, <http://dx.doi.org/10.2307/2290687>
- Ezzet, F. & Whitehead, J. (1991). A random effects model for ordinal responses from a crossover trial. *Statistics in Medicine* 10(6). 901-907. doi:10.1002/sim.4780100611, <http://dx.doi.org/10.1002/sim.4780100611>
- Goldstein, H. (1991). Nonlinear multilevel models with an application to discrete response data. *Biometrika*. 78(1), 45-51, doi:10.2307/2336894, <http://dx.doi.org/10.2307/2336894>

- Goldstein, H. (1995). *Multilevel Statistical Models*. 2nd Ed. London: Edward Arnold.
- Hedeker, D. & Gibbons, R.D. (1994). A random-effects ordinal regression model for multi-level analysis. *Biometrics*. 50(4), 933-944, doi:10.2307/2533433, <http://dx.doi.org/10.2307/2533433>
- Hedeker, D., Demirtas, H. & Mermelstein, R.J. (2009). A mixed ordinal location scale model for analysis of ecological momentary assessment (EMA) data. *Statistics and Its Interface*. 2(4), 391-401.
- Green, P. J. (1987). Penalized likelihood for general semi-parametric regression models. *Int. Statist Rev.* 55(3), 245-260, doi:10.2307/1403404, <http://dx.doi.org/10.2307/1403404>
- Hanley J.A. & McNeil B.J. (1982). The meaning and use of the area under a Receiver Operating Characteristic (ROC) Curve. *RADIOLOGY*.143(1), 29-36.
- Harrell, F.E., Lee, K.L., & Mark, D.B. (1996). Multivariate prognostic models: Issues in developing models, evaluating assumptions and adequacy, and measuring and reducing errors. *Stat Med*. 15(4), 361-387, doi:10.1002/(SICI)1097-0258(19960229)15:4<361::AID-SIM168>3.0.CO;2-4, [http://dx.doi.org/10.1002/\(SICI\)1097-0258\(19960229\)15:4<361::AID-SIM168>3.0.CO;2-4](http://dx.doi.org/10.1002/(SICI)1097-0258(19960229)15:4<361::AID-SIM168>3.0.CO;2-4)
- Liang, K. Y. & Zeger, S. L. (1986). Longitudinal data analysis using generalized linear models. *Biometrika*. 73(1), 13-22.
- Liang, K. Y. & Zeger, S. L. (1995). Inference based on estimating functions in the presence of nuisance parameters. *Statist. Sci.* 10(2), 158-173, doi:10.1214/ss/1177010028, <http://dx.doi.org/10.1214/ss/1177010028>
- Lipsitz, S.R., KIM, K. & Zhao, L. (1994). Analysis of repeated categorical data using generalized estimating equations. *Statist. Medic.* 13(11), 1149-1163, doi:10.1002/sim.4780131106, <http://dx.doi.org/10.1002/sim.4780131106>
- Liu, Q. & Pierce, D. A. (1994). A note on Gauss-Hermite quadrature. *Biometrika*. 81(3), 624-629, doi:10.2307/2337136, <http://dx.doi.org/10.2307/2337136>
- Kuss, O. & McLerran, D. (2007). A note on the estimation of the multinomial logistic model with correlated response in SAS. *Computer Method and Programs in Medicine* 87(3). 262-269, doi:10.1016/j.cmpb.2007.06.002, <http://dx.doi.org/10.1016/j.cmpb.2007.06.002>
- McCullagh, P. & Nelder, J.A. (1983). *Generalized Linear Models*. London. Chapman & Hall.
- McCulloch, C.E. & Searle, S. (2001). *Generalized, Linear, and Mixed Models*. New York. John Wiley & Sons.
- McCullagh, P. (1980). Regression models for ordinal data (with discussion). *Journal of the Royal Statistical Society*. B 42(2), 109-142.
- McFadden, D. (1974). *Conditional logit analysis of qualitative choice behavior in Frontiers in Econometrics*, ed. by P. Zarembka. New York. Academic Press, 105-142.
- Nelder, J.A. & Wedderburn, R.W.M. (1972). Generalized linear models. *Journal of the Royal Statistical Society*. A135(3), 370-384.
- Pinheiro, J. C. & Bate, D. M. (1995). Approximations to the log-likelihood function in the nonlinear mixed-effects models. *Journal of Computational and Graphical Statistics*. 4(1), 12-35, doi:10.2307/1390625, <http://dx.doi.org/10.2307/1390625>
- Pongsapakdee, V. & Sukgumphaphan, S. (2008). Three cumulative link models for ordinal responses: A review of methods and power comparisons. *Songklanakarin J. of Sci. Technol.* 30(6), 805-811.
- Searle, S. R., G. Casella, & C.E. McCulloch (1992). *Variance Components*. New York. John Wiley & Sons.
- Searle, S. R. (1997). *Linear Models*. Classic Edition. New York. John Wiley & Sons.
- Spiess, M. & Hamerle, A. (2000). A comparison of different methods for the estimation of regression models with correlated binary responses. *Computational Statistics and Data Analysis*. 33(4), 439-455, doi:10.1016/S0167-9473(99)00065-1, [http://dx.doi.org/10.1016/S0167-9473\(99\)00065-1](http://dx.doi.org/10.1016/S0167-9473(99)00065-1)
- Tang, T.C. & Chi, L.C. (2005). Predicting multilateral trade credit risks: Comparisons of logit and fuzzy logic models using ROC curve analysis. *Expert Systems with Applications*. 28(3), 547-556. doi:10.1016/j.eswa.2004.12.016, <http://dx.doi.org/10.1016/j.eswa.2004.12.016>
- Van, H. & Lee, C.S. (1990). Predictive values of statistical models. *Stat Med*. 9(11), 1303-1325, doi: 10.1002/sim.4780091109, <http://dx.doi.org/10.1002/sim.4780091109>

Waegeman, W., Baets. B.D. & Boullart, L. (2006). A comparison of different ROC measures for Ordinal regression. *Proceedings of the ICML 2006 workshop on ROC analysis in machine learning*, Pittsburg, USA.

Walker, S.H. & Duncan, D.B. (1967). Estimation of probability of an event as a function of several independent variables. *Biometrika*. 54(1), 167-179, doi:10.2307/2333860, <http://dx.doi.org/10.2307/2333860>

Zeger, S. L., Liang, K. Y. & Albert, P. S.( 1988). Models for longitudinal data: A generalized estimating equation approach. *Biometrics*. 44(4), 1049-1060, doi:10.2307/2531734, <http://dx.doi.org/10.2307/2531734>

Table 1. Area under ROC Curve (AUC) and Power of the test, classified by Intra-Cluster Correlation (ICC), Number of clusters (N), and Cluster size (n)

Intra-cluster correlation (ICC)	Number of Clusters (N)	Cluster size (n)	Area under ROC Curve (AUC)		Power of the test (Power)	
			Logit	Probit	Logit	Probit
0.05	50	3	0.71	0.703	0.475	0.428
		5	0.69	0.683	0.728	0.671
		7	0.686	0.681	0.872	0.827
	100	3	0.711	0.703	0.805	0.708
		5	0.692	0.685	0.964	0.918
		7	0.688	0.682	0.997	0.98
	250	3	0.712	0.702	0.998	0.927
		5	0.699	0.691	1	0.98
		7	0.692	0.689	1	1
0.2	50	3	0.823	0.819	0.428	0.426
		5	0.805	0.802	0.683	0.676
		7	0.789	0.787	0.838	0.839
	100	3	0.834	0.83	0.74	0.731
		5	0.807	0.804	0.952	0.949
		7	0.79	0.788	0.987	0.986
	250	3	0.842	0.838	0.991	0.989
		5	0.81	0.807	1	1
		7	0.792	0.79	1	1
0.4	50	3	0.897	0.895	0.354	0.355
		5	0.869	0.868	0.593	0.594
		7	0.852	0.851	0.761	0.762
	100	3	0.899	0.897	0.642	0.648
		5	0.869	0.868	0.903	0.903
		7	0.853	0.852	0.975	0.975
	250	3	0.9	0.898	0.971	0.968
		5	0.869	0.868	1	1
		7	0.855	0.854	1	1

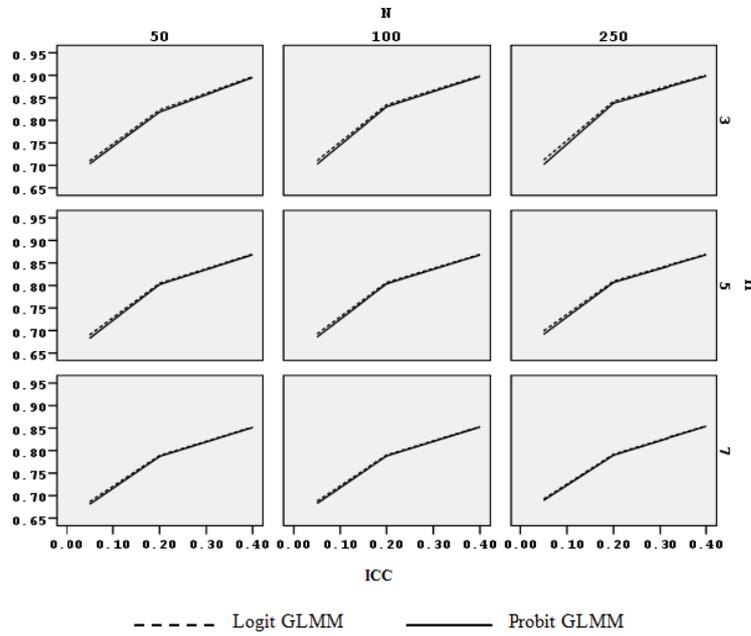


Figure 1. Area under ROC Curve (AUC) vs. Intra-Cluster Correlation (ICC) Classified by Number of cluster (N), Cluster size (n), and Models fitted.

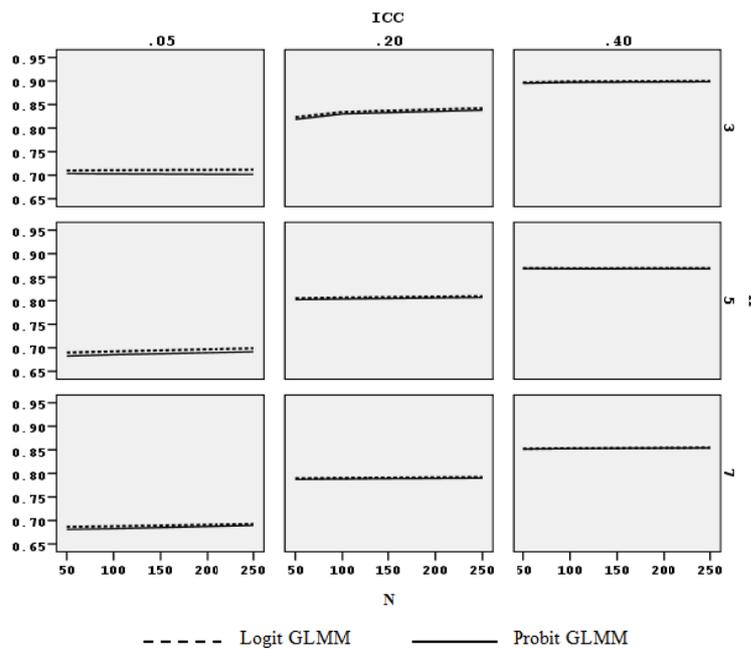


Figure 2. Area under ROC Curve (AUC) vs. Number of cluster (N) Classified by Intra-Cluster Correlation (ICC), Cluster size (n), and Models fitted.

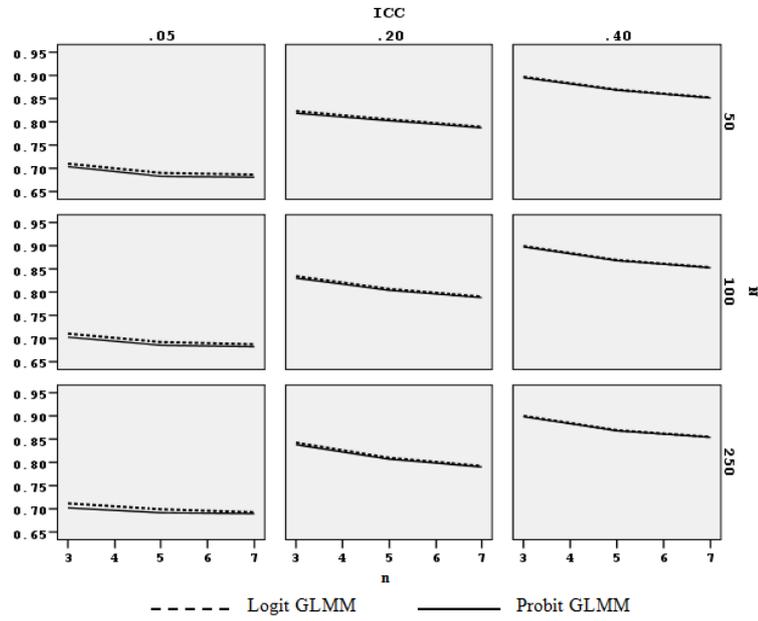


Figure 3. Area under ROC Curve (AUC) vs. Cluster size (n) Classified by Intra-Cluster Correlation (ICC), Number of cluster (N) and Models fitted.

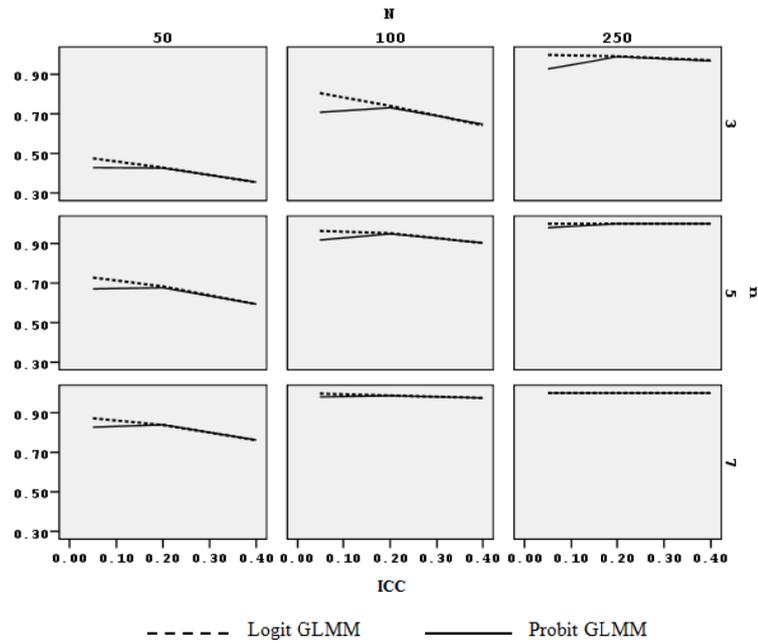


Figure 4. Power of the tests (Power) vs. Intra-Cluster Correlation (ICC) Classified by Number of cluster (N), Cluster size (n), and Models fitted.

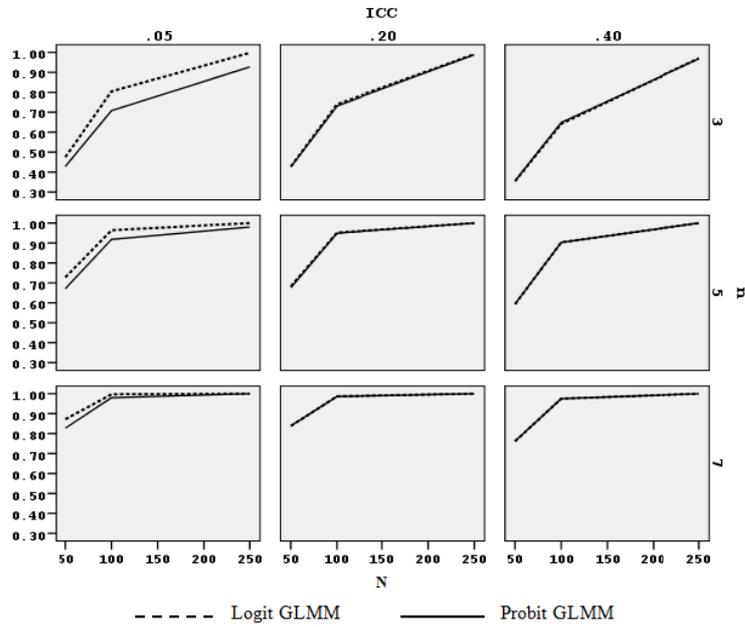


Figure 5. Power of the tests (Power) vs. Number of cluster (N) Classified by Intra-Cluster Correlation (ICC), Cluster size (n), and Models fitted.

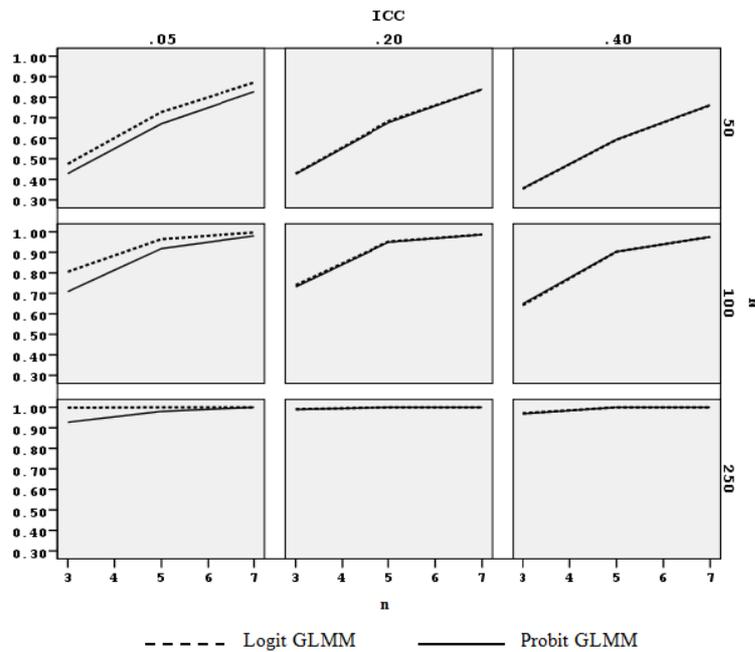


Figure 6. Power of the tests (Power) vs. Cluster size (n) Classified by Intra-Cluster Correlation (ICC), Number of cluster (N) and Models fitted.