Local Odds Ratio Is More Efficient Than Correlation Coefficient For Modeling Longitudinal Ordinal Data

Xinkai Zhou*[1], Ronghui Xu[2], and David Elashoff[1,3]

[1]Department of Medicine Statistics Core, University of California, Los Angeles
[2]Department of Mathematics and Department of Family and Preventive Medicine, University of California, San Diego
[3]Department of Biostatistics, University of California, Los Angeles

Abstract
While correlation coefficient is commonly used for parameterizing the "working" correlation structure in generalized estimating equations (GEE) for modeling longitudinal ordinal data using the proportional odds cumulative logit model, it is well known that its range is severely constrained as a result of the Fréchet bound. Although alternative parameterizations have been proposed, a direct comparison between them is lacking. Consequently, analysts usually fall back to the correlation coefficient method as the default option even though they are aware of its potential problems. To inform modeling choice, this paper conducted a simulation study and found that the correlation coefficient approach is not optimal in a wide range of scenarios. In fact, we found that the local odds ratio approach can achieve up to 30% efficiency gains (in a sense that will be defined in the article) compared to the correlation coefficient approach.

Key Words: Longitudinal Ordinal Data, Proportional Odds Cumulative Logit Model, Generalized Estimating Equations, Local Odds Ratio

1. Introduction

Parameter estimation for the proportional odds cumulative logit model in the longitudinal ordinal data setting has long been dominated by utilizing the generalized estimating equations (GEE) method (Liang and Zeger, 1986) where the association within subjects is parameterized by correlation coefficients (Clayton, 1992; Lipsitz et al., 1994; Parsons et al., 2009). While correlation coefficient can capture any association pattern between continuous variables, it may fail to do so for ordinal variables. In fact, its range is severely restricted by the marginal model. To mitigate this problem, Touloumis et al. (2013) proposed a brand new approach that has the potential to eradicate the problem and to drastically improve model parameter estimation. Specifically, they recognized α as a nuisance vector and parameterized it using local odds ratios, whose values are not restricted. Since the consistency of the GEE estimator does not depend on the correct specification of the "working" correlation structure, substituting local odds ratios for correlation coefficients has no effect on its asymptotic property. With that said, it is not clear how the local odds ratio approach performs at small to medium sample sizes, and whether it improves parameter estimation compared to the correlation coefficient method.

However, the comparison is not completely straightforward. For example, simulating correlated ordinal data sets can be a challenge. While a few methods have been proposed (Amatya and Demirtas, 2015; Barbiero and Ferrari, 2017), they require direct specification of the correlation structure between repeated ordinal

*Email: xinkaizhou@mednet.ucla.edu

responses. Since, as previously mentioned, the range of correlation coefficients between ordinal variables is severely limited, it is nearly impossible to find a correlation structure that is congruent to all marginal model specifications. The "independence" correlation structure ($diag(1, 1, \cdots, 1)$) would work, but is not interesting for the obvious reason. To circumvent the challenge, we simulated latent correlated continuous variables and cut them into ordinal ones. This workaround allowed us to make the intended comparison under correlation structures of various complexity.

Our simulation study demonstrates the clear advantage of the local odds ratio method over the traditional correlation coefficient approach, and provides valuable information that analysts can readily use to guide their modeling choice. The rest of the paper is organized as follows: Section 2 briefly reviews the model and the estimation methods under discussion; Section 3 presents the simulation method and the main results, which is followed by a discussion in Section 4. Details of the simulation results are left to Section 5.

2. A Brief Review of Theory

Let $Y_{it} \in \{1, 2, \cdots, I > 2\}$ be the ordinal response for subject i ($i \in \{1, 2, \cdots, N\}$) at time t ($t \in \{1, 2, \cdots, T_i\}$). For the purpose of modeling, we re-code each scalar Y_{it} as a binary vector $\mathbf{Y}_{it} = (Y_{it1}, Y_{it2}, \cdots, Y_{it(I-1)})'$ where $Y_{itj} = I(Y_{it} = j)$, and let $\mathbf{Y}_i = (\mathbf{Y}'_{i1}, \mathbf{Y}'_{i2}, \cdots, \mathbf{Y}'_{iT_i})'$. In other words, the $T_i \times 1$ vector of ordinal responses for subject i is represented as a $T_i(I-1) \times 1$ binary vector. Furthermore, let $\mathbf{x}_{it}$ be the $(I-1) \times p$ matrix of covariates that include variables of interest and category specific cut-points for subject i at time t, and let $\mathbf{X}_i = (\mathbf{x}'_{i1}, \cdots, \mathbf{x}'_{iT_i})'$. Finally, let $\pi_{itj} = \mathbb{E}(Y_{itj}|\mathbf{x}_{it}) = \Pr(Y_{itj} = 1|\mathbf{x}_{it})$, $\boldsymbol{\pi}_{it} = (\pi_{it1}, \pi_{it2}, \cdots, \pi_{it(I-1)})'$, and $\boldsymbol{\pi}_i = (\boldsymbol{\pi}'_{i1}, \boldsymbol{\pi}'_{i2}, \cdots, \boldsymbol{\pi}'_{iT_i})'$. Then the proportional odds cumulative logit model can be written as

$$\text{logit}(\mathbf{g}(\mathbb{E}(\mathbf{Y}_{it}|\mathbf{x}_{it}))) = \text{logit}(\mathbf{g}(\boldsymbol{\pi}_{it})) = \mathbf{x}_{it}\boldsymbol{\beta}, \tag{1}$$

where $\mathbf{g}(\boldsymbol{\pi}_{it}) = (\pi_{it1}, \pi_{it1} + \pi_{it2}, \cdots, \sum_{i=1}^{I-1} \pi_{itj})'$, and $\boldsymbol{\beta}$ is the vector of model parameters that we want to make inference on.

To estimate $\boldsymbol{\beta}$, the GEE method solves the following equations:

$$\boldsymbol{U}(\boldsymbol{\beta}, \boldsymbol{\alpha}) = \sum_{i=1}^{N} \boldsymbol{D}'_i \boldsymbol{V}_i^{-1} (\mathbf{Y}_i - \boldsymbol{\pi}_i) = \mathbf{0}, \tag{2}$$

where $\boldsymbol{D}_i = \partial \boldsymbol{\pi}_i / \partial \boldsymbol{\beta}$, $\boldsymbol{V}_i = \boldsymbol{V}_i(\boldsymbol{\beta}, \boldsymbol{\alpha})$ is the $T_i(I-1) \times T_i(I-1)$ "working" covariance matrix of $\mathbf{Y}_i$, and $\boldsymbol{\alpha}$ is a vector of nuisance parameters that captures the association between responses within the same subject at different times. Specifically,

$$\boldsymbol{V}_i(\boldsymbol{\beta}, \boldsymbol{\alpha}) = \boldsymbol{A}_i^{1/2} R_i(\boldsymbol{\alpha}) \boldsymbol{A}_i^{1/2}, \tag{3}$$

where $\boldsymbol{A}_i = diag(\boldsymbol{A}_{i1}, \cdots, \boldsymbol{A}_{iT_i})$, $\boldsymbol{A}_{it} = diag(\text{Var}(Y_{it1}), \cdots, \text{Var}(Y_{it(I-1)}))$, and $R_i(\boldsymbol{\alpha})$, commonly referred to as the "working" correlation structure, models the association pattern. Liang and Zeger (1986) showed that solving (2) yields consistent estimate for $\boldsymbol{\beta}$ under mild regularity conditions.

As mentioned in the introduction, the most commonly used parameterization for $R_i(\boldsymbol{\alpha})$ is through correlation coefficients. However, the range of correlation coefficient between binary variables can be severely restricted (Fréchet, 1940) (also see Touloumis et al., 2013, supplementary materials). For example, the correlation coefficient for two Bernoulli random variables with marginal probabilities 0.1 and

0.7 is bounded between -0.11 and 0.22. Consequently, the correlation coefficient parameterization is not ideal.

To circumvent this problem, Touloumis et al. proposed to parameterize $R_i(\boldsymbol{\alpha})$ with local odds ratios. Specifically, for each of the $L = \binom{T}{2}$ time pairs, where $T = \max\{T_1, \cdots, T_N\}$, an $I \times I$ marginalized contingency table is formed using all the available responses, and local odds ratios for all cut-points are calculated. This gives rise to the vector $\boldsymbol{\alpha} = (\theta_{1121}, \theta_{1121}, \cdots, \theta_{(T-1)(I-1)(T)(I-1)})$, where $\theta_{tjt'j'}$ is the local odds ratio at cut-point jj' for time pair tt'. To achieve parsimony, $\theta_{tjt'j'}$ is modeled as

$$\log \theta_{tjt'j'} = \phi_{tt'}(\mu_{tj}^{tt'} - \mu_{t(j+1)}^{tt'})(\mu_{t'j'}^{tt'} - \mu_{t'(j'+1)}^{tt'}), \tag{4}$$

which is a generalized version of the row-column effect model (Becker and Clogg, 1989). Here the parameter μ represents potentially time dependent category scores and $\phi_{tt'}$ partially controls time exchangeability. Touloumis et al. used the maximum likelihood method to estimate $\boldsymbol{\alpha}$, and showed how it leads to consistent estimate of $\boldsymbol{\beta}$. Furthermore, they recommended four association structures to use in practice:

- The uniform structure: $\log \theta_{tjt'j'} = \phi$;
- The category exchangeable structure: $\log \theta_{tjt'j'} = \phi_{tt'}$;
- The time exchangeable structure: $\log \theta_{tjt'j'} = \phi(\mu_j - \mu_{j+1})(\mu_{j'} - \mu_{j'+1})$;
- The row-column effect structure: $\log \theta_{tjt'j'} = \phi_{tt'}(\mu_j^{tt'} - \mu_{j+1}^{tt'})(\mu_{j'}^{tt'} - \mu_{j'+1}^{tt'})$;

We will examine their performance through simulation studies.

3. Simulation Study

We simulated 5-category ordinal responses measured on 4 occasions from the following model:

$$\text{logit}(\Pr(Y_{it} \leq j | x_{it})) = \beta_{0j} + \beta x_{it}. \tag{5}$$

Specifically, we set $I = 5$, $t = 1, ..., 4$, $N = 100, 500$, and $\boldsymbol{\beta} = (\beta_{01}, \beta_{02}, \beta_{03}, \beta_{04}, \beta)'$ $= (-1.5, -0.5, 0.5, 1.5, 1)'$. Moreover, $\mathbf{x}_i = (x_{i1}, \cdots, x_{i4})'$ was simulated from a tetra-variate normal distribution with mean vector $\mathbf{0}$, unit variances and a correlation matrix with off-diagonal elements equal to 0.8; $\boldsymbol{\epsilon}_i = (\epsilon_{i1}, \cdots, \epsilon_{i4})'$ was simulated from a tetra-variate normal distribution with mean vector $\mathbf{0}$, unit variances and correlation matrix $\boldsymbol{R}_\epsilon$. Once $\boldsymbol{\beta}$, $\mathbf{x}_i$ and $\boldsymbol{\epsilon}_i$ were generated, the ordinal response Y_{it} was obtained as follows:

$$Y_{it} = j \iff \beta_{0(j-1)} < \epsilon_{it} - \beta x_{it} \leq \beta_{0j}, \tag{6}$$

where $j = 1, ..., 5$, $\beta_{00} = -\infty$, and $\beta_{05} = \infty$. To evaluate the two methods under various correlation strength and complexity, we considered five specifications for $\boldsymbol{R}_\epsilon$. The simplest setting corresponds to the independence structure on the latent continuous scale, where the off-diagonal elements of $\boldsymbol{R}_\epsilon$ equal 0. Then, increasing in correlation strength, we set the off-diagonal elements to $0.15, 0.5, 0.85$, respectively. Finally, we considered the non-time-exchangeable structure: a Toeplitz matrix whose first row is $(1, 0.85, 0.5, 0.15)$, i.e.,

$$\begin{bmatrix} 1.00 & 0.85 & 0.50 & 0.15 \\ 0.85 & 1.00 & 0.85 & 0.50 \\ 0.50 & 0.85 & 1.00 & 0.85 \\ 0.15 & 0.50 & 0.85 & 1.00 \end{bmatrix}$$

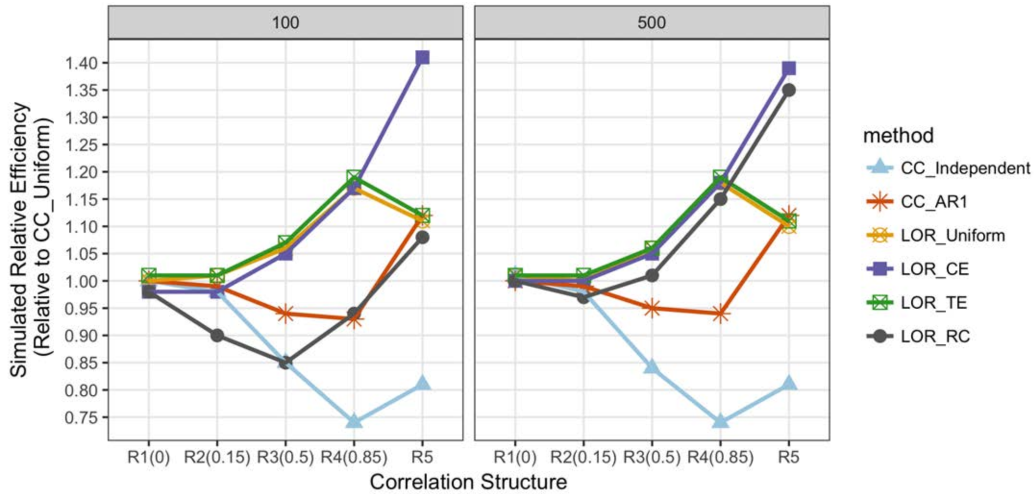

Figure 1: Comparing the simulated relative efficiency (SRE) for $N = 100, 500$ at $\boldsymbol{R}_\epsilon = \boldsymbol{R}_1, \boldsymbol{R}_2, ..., \boldsymbol{R}_5$ between the correlation coefficient (CC) method and the local odds ratio (LOR) method. SRE is defined as MSE($\hat{\beta}_{CC\_Uniform}$)/MSE($\hat{\beta}$). An SRE greater than 1 indicates that $\hat{\beta}$ is superior to $\hat{\beta}_{CC\_Uniform}$.

These five correlation structures are named $\boldsymbol{R}_1, \boldsymbol{R}_2, ..., \boldsymbol{R}_5$ henceforth.

The simulation study then proceeds as follows: at each sample size ($N = 100, 500$) and latent correlation structure ($\boldsymbol{R}_\epsilon = \boldsymbol{R}_1, \boldsymbol{R}_2, ..., \boldsymbol{R}_5$) setting, we repeated the following experiment $5,000$ times:

1. Simulate a data set: $(\boldsymbol{Y}_1, \cdots, \boldsymbol{Y}_N)$, $(\boldsymbol{X}_1, \cdots, \boldsymbol{X}_N)$;
2. Calculate the bias, standard error, and the 95% Wald confidence interval of β obtained from the following methods:
   - Correlation coefficient (CC) method with independent, uniform and AR1 working correlation specifications;
   - Local odds ratio (LOR) method with uniform, category exchangeable (CE), time exchangeable (TE) and row-column effect (RC) specifications.

To aggregate results from the $5,000$ iterations, we calculated the mean bias and standard error, the percentage of times that the confidence interval covered the true parameter value, and the simulated relative efficiency (SRE) which is defined as the ratio of the mean squared error of $\hat{\beta}_{\text{benchmark}}$ to $\hat{\beta}$. Bigger numbers suggest higher efficiency. Since the correlation coefficient method with the uniform working correlation ("CC_Uniform") specification is most widely used in practice, we chose it as the benchmark. R packages {SimCorMultRes, repolr, and multgee} were used to simulate data and fit models. SRE results are presented in Figure 1, further details of the simulation results are summarized in Tables 1-5 in Section 5.

From Figure 1, we can see that the local odds ratio approach compared favorably to the correlation coefficient approach. Specifically, when the latent correlation was truly uniform ($\boldsymbol{R}_1, ..., \boldsymbol{R}_4$), the SRE of the local odds ratio method with uniform, category exchangeable and time exchangeable specifications were close to or greater than 1, suggesting that it was similarly or more efficient than the "CC_Uniform" method. Moreover, the efficiency of the local odds ratio method increased monotonically as the latent correlation grew stronger. For example, the category exchangeable specification reached an SRE of 1.18 (or equivalently, 15% more efficient than the "CC_Uniform" method) under $\boldsymbol{R}_4$ when $N = 500$. The only exception to this general trend was the row-column effect specification at $N = 100$, where the method performed worse than the "CC_Uniform" method. This is likely because the sample size was too small to reliably estimate the numerous parameters in this specification. In other words, the row-column effect specification is more data hungry than the other specifications and should only be used on larger data sets. In fact, it showed noticeable performance improvement after the sample size increased to 500, reaching 1.15 in SRE at $\boldsymbol{R}_4$. In terms of bias, we can see from Table 1-4 that both uniform and time exchangeable specifications had smaller bias when the latent correlation was weak ($\boldsymbol{R}_1, \boldsymbol{R}_2$); when it became stronger ($\boldsymbol{R}_3, \boldsymbol{R}_4$), however, the category exchangeable and time exchangeable specifications tended to have smaller bias. The row-column effect specification, by contrast, showed consistent downward bias in all four scenarios, with a magnitude that was 3-5 times larger than most other methods.

On the other hand, when we compared between different specifications for the correlation coefficient method, we found that the independent and AR1 structures were less efficient than the uniform specification. This is perhaps why people primarily use the latter in practice, and justifies our choice of using it as the benchmark for comparison.

In addition to the superior SRE results at $\boldsymbol{R}_1, ..., \boldsymbol{R}_4$, the local odds ratio method also outperformed the correlation coefficient method under more complex latent correlation structures. For example, the category exchangeable specification reached an SRE of 1.4 at both sample sizes, amounting to a 30% efficiency boost. The row-column effect specification showed similar efficiency gain at $N = 500$.

Finally, comparing between the four specifications for the local odds ratio method, we noticed that the uniform, category exchangeable and time exchangeable structures performed similarly at $\boldsymbol{R}_1, ..., \boldsymbol{R}_4$. The row-column effect structure, on the other hand, was slightly behind even at $N = 500$. For the more complex latent structure $\boldsymbol{R}_5$, both the category exchangeable and row-column effect specifications worked well, whereas the uniform and time exchangeable structures seemed less adequate: their SREs were around 1.1 – much lower than that obtained from the category exchangeable structure.

4. Discussion

In this study, we have compared the efficiency of different methods for estimating the regression parameter β of proportional odds cumulative logit models for longitudinal ordinal data, and found that the local odds ratio method can achieve up to 30% efficiency gains compared to the widely used correlation coefficient method. Based on our results, we recommend using the category exchangeable local odds ratio method as the default modeling choice. To sum up, our results demonstrate the advantage of the local odds ratio parameterization, and provide a better understanding of how different parameterizations of the association structure can affect model parameter estimation. Analysts may find this study useful to their work. Specifically, they may decide to switch their default modeling choice from the correlation coefficient method to the local odds ratio method.

Since the local odds ratio method came out rather recently, few papers have examined its performance. Nooraee et al. (2014) compared the R package {multgee} with a few other software tools and focused on issues such as software stability and algorithm convergence rather than estimation efficiency. Touloumis et al. (2013) compared between different specifications within the local odds ratio framework, but not against the traditional correlation coefficient method, which is the focus of this study.

Future studies may extend the comparison to other classes of models and to nominal rather than ordinal categorical data; this would allow the evaluation of the full potential of the local odds ratio parameterization for modeling categorical data.

5. Supplementary Information

This section contains details of the simulation results. Specifically, we focused on the parameter estimation for β from model (5), which was set to 1 throughout the simulations, and aggregated the results from the $5,000$ iterations by summarizing the mean bias and standard error (SE), the percentage of times that β was covered by the 95% confidence interval (coverage probability, CP), and the simulated relative efficiency (SRE) which is defined as MSE($\hat{\beta}_{CC\_Uniform}$)/MSE($\hat{\beta}$).

	N = 100				N = 500			
Method	Bias	SE	CP	SRE	Bias	SE	CP	SRE
CC_Independent	0.0051	0.1045	0.94	1.00	0.0013	0.0465	0.95	1.01
CC_Uniform	0.0054	0.1048	0.95	1.00	0.0015	0.0467	0.95	1.00
CC_AR1	0.0054	0.1046	0.95	1.00	0.0016	0.0466	0.95	1.00
LOR_Uniform	0.0024	0.1041	0.94	1.00	0.0010	0.0465	0.95	1.01
LOR_CE	-0.0126	0.1032	0.93	0.98	-0.0020	0.0464	0.94	1.00
LOR_TE	0.0011	0.1039	0.94	1.01	0.0006	0.0464	0.95	1.01
LOR_RC	-0.0241	0.1014	0.93	0.98	-0.0042	0.0462	0.94	1.00

Table 1: Simulation results under the latent correlation matrix $\boldsymbol{R}_1$, which is uniform with off-diagonal elements equal to 0.

	N = 100				N = 500			
Method	Bias	SE	CP	SRE	Bias	SE	CP	SRE
CC_Independent	0.0069	0.1047	0.95	0.98	0.0015	0.0467	0.95	0.98
CC_Uniform	0.0070	0.1033	0.94	1.00	0.0018	0.0461	0.95	1.00
CC_AR1	0.0068	0.1038	0.95	0.99	0.0018	0.0463	0.95	0.99
LOR_Uniform	0.0045	0.1030	0.95	1.01	0.0012	0.0461	0.95	1.00
LOR_CE	-0.0107	0.1021	0.94	0.98	-0.0016	0.0460	0.94	1.00
LOR_TE	-0.0002	0.1026	0.94	1.01	0.0002	0.0460	0.94	1.01
LOR_RC	-0.0344	0.0995	0.91	0.90	-0.0081	0.0457	0.94	0.97

Table 2: Simulation results under the latent correlation matrix $\boldsymbol{R}_2$ (uniform with off-diagonal elements equal to 0.15).

	N = 100				N = 500			
Method	Bias	SE	CP	SRE	Bias	SE	CP	SRE
CC_Independent	0.0122	0.1098	0.95	0.85	0.0014	0.0490	0.95	0.84
CC_Uniform	0.0129	0.1011	0.95	1.00	0.0014	0.0450	0.95	1.00
CC_AR1	0.0123	0.1038	0.95	0.94	0.0013	0.0462	0.95	0.95
LOR_Uniform	0.0108	0.0984	0.95	1.06	0.0009	0.0438	0.95	1.06
LOR_CE	-0.0039	0.0974	0.94	1.05	-0.0019	0.0437	0.95	1.05
LOR_TE	0.0061	0.0979	0.95	1.07	0.0000	0.0437	0.95	1.06
LOR_RC	-0.0468	0.0941	0.88	0.85	-0.0097	0.0433	0.94	1.01

Table 3: Simulation results under the latent correlation matrix $\boldsymbol{R}_3$ (uniform with off-diagonal elements equal to 0.5).

	N = 100				N = 500			
Method	Bias	SE	CP	SRE	Bias	SE	CP	SRE
CC_Independent	0.0140	0.1219	0.95	0.74	0.0015	0.0546	0.95	0.74
CC_Uniform	0.0144	0.1047	0.95	1.00	0.0022	0.0466	0.95	1.00
CC_AR1	0.0134	0.1083	0.95	0.93	0.0018	0.0483	0.95	0.94
LOR_Uniform	0.0124	0.0962	0.95	1.17	0.0022	0.0428	0.95	1.18
LOR_CE	0.0004	0.0952	0.94	1.17	-0.0001	0.0427	0.95	1.18
LOR_TE	0.0095	0.0957	0.95	1.19	0.0016	0.0426	0.95	1.19
LOR_RC	-0.0448	0.0915	0.88	0.94	-0.0081	0.0422	0.94	1.15

Table 4: Simulation results under the latent correlation matrix $\boldsymbol{R}_4$ (uniform with off-diagonal elements equal to 0.85).

	N = 100				N = 500			
Method	Bias	SE	CP	SRE	Bias	SE	CP	SRE
CC_Independent	0.0104	0.1151	0.95	0.81	0.0014	0.0515	0.94	0.81
CC_Uniform	0.0096	0.1039	0.95	1.00	0.0017	0.0464	0.94	1.00
CC_AR1	0.0090	0.0982	0.95	1.12	0.0013	0.0439	0.94	1.12
LOR_Uniform	0.0072	0.0993	0.95	1.11	0.0015	0.0444	0.94	1.10
LOR_CE	-0.0037	0.0874	0.95	1.41	-0.0009	0.0395	0.94	1.39
LOR_TE	0.0026	0.0987	0.95	1.12	0.0006	0.0442	0.94	1.11
LOR_RC	-0.0458	0.0839	0.88	1.08	-0.0089	0.0387	0.93	1.35

Table 5: Simulation results under the latent correlation matrix $\boldsymbol{R}_5$ (a Toeplitz matrix whose first row is $(1, 0.85, 0.5, 0.15)$).

Acknowledgments

The research described was supported by NIH/National Center for Advancing Translational Science (NCATS) UCLA CTSI Grant Number UL1TR000124.

References

A. Amatya and H. Demirtas. Multiord: An r package for generating correlated ordinal data. *Communications in Statistics-Simulation and Computation*, 44(7): 1683–1691, 2015.

A. Barbiero and P. A. Ferrari. An r package for the simulation of correlated discrete variables. *Communications in Statistics-Simulation and Computation*, pages 1–18, 2017.

M. P. Becker and C. C. Clogg. Analysis of sets of two-way contingency tables using association models. *Journal of the American statistical association*, 84(405):142–151, 1989.

D. G. Clayton. Repeated ordinal measurements: a generalised estimating equation approach. *Medical Research Council Biostatistics Unit Technical Report. Cambridge, England*, 1992.

M. Fréchet. *Les probabilités associées à un système d'évènements comptables et dépendants: Evènements en nombre fini fixe.* Hermann, 1940.

K.-Y. Liang and S. L. Zeger. Longitudinal data analysis using generalized linear models. *Biometrika*, 73(1):13–22, 1986. ISSN 00063444. URL `http://www.jstor.org/stable/2336267`.

S. R. Lipsitz, K. Kim, and L. Zhao. Analysis of repeated categorical data using generalized estimating equations. *Statistics in medicine*, 13(11):1149–1163, 1994.

N. Nooraee, G. Molenberghs, and E. R. van den Heuvel. Gee for longitudinal ordinal data: comparing r-geepack, r-multgee, r-repolr, sas-genmod, spss-genlin. *Computational Statistics & Data Analysis*, 77:70–83, 2014.

N. R. Parsons, M. L. Costa, J. Achten, and N. Stallard. Repeated measures proportional odds logistic regression analysis of ordinal score data in the statistical software package r. *Computational statistics & data analysis*, 53(3):632–641, 2009.

A. Touloumis, A. Agresti, and M. Kateri. Gee for multinomial responses using a local odds ratios parameterization. *Biometrics*, 69(3):633–640, 2013.