Computational Statistics and Data Analysis 54 (2010) 1872–1880

Contents lists available at ScienceDirect

Computational Statistics and Data Analysis

journal homepage: www.elsevier.com/locate/csda

Goodness-of-fit tests for modeling longitudinal ordinal data

Kuo-Chin Lin *

Graduate Institute of Business and Management, Tainan University of Technology, Tainan, Taiwan

ARTICLE INFO

Article history:
Received 20 October 2009
Received in revised form 28 December 2009
Accepted 13 February 2010
Available online 26 February 2010

Keywords:
GEE
Goodness-of-fit
Logistic proportional odds model
Longitudinal ordinal data

ABSTRACT

Longitudinal studies involving categorical responses are extensively applied in many fields of research and are often fitted by the generalized estimating equations (GEE) approach and generalized linear mixed models (GLMMs). The assessment of model fit is an important issue for model inference. The purpose of this article is to extend Pan's (2002a) goodness-of-fit tests for GEE models with longitudinal binary data to the tests for logistic proportional odds models with longitudinal ordinal data. Two proposed methods based on Pearson chi-squared test and unweighted sum of residual squares are developed, and the approximate expectations and variances of the test statistics are easily computed. Four major variants of working correlation structures, independent, AR(1), exchangeable and unspecified, are considered to estimate the variances of the proposed test statistics. Simulation studies in terms of type I error rate and the power performance of the proposed tests are presented for various sample sizes. Furthermore, the approaches are demonstrated by two real data sets.

1. Introduction

Categorical variables commonly occur in the biomedical and social sciences, for measuring responses such as whether a patient recovers from a disease and the opinion on a new policy. Two major types of categorical response variables are nominal and ordinal. In this article, we focus on longitudinal ordinal data which comprise a series of ordinal responses measured repeatedly over time for each subject. One possible objective for analyzing longitudinal ordinal data is to explore the relationship between longitudinal ordinal responses and a set of time-stationary or time-dependent covariates. Longitudinal ordinal data can be fitted by the generalized estimating equations (GEE) approach and generalized linear mixed models (GLMMs). Liang and Zeger (1986) initially proposed the GEE method based on a quasi-likelihood function for the model parameter estimation in longitudinal binary data. Miller et al. (1993) modeled marginal means and used the inverse of Fisher's z transformation of pairwise correlations. Lipsitz et al. (1994) extended Liang and Zeger's (1986) method to the analysis of correlated categorical responses. Williamson et al. (1995) developed marginal means in cumulative logit and cumulative probit models based on a global odds ratio as the measure of association. Heagerty and Zeger (1996) and Fahrmeir and Pritscher (1996) adopted a proportional odds model for the marginal cumulative probabilities by a GEE approach for analyzing the repeated ordinal measurements. Kauermann (2000) used a nonparametric smoothing technique for modeling longitudinal ordinal data by varying coefficients. GLMMs expand ordinary regression by allowing nonnormal responses and including random effects. Liu and Hedeker (2006) provided a mixed-effects item response theory model for longitudinal multivariate ordinal data. A marginal approach based on GEE models has population average interpretations in model parameters, while a conditional approach based on GLMMs has subject-specific interpretations in model parameters. Before making inferences on model parameters, the assessment of model fit is an important issue.

* Tel.: +886 1 06 2389523; fax: +886 1 2545396.
E-mail address: t20053@mail.tut.edu.tw.

0167-9473/$ – see front matter © 2010 Elsevier B.V. All rights reserved.
doi:10.1016/j.csda.2010.02.013

In recent years, numerous goodness-of-fit tests have been vigorously developed for GEE models and GLMMs with categorical responses. However, most of these tests stressed on repeated binary responses. Tsiatis (1980) proposed a goodness-of-fit test for logistic regression models by partitioning the space of covariates. Barnhart and Williamson (1998) extended the method of Tsiatis to GEE fitted models. Horton et al. (1999) generalized the method of Hosmer and Lemeshow (1980) from ordinary logistic regression to GEE fitted models. Pan (2002a) developed two goodness-of-fit tests, a Pearson chi-squared type and the unweighted sum of residual squares, for GEE models with correlated binary data based on residuals. Lin and Myers (2006) compared type I error rates and the powers of the existing GEE goodness-of-fit tests using simulated data under different configurations. Williamson et al. (2003) utilized a kappa-like classification statistic for assessing GEE categorical response models. Lin et al. (2008) proposed a goodness-of-fit test for longitudinal binary data based on a nonparametric smoothing approach for GEE fitted models with both continuous and categorical covariates. Vonesh et al. (1996) provided a goodness-of-fit statistic for assessing the adequacy of generalized nonlinear mixed-effects models. Pan and Lin (2005) suggested goodness-of-fit tests of GLMMs by graphical and numerical approaches based on the cumulative sums of residuals over covariates or predicted values of the response variable.

In this article, two generalized methods of Pan's (2002a) tests for longitudinal ordinal data are proposed. The common tactics for analyzing ordinal responses mentioned by McCullagh (1980), Agresti (1999) and Clogg and Shihadeh (1994) are adjacent category models, continuation ratio models and proportional odds models. The cumulative logit model with proportional odds assumption is the most popular for the analysis of ordinal data. The proposed methods based on the Pearson chi-squared test and unweighted sum of residual squares for assessing the goodness-of-fit of proportional odds models for longitudinal ordinal data are developed in Section 2. Simulation studies are conducted for exploring the type I error rate and the power performance of the proposed test statistics in Section 3. In Section 4, two data sets cited by Hand and Crowder (1996) and Elliot et al. (1989) are employed to illustrate the proposed methods. Finally, the main findings and future research directions are discussed.

2. Proposed tests

Suppose that a longitudinal study consists of ordinal responses with $(J+1)$ categories and p-dimensional covariate vectors, $(Y_{it}, \mathbf{x}_{it})$, for $i = 1, \ldots, n$ and $t = 1, \ldots, n_i$, where Y_{it} denotes the observation for subject i at occasion t and the covariate vector $\mathbf{x}_{it}$ can be discrete or continuous. For simplicity we assume equal occasions, $n_i \equiv T$. Denote Y_{it} as a vector of J indicator variables, $\mathbf{y}_{it} = (y_{it}^{(1)}, \ldots, y_{it}^{(J)})'$ with $y_{it}^{(j)} = 1$ if response $Y_{it} = j$ and 0 otherwise. Let $\boldsymbol{\pi}_{it}$ and $\eta_{it}^{(j)}$ represent the vector of marginal probabilities and the marginal cumulative probabilities, respectively, where $\boldsymbol{\pi}_{it} = (\pi_{it}^{(1)}, \ldots, \pi_{it}^{(J)})'$ with $\pi_{it}^{(j)} = P(Y_{it} = j \mid \mathbf{x}_{it}) = P(y_{it}^{(j)} = 1 \mid \mathbf{x}_{it})$ and $\eta_{it}^{(j)} = P(Y_{it} \le j \mid \mathbf{x}_{it}) = \sum_{k=1}^{j} \pi_{it}^{(k)}$. It can be straightforwardly shown that $\mathrm{E}(\mathbf{y}_{it}) = \boldsymbol{\pi}_{it}$ and $\mathrm{Var}(\mathbf{y}_{it}) = \mathbf{V}_{it} = \mathrm{diag}(\boldsymbol{\pi}_{it}) - \boldsymbol{\pi}_{it}\boldsymbol{\pi}_{it}'$. The cumulative logit model with proportional odds assumption for describing the dependence of Y_{it} on $\mathbf{x}_{it}$ is given by

$$\mathrm{logit}(\eta_{it}^{(j)}) = \log\left(\frac{\eta_{it}^{(j)}}{1-\eta_{it}^{(j)}}\right) = \lambda_j + \mathbf{x}_{it}'\boldsymbol{\beta} = \zeta_{it}^{(j)}, \quad j = 1, \ldots, J, \tag{1}$$

where the intercepts $\lambda_1, \ldots, \lambda_J$ satisfy $\lambda_1 \le \cdots \le \lambda_J$, $\boldsymbol{\beta}$ is the vector of regression coefficients with $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_p)'$, and $\zeta_{it}^{(j)}$ is the jth element of a J-dimensional linear predictor $\boldsymbol{\zeta}_{it} = (\zeta_{it}^{(1)}, \ldots, \zeta_{it}^{(J)})'$. Since $\pi_{it}^{(1)} = \eta_{it}^{(1)}$ and

$$\pi_{it}^{(j)} = \eta_{it}^{(j)} - \eta_{it}^{(j-1)} = \frac{\exp(\zeta_{it}^{(j)})}{1+\exp(\zeta_{it}^{(j)})} - \frac{\exp(\zeta_{it}^{(j-1)})}{1+\exp(\zeta_{it}^{(j-1)})}, \quad \text{for } j = 2, \ldots, J, \tag{2}$$

the linear predictor $\boldsymbol{\zeta}_{it}$ can be rewritten as $\boldsymbol{\zeta}_{it} = \mathbf{Z}_{it}'\boldsymbol{\theta}$ with the parameter vector $\boldsymbol{\theta} = (\lambda_1, \ldots, \lambda_J, \boldsymbol{\beta}')'$ and the $J \times (J+p)$ design matrix

$$\mathbf{Z}_{it}' = \begin{bmatrix} 1 & & & \mathbf{x}_{it}' \\ & \ddots & & \vdots \\ & & 1 & \mathbf{x}_{it}' \end{bmatrix}.$$

A J-dimensional link function g connects $\boldsymbol{\pi}_{it}$ and the linear predictor $\mathbf{Z}_{it}'\boldsymbol{\theta}$ as $\boldsymbol{\pi}_{it} = g^{-1}(\mathbf{Z}_{it}'\boldsymbol{\theta})$.

Denote the responses, the marginal probabilities and the design matrix for subject i as $\mathbf{Y}_i = (\mathbf{y}_{i1}', \ldots, \mathbf{y}_{iT}')'$, $\boldsymbol{\pi}_i = (\boldsymbol{\pi}_{i1}', \ldots, \boldsymbol{\pi}_{iT}')'$, and $\mathbf{Z}_i = (\mathbf{Z}_{i1}, \ldots, \mathbf{Z}_{iT})'_{TJ\times(J+p)}$, respectively. The multivariate generalized estimating equations proposed by Lipsitz et al. (1994) and Liang and Zeger (1986) for estimating $\boldsymbol{\theta}$ is the solution to

$$\sum_{i=1}^{n} \mathbf{D}_i'\mathbf{V}_i^{-1}(\mathbf{Y}_i - \boldsymbol{\pi}_i) = 0, \tag{3}$$

where $\mathbf{D}_i = \partial\boldsymbol{\pi}_i/\partial\boldsymbol{\theta} = (\mathbf{D}_{i1}', \ldots, \mathbf{D}'iT)'$ with the jth row vector of $\mathbf{D}_{it}$ expressed by $(\partial\pi_{it}^{(j)}/\partial\lambda_1, \ldots, \partial\pi_{it}^{(j)}/\partial\beta_p)$ for $j = 1, \ldots, J$; $t = 1, \ldots, T$, and $\mathbf{V}_i = \mathbf{A}_i^{1/2}\mathbf{R}_i(\boldsymbol{\alpha})\mathbf{A}_i^{1/2}$. Here $\mathbf{A}_i = \mathrm{diag}(\mathbf{A}_{i1}, \ldots, \mathbf{A}_{iT})$ with diagonal block $\mathbf{A}_{it} = \mathrm{diag}(\pi_{it}^{(1)}(1 -$

$\pi_{it}^{(1)}), \ldots, \pi_{it}^{(J)}(1-\pi_{it}^{(J)}))$, the 'working' correlation matrix $\mathbf{R}_i(\boldsymbol{\alpha})$ is the correlation of $\mathbf{Y}_i$, and $\boldsymbol{\alpha}$ is a vector of parameters involved in the working correlation structure. In general, $\mathbf{V}_i \neq \mathrm{Var}(\mathbf{Y}_i)$. Liang and Zeger (1986) proposed a 'working' correlation matrix to gain efficiency in estimating $\boldsymbol{\theta}$. The solution to (3) is a consistent estimate of $\boldsymbol{\theta}$ for a variety of settings of the $TJ \times TJ$ 'working' correlation matrices. Let $\mathbf{R}_i(\boldsymbol{\alpha})$ be expressed as follows:

$$\mathbf{R}_i(\boldsymbol{\alpha}) = \begin{bmatrix} \mathbf{M}_{i1} & \mathbf{B}_{i12} & \cdots & \mathbf{B}_{i1T} \\ \mathbf{B}_{i21} & \mathbf{M}_{i2} & \cdots & \mathbf{B}_{i2T} \\ \vdots & & \ddots & \vdots \\ \mathbf{B}_{iT1} & \mathbf{B}_{iT2} & \cdots & \mathbf{M}_{iT} \end{bmatrix}_{TJ\times TJ},$$

where the tth diagonal block $\mathbf{M}_{it} = \mathbf{A}_{it}^{-1/2}\mathbf{V}_{it}\mathbf{A}_{it}^{-1/2}$ denoted throughout by $\boldsymbol{\pi}_i$ for $t = 1, \ldots, T$, and the off-diagonal matrix $\mathbf{B}_{ist} = \mathbf{A}_{is}^{-1/2}E[(\mathbf{y}_{is} - \boldsymbol{\pi}_{is})(\mathbf{y}_{it} - \boldsymbol{\pi}_{it})']\mathbf{A}_{it}^{-1/2}$ parameterized by $\boldsymbol{\alpha}$ for $s \neq t$. The correlations of longitudinal ordinal data include two parts. One is the correlation between the repeated responses for subject i specified by $\mathbf{R}_i(\boldsymbol{\alpha})$ and the other is the correlation between ordinal categories specified by $\mathbf{B}_{ist}$. The estimation of $\boldsymbol{\theta}$ based on the Fisher scoring algorithm and iterative proportional fitting can be obtained by the geepack package in R software. Once the estimate of the parameter vector $\hat{\boldsymbol{\theta}}$ is derived, the standardized residual vector, $\hat{\mathbf{e}}_i = (\hat{\mathbf{e}}_{i1}', \ldots, \hat{\mathbf{e}}_{iT}')'$, can be computed with the element vector $\hat{\mathbf{e}}_{it} = \hat{\mathbf{A}}_{it}^{-1/2}(\mathbf{y}_{it} - \hat{\boldsymbol{\pi}}_{it})$, where $\hat{\boldsymbol{\pi}}_{it} = g^{-1}(\mathbf{Z}_{it}'\hat{\boldsymbol{\theta}})$ for $i = 1, \ldots, n$; $t = 1, \ldots, T$.

In this article, four 'working' correlation matrices, independence ($\mathbf{R}_i(\boldsymbol{\alpha}) = \mathbf{I}$, the identity matrix), AR(1) structure, exchangeable structure ($\mathbf{B}_{ist} = \mathbf{B}$) and unspecified structure ($\mathbf{B}_{ist} = \mathbf{B}_{st}$), are considered; namely

$$\mathbf{R}_{ind,i}(\boldsymbol{\alpha}) = \mathbf{I}_{TJ\times TJ}, \qquad \mathbf{R}_{ar(1),i}(\boldsymbol{\alpha}) = \begin{bmatrix} \mathbf{M}_{i1} & \mathbf{L} & \mathbf{L}^2 & \cdots & \mathbf{L}^{T-1} \\ \mathbf{L} & \mathbf{M}_{i2} & \mathbf{L} & \cdots & \mathbf{L}^{T-2} \\ \mathbf{L}^2 & \mathbf{L} & \mathbf{M}_{i3} & \cdots & \mathbf{L}^{T-3} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \mathbf{L}^{T-1} & \mathbf{L}^{T-2} & \mathbf{L}^{T-3} & \cdots & \mathbf{M}_{iT} \end{bmatrix},$$

$$\mathbf{R}_{ex,i}(\boldsymbol{\alpha}) = \begin{bmatrix} \mathbf{M}_{i1} & \mathbf{B} & \cdots & \mathbf{B} \\ \mathbf{B} & \mathbf{M}_{i2} & \cdots & \mathbf{B} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{B} & \mathbf{B} & \cdots & \mathbf{M}_{iT} \end{bmatrix} \quad \text{and} \quad \mathbf{R}_{un,i}(\boldsymbol{\alpha}) = \begin{bmatrix} \mathbf{M}_{i1} & \mathbf{B}_{12} & \cdots & \mathbf{B}_{1T} \\ \mathbf{B}_{21} & \mathbf{M}_{i2} & \cdots & \mathbf{B}_{2T} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{B}_{T1} & \mathbf{B}_{T2} & \cdots & \mathbf{M}_{iT} \end{bmatrix}$$

for $i = 1, \ldots, n$ as well as the corresponding off-diagonal matrices are estimated by

$$\hat{\mathbf{L}} = \frac{\sum_{t=1}^{T-1}\sum_{i=1}^{n}\hat{\mathbf{e}}_{it}\hat{\mathbf{e}}_{i,t+1}'}{(T-1)(n-p)}, \qquad \hat{\mathbf{B}} = \frac{\sum_{i=1}^{n}\sum_{s<t}\hat{\mathbf{e}}_{is}\hat{\mathbf{e}}_{it}'}{\left[\frac{1}{2}\sum_{i=1}^{n}T(T-1)\right] - p},$$

$$\hat{\mathbf{B}}_{st} = \frac{\sum_{i=1}^{n}\hat{\mathbf{e}}_{is}\hat{\mathbf{e}}_{it}'}{n-p} \quad \text{for } s \neq t = 1, \ldots, T.$$

Then the four 'working' covariance matrices, independent, AR(1), exchangeable and unspecified, are denoted by

(a) $\mathrm{Var}(\mathbf{Y})_{ind} = \mathbf{A}, \mathbf{A} = \mathrm{diag}(\mathbf{A}_1, \ldots, \mathbf{A}_n)$,
(b) $\mathrm{Var}(\mathbf{Y})_{ar(1)} = \mathbf{A}^{1/2}\mathbf{R}_{ar(1)}(\boldsymbol{\alpha})\mathbf{A}^{1/2}, \mathbf{R}_{ar(1)}(\boldsymbol{\alpha}) = \mathrm{diag}(\mathbf{R}_{ar(1),1}(\boldsymbol{\alpha}), \ldots, \mathbf{R}_{ar(1),n}(\boldsymbol{\alpha}))$,
(c) $\mathrm{Var}(\mathbf{Y})_{ex} = \mathbf{A}^{1/2}\mathbf{R}_{ex}(\boldsymbol{\alpha})\mathbf{A}^{1/2}, \mathbf{R}_{ex}(\boldsymbol{\alpha}) = \mathrm{diag}(\mathbf{R}_{ex,1}(\boldsymbol{\alpha}), \ldots, \mathbf{R}_{ex,n}(\boldsymbol{\alpha}))$ and
(d) $\mathrm{Var}(\mathbf{Y})_{un} = \mathbf{A}^{1/2}\mathbf{R}_{un}(\boldsymbol{\alpha})\mathbf{A}^{1/2}, \mathbf{R}_{un}(\boldsymbol{\alpha}) = \mathrm{diag}(\mathbf{R}_{un,1}(\boldsymbol{\alpha}), \ldots, \mathbf{R}_{un,n}(\boldsymbol{\alpha}))$, respectively.

Two proposed goodness-of-fit test statistics for assessing the proportional odds model with longitudinal ordinal data, Pearson chi-squared statistic and unweighted sum of residual squares, are generalizations of the test statistics developed by Pan (2002a) for the logistic regression model with correlated binary data. The proposed Pearson chi-squared statistic is expressed by

$$G = \sum_{i=1}^{n}\sum_{t=1}^{T}\sum_{j=1}^{J}\frac{(y_{it}^{(j)} - \hat{\pi}_{it}^{(j)})^2}{\hat{\pi}_{it}^{(j)}(1-\hat{\pi}_{it}^{(j)})}. \tag{4}$$

The test statistic G can be rewritten in matrix form as follows:

$$\begin{aligned} G &= \hat{\mathbf{e}}'\hat{\mathbf{e}} \\ &= (\hat{\mathbf{A}}^{-1/2}(\mathbf{Y} - \hat{\boldsymbol{\pi}}))'(\hat{\mathbf{A}}^{-1/2}(\mathbf{Y} - \hat{\boldsymbol{\pi}})) \\ &= (\mathbf{Y} + \hat{\boldsymbol{\pi}} - 1)'\hat{\mathbf{A}}^{-1}(\mathbf{Y} - \hat{\boldsymbol{\pi}}) + (1 - 2\hat{\boldsymbol{\pi}})'\hat{\mathbf{A}}^{-1}(\mathbf{Y} - \hat{\boldsymbol{\pi}}) \\ &= nTJ + (1 - 2\hat{\boldsymbol{\pi}})'\hat{\mathbf{A}}^{-1}(\mathbf{Y} - \hat{\boldsymbol{\pi}}), \end{aligned}$$

where $\hat{\mathbf{e}}' = (\hat{\mathbf{e}}'_1, \ldots, \hat{\mathbf{e}}'_n)$, $\mathbf{Y}' = (\mathbf{Y}'_1, \ldots, \mathbf{Y}'_n)$, $\hat{\mathbf{A}} = \mathrm{diag}(\hat{\mathbf{A}}_1, \ldots, \hat{\mathbf{A}}_n)$ and $\hat{\boldsymbol{\pi}}' = (\hat{\boldsymbol{\pi}}'_1, \ldots, \hat{\boldsymbol{\pi}}'_n)$. In particular, when $J = 1$ for binary responses, our test statistic G can be reduced to Pan's statistic (2002a).

We apply the approximation of $\mathbf{Y} - \hat{\boldsymbol{\pi}}$ referred to by Pan (2002a) and Lin et al. (2008),

$$\mathbf{Y} - \hat{\boldsymbol{\pi}} \approx (\mathbf{I} - \mathbf{H})(\mathbf{Y} - \boldsymbol{\pi}),$$

where $\approx$ denotes asymptotic equivalence, $\mathbf{H} = \mathbf{AZ}(\mathbf{Z}'\mathbf{AV}^{-1}\mathbf{AZ})^{-1}\mathbf{Z}'\mathbf{AV}^{-1}$ with $\mathbf{Z} = (\mathbf{Z}'_1, \ldots, \mathbf{Z}'_n)$ being a $(J + p) \times (nTJ)$ matrix, $\mathbf{V} = \mathrm{diag}(\mathbf{V}_1, \ldots, \mathbf{V}_n)$ and $\mathbf{I}$ is an identity matrix. Hence,

$$G \approx nTJ + (1 - 2\hat{\boldsymbol{\pi}})'\hat{\mathbf{A}}^{-1}(\mathbf{I} - \mathbf{H})(\mathbf{Y} - \boldsymbol{\pi}). \tag{5}$$

Based on Pan's framework, the term $(1 - 2\hat{\boldsymbol{\pi}})'\hat{\mathbf{A}}^{-1}$ is treated as fixed. The approximate expectation and variance of test statistic G are acquired by

$$\widehat{E(G)} = nTJ, \quad \text{and} \quad \widehat{\mathrm{Var}(G)} = (1 - 2\hat{\boldsymbol{\pi}})'\hat{\mathbf{A}}^{-1}(\mathbf{I} - \hat{\mathbf{H}})\widehat{\mathrm{Var}(\mathbf{Y})}(\mathbf{I} - \hat{\mathbf{H}})'\hat{\mathbf{A}}^{-1}(1 - 2\hat{\boldsymbol{\pi}}),$$

where $\hat{\mathbf{H}} = \hat{\mathbf{A}}\mathbf{Z}(\mathbf{Z}'\hat{\mathbf{A}}\hat{\mathbf{V}}^{-1}\hat{\mathbf{A}}\mathbf{Z})^{-1}\mathbf{Z}'\hat{\mathbf{A}}\hat{\mathbf{V}}^{-1}$.

The second proposed statistic based on the unweighted sum of residual squares is defined by

$$U = \sum_{i=1}^{n}\sum_{t=1}^{T}\sum_{j=1}^{J}(y_{it}^{(j)} - \hat{\pi}_{it}^{(j)})^2. \tag{6}$$

The form of test statistic U was discussed by Hosmer et al. (1997) and Pan (2002a). For simplicity we rewrite U in matrix form:

$$\begin{aligned} U &= (\mathbf{Y} - \hat{\boldsymbol{\pi}})'(\mathbf{Y} - \hat{\boldsymbol{\pi}}) \\ &= (\mathbf{Y} + \hat{\boldsymbol{\pi}} - 1)'(\mathbf{Y} - \hat{\boldsymbol{\pi}}) + (1 - 2\hat{\boldsymbol{\pi}})'(\mathbf{Y} - \hat{\boldsymbol{\pi}}). \end{aligned}$$

Similarly, the approximate expectation and variance of test statistic U are given by

$$\widehat{E(U)} = \hat{\boldsymbol{\pi}}'(1 - \hat{\boldsymbol{\pi}}), \quad \text{and} \quad \widehat{\mathrm{Var}(U)} = (1 - 2\hat{\boldsymbol{\pi}})'(\mathbf{I} - \hat{\mathbf{H}})\widehat{\mathrm{Var}(\mathbf{Y})}(\mathbf{I} - \hat{\mathbf{H}})'(1 - 2\hat{\boldsymbol{\pi}}).$$

The estimated $\mathrm{Var}(\mathbf{Y})$ can be replaced by $\widehat{\mathrm{Var}(\mathbf{Y})}_{ind}$, $\widehat{\mathrm{Var}(\mathbf{Y})}_{ar(1)}\widehat{\mathrm{Var}(\mathbf{Y})}_{ex}$ or $\widehat{\mathrm{Var}(\mathbf{Y})}_{un}$. For large samples, G and U both approximate normal distributions based on the asymptotic approach with an increasing number of cells. However, when the model only contains discrete covariates, the asymptotic normality of test statistics may not hold due to the fixed-cells asymptotics, referred to by Osius and Rojek (1992).

3. Simulation study

To evaluate the performance of the proposed tests for the proportional odds model fit in terms of type I error rate and the power, the simulated longitudinal ordinal data are generated from the following models:

$$\text{Model 1: } \mathrm{logit}(\eta_{it}^{(j)}) = \lambda_{1j} + 0.25D_{1i} + 1.5X_{1it} + \beta D_{1i}X_{1it},$$

$$\text{Model 2: } \mathrm{logit}(\eta_{it}^{(j)}) = \lambda_{2j} - 0.5D_{2i} + X_{1it} + \beta X_{1it}^2,$$

where the monotone difference intercepts are assigned by $\lambda_1 = (-0.5, 0, 0.5)$ and $\lambda_2 = (0.25, 0.5, 0.75)$, respectively, $D_{1i} = 0$ for $i \le n/2$ and 1 otherwise, X_{1it} follows a uniform distribution on $(-1, 1)$ and D_{2i} follows a Bernoulli distribution with probability 0.5 for $i = 1, \ldots, n$, $t = 1, 2, 3$, $j = 1, 2, 3$ and $n = (100, 500, 1000)$. Here D_{1i} and D_{2i} are both time-stationary covariates and X_{1it} is a time-dependent covariate. The pairwise correlations between the observations at the three occasions within a subject are assumed to be 0.5. The covariance matrix of the uniform random variables $(X_{1i1}, X_{1i2}, X_{1i3})$ is denoted by

$$\boldsymbol{\Sigma} = \frac{1}{3}\begin{bmatrix} 1 & 0.5 & 0.5 \\ 0.5 & 1 & 0.5 \\ 0.5 & 0.5 & 1 \end{bmatrix}.$$

The simulated correlated random numbers (u_1, u_2) based on $\boldsymbol{\Sigma}$ are shown in Fig. 1 and are generated by $u_1 \sim \mathrm{Unif}(-1, 1)$ and $u_2 = \mathbf{C}\Lambda^{1/2}u_1$, where $\mathbf{C} = (c_1, c_2, c_3)$ and $\Lambda = (a_1, a_2, a_3)$ are the eigenvectors and the eigenvalues of $\boldsymbol{\Sigma}$, respectively. Then the longitudinal ordinal data are generated from the rMultinom function in the Hmisc package of R software when the marginal probabilities are computed by Eq. (2). The coefficient β is considered from 0 to 5 in steps of 1 to compute the empirical type I error rate $(H_0 : \beta = 0)$ and the powers $(\beta = 1, 2, 3, 4, 5)$ of the proposed tests for detecting the interaction term in Model 1 and the squared term of a continuous covariate in Model 2. The simulated results for 1000 replications are shown in Table 1.

The data presented in Table 1 indicate that for all correlation structures the empirical type I error rates of tests G and U by their asymptotic normal distribution are quite close to the 5% level of significance, except the distribution of G has slightly heavier tails and the distribution of U has slightly lighter tails under the independent structure in Model 2. Obviously the

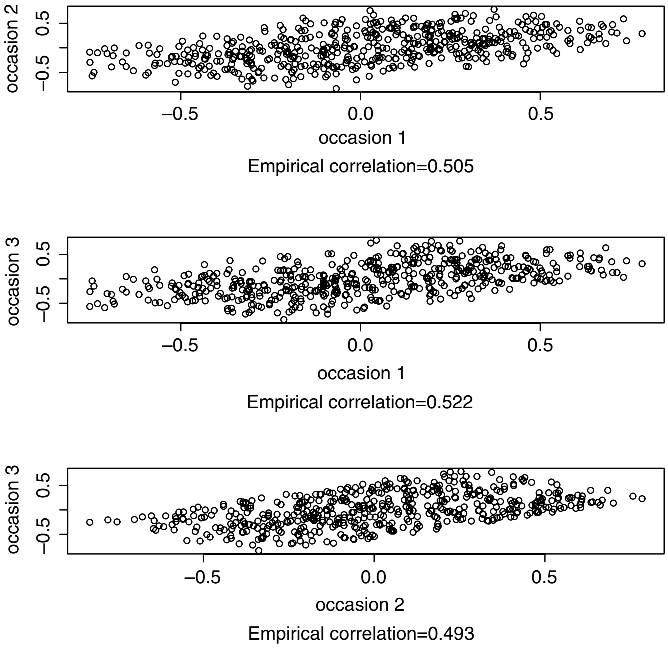

Fig. 1. The pairwise scatter plots of random numbers for three occasions.

Table 1
The empirical type I error rates and powers of proposed tests G and U under four variant working correlation structures for $n = 100, 500, 1000$ and $\alpha = 0.05$ in Models 1 and 2.

β	Independent			AR(1)			Exchangeable			Unspecified		
	100	500	1000	100	500	1000	100	500	1000	100	500	1000
Model 1												
G												
0	0.052	0.055	0.058	0.042	0.059	0.051	0.052	0.050	0.052	0.062	0.058	0.054
1	0.396	0.512	0.598	0.315	0.345	0.381	0.352	0.396	0.452	0.213	0.265	0.304
2	0.545	0.612	0.687	0.490	0.516	0.539	0.518	0.531	0.564	0.367	0.419	0.435
3	0.699	0.764	0.803	0.607	0.626	0.644	0.620	0.667	0.701	0.557	0.573	0.598
4	0.761	0.821	0.894	0.717	0.840	0.919	0.733	0.758	0.798	0.648	0.739	0.765
5	0.836	0.897	0.945	0.799	0.906	0.939	0.788	0.824	0.865	0.734	0.786	0.806
U												
0	0.050	0.061	0.053	0.047	0.053	0.058	0.041	0.049	0.051	0.044	0.049	0.045
1	0.767	0.795	0.825	0.698	0.712	0.721	0.715	0.724	0.754	0.670	0.681	0.698
2	0.869	0.887	0.906	0.823	0.871	0.901	0.838	0.842	0.857	0.797	0.825	0.849
3	0.949	0.967	0.988	0.943	0.955	0.968	0.939	0.945	0.960	0.936	0.942	0.957
4	0.987	0.994	1.000	0.981	0.996	0.999	0.986	0.996	0.999	0.978	0.992	0.999
5	0.996	1.000	1.000	1.000	1.000	1.000	0.995	0.999	0.999	0.994	0.999	1.000
Model 2												
G												
0	0.068	0.065	0.061	0.046	0.054	0.047	0.057	0.053	0.054	0.044	0.046	0.049
1	0.295	0.315	0.328	0.191	0.221	0.241	0.208	0.218	0.287	0.094	0.122	0.135
2	0.422	0.436	0.450	0.296	0.316	0.331	0.329	0.369	0.414	0.131	0.215	0.234
3	0.518	0.529	0.541	0.399	0.460	0.483	0.491	0.528	0.564	0.275	0.299	0.315
4	0.636	0.659	0.673	0.511	0.521	0.535	0.570	0.604	0.687	0.327	0.342	0.360
5	0.695	0.719	0.728	0.606	0.687	0.702	0.638	0.691	0.786	0.479	0.499	0.516
U												
0	0.036	0.038	0.046	0.042	0.052	0.047	0.046	0.053	0.055	0.039	0.051	0.053
1	0.795	0.811	0.828	0.683	0.701	0.721	0.730	0.788	0.824	0.631	0.648	0.659
2	0.801	0.815	0.837	0.716	0.779	0.790	0.752	0.821	0.851	0.642	0.657	0.667
3	0.841	0.859	0.867	0.770	0.812	0.823	0.768	0.833	0.861	0.697	0.710	0.721
4	0.871	0.889	0.893	0.823	0.850	0.888	0.834	0.861	0.884	0.712	0.735	0.758
5	0.905	0.928	0.941	0.890	0.914	0.930	0.886	0.901	0.926	0.858	0.870	0.891

Table 2
The empirical type I error rates and powers of proposed tests G and U under four variant working correlation structures for $n = 100, 500, 1000$ and $\alpha = 0.05$ in Model 3.

β	Independent			AR(1)			Exchangeable			Unspecified		
	100	500	1000	100	500	1000	100	500	1000	100	500	1000
Model 3												
G												
0	0.068	0.058	0.062	0.069	0.060	0.071	0.054	0.052	0.057	0.059	0.056	0.055
1	0.384	0.405	0.418	0.308	0.326	0.333	0.340	0.350	0.364	0.252	0.269	0.280
2	0.476	0.487	0.495	0.394	0.435	0.459	0.416	0.421	0.440	0.321	0.342	0.358
3	0.564	0.576	0.588	0.456	0.468	0.484	0.504	0.524	0.551	0.379	0.391	0.412
4	0.646	0.660	0.678	0.558	0.628	0.640	0.585	0.637	0.658	0.481	0.528	0.534
5	0.737	0.746	0.754	0.634	0.691	0.726	0.677	0.718	0.748	0.516	0.542	0.554
U												
0	0.063	0.062	0.074	0.071	0.072	0.068	0.056	0.052	0.058	0.057	0.053	0.053
1	0.851	0.864	0.875	0.793	0.815	0.830	0.788	0.799	0.815	0.724	0.742	0.756
2	0.859	0.866	0.879	0.841	0.859	0.876	0.815	0.825	0.834	0.812	0.826	0.843
3	0.911	0.931	0.948	0.872	0.884	0.909	0.866	0.875	0.880	0.835	0.852	0.867
4	0.943	0.957	0.968	0.919	0.968	0.986	0.910	0.935	0.964	0.916	0.928	0.941
5	0.987	0.992	0.999	0.970	0.990	0.997	0.970	0.988	0.994	0.968	0.978	0.989

powers of the proposed tests G and U increase as the sample size increase. The powers of test U dominate over those of test G for all correlation structures in Models 1 and 2. This coincides with the result discussed in Hosmer et al. (1997) that test U has better power performance. Table 1 also shows that the powers for detecting the interaction term in Model 1 are larger than those for detecting the squared term in Model 2.

In many clinical trials the data go along with discrete covariates and repeated measurements. For example, a randomized double-blind clinical trial compared an active hypnotic drug with a placebo in patients who have insomnia problems. A longitudinal ordinal data set consisting of ordinal responses and discrete covariates over time is generated from the following proportional odds model:

$$\text{Model 3: } \operatorname{logit}(\eta_{it}^{(j)}) = \lambda_{3j} - 0.1D_{2i} + 0.1D_{3i} + 0.3t + \beta_3 D_{1i}t,$$

where $\lambda_3 = (-0.5, 0, 0.5)$, D_{1i} and D_{2i} are defined as above and D_{3i} follows a multinomial distribution with probability 0.25, 0.45 and 0.30 for $i = 1, \ldots, n$ and $n = (100, 500, 1000)$. The probability of Bernoulli distribution for D_{2i} is set at 0.5 as the patients were randomly assigned to the active group or the placebo group. D_{3i} represents the baseline measurement and t denotes the follow-up visit for $t = 1, 2, 3$. To detect the interaction effect of treatment and time on the ordinal response, the coefficient β_3 is considered from 0 to 0.5 in steps of 0.1 to compute the empirical type I error rates and powers of the proposed tests. The results for 1000 simulations are summarized in Table 2. It reveals that the empirical type I error rates are a little larger than the specified level of significance 0.05 for all correlation structures, and the distributions of test statistics G and U have slightly heavier tails. In addition, test U has better power performance than test G for all configurations in Model 3.

4. Examples

To illustrate the application of the proposed tests for assessing the model fit with longitudinal ordinal responses and continuous as well as categorical covariates, two data sets from pill palatability and marijuana studies are discussed in the following.

4.1. Pill palatability

The pill study cited by Hand and Crowder (1996) was used to investigate the palatability of circular oral treatments for dogs related to two time-stationary covariates (baseline age of the dog and gender of the dog) and one time-dependent covariate (the drug formulations vary with the days). Each drug has four formulations and each dog receives each formulation twice through the eight occasions. The response variable is the palatability recorded by ordinal score as 1 = readily accepted, 2 = accept or reluctantly accept, and 3 = accepted in food or refused. The main interest of this study is to examine the effect of formulation on the response, and a proportional odds model for the longitudinal ordinal outcomes is considered as follows:

$$\operatorname{logit}(\eta_{it}^{(j)}) = \lambda_j + \beta_1 X_{1i} + \beta_2 X_{2i} + \beta_3 X_{3it}, \quad \text{for } i = 1, \ldots, 32;\ t = 1, \ldots, 8;\ j = 1, 2,$$

where X_1 and X_2 denote the gender of the dog and baseline age of the dog, respectively, and X_3 is the drug formulation.

The values of goodness-of-fit test statistics G and U and their means, variances and p-values under independent, AR(1), exchangeable and unspecified correlation structures as well as parameter estimates are shown in Table 3. The p-values

Table 3
The values of goodness-of-fit test statistics, their estimated means and variances as well as p-values and parameter estimates for the pill palatability data.

	Independent	AR(1)	Exchangeable	Unspecified
G	539.024	539.024	539.024	539.024
$\widehat{\mathrm{E}(G)}$	512.000	512.000	512.000	512.000
$\widehat{\mathrm{Var}(G)}$	54.371	143.660	227.831	24527.510
p-value	0.000	0.024	0.073	0.863
U	108.841	108.841	108.841	108.841
$\widehat{\mathrm{E}(U)}$	104.573	104.573	104.573	104.573
$\widehat{\mathrm{Var}(U)}$	1.637	4.743	7.775	685.513
p-value	0.001	0.051	0.126	0.871
	$\hat{\beta}$	SE	Wald stat.	p-value
Intercept 1	−0.196	0.626	0.098	0.754
Intercept 2	0.785	0.624	1.585	0.208
Gender	−0.738	0.656	1.266	0.261
Age	−0.047	0.073	0.411	0.521
Formulation	0.281	0.084	11.151	0.0008

of the proposed tests G and U are both greater than 0.05 under exchangeable and unspecified correlation structures, and confirm the adequate fit of the current model, while they are less than 0.05 under independence for both tests and under AR(1) for test G. In longitudinal studies, some researchers may prefer using independent, AR(1), exchangeable or unspecified correlation structures. The four variant correlation structures are commonly used for describing the association of responses. The testing results may produce different conclusions. When the goodness-of-fit test rejects the current model, it conveys a message of model inadequacy and the model may need to be fixed by utilizing some other measures. In contrast, when the goodness-of-fit test does not reject the current model, this does not guarantee that the model fits well and it requires one to use more techniques to confirm the model adequacy. If the adequate model holds, the estimate of the model parameter for drug formulation is $\beta_3 = 0.281$ (SE $= 0.084$), and the Wald test statistic of 11.151 (p-value $= 0.0008$) demonstrates the effect of formulation to be convincingly confirmed.

4.2. Marijuana study

The second example, taken from a study of marijuana use for teenagers, was analyzed by Lang et al. (1999) and Vermunt and Hagenaars (2004), where 237 thirteen-year-old children (117 boys and 120 girls) who had used marijuana in the past were observed over a period of five consecutive years. This data set was originally from the US National Youth Survey (Elliot et al., 1989) in a longitudinal study of delinquency and drug use. The response variable was a trichotomous ordinal variable (1, never; 2, not more than once a month; 3, more than once a month), and two covariates were annual time (1976–1980) and gender of children. The study goal was to investigate the effects of time and gender on marijuana use for teenagers from five annual waves. Vermunt and Hagenaars (2004) considered modeling this longitudinal ordinal data as an additive model,

$$\mathrm{logit}(\eta_{it}^{(j)}) = \lambda_j + \beta_1 t + \beta_2 X_i,$$

where t is the annual wave for $t = 1, \ldots, 5$, the gender variable $X_i = 0$ for boys; $i = 1, \ldots, 117$ and $X_i = 1$ for girls; $i = 118, \ldots, 237$.

The values of goodness-of-fit test statistics G and U, their means and variances as well as p-values and parameter estimates under four correlation structures are displayed in Table 4. The p-values of tests G and U are all greater than 0.05 except for the test U under independence and AR(1) correlation structures, leading to the goodness-of-fit of the additive model. A more complicated model containing the time effect differing by gender, $\mathrm{logit}(\eta_{it}^{(j)}) = \lambda_j + \beta_1 t + \beta_2 X_i + \beta_3 t X_i$, was also considered by Vermunt and Hagenaars (2004). The p-values of test G for the model with the interaction term are 0.051, 0.159, 0.171 and 0.242 (not shown) under independent, AR(1), exchangeable and unspecified correlation structures, respectively, while the p-values of test U are 2E−5, 0.018, 0.046 and 0.165 under independent, AR(1), exchangeable and unspecified correlation structures, respectively. The result of goodness-of-fit test for the interaction model reaches the same conclusion as the additive model, namely that the model fit is appropriate. However, the inference of the interaction term is not significant (p-value $= 0.449$). Consequently, the additive model with two main effects of time and gender for the marijuana data is a prior consideration.

5. Conclusion and discussion

The main purpose of this article is to propose goodness-of-fit tests based on the Pearson chi-squared statistic (G) and unweighted sum of residual squares (U) for modeling longitudinal ordinal data. The proposed tests can be regarded as an extension of Pan's statistics, which is easy to compute and able to check the model fit with both continuous and categorical covariates. Simulated results reveal that the proposed test statistic based on unweighted sum of residual squares has better

Table 4
The values of goodness-of-fit test statistics, their estimated means and variances as well as p-values and parameter estimates for the marijuana study.

	Independent	AR(1)	Exchangeable	Unspecified
G	2494.703	2494.703	2494.703	2494.703
$\widehat{\mathrm{E}(G)}$	2370.000	2370.000	2370.000	2370.000
$\widehat{\mathrm{Var}(U)}$	4424.517	12581.030	13149.130	31800.200
p-value	0.061	0.266	0.277	0.484
U	359.231	359.231	359.231	359.231
$\widehat{\mathrm{E}(U)}$	343.418	343.418	343.418	343.418
$\widehat{\mathrm{Var}(U)}$	14.208	46.072	65.667	413.831
p-value	0.000	0.020	0.051	0.437
	$\hat{\beta}$	SE	Wald stat.	p-value
Intercept 1	3.112	0.302	106.373	0.000
Intercept 2	4.213	0.330	163.149	0.000
Time	−0.850	0.277	9.393	0.002
Gender	−0.493	0.059	69.666	0.000

power performance than the test based on the Pearson chi-squared type statistic for large sample sizes. To demonstrate the application of the proposed approaches, studies of pill palatability and marijuana use are discussed.

The idea of our goodness-of-fit tests is to utilize GEE methodology for repeated ordinal data referred to by Lipsitz et al. (1994), and combine it with the sum of residual squares. Alternative competing classes for modeling repeated ordinal data based on the marginal pairwise global odds ratios model, transitional models and random-effect growth models can be found in Williamson et al. (1995), Heagerty and Zeger (1996), Vermunt and Hagenaars (2004), and Lee and Daniels (2007, 2008). In this article, four major variants of working covariances as the measure of association between the responses at different occasions for each subject are considered. Williamson et al. (1995) provided the methodology for the measure of association by using the global odds ratio approach. The test statistics for assessing the goodness-of-fit of proportional odds models using a GEE approach with global odds ratio as the measure of association can be future work. Another class of model checking tests for correlated discrete data, conditional moment tests (CMTs), can cover some of other tests in the literature by defining a particular function vector corresponding to the moment function vector referred to by Pan (2002b). The basic idea of CMTs is to test the closeness to zero of the sample moment for assessing the fitness of the specified model.

Other methods utilizing nonparametric smoothing may be of interest. For example, Lin et al. (2008) developed a goodness-of-fit test for longitudinal binary data using multivariate local polynomial smoothing of the standardized residuals obtained after fitting the null model. The advantage of a goodness-of-fit test based on nonparametric smoothing residuals with asymptotic scaled chi-squared distribution is a global measure statistic to amend the inappropriateness of the fixed-cells asymptotics dilemma. However, a serious drawback of the test statistic based on a smoothing technique is that it requires tedious computation. This approach can be applied to longitudinal ordinal data and should be an interesting topic of the ongoing research.

Acknowledgements

The author is grateful to the editors and reviewers for their valuable comments and constructive suggestions on the original draft, which led to substantial improvements in this paper. The research was supported in part by National Science Council project NSC-98-2118-M-165-001.

References

Agresti, A., 1999. Modelling ordered categorical data: Recent advances and future challenges. Statistics in Medicine 18, 2191–2207.
Barnhart, H.X., Williamson, J.M., 1998. Goodness-of-fit tests for GEE modeling with binary responses. Biometrics 54, 720–729.
Clogg, C.C., Shihadeh, E.S., 1994. Statistical Models for Ordinal Data. Thousand Oaks.
Elliot, D.S., Huizinga, D., Menard, S., 1989. Multiple Problem Youth: Delinquence, Substance use and Mental Health Problems. Springer-Verlag, New York.
Fahrmeir, L., Pritscher, L., 1996. Regression analysis of forest damage by marginal models for correlated ordinal responses. Journal of Environmental and Ecological Statistics 3, 257–268.
Hand, D., Crowder, M., 1996. Practical Longitudinal Data Analysis. Chapmam & Hall, New York.
Heagerty, P.J., Zeger, S.L., 1996. Marginal regression models for clustered ordinal measurements. Journal of the American Statistical Association 91, 1024–1036.
Horton, N.J., Bebchuk, J.D., Jones, C.L., Lipsitz, S.R., Catalano, P.J., Zahner, G.P., Fitzmaurice, G.M., 1999. Goodness of fit for GEE: An example with mental health service utilization. Statistics in Medicine 18, 213–222.
Hosmer, D.W., Hosmer, T., LeCessie, S.L., Lemeshow, S., 1997. A comparison of goodness-of-fit tests for the logistic regression model. Statistics in Medicine 16, 965–980.
Hosmer, D.W., Lemeshow, S.L., 1980. A goodness-of-fit test for the multiple logistic regression model. Communications in Statistics A10, 1043–1069.
Kauermann, G., 2000. Modeling longitudinal data with ordinal response by varying coefficients. Biometrics 56, 692–698.
Lang, J.B., McDonald, J.W., Smith, P.W.F., 1999. Association modeling of multivariate categorical responses: A maximum likelihood approach. Journal of the American Statistical Association 94, 1161–1171.
Lee, K., Daniels, M., 2007. A Class of Markov Models for Longitudinal Ordinal Data. Biometrics 63, 1060–1067.

Lee, K., Daniels, M., 2008. Marginalized models for longitudinal ordinal data with application to quality of life studies. Statistics in Medicine 27, 4359–4380.
Liang, K.Y., Zeger, S.L., 1986. Longitudinal data analysis using generalized linear models. Biometrika 73, 13–22.
Lin, K.C., Chen, Y.J., Shyr, Y., 2008. A nonparametric smoothing method for assessing GEE models with longitudinal binary data. Statistics in Medicine 27, 4428–4439.
Lin, H.Y., Myers, L., 2006. Power and type I error rates of goodness-of-fit statistics for binomial generalized estimating equations (GEE) models. Computational Statistics & Data Analysis 50, 3432–3448.
Lipsitz, S.R., Kim, K., Zhao, L., 1994. Analysis of repeated categorical data using generalized estimating equations. Statistics in Medicine 13, 1149–1163.
Liu, L.C., Hedeker, D., 2006. A mixed-effects regression model for longitudinal multivariate ordinal data. Biometrics 62, 261–268.
McCullagh, P., 1980. Regression models for ordinal data. Journal of the Royal Statistical Society, Series B 42, 109–142.
Miller, M.E., Davis, C.S., Landis, J.R., 1993. The analysis of longitudinal polytomous data: Generalized estimating equations and connections with weighted least squares. Biometrics 49, 1033–1044.
Osius, G., Rojek, D., 1992. Normal goodness-of-fit tests for multinomial models with large degrees of freedom. Journal of the American Statistical Association 87, 1145–1152.
Pan, W., 2002a. Goodness-of-fit tests for GEE with correlated binary data. Scandinavian Journal of Statistics 29, 101–110.
Pan, W., 2002b. Application of conditional moment tests to model checking for generalized linear models. Biostatistics 3, 267–276.
Pan, Z., Lin, D.Y., 2005. Goodness-of-fit methods for generalized linear mixed models. Biometrics 61, 1000–1009.
Tsiatis, A.A., 1980. A note on a goodness-of-fit test for the logistic regression model. Biometrika 67, 250–251.
Vermunt, J.K., Hagenaars, J.A., 2004. Ordinal longitudinal data analysis. In: Hauspie, R.C., Cameron, N., Molinari, L. (Eds.), Methods in Human Growth Research. Cambridge University Press.
Vonesh, E.F., Chinchilli, V.M., Pu, K., 1996. Goodness-of-fit in generalized nonlinear mixed-effects models. Biometrics 52, 572–587.
Williamson, J.M., Kim, K., Lipsitz, S.R., 1995. Analyzing bivariate ordinal data using a global odds ratio. Journal of the American Statistical Association 90, 1432–1437.
Williamson, J.M., Lin, H.M., Barnhart, H.X., 2003. A classification statistic for GEE categorical response models. Journal of Data Science 1, 149–165.