



---

*Research article***Consistency of the true correlation matrix estimator in generalized estimating equations for high-dimensional longitudinal data****Xianbin Chen\*** and Donghai Peng

Innovation and Entrepreneurship College, Zhongshan Polytechnic, Zhongshan, 528400, China

\* **Correspondence:** Email: xian\_bin\_chen@163.com.

**Abstract:** In this study, we consider the estimator of true correlation matrix within generalized estimating equations for high-dimensional longitudinal data. More precisely, a modified unit correlation matrix structure is constructed, and initial estimator during the iterative procedure is derived. Under mild regularity conditions, we establish the asymptotic consistency of the true correlation matrix estimator by techniques including the existence theorem of roots and Markov's inequality. Finally, we also perform some Monte Carlo studies to illustrate the theoretical analysis. The main result in this paper extends existing high-dimensional GEE theory to the multinomial response variable case.

**Keywords:** generalized estimating equations; true correlation matrix; consistency; multinomial response variable; high-dimensional longitudinal data

**Mathematics Subject Classification:** 62E20, 62F12, 62J12

---

**1. Introduction**

Longitudinal data arise frequently in biomedical and social science research, where the same subjects are measured repeatedly over time. The generalized estimating equation (GEE) method of Liang and Zeger [1] is now the standard approach for analyzing such data. A key ingredient of GEE is the working correlation matrix, which describes the within-subject dependence among repeated observations. The regression coefficient estimator remains consistent and asymptotically normal even when this matrix is misspecified, but misspecification costs efficiency. To attain optimal efficiency, the working correlation matrix must consistently estimate the true underlying correlation structure, and this concern has driven much of the methodological development in GEE.

In the original formulation of Liang and Zeger [1], the working correlation matrix  $R(\tau)$  is indexed by a finite-dimensional nuisance parameter  $\tau$ , and the asymptotic theory requires  $\sqrt{n}$ -consistency of  $\hat{\tau}$ . This parameterized view dominated the early literature. Xie and Yang [2] and Balan and Schiopu-

Kratina [3] studied asymptotic properties under large cluster sizes, and Fahrmeir and Kaufmann [4] established consistency and asymptotic normality for generalized linear models. Zeger and Liang [5] extended the framework to discrete and continuous outcomes. More recent work along these lines includes Goshio et al. [6] on bias-adjusted GEE for sparse binary data and Yi, Zeng and Cook [7] on robust pairwise likelihood for incomplete longitudinal data.

The correlation coefficient is not always a convenient parameterization, particularly for binary outcomes where correlations are bounded and hard to estimate. Several alternatives were explored. Prentice [8] added a second set of estimating equations for pairwise correlations, giving rise to GEE2. Lipsitz, Laird and Harrington [9] used the odds ratio as the association measure, and Carey, Zeger and Diggle [10] developed alternating logistic regressions. Qu, Lindsay and Li [11] proposed the quadratic inference function, which represents  $R^{-1}(\tau)$  as a linear combination of basis matrices and avoids estimating the correlation parameter directly. Pan [12] introduced the QIC criterion for selecting among working structures. Other parameterizations and selection criteria were studied by Lange et al. [13], Shults and Ratcliffe [14], and Fu, Hao and Wang [15].

High-dimensional covariates changed the picture. When  $p_n$  grows with  $n$ , the  $\sqrt{n}$ -consistency requirement on  $\hat{\tau}$  is no longer the natural regime. Wang [16] bypassed the nuisance parameter entirely: rather than estimating  $\tau$ , Wang constructed a fixed positive-definite matrix  $\bar{R}$  and assumed that

$$\|\hat{R}^{-1} - \bar{R}^{-1}\| = O_p(\sqrt{p_n/n}),$$

even as  $p_n$  diverges. This result laid the foundation for high-dimensional GEE theory. Chen and Yin [17] subsequently established the existence, consistency and asymptotic normality of the general GEE estimator in this framework. Since then, high-dimensional GEE has developed along several directions. Penalized variable selection was studied by Wang et al. [18], Green et al. [19], Kamruzzaman et al. [20], and Zhou et al. [21]. Inference and testing were addressed by Xia and Shojaie [22], Fang, Ning and Li [23], and Dumitrescu et al. [24]. Robust estimation of the correlation matrix itself was considered by Lu, Xue and Hu [25], Lu, Xue and Cai [26], and Tang and Wang [27].

However, Wang's [16] consistency result, is restricted to binary responses. In fact, polytomous ordinal outcomes are common in two broad settings. The first involves high-dimensional longitudinal clinical follow-up: In chronic disease cohort studies, patients are assessed repeatedly, and outcomes such as disease severity (mild/moderate/severe) or treatment response are recorded along with a large set of biomarker and genomic covariates [28,29]. The second involves consumer panel data: Respondents are tracked across multiple survey waves, and satisfaction or preference ratings are measured alongside a high-dimensional bank of questionnaire items [30]. In both settings, the within-subject correlation structure is unlikely to follow a simple exchangeable or AR-1 form. A misspecified working correlation matrix then reduces the efficiency of regression coefficient estimates. By providing a consistent estimator of the true correlation matrix for multi-category longitudinal data with diverging covariates, our method directly improves GEE inference precision. Existing GEE methods for polytomous longitudinal data, such as Miller, Davis and Landis [31] and Heagerty and Zeger [32], operate in the fixed-dimensional regime and do not address consistency of the correlation matrix estimator when  $p_n$  diverges.

This paper extends Wang's [16] consistency result to multinomial responses. We use a two-step logit model for a three-category ordinal response. The extension involves more than a change of notation. The response vector doubles in dimension, the design matrix is block diagonal, and the

initial estimating equation differs structurally from Eq (3.2) in Wang [16]. The proof relies on a block-matrix decomposition (Lemma 4.2) that do not arise in the binary case. Under regularity conditions, we prove the asymptotic consistency of the true correlation matrix. Finally, the performance of the established theory is illustrated using Monte Carlo simulations.

The remainder of this paper is organized as follows: Section 2 introduces the two-step logit model and derives the corresponding generalized estimating equations. Section 3 presents the asymptotic properties of the true correlation matrix estimator. Section 4 provides a detailed proof of the main result. Section 5 presents Monte Carlo simulations to verify the theoretical analysis. Section 6 gives a brief conclusion.

**Notation:** Throughout this paper, superscript “ $'$ ” always denotes the transpose of a vector or a matrix.  $\text{Tr}(\mathbf{A})$  is the trace of a matrix  $\mathbf{A}$ . Moreover, we use  $\|\cdot\|$  to denote the Frobenius norm. For a matrix  $\mathbf{D}$ ,  $\|\mathbf{D}\| = [\text{Tr}(\mathbf{D}\mathbf{D}')]^{1/2}$ . Particularly, for a vector  $\mathbf{x}$ ,  $\|\mathbf{x}\| = \|\mathbf{x}\|_2 = (x_1^2 + \cdots + x_n^2)^{1/2}$ . For any vector  $\mathbf{v}$ ,  $\text{diag}(\mathbf{v})$  represents a diagonal matrix whose diagonal elements are the elements of  $\mathbf{v}$ .

## 2. Model description

### 2.1. Preliminaries

Suppose  $(Y_{ij}, \mathbf{X}_{ij})$  are observations for the  $j$ th measurement on the  $i$ th subject,  $i = 1, 2, \dots, n$ , and  $j = 1, 2, \dots, m$ , where  $Y_{ij}$  is a response variable,  $\mathbf{X}_{ij}$  is a  $p_n$ -dimensional vector of covariates. Let  $\mathbf{Y}_i = (Y_{i1}, \dots, Y_{im})'$  denote the vector of responses for the  $i$ th subject, and assume  $\mathbf{X}_i = (\mathbf{X}_{i1}, \dots, \mathbf{X}_{im})'$  be the associated  $m \times p_n$  matrix of covariates. The dimension of the covariates  $p_n$  is allowed to depend on the number of subjects  $n$ . Note that, the observations are independent if they are from different subjects, but are correlated if they are from the same subject.

Denote the marginal expectation and variance of  $Y_{ij}$  by  $\pi_{ij}(\boldsymbol{\beta}_n) = E_{\boldsymbol{\beta}_n}(Y_{ij})$  and  $A_{ij}(\boldsymbol{\beta}_n) = \text{Var}_{\boldsymbol{\beta}_n}(Y_{ij})$ . Let  $\boldsymbol{\pi}_i(\boldsymbol{\beta}_n) = (\pi_{i1}(\boldsymbol{\beta}_n), \dots, \pi_{im}(\boldsymbol{\beta}_n))'$  and  $\mathbf{A}_i(\boldsymbol{\beta}_n) = \text{diag}(A_{i1}(\boldsymbol{\beta}_n), \dots, A_{im}(\boldsymbol{\beta}_n))$ . According to [1], the optimal estimating equations for  $\boldsymbol{\beta}_n$  is given by

$$\sum_{i=1}^n \left( \frac{\partial \boldsymbol{\pi}_i'}{\partial \boldsymbol{\beta}_n} \right) \mathbf{V}_i^{-1}(\boldsymbol{\beta}_n, \boldsymbol{\tau}) (\mathbf{Y}_i - \boldsymbol{\pi}_i(\boldsymbol{\beta}_n)) = \mathbf{0},$$

where  $\mathbf{V}_i(\boldsymbol{\beta}_n, \boldsymbol{\tau})$  is a working covariance matrix of  $\mathbf{Y}_i$ . In applications, the true correlation matrix of  $\mathbf{Y}_i$ , denoted by  $\mathbf{R}_0$ , is unknown. The GEE procedure adopts a working covariance matrix, which is specified through a working correlation matrix  $\mathbf{R}(\boldsymbol{\tau})$ :  $\mathbf{V}_i(\boldsymbol{\beta}_n, \boldsymbol{\tau}) = \mathbf{A}_i^{1/2}(\boldsymbol{\beta}_n) \mathbf{R}(\boldsymbol{\tau}) \mathbf{A}_i^{1/2}(\boldsymbol{\beta}_n)$ , where  $\boldsymbol{\tau}$  is a finite dimensional parameter. It should be noted that, if  $\mathbf{R}(\boldsymbol{\tau})$  is equal to the true correlation matrix  $\mathbf{R}_0$ , then  $\mathbf{V}_i(\boldsymbol{\beta}_n, \boldsymbol{\tau}) = \text{Cov}(\mathbf{Y}_i)$  at the true parameter  $\boldsymbol{\beta}_{n0}$ . This motivates the need for consistent estimation of  $\mathbf{R}_0$ , which is the primary focus of this paper.

### 2.2. Two-step logit model

Without loss of generality, consider a response variable  $Y$  with three categories, denoted as  $S = \{1, 2, 3\}$ . Suppose  $S$  is partitioned into two subsets:  $S_1 = \{1\}$  and  $S_2 = \{2, 3\}$ . The model is constructed in two stages as follows.

**Stage 1:** Define the probability that  $Y$  belongs to  $S_1$  as

$$P(Y \in S_1 | \mathbf{X}) = \frac{\exp(\mathbf{X}'\boldsymbol{\beta}_1)}{1 + \exp(\mathbf{X}'\boldsymbol{\beta}_1)},$$

where  $\mathbf{X}$  is a covariate vector and  $\boldsymbol{\beta}_1$  is the corresponding regression parameter vector.

**Stage 2:** Conditional on  $Y \in S_2$ , define the probability that  $Y = 2$  as

$$P(Y = 2 | Y \in S_2, \mathbf{X}) = \frac{\exp(\mathbf{X}'\boldsymbol{\beta}_2)}{1 + \exp(\mathbf{X}'\boldsymbol{\beta}_2)},$$

where  $\boldsymbol{\beta}_2$  is another regression parameter vector. From these definitions, the marginal probabilities for each category are derived as follows:

$$\begin{aligned} P(Y = 1 | \mathbf{X}) &= \frac{\exp(\mathbf{X}'\boldsymbol{\beta}_1)}{1 + \exp(\mathbf{X}'\boldsymbol{\beta}_1)}, \\ P(Y = 2 | \mathbf{X}) &= \frac{\exp(\mathbf{X}'\boldsymbol{\beta}_2)}{[1 + \exp(\mathbf{X}'\boldsymbol{\beta}_2)][1 + \exp(\mathbf{X}'\boldsymbol{\beta}_1)]}, \\ P(Y = 3 | \mathbf{X}) &= \frac{1}{[1 + \exp(\mathbf{X}'\boldsymbol{\beta}_2)][1 + \exp(\mathbf{X}'\boldsymbol{\beta}_1)]}. \end{aligned}$$

This model is referred to as the two-step logit model. Further details are comprehensively provided by Morawitz and Tutz [33].

### 2.3. Two-step logit model in GEE

Suppose  $(\bar{y}_{ij}, \bar{\mathbf{X}}_{ij})$  represents the observed data for the  $j$ th measurement of the  $i$ th subject, where  $i = 1, \dots, n$ ,  $j = 1, \dots, m$ .  $\bar{y}_{ij}$  is the response, and  $\bar{\mathbf{X}}_{ij}$  is a  $\bar{p}_n$ -dimensional covariate vector. Define the category probabilities as

$$\pi_{ijr} = P(\bar{y}_{ij} = r | \bar{\mathbf{X}}_{ij}), \quad r = 1, 2, 3.$$

Since  $\pi_{ij1} + \pi_{ij2} + \pi_{ij3} = 1$ , it suffices to model  $\pi_{ij1}$  and  $\pi_{ij2}$ . Additional notation is introduced

$$Y_{ij1} = \begin{cases} 1, & \bar{y}_{ij} = 1, \\ 0, & \bar{y}_{ij} = 2 \text{ or } 3, \end{cases} \quad Y_{ij2} = \begin{cases} 1, & \bar{y}_{ij} = 2, \\ 0, & \bar{y}_{ij} = 1 \text{ or } 3. \end{cases}$$

Let

$$\mathbf{Y}_i = (\mathbf{Y}'_{i1}, \dots, \mathbf{Y}'_{im})', \quad \mathbf{Y}_{ij} = (Y_{ij1}, Y_{ij2})', \quad \mathbf{X}_i = (\mathbf{X}_{i1}, \dots, \mathbf{X}_{im})',$$

where the design matrix is

$$\mathbf{X}'_{ij} = \begin{pmatrix} (1, \bar{\mathbf{X}}'_{ij}) & \mathbf{0} \\ \mathbf{0} & (1, \bar{\mathbf{X}}'_{ij}) \end{pmatrix}_{2 \times 2(1 + \bar{p}_n)}.$$

The parameter vector of interest is  $\boldsymbol{\beta}_n = (\boldsymbol{\beta}_{01}, \boldsymbol{\beta}'_1, \boldsymbol{\beta}_{02}, \boldsymbol{\beta}'_2)'$ , which has dimension  $p_n = 2(1 + \bar{p}_n)$ . It is not hard to show that there exists a 2-dimensional link function  $g$  such that  $g(\boldsymbol{\pi}_{ij}) = \mathbf{X}'_{ij}\boldsymbol{\beta}_n$  (see, for example, [34, pp. 63–95]). Furthermore, the expectation of  $\mathbf{Y}_i$  is

$$\boldsymbol{\pi}_i(\boldsymbol{\beta}_n) = E_{\boldsymbol{\beta}_n}(\mathbf{Y}_i) = (\boldsymbol{\pi}'_{i1}(\boldsymbol{\beta}_n), \dots, \boldsymbol{\pi}'_{im}(\boldsymbol{\beta}_n))', \quad \boldsymbol{\pi}_{ij}(\boldsymbol{\beta}_n) = (\pi_{ij1}, \pi_{ij2})'.$$

The variance of  $Y_i$  is given by the diagonal matrix

$$A_i(\beta_n) = \text{diag}(\text{Var}(Y_{i11}), \text{Var}(Y_{i12}), \dots, \text{Var}(Y_{im1}), \text{Var}(Y_{im2})).$$

Let  $\hat{R}$  is the working correlation matrix, the GEE for the two-step logit model as follows:

$$S_n(\beta_n) = \sum_{i=1}^n X_i' H_i(\beta_n) A_i^{-1/2}(\beta_n) \hat{R}^{-1} A_i^{-1/2}(\beta_n) (Y_i - \pi_i(\beta_n)) = \mathbf{0}, \quad (1)$$

where  $H_i(\beta_n) = \text{diag}(H_{i1}(\beta_n), \dots, H_{im}(\beta_n))$ ,  $H_{ij}(\beta_n) = \partial \pi_{ij}(\beta_n) / \partial t_{ij}$ , with  $t_{ij} = (t_{ij1}, t_{ij2})' = X_{ij}' \beta_n$ . Note that, the explicit forms of the marginal expectation and variance are

$$\begin{aligned} \pi_{ij}'(\beta_n) &= \left( \frac{e^{t_{ij1}}}{1 + e^{t_{ij1}}}, \frac{e^{t_{ij2}}}{(1 + e^{t_{ij1}})(1 + e^{t_{ij2}})} \right), \\ A_{ij}(\beta_n) &= \text{diag}(\pi_{ij1}(\beta_n)(1 - \pi_{ij1}(\beta_n)), \pi_{ij2}(\beta_n)(1 - \pi_{ij2}(\beta_n))). \end{aligned} \quad (2)$$

We also consider a modified estimating equations

$$\tilde{S}_n(\beta_n) = \sum_{i=1}^n X_i' H_i(\beta_n) A_i^{-1}(\beta_n) (Y_i - \pi_i(\beta_n)) = \mathbf{0}, \quad (3)$$

where  $\tilde{\beta}_n$  denotes the solution to Eq (3), which will be used to establish the asymptotic properties in the subsequent sections.

**General  $k$ -class extension.** Let  $Y_{ij}$  denote ordinal responses with  $k$  categories, and  $\bar{X}_{ij}$  represent  $\bar{p}_n$ -dimensional covariate vectors for cluster  $i$  at occasion  $j$  for  $i = 1, \dots, n$  and  $j = 1, \dots, m_i$ . Define  $\mathbf{y}_{ij}$  as a vector of  $q(=k-1)$ -dimensional variables, where  $\mathbf{y}_{ij} = (y_{ij1}, \dots, y_{ijq})'$  with  $y_{ijr} = 1$  if response  $Y_{ij} = r$  and 0 else. Besides, let  $\pi_{ij}$  and  $\eta_{ijr}$  be the vector of marginal probabilities and marginal cumulative probabilities, respectively, where  $\pi_{ij} = (\pi_{ij1}, \dots, \pi_{ijq})'$  with  $\pi_{ijr} = P(Y_{ij} = r | \mathbf{x}_{ij}) = P(y_{ijr} = 1 | \mathbf{x}_{ij})$ , and  $\eta_{ijr} = P(Y_{ij} \leq r | \mathbf{x}_{ij}) = \sum_{\ell=1}^r \pi_{ij\ell}$ .

Furthermore, the cumulative logit model for describing the dependence of  $Y_{ij}$  on  $\mathbf{x}_{ij}$  is given by

$$\text{logit}(\eta_{ijr}) = \log \frac{\eta_{ijr}}{1 - \eta_{ijr}} = \phi_r + \mathbf{x}_{ij}' \boldsymbol{\delta} = \gamma_{ijr}, \quad r = 1, \dots, q,$$

where the intercepts  $\phi_1, \dots, \phi_q$  satisfy  $\phi_1 \leq \dots \leq \phi_q$ ,  $\boldsymbol{\delta}$  is the vector of regression coefficients, and  $\gamma_{ijr}$  is the  $r$ th element of a  $q$ -dimensional linear predictor  $\boldsymbol{\gamma}_{ij} = (\gamma_{ij1}, \dots, \gamma_{ijq})'$ . Note that  $\pi_{ij1} = \eta_{ij1}$  and

$$\pi_{ijr} = \frac{\exp(\gamma_{ijr})}{1 + \exp(\gamma_{ijr})} - \frac{\exp(\gamma_{ij,r-1})}{1 + \exp(\gamma_{ij,r-1})}, \quad r = 2, \dots, q.$$

Then, the linear predictor  $\boldsymbol{\gamma}_{ij}$  can be represented as  $\boldsymbol{\gamma}_{ij} = X_{ij}' \beta_n$ , where  $\beta_n = (\phi_1, \dots, \phi_q, \boldsymbol{\delta}')'$  is the  $p_n$ -dimensional parameter vector of interest, and

$$X_{ij}' = \begin{pmatrix} 1 & 0 & \cdots & 0 & \bar{X}_{ij}' \\ 0 & 1 & \cdots & 0 & \bar{X}_{ij}' \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & 1 & \bar{X}_{ij}' \end{pmatrix}_{q \times (q + \bar{p}_n)}$$

is the  $q \times p_n$  design matrix. It is not hard to show that there exists a  $q$ -dimensional link function  $\mathbf{g}$  such that  $\mathbf{g}(\pi_{ij}) = \mathbf{X}'_{ij}\boldsymbol{\beta}_n$  (see Fahrmeir and Tutz [34], pp.73-84).

Compared with the three-class ( $q = 2$ ) case, the response dimension expands from  $2m$  to  $qm$ , the design matrix expands from  $2 \times 2(1 + \bar{p}_n)$  to  $q \times (q + \bar{p}_n)$ . Since  $q$  and  $m$  are fixed, the block-matrix proof in Lemma 4.2 extends naturally. Besides, the Theorem 3.1 remains valid for arbitrary  $k \geq 3$ .

### 3. Main results

To prove consistency for the true correlation matrix  $\mathbf{R}_0$ , we shall impose some regularity conditions as follows:

(A1)  $\sup_{i,j} \|\mathbf{X}_{ij}\| = O(\sqrt{p_n})$ ;

(A2) the unknown parameter  $\boldsymbol{\beta}_n$  belongs to a compact subset  $\mathbf{B} \subseteq R^{p_n}$ , the true parameter value  $\boldsymbol{\beta}_{n0}$  lies in the interior of  $\mathbf{B}$  and there exist positive constants  $c_1, c_2$  such that

$$0 < c_1 \leq \pi_{ij1}(\boldsymbol{\beta}_{n0}), \pi_{ij2}(\boldsymbol{\beta}_{n0}) \leq c_2 < 1, \forall i, j;$$

Note that condition (A2) imposes bounds on marginal probabilities, which correspond to the expectations of our binary-indicator responses. Since these responses only take values 0 and 1, their variances are fully determined by the marginal probabilities, so no extra conditions on responses are needed.

(A3) there exist positive constants,  $c_3, c_4$ , such that

$$c_3 \leq \lambda_{\min}(n^{-1} \sum_{i=1}^n \mathbf{X}'_i \mathbf{X}_i) \leq \lambda_{\max}(n^{-1} \sum_{i=1}^n \mathbf{X}'_i \mathbf{X}_i) \leq c_4$$

where  $\lambda_{\min}$  (resp.  $\lambda_{\max}$ ) denotes the minimum (resp. maximum) eigenvalue of a matrix;

(A4)  $p_n^2/n \rightarrow 0$  ( $n \rightarrow \infty$ ).

Now let us give the main theorem in this paper.

**Theorem 3.1.** *Under assumptions (A1)-(A4), then we have*

$$\|\hat{\mathbf{R}}^{-1} - \mathbf{R}_0^{-1}\| = O_p(\sqrt{p_n/n}),$$

where  $\hat{\mathbf{R}} = \frac{1}{n} \sum_{i=1}^n \mathbf{A}_i^{-1/2}(\tilde{\boldsymbol{\beta}}_n)(\mathbf{Y}_i - \boldsymbol{\pi}_i(\tilde{\boldsymbol{\beta}}_n))(\mathbf{Y}_i - \boldsymbol{\pi}_i(\tilde{\boldsymbol{\beta}}_n))' \mathbf{A}_i^{-1/2}(\tilde{\boldsymbol{\beta}}_n)$ .

**Remark 1.** Conditions (A1)–(A4) are identical to those in Wang [16].

**Remark 2.** As a special case for clustered binary data, the generalized estimating equations in (1) reduce to Eq (3.1) presented by [16].

**Remark 3.** Our main result generalizes that of [16] from a binary to a multinomial response and can be naturally extended to the case where the  $k$  response categories are partitioned into  $t$  subsets.

#### 4. Proof of main results

Throughout this section, let  $c, c_1, c_2, \dots$  denote generic positive constants that may vary from line to line. For notational simplicity, when functions of  $\beta_n$  are evaluated at the true parameter value  $\beta_{n0}$ , we omit the argument; for instance, we write  $S_n = S_n(\beta_{n0})$ ,  $\tilde{S}_n = \tilde{S}_n(\beta_{n0})$ , etc.

Our proofs adapt and extend some key ideas from Wang [16]. To establish Theorem 3.1, we first present two essential lemmas.

**Lemma 4.1.** *Let  $C$  be an open and bounded set in  $\mathbb{R}^n$ , with closure  $\bar{C}$  and boundary  $\partial C$ . Suppose  $F : \bar{C} \subset \mathbb{R}^n \rightarrow \mathbb{R}^n$  is a continuous function satisfying*

$$(\mathbf{x} - \mathbf{x}^0)' F(\mathbf{x}) \leq 0$$

*for some  $\mathbf{x}^0 \in C$  and all  $\mathbf{x} \in \partial C$ . Then the equation  $F(\mathbf{x}) = \mathbf{0}$  has a solution in  $\bar{C}$ .*

*Proof.* For a detailed proof, the reader could refer to the Theorem 4.6.3 of Ortega and Rheinboldt [35].

**Lemma 4.2.** *Under assumptions (A1)–(A4), Eq (3) have a solution  $\tilde{\beta}_n$  such that*

$$\|\tilde{\beta}_n - \beta_{n0}\| = O_p(\sqrt{p_n/n}).$$

*Proof.* Combining (A2) with Lemma 4.1, it suffices to show:  $\forall \varepsilon > 0$ , there exists a constant  $\Delta > 0$  such that for all  $n$  sufficiently large,

$$P \left\{ \sup_{\|\beta_n - \beta_{n0}\| = \Delta \sqrt{p_n/n}} (\beta_n - \beta_{n0})' \tilde{S}_n(\beta_n) < 0 \right\} \geq 1 - \varepsilon. \quad (4)$$

Firstly, an application of Taylor expansion yields that

$$\begin{aligned} & (\beta_n - \beta_{n0})' \tilde{S}_n(\beta_n) \\ &= (\beta_n - \beta_{n0})' \tilde{S}_n + (\beta_n - \beta_{n0})' F_n(\beta_n^*) (\beta_n - \beta_{n0}) \\ &\triangleq I_{n1} + I_{n2}, \end{aligned} \quad (5)$$

where  $F_n(\beta_n^*) = \partial \tilde{S}_n(\beta_n) / \partial \beta_n' |_{\beta_n = \beta_n^*}$  and  $\beta_n^*$  lies between  $\beta_n$  and  $\beta_{n0}$ . Next, we will evaluate  $I_{n1}$  and  $I_{n2}$  respectively. For  $I_{n1}$ , it follows from (A1), (A2) and (2) that

$$\begin{aligned} E\|\tilde{S}_n\|^2 &= E \left[ \sum_{i=1}^n (Y_i - \pi_i)' A_i^{-1} H_i' X_i X_i' H_i A_i^{-1} (Y_i - \pi_i) \right] \\ &\leq E \left[ \sum_{i=1}^n \lambda_{\max}(X_i X_i') \cdot \|H_i\|^2 \cdot \|A_i^{-1}\|^2 \cdot \|Y_i - \pi_i\|^2 \right] \\ &\leq c \cdot \text{Tr} \left( \sum_{i=1}^n X_i X_i' \right) = O(np_n), \end{aligned}$$

which implies  $\|\tilde{S}_n\| = O_p(\sqrt{np_n})$ . It is easy to derive that

$$\sup_{\|\beta_n - \beta_{n0}\| = \Delta \sqrt{p_n/n}} |I_{n1}| \leq \Delta \sqrt{p_n/n} \cdot \|\tilde{S}_n\| \leq \Delta O_p(p_n). \quad (6)$$

For  $I_{n2}$ , define  $\mathbf{H}_{ij}\mathbf{A}_{ij}^{-1} = (\mathbf{w}_{ij1}, \mathbf{w}_{ij2})$ , then  $\mathbf{H}_i\mathbf{A}_i^{-1}(\mathbf{Y}_i - \boldsymbol{\pi}_i)$  can be written as a column vector:

$$(\mathbf{w}'_{i11}(y_{i11} - \pi_{i11}) + \mathbf{w}'_{i12}(y_{i12} - \pi_{i12}), \dots, \mathbf{w}'_{im1}(y_{im1} - \pi_{im1}) + \mathbf{w}'_{im2}(y_{im2} - \pi_{im2}))'.$$

Denote  $\dot{\mathbf{w}}_{ijs} = \partial \mathbf{w}_{ijs} / \partial t'_{ij}$ , and let  $\mathbf{W}_{ij}^{(s)}$  be a block-diagonal matrix with the  $j$ -th block equal to  $\dot{\mathbf{w}}_{ijs}$  and 0 elsewhere, for  $s = 1, 2$ . It is not hard to get that

$$\begin{aligned} \mathbf{F}_n(\boldsymbol{\beta}_n) &= \frac{\partial \tilde{\mathbf{S}}_n(\boldsymbol{\beta}_n)}{\partial \boldsymbol{\beta}'_n} \\ &= - \sum_{i=1}^n \mathbf{X}'_i \mathbf{H}_i(\boldsymbol{\beta}_n) \mathbf{A}_i^{-1}(\boldsymbol{\beta}_n) \mathbf{H}'_i(\boldsymbol{\beta}_n) \mathbf{X}_i \\ &\quad + \sum_{i=1}^n \sum_{j=1}^m \sum_{s=1}^2 \mathbf{X}'_i \mathbf{W}_{ij}^{(s)}(\boldsymbol{\beta}_n) [Y_{ijs} - \pi_{ijs}(\boldsymbol{\beta}_n)] \mathbf{X}_i \\ &\triangleq I_{n21} + I_{n22}. \end{aligned} \quad (7)$$

Utilizing (A1) and (A4), we immediately get that

$$\sup_{\|\boldsymbol{\beta}_n - \boldsymbol{\beta}_{n0}\| \leq \Delta \sqrt{p_n/n}} |\mathbf{X}'_{ij}(\boldsymbol{\beta}_n - \boldsymbol{\beta}_{n0})| \rightarrow 0.$$

Moreover, from (A2) and (2) we have

$$0 < c_1 \leq \pi_{ij1}(\boldsymbol{\beta}_n), \pi_{ij2}(\boldsymbol{\beta}_n) \leq c_2 < 1.$$

It is easily seen that

$$\inf_{\|\boldsymbol{\beta}_n - \boldsymbol{\beta}_{n0}\| \leq \Delta \sqrt{p_n/n}} \lambda_{\min} [\mathbf{H}_i(\boldsymbol{\beta}_n) \mathbf{A}_i^{-1}(\boldsymbol{\beta}_n) \mathbf{H}'_i(\boldsymbol{\beta}_n)] > 0. \quad (8)$$

Using (A3) and (8), we can get that

$$\sup_{\|\boldsymbol{\beta}_n - \boldsymbol{\beta}_{n0}\| \leq \Delta \sqrt{p_n/n}, \|\mathbf{a}_n\|=1} \mathbf{a}'_n I_{n21} \mathbf{a}_n < -cn. \quad (9)$$

For  $I_{n22}$ , one can easily derive that

$$\begin{aligned} I_{n22} &= \sum_{i=1}^n \sum_{j=1}^m \sum_{s=1}^2 \mathbf{X}'_i \mathbf{W}_{ij}^{(s)}(\boldsymbol{\beta}_n) [Y_{ijs} - \pi_{ijs}(\boldsymbol{\beta}_n)] \mathbf{X}_i \\ &= \sum_{s=1}^2 \sum_{j=1}^m \sum_{i=1}^n \mathbf{X}'_i \mathbf{W}_{ij}^{(s)} [Y_{ijs} - \pi_{ijs}] \mathbf{X}_i \\ &\quad + \sum_{i=1}^n \sum_{j=1}^m \sum_{s=1}^2 \mathbf{X}'_i [\mathbf{W}_{ij}^{(s)}(\boldsymbol{\beta}_n) - \mathbf{W}_{ij}^{(s)}] [Y_{ijs} - \pi_{ijs}(\boldsymbol{\beta}_n)] \mathbf{X}_i \\ &\quad + \sum_{i=1}^n \sum_{j=1}^m \sum_{s=1}^2 \mathbf{X}'_i [\mathbf{W}_{ij}^{(s)}(\boldsymbol{\beta}_n) - \mathbf{W}_{ij}^{(s)}] [\pi_{ijs} - \pi_{ijs}(\boldsymbol{\beta}_n)] \mathbf{X}_i \\ &\quad + \sum_{i=1}^n \sum_{j=1}^m \sum_{s=1}^2 \mathbf{X}'_i \mathbf{W}_{ij}^{(s)} [\pi_{ijs} - \pi_{ijs}(\boldsymbol{\beta}_n)] \mathbf{X}_i \\ &\triangleq J_{nja} + J_{njb} + J_{njc} + J_{njd}. \end{aligned}$$

According to (A2) and (2), we have

$$\|\mathbf{H}_{ij}(\boldsymbol{\beta}_n)\| < c, \quad \|\partial \mathbf{H}_{ij}(\boldsymbol{\beta}_n)/\partial \mathbf{t}'_{ij}\| < c,$$

$$\|\mathbf{A}_{ij}(\boldsymbol{\beta}_n)\| < c, \quad \|\partial \mathbf{A}_{ij}(\boldsymbol{\beta}_n)/\partial \mathbf{t}'_{ij}\| < c.$$

Since  $\mathbf{W}_{ij}^{(s)}$  is a  $2m \times 2m$  matrix,  $s = 1, 2$ , we obtain

$$\|\mathbf{W}_{ij}^{(1)}\| < c, \quad \|\mathbf{W}_{ij}^{(2)}\| < c, \quad \forall i, j$$

which, together with (A1), give

$$E\|J_{nja}\|^2 \leq np_n^2.$$

It follows from Markov's inequality that

$$\|J_{nja}\| = O_p(\sqrt{np_n}) = o_p(n). \quad (10)$$

By (A1) and (A4), we have

$$\sup_{\|\boldsymbol{\beta}_n - \boldsymbol{\beta}_{n0}\| \leq \Delta \sqrt{p_n/n}} \|\mathbf{W}_{ij}^{(1)}(\boldsymbol{\beta}_n) - \mathbf{W}_{ij}^{(1)}\| \leq c \cdot \|\mathbf{X}_{ij}\| \cdot \|\boldsymbol{\beta}_n - \boldsymbol{\beta}_{n0}\| = o(1). \quad (11)$$

From (A3) and (11), we obtain

$$\sup_{\|\boldsymbol{\beta}_n - \boldsymbol{\beta}_{n0}\| \leq \Delta \sqrt{p_n/n}, \|\mathbf{a}_n\|=1} \mathbf{a}'_n J_{njb} \mathbf{a}_n = o(n). \quad (12)$$

By the same arguments,

$$\begin{aligned} \sup_{\|\boldsymbol{\beta}_n - \boldsymbol{\beta}_{n0}\| \leq \Delta \sqrt{p_n/n}, \|\mathbf{a}_n\|=1} \mathbf{a}'_n J_{njb} \mathbf{a}_n &= o(n), \\ \sup_{\|\boldsymbol{\beta}_n - \boldsymbol{\beta}_{n0}\| \leq \Delta \sqrt{p_n/n}, \|\mathbf{a}_n\|=1} \mathbf{a}'_n J_{njd} \mathbf{a}_n &= o(n). \end{aligned} \quad (13)$$

(10), (12) and (13) give the result that

$$\sup_{\|\boldsymbol{\beta}_n - \boldsymbol{\beta}_{n0}\| \leq \Delta \sqrt{p_n/n}, \|\mathbf{a}_n\|=1} \mathbf{a}'_n I_{n22} \mathbf{a}_n = o(n). \quad (14)$$

In fact,

$$P \left\{ \sup_{\|\mathbf{a}_n\|=1} \mathbf{a}'_n \mathbf{F}_n(\boldsymbol{\beta}_n^*) \mathbf{a}_n \leq -cn \right\} \rightarrow 1.$$

with the help of (7), (9) and (14). This implies that

$$P \left\{ \sup_{\|\boldsymbol{\beta}_n - \boldsymbol{\beta}_{n0}\| = \Delta \sqrt{p_n/n}} (\boldsymbol{\beta}_n - \boldsymbol{\beta}_{n0})' \mathbf{F}_n(\boldsymbol{\beta}_n^*) (\boldsymbol{\beta}_n - \boldsymbol{\beta}_{n0}) \leq -c\Delta^2 p_n \right\} \rightarrow 1. \quad (15)$$

From (6), (15) and the arbitrariness of  $\Delta$ , we can choose  $\Delta$  sufficiently large so that (4) holds.

*Proof of Theorem 3.1.* It follows from Chebyshev's inequality,

$$\|\mathbf{R}^* - \mathbf{R}_0\| = O_p(1/\sqrt{n}). \quad (16)$$

where

$$\mathbf{R}^* = \frac{1}{n} \sum_{i=1}^n \mathbf{A}_i^{-1/2} (\mathbf{Y}_i - \boldsymbol{\pi}_i) (\mathbf{Y}_i - \boldsymbol{\pi}_i)' \mathbf{A}_i^{-1/2}.$$

It is clear that  $E\mathbf{R}^* = \mathbf{R}_0$ . Consider the  $(k, j)$ -th entry  $R_{kj}^*$ . By the Markov inequality and the moment conditions  $\text{Var}(R_{kj}^*) = O(1/n)$ , for any  $\varepsilon > 0$ , there exists  $M = M(\varepsilon) > 0$  such that

$$P\left(\sqrt{n}|R_{kj}^* - R_{0,kj}| \geq M\right) \leq \frac{n E|R_{kj}^* - R_{0,kj}|^2}{M^2} = \frac{n \text{Var}(R_{kj}^*)}{M^2} < \varepsilon.$$

Hence  $\sqrt{n}(R_{kj}^* - R_{0,kj}) = O_p(1)$  for every  $(k, j)$ . It follows that  $\|\mathbf{R}^* - \mathbf{R}_0\| = O_p(1/\sqrt{n})$ .

We now prove  $\|\hat{\mathbf{R}} - \mathbf{R}^*\| = O_p(\sqrt{p_n/n})$ . Note that

$$\|\hat{\mathbf{R}} - \mathbf{R}^*\|^2 = \sum_{k=1}^m \sum_{j=1}^m \sum_{s=1}^2 [\hat{R}_{kjs} - R_{kjs}^*]^2, \quad (17)$$

where

$$\hat{R}_{kjs} - R_{kjs}^* = \frac{1}{n} \sum_{i=1}^n \left[ \frac{(Y_{iks} - \bar{\pi}_{iks})(Y_{ijs} - \bar{\pi}_{ijs})}{\sqrt{\tilde{A}_{iks}} \sqrt{\tilde{A}_{ijs}}} - \frac{(Y_{iks} - \pi_{iks}^0)(Y_{ijs} - \pi_{ijs}^0)}{\sqrt{A_{iks}^0} \sqrt{A_{ijs}^0}} \right],$$

with  $\pi_{iks}^0 = \pi_{iks}(\boldsymbol{\beta}_{n0})$ ,  $\bar{\pi}_{iks} = \pi_{iks}(\tilde{\boldsymbol{\beta}}_n)$ ,  $A_{iks}^0 = \pi_{iks}^0(1 - \pi_{iks}^0)$ , and  $\tilde{A}_{iks} = \bar{\pi}_{iks}(1 - \bar{\pi}_{iks})$ . We then obtain

$$\begin{aligned} |\hat{R}_{kjs} - R_{kjs}^*| &\leq \left| \frac{1}{n} \sum_{i=1}^n \frac{(Y_{iks} - \bar{\pi}_{iks})(Y_{ijs} - \bar{\pi}_{ijs}) - (Y_{iks} - \pi_{iks}^0)(Y_{ijs} - \pi_{ijs}^0)}{\sqrt{A_{iks}^0} \sqrt{A_{ijs}^0}} \right| \\ &\quad + \left| \frac{1}{n} \sum_{i=1}^n \frac{(Y_{iks} - \bar{\pi}_{iks})(Y_{ijs} - \bar{\pi}_{ijs})}{\sqrt{A_{iks}^0} \sqrt{A_{ijs}^0}} \delta_{ijks} \right| \\ &\triangleq I_{kjs,1} + I_{kjs,2}, \end{aligned} \quad (18)$$

where  $\delta_{ijks} = [A_{iks}^0 A_{ijs}^0]^{1/2} [\tilde{A}_{iks} \tilde{A}_{ijs}]^{-1/2} - 1$ . From (17) and (18), we can get

$$\|\hat{\mathbf{R}} - \mathbf{R}^*\|^2 \leq 2 \sum_{k=1}^m \sum_{j=1}^m \sum_{s=1}^2 I_{kjs,1}^2 + 2 \sum_{k=1}^m \sum_{j=1}^m \sum_{s=1}^2 I_{kjs,2}^2 \triangleq I_{n1} + I_{n2}. \quad (19)$$

By triangle inequality, it is not difficult to obtain

$$\begin{aligned}
 I_{kjs,1} &\leq \frac{1}{n} \sum_{i=1}^n \frac{|(\pi_{iks}^0 - \bar{\pi}_{iks})(\pi_{ijs}^0 - \bar{\pi}_{ijs})|}{\sqrt{A_{iks}^0} \sqrt{A_{ijs}^0}} \\
 &\quad + \frac{1}{n} \sum_{i=1}^n \frac{|(\pi_{iks}^0 - \bar{\pi}_{iks})(Y_{ijs} - \pi_{ijs}^0)|}{\sqrt{A_{iks}^0} \sqrt{A_{ijs}^0}} \\
 &\quad + \frac{1}{n} \sum_{i=1}^n \frac{|(\pi_{ijs}^0 - \bar{\pi}_{ijs})(Y_{iks} - \pi_{iks}^0)|}{\sqrt{A_{iks}^0} \sqrt{A_{ijs}^0}} \\
 &\triangleq I_{kjs,11} + I_{kjs,12} + I_{kjs,13}.
 \end{aligned} \tag{20}$$

Using (19) and (20), we can derive

$$\begin{aligned}
 I_{n1} &\leq 6 \sum_{k=1}^m \sum_{j=1}^m \sum_{s=1}^2 I_{kjs,11}^2 + 6 \sum_{k=1}^m \sum_{j=1}^m \sum_{s=1}^2 I_{kjs,12}^2 + 6 \sum_{k=1}^m \sum_{j=1}^m \sum_{s=1}^2 I_{kjs,13}^2 \\
 &\triangleq I_{n11} + I_{n12} + I_{n13}.
 \end{aligned} \tag{21}$$

By Cauchy-Schwarz inequality, we have

$$I_{kjs,11}^2 \leq \left[ \frac{1}{n} \sum_{i=1}^n \frac{(\pi_{iks}^0 - \bar{\pi}_{iks})^2}{\pi_{iks}^0 (1 - \pi_{iks}^0)} \right] \cdot \left[ \frac{1}{n} \sum_{i=1}^n \frac{(\pi_{ijs}^0 - \bar{\pi}_{ijs})^2}{\pi_{ijs}^0 (1 - \pi_{ijs}^0)} \right].$$

Hence,

$$I_{n11} \leq 6 \left[ \frac{1}{n} \sum_{i=1}^n \sum_{k=1}^m \sum_{s=1}^2 \frac{(\pi_{iks}^0 - \bar{\pi}_{iks})^2}{\pi_{iks}^0 (1 - \pi_{iks}^0)} \right] \cdot \left[ \frac{1}{n} \sum_{i=1}^n \sum_{j=1}^m \sum_{s=1}^2 \frac{(\pi_{ijs}^0 - \bar{\pi}_{ijs})^2}{\pi_{ijs}^0 (1 - \pi_{ijs}^0)} \right]. \tag{22}$$

Let  $\lambda_s$  be the unit vector with 1 in the  $s$ -th entry and 0 elsewhere. We have

$$\frac{\partial \pi_{iks}}{\partial \beta'_n} = \lambda'_s \mathbf{H}_{ik}(\beta_n) \mathbf{X}'_{ik}, \quad s = 1, 2. \tag{23}$$

For some  $\bar{\beta}_n$  between  $\tilde{\beta}_n$  and  $\beta_{n0}$ , It follows from (23) and (A3) that

$$\begin{aligned}
 &\frac{1}{n} \sum_{i=1}^n \sum_{k=1}^m \sum_{s=1}^2 \frac{(\pi_{iks}^0 - \bar{\pi}_{iks})^2}{\pi_{iks}^0 (1 - \pi_{iks}^0)} \\
 &= \frac{1}{n} \sum_{i=1}^n \sum_{k=1}^m \sum_{s=1}^2 \frac{(\tilde{\beta}_n - \beta_{n0})' \mathbf{X}_{ik} \mathbf{H}'_{ik}(\bar{\beta}_n) \lambda_s \lambda'_s \mathbf{H}_{ik}(\bar{\beta}_n) \mathbf{X}'_{ik} (\tilde{\beta}_n - \beta_{n0})}{\pi_{ik}^0 (1 - \pi_{ik}^0)} \\
 &\leq \frac{1}{n} c \cdot (\tilde{\beta}_n - \beta_{n0})' \left( \sum_{i=1}^n \mathbf{X}'_i \mathbf{X}_i \right) (\tilde{\beta}_n - \beta_{n0}) \\
 &\leq \frac{1}{n} c \cdot \|\tilde{\beta}_n - \beta_{n0}\|^2 \cdot \lambda_{\max} \left( \sum_{i=1}^n \mathbf{X}'_i \mathbf{X}_i \right) \\
 &= O(\|\tilde{\beta}_n - \beta_{n0}\|^2).
 \end{aligned} \tag{24}$$

From (22), (24) and Lemma 4.2, we get  $I_{n11} = O_p(p_n^2/n^2)$ . Similarly,

$$I_{n12} = O_p(p_n/n), \quad I_{n13} = O_p(p_n/n).$$

Hence we have

$$I_{n1} = O_p(p_n/n). \quad (25)$$

To evaluate  $\delta_{ijks}$ , define  $g(\beta_n) = [A_{iks}^0 A_{ijs}^0]^{1/2} [A_{iks}(\beta_n) A_{ijs}(\beta_n)]^{-1/2}$ , then  $\delta_{ijks} = g(\tilde{\beta}_n) - g(\beta_{n0})$ . By mean value theorem,

$$\|\delta_{ijks}\| = \|g(\tilde{\beta}_n) - g(\beta_{n0})\| \leq c_1 \cdot \|X'_{ik}(\tilde{\beta}_n - \beta_{n0})\| + c_2 \cdot \|X'_{ij}(\tilde{\beta}_n - \beta_{n0})\|. \quad (26)$$

Using Cauchy-Schwarz inequality, (26), (A3) and Lemma 4.2, we obtain

$$\begin{aligned} I_{n2} &\leq 2 \sum_{k=1}^m \sum_{j=1}^m \sum_{s=1}^2 \left[ \frac{1}{n} \sum_{i=1}^n \frac{(Y_{iks} - \bar{\pi}_{iks})^2 (Y_{ijs} - \bar{\pi}_{ijs})^2}{\pi_{iks}^0 (1 - \pi_{iks}^0) \pi_{ijs}^0 (1 - \pi_{ijs}^0)} \right] \left[ \frac{1}{n} \sum_{i=1}^n \delta_{ijks}^2 \right] \\ &\leq 2 \cdot \frac{c}{n} \sum_{i=1}^n \sum_{k=1}^m \sum_{j=1}^m \sum_{s=1}^2 \delta_{ijks}^2 \\ &\leq 4 \cdot \frac{c}{n} \sum_{i=1}^n \sum_{k=1}^m \sum_{j=1}^m \sum_{s=1}^2 (\tilde{\beta}_n - \beta_{n0})' \left[ c_1^2 X_{ik} X'_{ik} + c_2^2 X_{ij} X'_{ij} \right] (\tilde{\beta}_n - \beta_{n0}) \\ &\leq 4c \cdot \|\tilde{\beta}_n - \beta_{n0}\|^2 \cdot \left[ \lambda_{\max}(n^{-1} \sum_{i=1}^n X_i' X_i) + \lambda_{\max}(n^{-1} \sum_{i=1}^n X_i' X_i) \right] \\ &= O_p(p_n/n). \end{aligned} \quad (27)$$

From (19), (25) and (27), we conclude

$$\|\hat{\mathbf{R}} - \mathbf{R}^*\| = O_p(\sqrt{p_n/n}). \quad (28)$$

Combining (16) and (28), we obtain

$$\|\hat{\mathbf{R}} - \mathbf{R}_0\| \leq \|\hat{\mathbf{R}} - \mathbf{R}^*\| + \|\mathbf{R}^* - \mathbf{R}_0\| = O_p(\sqrt{p_n/n}). \quad (29)$$

Finally, since

$$\|\hat{\mathbf{R}}^{-1} - \mathbf{R}_0^{-1}\| = \|\hat{\mathbf{R}}^{-1}(\hat{\mathbf{R}} - \mathbf{R}_0)\mathbf{R}_0^{-1}\|,$$

Theorem 3.1 follows from (29).

## 5. Monte Carlo simulations

In this section, we conduct Monte Carlo simulations to verify the theoretical result established in Theorem 1. The data are generated from a two-step logit model for multinomial responses with three categories, as described in Section 2. For each subject  $i = 1, \dots, n$ , we generate  $m = 4$  repeated

measurements. The covariate vector  $\tilde{X}_{ij} \in \mathbb{R}^{\bar{p}_n}$  is drawn from a multivariate normal distribution with mean  $\mathbf{0}$  and covariance matrix  $\Sigma$ , where  $\Sigma_{kl} = 0.5^{|k-l|}$ . The true parameter vector is set as

$$\beta_{n0} = (0.5, 0.2 \cdot \mathbf{1}'_{\bar{p}_n}, -0.5, -0.2 \cdot \mathbf{1}'_{\bar{p}_n})',$$

where  $\mathbf{1}_{\bar{p}_n}$  denotes a  $\bar{p}_n$ -dimensional vector of ones. The true correlation matrix  $\mathbf{R}_0$  is specified as an exchangeable structure with correlation coefficient  $\rho$ , and we consider three levels of within-subject correlation:  $\rho = 0.3, 0.5, 0.8$ . To evaluate the asymptotic behavior under increasing dimensionality, we consider three sample size-dimension configurations: (i)  $n = 200$ ,  $\bar{p}_n = 5$ , hence  $p_n = 2(1 + \bar{p}_n) = 12$ ; (ii)  $n = 400$ ,  $\bar{p}_n = 10$ , hence  $p_n = 22$ ; (iii)  $n = 800$ ,  $\bar{p}_n = 20$ , hence  $p_n = 42$ . These settings satisfy the conditions as required by Section 2.3.

To obtain the initial estimator  $\tilde{\beta}_n$ , we iterate a modified Fisher scoring algorithm for the estimating equations in (3), following the procedure in Liang and Zeger [1]. The estimated correlation matrix is then constructed as

$$\hat{\mathbf{R}} = \frac{1}{n} \sum_{i=1}^n A_i^{-1/2}(\tilde{\beta}_n)(Y_i - \pi_i(\tilde{\beta}_n))(Y_i - \pi_i(\tilde{\beta}_n))' A_i^{-1/2}(\tilde{\beta}_n).$$

We measure the estimation error by the Frobenius norm of the inverse-matrix:

$$\text{Error} = \|\hat{\mathbf{R}}^{-1} - \mathbf{R}_0^{-1}\|.$$

All results are based on 500 independent replications.

The average normalized errors and their standard deviations over 500 replications are reported in Tables 1–3 for  $\rho = 0.3, 0.5, 0.8$ , respectively. According to these tables, we can see that when the covariate dimension  $p_n$  grows at a suitable rate as the sample size  $n$ , the accuracy of the correlation matrix estimator is satisfactory for all three correlation levels. To verify that the theoretical result is not specific to the exchangeable correlation structure, we repeated the simulations under an AR-1 true correlation matrix  $(R_0)_{jk} = \rho^{|j-k|}$  with  $\rho = 0.5$ . The results are reported in Appendix A.

**Table 1.** Average error ( $\|\hat{\mathbf{R}}^{-1} - \mathbf{R}_0^{-1}\|$ ) for  $\rho = 0.3$  over 500 replications.

| $n$ | $\bar{p}_n$ | $p_n$ | Error  | SD     |
|-----|-------------|-------|--------|--------|
| 200 | 5           | 12    | 0.0104 | 0.0029 |
| 400 | 10          | 22    | 0.0053 | 0.0019 |
| 800 | 20          | 42    | 0.0029 | 0.0008 |

**Table 2.** Average error ( $\|\hat{\mathbf{R}}^{-1} - \mathbf{R}_0^{-1}\|$ ) for  $\rho = 0.5$  over 500 replications.

| $n$ | $\bar{p}_n$ | $p_n$ | Error  | SD     |
|-----|-------------|-------|--------|--------|
| 200 | 5           | 12    | 0.0109 | 0.0030 |
| 400 | 10          | 22    | 0.0056 | 0.0020 |
| 800 | 20          | 42    | 0.0030 | 0.0009 |

**Table 3.** Average error ( $\|\hat{\mathbf{R}}^{-1} - \mathbf{R}_0^{-1}\|$ ) for  $\rho = 0.8$  over 500 replications.

| $n$ | $\bar{p}_n$ | $p_n$ | Error  | SD     |
|-----|-------------|-------|--------|--------|
| 200 | 5           | 12    | 0.0117 | 0.0032 |
| 400 | 10          | 22    | 0.0060 | 0.0022 |
| 800 | 20          | 42    | 0.0032 | 0.0009 |

## 6. Conclusions

This paper investigates the true correlation matrix estimator in generalized estimating equations for high-dimensional longitudinal multinomial data. Under mild regularity conditions, we establish its asymptotic consistency with a convergence rate of  $O_p(\sqrt{p_n/n})$ , extending the classical result from binary responses to the multinomial setting. The proof employs a block-matrix decomposition technique to accommodate the expanded response dimension, and the conclusion naturally generalizes to ordinal outcomes with an arbitrary number of categories. Finally, Monte Carlo simulations under both exchangeable and AR-1 correlation structures verify our theoretical findings. Future research may extend to scenarios with missing observations, penalized variable selection, and more general categorical response models.

## Appendix A. Simulation results under the AR-1 correlation structure

**Table A.1.** Average error ( $\|\hat{\mathbf{R}}^{-1} - \mathbf{R}_0^{-1}\|$ ) with  $\rho = 0.3$  over 500 replications.

| $n$ | $\bar{p}_n$ | $p_n$ | Error  | SD     |
|-----|-------------|-------|--------|--------|
| 200 | 5           | 12    | 0.0104 | 0.0029 |
| 400 | 10          | 22    | 0.0053 | 0.0019 |
| 800 | 20          | 42    | 0.0029 | 0.0009 |

**Table A.2.** Average error ( $\|\hat{\mathbf{R}}^{-1} - \mathbf{R}_0^{-1}\|$ ) with  $\rho = 0.5$  over 500 replications.

| $n$ | $\bar{p}_n$ | $p_n$ | Error  | SD     |
|-----|-------------|-------|--------|--------|
| 200 | 5           | 12    | 0.0109 | 0.0030 |
| 400 | 10          | 22    | 0.0056 | 0.0020 |
| 800 | 20          | 42    | 0.0030 | 0.0009 |

**Table A.3.** Average error ( $\|\hat{\mathbf{R}}^{-1} - \mathbf{R}_0^{-1}\|$ ) with  $\rho = 0.8$  over 500 replications.

| $n$ | $\bar{p}_n$ | $p_n$ | Error  | SD     |
|-----|-------------|-------|--------|--------|
| 200 | 5           | 12    | 0.0117 | 0.0033 |
| 400 | 10          | 22    | 0.0060 | 0.0022 |
| 800 | 20          | 42    | 0.0032 | 0.0010 |

Note: All settings are identical to those in Section 5 except that the true correlation matrix  $\mathbf{R}_0$  follows the AR-1 structure with  $(\mathbf{R}_0)_{jk} = \rho^{|j-k|}$ ,  $j, k = 1, \dots, m = 4$ .

## Author contributions

Xianbin Chen: Conceptualization, methodology, formal analysis, writing – original draft; Donghai Peng: Writing – review & editing, validation. All authors have read and approved the final version of the manuscript for publication.

## Use of Generative-AI tools declaration

The authors declare they have not used Artificial Intelligence (AI) tools in the creation of this article.

## Acknowledgments

This work is partly supported by the University-Level High-Level Talent Program (KYG2401).

## Conflict of interest

The authors declare no conflict of interest.

## References

1. K. Y. Liang, S. L. Zeger, Longitudinal data analysis using generalized linear models, *Biometrika*, **73** (1986), 13–22. <http://dx.doi.org/10.1093/biomet/73.1.13>
2. M. G. Xie, Y. N. Yang, Asymptotics for generalized estimating equations with large cluster sizes, *Ann. Stat.*, **31** (2003), 310–347. <http://dx.doi.org/10.1214/aos/1043351255>
3. R. M. Balan, I. Schiopu-Kratina, Asymptotic results with generalized estimating equations for longitudinal data, *Ann. Stat.*, **33** (2005), 522–541. <http://dx.doi.org/10.1214/009053604000001255>
4. L. Fahrmeir, H. Kaufmann, Consistency and asymptotic normality of the maximum likelihood estimator in generalized linear models, *Ann. Stat.*, **13** (1985), 342–368. <http://dx.doi.org/10.1214/aos/1176346597>
5. S. L. Zeger, K. Y. Liang, Longitudinal data analysis for discrete and continuous outcomes, *Biometrics*, **42** (1986), 121–130. <http://dx.doi.org/10.2307/2531248>
6. M. Gosho, R. Ishii, H. Noma, K. Maruo, A comparison of bias-adjusted generalized estimating equations for sparse binary data in small-sample longitudinal studies, *Stat. Med.*, **42** (2023), 2711–2727. <http://dx.doi.org/10.1002/sim.9744>
7. G. Y. Yi, L. Zeng, R. J. Cook, A robust pairwise likelihood method for incomplete longitudinal binary data arising in clusters, *Can. J. Stat.*, **39** (2011), 34–51. <http://dx.doi.org/10.1002/cjs.10089>
8. R. L. Prentice, Correlated binary regression with covariates specific to each binary observation, *Biometrics*, **44** (1988), 1033–1048. <http://dx.doi.org/10.2307/2531726>
9. S. R. Lipsitz, N. M. Laird, D. P. Harrington, Generalized estimating equations for correlated binary data: Using the odds ratio as a measure of association, *Biometrika*, **78** (1991), 153–160. <http://dx.doi.org/10.1093/biomet/78.1.153>

10. V. Carey, S. L. Zeger, P. Diggle, Modelling multivariate binary data with alternating logistic regressions, *Biometrika*, **80** (1993), 517–526. <http://dx.doi.org/10.1093/biomet/80.3.517>
11. A. Qu, B. G. Lindsay, B. Li, Improving generalized estimating equations using quadratic inference functions, *Biometrika*, **87** (2000), 823–836. <http://dx.doi.org/10.1093/biomet/87.4.823>
12. W. Pan, Akaike's information criterion in generalized estimating equations, *Biometrics*, **57** (2001), 120–125. <http://dx.doi.org/10.1111/j.0006-341X.2001.00120.x>
13. N. Lange, R. J. A. Little, J. M. G. Taylor, Robust statistical modeling using repeated measures data, *JASA*, **84** (1989), 97–101. <http://dx.doi.org/10.1080/01621459.1989.10478743>
14. J. Shults, S. J. Ratcliffe, The mean square error of estimates of the regression parameter in GEE under misspecified working correlation, *Biometrical J.*, **46** (2004), 406–417. <http://dx.doi.org/10.1002/bimj.200310033>
15. X. Fu, L. Hao, Y. Wang, Alternative parameterizations of the working correlation matrix in GEE, *Stat. Comput.*, **32** (2022), 73. <http://dx.doi.org/10.1007/s11222-022-10136-3>
16. L. Wang, GEE analysis of clustered binary data with diverging number of covariates, *Ann. Stat.*, **39** (2011), 389–417. <http://dx.doi.org/10.1214/10-AOS838>
17. X. B. Chen, J. L. Yin, Asymptotics of the general GEE estimator for high-dimensional longitudinal data, *Commun. Stat.-Theor. M.*, **53** (2024), 5041–5056. <http://dx.doi.org/10.1080/03610926.2023.2205045>
18. L. Wang, J. H. Zhou, A. N. Qu, Penalized generalized estimating equations for high-dimensional longitudinal data analysis, *Biometrics*, **68** (2012), 353–360. <http://dx.doi.org/10.1111/j.1541-0420.2011.01678.x>
19. H. Green, J. S. Barry, S. Wang, Penalized GEE for high-dimensional longitudinal data with divergent cluster sizes, *Biometrics*, **76** (2020), 1226–1238. <http://dx.doi.org/10.1111/biom.13247>
20. M. Kamruzzaman, B. C. Sutradhar, A. I. McLeod, Penalized generalized estimating equations for high-dimensional longitudinal data, *J. Multivariate Anal.*, **184** (2021), 104746. <http://dx.doi.org/10.1016/j.jmva.2021.104746>
21. W. Zhou, J. Liu, W. Sun, Variable selection for high-dimensional longitudinal data via penalized GEE, *Stat. Sin.*, **20** (2010), 1719–1735. <http://dx.doi.org/10.5705/ss.2008.091>
22. L. Xia, A. Shojaie, Statistical inference for high-dimensional generalized estimating equations, *Biostatistics*, **27** (2026), kxag013. <http://dx.doi.org/10.1093/biostatistics/kxag013>
23. E. X. Fang, Y. Ning, R. Li, Test of significance for high-dimensional longitudinal data, *Ann. Stat.*, **48** (2020), 2622–2645. <http://dx.doi.org/10.1214/19-AOS1900>
24. I. Dumitrescu, Y. Liu, L. Wang, Statistical inference for high-dimensional GEE, *Ann. Stat.*, **49** (2021), 1893–1917. <http://dx.doi.org/10.1214/20-AOS1997>
25. T. Lu, L. Xue, J. Hu, Robust estimation of the correlation matrix in high-dimensional GEE with missing observations, *J. Am. Stat. Assoc.*, **115** (2020), 1406–1419. <http://dx.doi.org/10.1080/01621459.2019.1660170>
26. T. Lu, L. Xue, J. Cai, Robust estimation of the correlation matrix in high-dimensional GEE, *Stat. Sin.*, **31** (2021), 1287–1310. <http://dx.doi.org/10.5705/ss.202019.0325>

27. R. Tang, Y. Wang, Robust estimation of within-subject covariance matrix for high-dimensional longitudinal data, *Biometrics*, **78** (2022), 1023–1035. <http://dx.doi.org/10.1111/biom.13492>
28. S. R. Lipsitz, K. Kim, L. Zhao, Analysis of repeated categorical data using generalized estimating equations, *Stat. Med.*, **13** (1994), 1149–1163. <http://dx.doi.org/10.1002/sim.4780131107>
29. A. Touloumis, A. Agresti, M. Kateri, GEE for longitudinal ordinal data: Comparing R-geepack, R-multgee, R-repolr, SAS-GENMOD, SPSS-GENLIN, *Comput. Stat. Data Anal.*, **67** (2013), 132–144. <http://dx.doi.org/10.1016/j.csda.2013.06.003>
30. A. Agresti, *Categorical data analysis*, 2 Eds., Hoboken: Wiley, 2002.
31. M. E. Miller, C. S. Davis, J. R. Landis, The analysis of longitudinal polytomous outcomes with application to neurologic safety studies, *Stat. Med.*, **12** (1993), 1873–1880. <http://dx.doi.org/10.1002/sim.4780121905>
32. P. J. Heagerty, S. L. Zeger, Marginal regression models for clustered ordinal measurements with application to longitudinal CD4 cell counts, *Biometrics*, **52** (1996), 1030–1044. <http://dx.doi.org/10.2307/2533045>
33. B. Morawitz, G. Tutz, Alternative parameterizations in business tendency surveys, *Z. Oper. Res.*, **34** (1990), 143–156. <http://dx.doi.org/10.1007/BF01415850>
34. L. Fahrmeir, G. Tutz, *Multivariate statistical modelling based on generalized linear models*, New York: Springer, 1994, 63–95.
35. J. Ortega, W. Rheinboldt, *Iterative solution of nonlinear equations in several variables*, New York: Elsevier Inc, Academic Press, 1970.



AIMS Press

©2026 the Author(s), licensee AIMS Press. This is an open access article distributed under the terms of the Creative Commons Attribution License (<https://creativecommons.org/licenses/by/4.0>)