Analysis of longitudinal data with missing responses adjusted by inverse probability weights

Dina Jankovic

Thesis submitted to the Faculty of Graduate and Postdoctoral Studies in partial fulfillment of the requirements for the degree of Master of Science in Mathematics[1]

Department of Mathematics and Statistics
Faculty of Science
University of Ottawa

[1]The M.Sc. program is a joint program with Carleton University, administered by the Ottawa-Carleton Institute of Mathematics and Statistics

Abstract

We propose a new method for analyzing longitudinal data which contain responses that are missing at random. This method consists in solving the generalized estimating equation (GEE) of [7] in which the incomplete responses are replaced by values adjusted using the inverse probability weights proposed in [14]. We show that the root estimator is consistent and asymptotically normal, essentially under some conditions on the marginal distribution and the surrogate correlation matrix as those presented in [12] in the case of complete data, and under minimal assumptions on the missingness probabilities. This method is applied to a real-life dataset taken from [10], which examines the incidence of respiratory disease in a sample of 250 pre-school age Indonesian children which were examined every 3 months for 18 months, using as covariates the age, gender, and vitamin A deficiency.

Dedications

I dedicate this thesis to my family and friends, whose unselfish love and tremendous support over many years laid the foundations for the discipline and efforts to complete this work successfully.

Acknowledgement

I would like to thank my supervisor Raluca Balan for her guidance and patience during my graduate studies. She is an important part of my success. I would also like to thank the professors, graduate students, and other employees of the Department of Mathematics and Statistics for their immense support, including Diane Demers, Mayada El-Maalouf, Mateja Sajna, Benoit Dionne, Julie Hudson, and Gustavo Valente. Finally, I would like to thank my family and friends in Europe and North America for their encouragement throughout my educational career.

Contents

Chapter 1

Introduction

Longitudinal data are presented when following particular individuals over prolonged periods of time, often years or even decades. A dataset is longitudinal if it tracks the same type of information on the same subjects at multiple time points. For instance, a longitudinal dataset can contain information about specific students, their test results and other achievements in ten successive years. The primary advantage of longitudinal data over cross-sectional data is that they can measure change. However, the complexity of their analysis is a big challenge for statisticians. More details about longitudinal data are provided in [6].

In statistical analysis, it is important to analyze the relationship between a response variable (for instance, the presence of diabetes) and multiple continuous or discrete predictor variables that may be more or less influential (for instance, weight, height, race, blood pressure). A commonly used method introduced by Liang and Zeger in [7] is to assume that the marginal distribution of each response follows a Generalized Linear Model (GLM) with regression parameter β, while the correlation between the responses is modeled by a surrogate correlation matrix which depends on another parameter α. More details on GLMs can be found in [8]. The goal of this method is to solve the Generalized Estimating Equation (GEE) and obtain a consistent estimator β.

Building upon earlier work of [4] and [16] for estimating equations for classical datasets, the article [12] contains a thorough analysis of the asymptotic properties of the GEE estimator, including the case when the number of observations made on each individual (called the *cluster size*) goes to infinity. Similar theoretical investigations were pursed in [2] for fixed cluster size, for an estimator defined as the root of a pseudo-likelihood equation, which contains an estimator of the correlation matrix based on the data.

The complexity of longitudinal data analysis increases in the presence of incomplete observations. Several methods for dealing with longitudinal data which contain missing responses (or missing covariates, or both) have been proposed by various authors. We refer the reader to [5], [11], [13], [14], [15] for a sample of relevant references. Our goal is to adapt the GEE method of [7], [12] to the case when the responses are *missing at random* (a term whose meaning will be explained in Chapter 2 below). For this, we will replace the incomplete responses by values adjusted using the inverse probability weights proposed in [14]. Under minimal assumptions on the missingness probabilities, we will show that the root estimator of β is consistent and asymptotically normal, under the same conditions on the marginal distribution and the surrogate correlation matrix as in [12].

In Chapter 2, we provide a new method for estimating the regression parameter β in the marginal GLM proposed by Liang and Zeger in [7] for longitudinal data with missing responses. In Section 2.1, we introduce our framework with several examples, as well as the GEE with an explanation of the MAR (Missing at Random Assumption) mechanism, together with a discussion of its properties. In Section 2.2, we show that the GEE has a solution under certain conditions, and the solution is a consistent estimator of β. The proofs are very similar to those in [12]. In Section 2.3, we show that the estimator introduced in the previous section is asymptotically normal, under certain conditions. We introduce the condition (CC), which involves the derivative of the GEE. We refer to [3] for the use of the Central Limit Theorem for triangular arrays. In Section 2.4, we provide sufficient conditions which ensure that Condition (CC) holds. We refer the reader to [12] and Appendix A for the used results.

In Chapter 3, we give an application of our method on a real-life example, with a dataset taken from the Indonesian Children's Health Study (see [10]). We consider a marginal logistic regression model and we introduce the predictor variables. The missing values are created artificially, since the original data does not contain any missing values. In Section 3.1, we apply our method to the dataset and obtain two estimates (with their precisions) for β, since we consider two cases of the conditional correlation matrix when solving the GEE. In Section 3.2, we apply to the same dataset the Generalized Method of Moments, and we come up with very small p-values for all coefficients.

Appendix A contains some auxiliary results used in Chapter 2. Appendix A.1 contains some elementary properties of matrices, while Appendix A.2 contains several asymptotic results, which are used in the proofs of several lemmas in Section 2.4. Appendix B includes the R code used to obtain the results provided in Chapter 3. Comments about certain parts of the code are also provided.

The results have been summarized and submitted for publication in the companion article [1].

Chapter 2

Theoretical Results

In this chapter, we develop a new method for estimating the regression parameter in the marginal generalized linear model proposed by Liang and Zeger in [7] for longitudinal data, in the presence of missing responses. Using this method, we identify some sufficient conditions for the existence, consistency and asymptotic normality of an estimator $\hat{\beta}_n$ of β.

We denote by β_0 the true value of β. We make the usual convention of omitting to write β_0 when it appears in the argument of a function; for instance, instead of $g_n(\beta_0)$ we write simply g_n.

2.1 The Estimating Equation

We consider n individuals whose measurements are recorded on m occasions. For each $i = 1, \ldots, n$ and $j = 1, \ldots, m$, we denote by Y_{ij} the response of individual i at time j. Some of these responses may be missing. We let

$$I_{ij} = \begin{cases} 1, & \text{if } Y_{ij} \text{ is observed} \\ 0, & \text{if } Y_{ij} \text{ is missing} \end{cases}$$

Moreover, for each $i = 1, \ldots, n$, and $j = 1, \ldots, m$, we let $\mathbf{X}_{ij} = (X_{ij}^{(1)}, \ldots, X_{ij}^{(p)})^T$ be the p-dimensional vector of covariates for individual i at time j. We assume that X_{ij} is *random*. We consider the matrix

$$\mathbf{X}_i = \begin{bmatrix} X_{i1}^T \\ \vdots \\ X_{im}^T \end{bmatrix} = \begin{bmatrix} X_{i1}^{(1)} & \ldots & X_{i1}^{(p)} \\ \vdots & \ddots & \vdots \\ X_{im}^{(1)} & \ldots & X_{im}^{(p)} \end{bmatrix}$$

Note that $\mathbf{X}_i$ is an $m \times p$ matrix which contains all the covariates of the i^{th} individual.

We let $\mathbf{Y}_i = (Y_{i1}, \ldots, Y_{im})^T$ be the vector of responses of the i^{th} individual, $\mathbf{I}_i = (I_{i1}, \ldots, I_{im})^T$ be the vector of missingness indicators for this individual.

We assume that $\{(\mathbf{Y}_i, \mathbf{X}_i, \mathbf{I}_i)\}_{i \geq 1}$ are independent and identically distributed, and there exists a one-to-one differentiable function μ on $\mathbb{R}$ such that

$$\mu_{ij}(\beta) := E(Y_{ij}|\mathbf{X}_i) = \mu(\mathbf{X}_{ij}^T\beta),$$
$$\sigma_{ij}^2(\beta) := \mathrm{Var}(Y_{ij}|\mathbf{X}_i) = \mu'(\mathbf{X}_{ij}^T\beta)$$

for a p-dimensional parameter β. The inverse g of the function μ is called the *link function.* We denote $\mu_i(\beta) = (\mu_{i1}(\beta), \ldots, \mu_{im}(\beta))^T$ and

$$\mathbf{D}_i(\beta) = \frac{\partial \mu_i(\beta)}{\partial \beta^T} = \begin{bmatrix} \frac{\partial \mu_{i1}}{\partial \beta^T}(\beta) \\ \vdots \\ \frac{\partial \mu_{im}}{\partial \beta^T}(\beta) \end{bmatrix}$$

Note that $\frac{\partial \mu_{ij}}{\partial \beta^T}(\beta) = \mathbf{X}_{ij}^T \mu'(\mathbf{X}_{ij}^T\beta)$ and hence $\mathbf{D}_i(\beta) = \mathbf{A}_i(\beta)\mathbf{X}_i$, where $\mathbf{A}_i(\beta)$ is the diagonal matrix with entries $\mu'(\mathbf{X}_{i1}^T\beta), \ldots, \mu'(\mathbf{X}_{im}^T\beta)$:

$$\mathbf{A}_i(\beta) = \begin{bmatrix} \mu'(\mathbf{X}_{i1}^T\beta) & \ldots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \ldots & \mu'(\mathbf{X}_{im}^T\beta) \end{bmatrix}$$

Example 2.1.1. (Normal Linear Regression for Quantitative Responses) When the responses Y_{ij} represent quantitative measurements, we may assume that Y_{ij} has an normal distribution with mean $\mu_{ij}(X_{ij}^T\beta)$ and known variance $\sigma_{ij}^2(\beta) = \phi$, for a nuisance parameter ϕ which is estimated separately. We assume for simplicity that $\phi = 1$. In this case, $\mu(x) = x$ and $\mu'(x) = 1$. The link function is $g(x) = x$.

Example 2.1.2. (Log-linear Regression for Count-type Responses) When the responses Y_{ij} represent count-type measurements, we may assume that Y_{ij} has a Poisson distribution with mean $\mu_{ij}(\beta) = \exp(X_{ij}^T\beta)$. In this case, $\mu(x) = e^x$ and $\mu'(x) = e^x$. The link function is $g(x) = \log x$ for $x > 0$.

Example 2.1.3. (Logistic Regression for Binary Responses) When the responses Y_{ij} represent binary measurements, we may assume that Y_{ij} has a Bernoulli distribution with mean $\mu_{ij}(\beta) = \frac{\exp(X_{ij}^T\beta)}{1+\exp(X_{ij}^T\beta)}$. In this case, $\mu(x) = \frac{e^x}{1+e^x}$ and $\mu'(x) = \frac{e^x}{(1+e^x)^2}$. The link function is $g(x) = \log \frac{x}{1-x} =: \mathrm{logit}(x)$ for $x \in (0, 1)$.

Let $\Sigma_i(\beta)$ be the conditional covariance matrix of $\mathbf{Y}_i$ given $\mathbf{X}_i$, i.e.

$$\Sigma_i(\beta) = \mathrm{Cov}(\mathbf{Y}_i|\mathbf{X}_i) = \mathrm{E}[(\mathbf{Y}_i - \mu_i(\beta))(\mathbf{Y}_i - \mu_i(\beta))^T|\mathbf{X}_i].$$

We denote by $\sigma_{i,jk}(\beta)$ the element of $\Sigma_i(\beta)$ situated on row j and column k, for $1 \leq j, k \leq m$. Then

$$\sigma_{i,jk}(\beta) = \mathrm{E}[\varepsilon_{ij}(\beta)\varepsilon_{ik}(\beta)|\mathbf{X}_i]$$

where $\varepsilon_{ij}(\beta) = Y_{ij} - \mu_{ij}(\beta)$, for all $j = 1, \ldots, m$. In particular, $\sigma_{i,jj}(\beta) = \sigma_{ij}^2(\beta) = \mu'(X_{ij}^T\beta)$. Note that

$$\Sigma_i(\beta) = \mathbf{A}_i(\beta)^{1/2}\overline{\mathbf{R}}_i\mathbf{A}_i(\beta)^{1/2}$$

where $\overline{\mathbf{R}}_i$ is the conditional correlation matrix of $\mathbf{Y}_i$ given $\mathbf{X}_i$, i.e. $\overline{\mathbf{R}}_i = (\overline{r}_{i,jk})_{1\leq j,k\leq m}$, and

$$\overline{r}_{i,jk} = \mathrm{Corr}(Y_{ij}, Y_{ik}|\mathbf{X}_i) = \frac{\sigma_{i,jk}(\beta)}{\sigma_{ij}(\beta)\sigma_{ik}(\beta)}.$$

In this framework, it is assumed that $\overline{\mathbf{R}}_i$ does not depend on β.

The method proposed by Liang and Zeger in [7] consists in replacing the *unknown* correlation matrix $\overline{\mathbf{R}}_i$ by a known correlation matrix $\mathbf{R}_i(\alpha)$ (possibly depending on another parameter α which is estimated separately), and solving the Generalized Estimated Equation (GEE):

$$\sum_{i=1}^{n} \mathbf{D}_i^T(\beta)\mathbf{V}_i^{-1}(\beta, \alpha)(\mathbf{Y}_i - \mu_i(\beta)) = 0, \tag{2.1.1}$$

where $\mathbf{V}_i(\beta) = \mathbf{A}_i(\beta)^{1/2}\mathbf{R}_i(\alpha)\mathbf{A}_i(\beta)^{1/2}$ is an approximation of the unknown covariance matrix $\Sigma_i(\beta)$. The case $\mathbf{R}_i(\alpha) = \mathbf{I}$ for all $i = 1, \ldots, n$ is called *working independence.* In this case, equation (2.1.1) becomes

$$\sum_{i=1}^{n}\sum_{j=1}^{m} \mathbf{X}_{ij}(Y_{ij} - \mu(\mathbf{X}_{ij}^T\beta)) = 0. \tag{2.1.2}$$

In 2003, Xie and Yang proved rigorously in [12] that equation (2.1.1) has a root $\widehat{\beta}_n$ which is a consistent estimator of β, and derived the asymptotic normality of this estimator. The goal of this chapter is to develop a method similar to that of [12] which can be applied when some of the responses are missing.

We assume that the following *Missing at Random Assumption* (MAR) holds: for any $i = 1, \ldots, n$, $\mathbf{Y}_i$ is conditionally independent of $\mathbf{I}_i$ given $\mathbf{X}_i$. For any $i = 1, \ldots, n$ and $j = 1, \ldots, m$, we let

$$\pi_{ij} = P(I_{ij} = 1|\mathbf{X}_i, \mathbf{Y}_i) = P(I_{ij} = 1|\mathbf{X}_i)$$

and we consider the *inverse probability weighted response*

$$Y^*_{ij} = \frac{Y_{ij} I_{ij}}{\pi_{ij}}.$$

Note that I_{ij} is a variable which takes only the values 0 and 1, and hence $\mathrm{E}(I_{ij}|\mathbf{X}_i) = P(I_{ij} = 1|\mathbf{X}_i) = \pi_{ij}$. Moreover, $\mathrm{E}(I^2_{ij}|\mathbf{X}_i) = P(I_{ij} = 1|\mathbf{X}_i) = \pi_{ij}$.

Let $\mathbf{Y}^*_i = (Y^*_{i1}, \ldots, Y^*_{im})^T$ be the vector of weighted responses for the i^{th} individual. Note that

$$Y^*_{ij} - Y_{ij} = \Big(\frac{I_{ij}}{\pi_{ij}} - 1\Big) Y_{ij}. \tag{2.1.3}$$

We denote $\varepsilon^*_i(\beta) = (\varepsilon^*_{i1}(\beta), \ldots, \varepsilon^*_{im}(\beta))^T$, where $\varepsilon^*_{ij}(\beta) = Y^*_{ij} - \mu_{ij}(\beta)$, $j = 1, \ldots, m$.

Lemma 2.1.4. *For each $i = 1, \ldots, n$ and $j = 1, \ldots, m$, $\mathrm{E}(Y^*_{ij}|\boldsymbol{X}_i) = \mu_{ij}(\beta)$. Therefore*

$$\mathrm{E}(\varepsilon^*_i(\beta)|\boldsymbol{X}_i) = 0 \ for\ all\ i \geq 1.$$

Proof: Using (2.1.3) and double conditioning, we have:

$$\begin{aligned}
\mathrm{E}(Y^*_{ij}|\mathbf{X}_i) &= \mathrm{E}(Y^*_{ij} - Y_{ij}|\mathbf{X}_i) + \mathrm{E}(Y_{ij}|\mathbf{X}_i) \\
&= \mathrm{E}\Big[\Big(\frac{I_{ij}}{\pi_{ij}} - 1\Big) Y_{ij}|\mathbf{X}_i\Big] + \mu_{ij}(\beta) \\
&= \mathrm{E}\Big[Y_{ij} \mathrm{E}\Big[\frac{I_{ij}}{\pi_{ij}} - 1|\mathbf{X}_i, \mathbf{Y}_i\Big]|\mathbf{X}_i\Big] + \mu_{ij}(\beta) \\
&= \mu_{ij}(\beta),
\end{aligned}$$

where for the last line, we used the fact that, by the MAR assumption,

$$\mathrm{E}\Big[\frac{I_{ij}}{\pi_{ij}} - 1|\mathbf{Y}_i, \mathbf{X}_i\Big] = \mathrm{E}\Big[\frac{I_{ij}}{\pi_{ij}} - 1|\mathbf{X}_i\Big] = \frac{1}{\pi_{ij}} \mathrm{E}(I_{ij}|\mathbf{X}_i) - 1 = 0. \tag{2.1.4}$$

This concludes the proof. ∎

For each $i = 1, \ldots, n$ and $j, k = 1, \ldots, m$, we let $q_{i,jk}$ be the conditional probability that the responses of the i^{th} individual are missing on both j^{th} and k^{th} occasions, given the covariate matrix $\mathbf{X}_i$:

$$q_{i,jk} = P(I_{ij} = 1, I_{ik} = 1|\mathbf{X}_i).$$

In the next lemma, we compute the conditional covariance matrix of $\varepsilon^*_i(\beta)$.

Lemma 2.1.5. *The conditional covariance matrix of* $\varepsilon_i^*(\beta)$ *given* $\boldsymbol{X}_i$ *is*

$$\Sigma_i^*(\beta) := \mathrm{E}[\varepsilon_i^*(\beta)\varepsilon_i^*(\beta)^T|\boldsymbol{X}_i] = (\sigma_{i,jk}^*(\beta))_{1\leq j,k\leq m},$$

where

$$\sigma_{i,jk}^*(\beta) = \sigma_{i,jk}(\beta) + \Big(\tfrac{1}{\pi_{ij}\pi_{ik}}q_{i,jk} - 1\Big)\Big(\sigma_{i,jk}(\beta) + \mu_{ij}(\beta)\mu_{ik}(\beta)\Big).$$

Proof: For any $j, k = 1, \ldots, m$ fixed,

$$\begin{aligned}\sigma_{i,jk}^*(\beta) &= \mathrm{E}[\varepsilon_{ij}^*(\beta)\varepsilon_{ik}^*(\beta)|\mathbf{X}_i]\\ &= \mathrm{E}[(Y_{ij}^* - \mu_{ij}(\beta))(Y_{ik}^* - \mu_{ik}(\beta))|\mathbf{X}_i]\\ &= \mathrm{E}[(Y_{ik}^* - Y_{ij} + Y_{ij} - \mu_{ik}(\beta))(Y_{ik}^* - Y_{ik} + Y_{ik} - \mu_{ik}(\beta))|\mathbf{X}_i]\\ &= \mathrm{E}[(Y_{ij}^* - Y_{ij})(Y_{ik}^* - Y_{ik})|\mathbf{X}_i] + \mathrm{E}[(Y_{ij}^* - Y_{ij})(Y_{ik} - \mu_{ik}(\beta))|\mathbf{X}_i]\\ &+ \mathrm{E}[(Y_{ik}^* - Y_{ik})(Y_{ij} - \mu_{ij}(\beta))|\mathbf{X}_i] + \mathrm{E}[(Y_{ij} - \mu_{ij}(\beta))(Y_{ik} - \mu_{ik}(\beta))|\mathbf{X}_i].\end{aligned} \tag{2.1.5}$$

We treat separately the four terms. First we show that the second term is equal to 0:

$$\begin{aligned}\mathrm{E}[(Y_{ij}^* - Y_{ij})(Y_{ik} - \mu_{ik}(\beta))|\mathbf{X}_i] &= \mathrm{E}\Big[\Big(\frac{I_{ij}}{\pi_{ij}} - 1\Big)Y_{ij}(Y_{ik} - \mu_{ij}(\beta))|\mathbf{X}_i\Big]\\ &= \mathrm{E}\Big[Y_{ij}(Y_{ik} - \mu_{ij}(\beta))\mathrm{E}\Big[\frac{I_{ij}}{\pi_{ij}} - 1|\mathbf{X}_i, \mathbf{Y}_i\Big]|\mathbf{X}_i\Big]\\ &= 0,\end{aligned}$$

where we used (2.1.4) for the last equality. Similarly, it can be shown that the third term in (2.1.5) is also equal to 0. Note that the fourth term in (2.1.5) is

$$\mathrm{E}[(Y_{ij} - \mu_{ij}(\beta))(Y_{ik} - \mu_{ik}(\beta))|\mathbf{X}_i] = \mathrm{E}[\varepsilon_{ij}(\beta)\varepsilon_{ik}(\beta)|\mathbf{X}_i] = \sigma_{i,jk}(\beta).$$

We infer from here that

$$\sigma_{i,jk}^*(\beta) = \mathrm{E}[(Y_{ij}^* - Y_{ij})(Y_{ik}^* - Y_{ik})|\mathbf{X}_i] + \sigma_{i,jk}(\beta). \tag{2.1.6}$$

Note that

$$\begin{aligned}\mathrm{E}[(Y_{ij}^* - Y_{ij})(Y_{ik}^* - Y_{ik})|\mathbf{X}_i] &= \mathrm{E}\Big[\Big(\frac{I_{ij}}{\pi_{ij}} - 1\Big)\Big(\frac{I_{ik}}{\pi_{ik}} - 1\Big)Y_{ij}Y_{ik}|\mathbf{X}_i\Big]\\ &= \mathrm{E}\Big[Y_{ij}Y_{ik}\mathrm{E}\Big[\Big(\frac{I_{ij}}{\pi_{ij}} - 1\Big)\Big(\frac{I_{ik}}{\pi_{ik}} - 1\Big)|\mathbf{X}_i, \mathbf{Y}_i\Big]|\mathbf{X}_i\Big].\end{aligned} \tag{2.1.7}$$

We compute separately the inner conditional expectation. By the MAR assumption,

$$\mathrm{E}\Big[\Big(\frac{I_{ij}}{\pi_{ij}} - 1\Big)\Big(\frac{I_{ik}}{\pi_{ik}} - 1\Big)|\mathbf{X}_i, \mathbf{Y}_i\Big]$$

$$= \frac{1}{\pi_{ij}\pi_{ik}}\mathrm{E}(I_{ij}I_{jk}|\mathbf{X}_i, \mathbf{Y}_i) - \frac{1}{\pi_{ij}}\mathrm{E}(I_{ij}|\mathbf{X}_i, \mathbf{Y}_i) - \frac{1}{\pi_{ik}}\mathrm{E}(I_{ik}|\mathbf{X}_i, \mathbf{Y}_i) + 1$$
$$= \frac{1}{\pi_{ij}\pi_{ik}}\mathrm{E}(I_{ij}I_{jk}|\mathbf{X}_i) - \frac{1}{\pi_{ij}}\mathrm{E}(I_{ij}|\mathbf{X}_i) - \frac{1}{\pi_{ik}}\mathrm{E}(I_{ik}|\mathbf{X}_i) + 1,$$

using the fact that $\pi_{ij} = \mathrm{E}(I_{ij}|\mathbf{X}_i) = h_j(\mathbf{X}_i)$ is a function of $\mathbf{X}_i$. Since $I_{ij}I_{ik}$ is a variable which takes values 0 and 1, we have

$$\mathrm{E}(I_{ij}I_{ik}|\mathbf{X}_i) = P(I_{ij}I_{ik} = 1|\mathbf{X}_i) = P(I_{ij} = 1, I_{ik} = 1|\mathbf{X}_i) = q_{i,jk}.$$

As we noticed before,

$$\tfrac{1}{\pi_{ij}}\mathrm{E}(I_{ij}|\mathbf{X}_i) = 1 \text{ and } \tfrac{1}{\pi_{ik}}\mathrm{E}(I_{ik}|\mathbf{X}_i) = 1.$$

Therefore,

$$\mathrm{E}\Big[\Big(\tfrac{I_{ij}}{\pi_{ij}} - 1\Big)\Big(\tfrac{I_{ij}}{\pi_{ij}} - 1\Big)|\mathbf{X}_i, \mathbf{Y}_i\Big] = \tfrac{1}{\pi_{ij}\pi_{ik}}q_{i,jk} - 1.$$

Coming back to (2.1.7), we obtain

$$\begin{aligned}\mathrm{E}[(Y^*_{ij} - Y_{ij})(Y^*_{ik} - Y_{ik})|\mathbf{X}_i] &= \mathrm{E}\Big[Y_{ij}Y_{ik}\Big(\frac{1}{\pi_{ij}\pi_{ik}}q_{i,jk} - 1\Big)|\mathbf{X}_i\Big] \\ &= \Big(\frac{1}{\pi_{ij}\pi_{ik}}q_{i,jk} - 1\Big)\mathrm{E}(Y_{ij}Y_{ik}|\mathbf{X}_i) \\ &= \Big(\frac{1}{\pi_{ij}\pi_{ik}}q_{i,jk} - 1\Big)(\sigma_{i,jk} + \mu_{ij}(\beta)\mu_{ik}(\beta)), \qquad (2.1.8)\end{aligned}$$

where for the last equality, we used the fact that

$$\begin{aligned}\mathrm{E}(Y_{ij}Y_{ik}|\mathbf{X}_i) &= \mathrm{E}[(Y_{ij} - \mu_{ij}(\beta) + \mu_{ij}(\beta))(Y_{ik} - \mu_{ik}(\beta) + \mu_{ik}(\beta))|\mathbf{X}_i] \\ &= \mathrm{E}[(Y_{ij} - \mu_{ij}(\beta))(Y_{ik} - \mu_{ik}(\beta))|\mathbf{X}_i] \\ &+ \mu_{ik}(\beta)\mathrm{E}(Y_{ij} - \mu_{ij}(\beta)|\mathbf{X}_i) + \mu_{ij}(\beta)\mathrm{E}(Y_{ik} - \mu_{ik}(\beta)|\mathbf{X}_i) \\ &+ \mu_{ij}(\beta)\beta_{ik}(\beta) = \sigma_{ij}(\beta)) + \mu_{ij}(\beta)\mu_{ik}(\beta),\end{aligned}$$

since $\mathrm{E}(Y_{ij} - \mu_{ij}(\beta)|\mathbf{X}_i) = 0$, and $\mathrm{E}(Y_{ik} - \mu_{ik}(\beta)|\mathbf{X}_i) = 0$. The result follows from relations (2.1.6) and (2.1.8). ∎

Note that $q_{i,jj} = P(I_{ij} = 1|\mathbf{X}_i) = \pi_{ij}$, so the conditional marginal variance of Y^*_{ij} given $\mathbf{X}_i$ is

$$\sigma^*_{i,jj}(\beta) = \sigma^2_{ij}(\beta) + \Big(\frac{1}{\pi_{ij}} - 1\Big)(\sigma^2_{ij}(\beta) + \mu^2_{ij}(\beta))$$

$$= \mu'(\mathbf{X}_{ij}^T\beta) + \Big(\frac{1}{\pi_{ij}} - 1\Big)\Big(\mu'(\mathbf{X}_{ij}^T\beta) + \mu^2(\mathbf{X}_{ij}^T\beta)). \quad (2.1.9)$$

Let $\mathbf{A}_i^*(\beta) = \text{diag}[\sigma_{i,11}^*(\beta), \dots, \sigma_{i,mm}^*(\beta)]$, i.e. $\mathbf{A}_i^*(\beta)$ is the diagonal matrix with elements $\sigma_{i,11}^*(\beta), \dots, \sigma_{i,mm}^*(\beta)$. We let $\mathbf{R}_i(\alpha)$ be an arbitrary correlation matrix which depends on a parameter α. Typical examples of such matrices are

$$\mathbf{R}_i(\alpha) = \begin{bmatrix} 1 & \alpha & \dots & \alpha \\ \alpha & 1 & \dots & \alpha \\ \vdots & \vdots & \ddots & \vdots \\ \alpha & \alpha & \dots & 1 \end{bmatrix}$$

and

$$\mathbf{R}_i(\alpha) = \begin{bmatrix} 1 & \alpha & \dots & \alpha^m \\ \alpha & 1 & \dots & \alpha^{m-1} \\ \vdots & \vdots & \ddots & \vdots \\ \alpha^m & \alpha^{m-1} & \dots & 1 \end{bmatrix}$$

We define

$$\mathbf{V}_i^*(\beta, \alpha) = \mathbf{A}_i^*(\beta)^{1/2}\mathbf{R}_i(\alpha)\mathbf{A}_i^*(\beta)^{1/2}. \quad (2.1.10)$$

We are interested in solving the equation

$$\mathbf{g}_n(\beta) := \sum_{i=1}^{n} \mathbf{D}_i^T(\beta)\mathbf{V}_i^*(\beta, \alpha)^{-1}(\mathbf{Y}_i^* - \mu_i(\beta)) = 0. \quad (2.1.11)$$

Note that equation (2.1.11) is the analogue of equation (3) of [12] for the case of missing responses which are adjusted using the inverse probability weights.

We have:

$$\mathbf{g}_n(\beta) = \sum_{i=1}^{n} \mathbf{X}_i^T\mathbf{F}_i(\beta)\mathbf{R}_i(\alpha)^{-1}\mathbf{A}_i^*(\beta)^{-1/2}\varepsilon_i^*(\beta), \quad (2.1.12)$$

where $\mathbf{F}_i(\beta) = \mathbf{A}_i(\beta)\mathbf{A}_i^*(\beta)^{-1/2}$ is the diagonal matrix with entries $f_{ij}(\beta), 1 \le j \le m$, given by:

$$f_{ij}(\beta) = \frac{\sigma_{ij}^2(\beta)}{\sqrt{\sigma_{i,jj}^*(\beta)}} = \frac{\mu'(\mathbf{X}_{ij}^T\beta)}{\sqrt{\mu'(\mathbf{X}_{ij}^T\beta) + (\frac{1}{\pi_{ij}} - 1)(\mu'(\mathbf{X}_{ij}^T\beta) + \mu^2(\mathbf{X}_{ij}^T\beta))}} \quad (2.1.13)$$

The following result gives the mean and the covariance matrix of the estimating function $\mathbf{g}_n(\beta)$. Note that the expected value of a matrix $\mathbf{A} = (A_{jk})_{1\le j,k\le m}$ whose elements are random variables A_{jk} is, by definition, the matrix $\text{E}(\mathbf{A}) = \text{E}(A_{jk})_{1\le j,k\le m}$.

Lemma 2.1.6. *$\boldsymbol{g}_n(\beta)$ is an unbiased estimating function, i.e.* $\mathrm{E}[\boldsymbol{g}_n(\beta)] = 0$ *for all* β. *The covariance matrix of* $\boldsymbol{g}_n(\beta)$ *is*

$$\boldsymbol{M}_n(\beta) := \mathrm{E}[\boldsymbol{g}_n(\beta)\boldsymbol{g}_n(\beta)^T] = \mathrm{E}[\boldsymbol{M}_n^*(\beta)],$$

where

$$\boldsymbol{M}_n^*(\beta) = \sum_{i=1}^n \boldsymbol{D}_i^T(\beta)\,\boldsymbol{V}_i^*(\beta,\alpha)^{-1}\Sigma_i^*(\beta)\,\boldsymbol{V}_i^*(\beta,\alpha)^{-1}\boldsymbol{D}_i(\beta).$$

Proof: The first statement follows by Lemma 2.1.4 since

$$\mathrm{E}[\mathbf{g}_n(\beta)] = \sum_{i=1}^n \mathrm{E}[\mathbf{D}_i^T(\beta)\mathbf{V}_i^*(\beta,\alpha)^{-1}\mathrm{E}[\varepsilon_i^*(\beta)|\mathbf{X}_i]] = 0. \tag{2.1.14}$$

We proceed now with the calculation of the covariance matrix of $\mathbf{g}_n(\beta)$. Note that

$$\mathbf{g}_n(\beta)\mathbf{g}_n(\beta)^T = \sum_{i=1}^n\sum_{l=1}^n \mathbf{D}_i^T(\beta)\mathbf{V}_i^*(\beta,\alpha)^{-1}\varepsilon_i^*(\beta)\varepsilon_l^*(\beta)^T\mathbf{V}_l^*(\beta,\alpha)^{-1}\mathbf{D}_l(\beta).$$

Since $\pi_{ij} = P(I_{ij}=1|\mathbf{X}_i) = h_j(\mathbf{X}_i)$ for a function h_j,

$$\varepsilon_{ij}^*(\beta) = Y_{ij}^* - \mu(X_{ij}^T\beta) = \tfrac{Y_{ij}I_{ij}}{\pi_{ij}} - \mu(X_{ij}^T\beta) = \Phi_j(Y_{ij},\mathbf{X}_i,I_{ij})$$

for a certain function Φ_j. Since $\{(\mathbf{Y}_i,\mathbf{X}_i,\mathbf{I}_i)\}_{i\geq 1}$ are independent, it follows that $\{\varepsilon_i^*(\beta)\}_{i\geq 1}$ are independent. The same argument shows that $\{\mathbf{D}_i^T(\beta)\mathbf{V}_i^*(\beta,\alpha)^{-1}\varepsilon_i^*(\beta)\}_{i\geq 1}$ are independent. Note that

$$\begin{aligned}\mathrm{E}[\mathbf{D}_i^T(\beta)\mathbf{V}_i^*(\beta,\alpha)^{-1}\varepsilon_i^*(\beta)] &= \mathrm{E}[\mathrm{E}[\mathbf{D}_i^T(\beta)\mathbf{V}_i^*(\beta,\alpha)^{-1}\varepsilon_i^*(\beta)|\mathbf{X}_i]]\\ &= \mathrm{E}[\mathbf{D}_i^T(\beta)\mathbf{V}_i^*(\beta,\alpha)^{-1}\mathrm{E}[\varepsilon_i^*(\beta)|\mathbf{X}_i]]\\ &= 0,\end{aligned}$$

where for the last equality we used Lemma 2.1.4. Therefore if $i \neq l$,

$$\mathrm{E}[\mathbf{D}_i^T(\beta)\mathbf{V}_i^*(\beta,\alpha)^{-1}\varepsilon_i^*(\beta)\varepsilon_l^*(\beta)\mathbf{V}_l^*(\beta,\alpha)^{-1}\mathbf{D}_l(\beta)] = 0.$$

Coming back to the calculation of $\mathrm{E}[g_n(\beta)g_n(\beta)^T]$, we obtain using conditioning again

$$\begin{aligned}\mathrm{E}[g_n(\beta)g_n(\beta)^T] &= \sum_{i=1}^n \mathrm{E}[\mathrm{E}[\mathbf{D}_i^T(\beta)\mathbf{V}_i^*(\beta,\alpha)^{-1}\varepsilon_i^*(\beta)\varepsilon_i^*(\beta)^T\mathbf{V}_i^*(\beta,\alpha)^{-1}\mathbf{D}_i(\beta)|\mathbf{X}_i]]\\ &= \sum_{i=1}^n \mathrm{E}[\mathbf{D}_i^T(\beta)\mathbf{V}_i^*(\beta,\alpha)^{-1}\mathrm{E}[\varepsilon_i^*(\beta)\varepsilon_i^*(\beta)^T|\mathbf{X}_i]\mathbf{V}_i^*(\beta,\alpha)^{-1}\mathbf{D}_i(\beta)]\end{aligned}$$

$$= \sum_{i=1}^{n} \mathrm{E}[\mathbf{D}_i^T(\beta)\mathbf{V}_i^*(\beta,\alpha)^{-1}\Sigma_i^*(\beta)\mathbf{V}_i^*(\beta,\alpha)^{-1}\mathbf{D}_i(\beta)]$$
$$= \mathrm{E}[\mathbf{M}_n^*(\beta)].$$

This finishes the proof. ∎

Remark 2.1.7. Using the fact that $\mathbf{D}_i(\beta) = \mathbf{A}_i(\beta)\mathbf{X}_i$, $\mathbf{V}_i^*(\beta,\alpha) = \mathbf{A}_i^*(\beta)^{1/2}\mathbf{R}_i(\alpha)\mathbf{A}_i^*(\beta)^{1/2}$, and the fact that the covariance matrix of $\varepsilon_i^*(\beta)$ can be expressed as

$$\Sigma_i^*(\beta) = \mathbf{A}_i^*(\beta)^{1/2}\mathbf{R}_i^*\mathbf{A}_i^*(\beta)^{1/2}, \tag{2.1.15}$$

where $\mathbf{R}_i^*$ is the correlation matrix of $\varepsilon_i^*(\beta)$, we obtain the following alternative formula for $\mathbf{M}_n^*(\beta)$:

$$\mathbf{M}_n^*(\beta) = \sum_{i=1}^{n} \mathbf{X}_i^T\mathbf{A}_i(\beta)\mathbf{A}_i^*(\beta)^{-1/2}\mathbf{R}_i(\alpha)^{-1}\mathbf{R}_i^*\mathbf{R}_i(\alpha)^{-1}\mathbf{A}_i^*(\beta)^{-1/2}\mathbf{A}_i(\beta)\mathbf{X}_i.$$

We denote $\tau_n = \max_{i\le n}\{\lambda_{\max}(\mathbf{R}_i(\alpha)^{-1/2}\mathbf{R}_i^*\mathbf{R}_i(\alpha)^{-1/2})\} = \max_{i\le n}\{\lambda_{\max}(\mathbf{R}_i^{-1}(\alpha)\mathbf{R}_i^*)\}$. By Theorem A.1.5, $\mathbf{M}_n^*(\beta) \le \tau_n\mathbf{H}_n^*(\beta)$, where

$$\mathbf{H}_n^*(\beta) = \sum_{i=1}^{n} \mathbf{X}_i^T\mathbf{A}_i(\beta)\mathbf{A}_i^*(\beta)^{-1/2}\mathbf{R}_i(\alpha)^{-1}\mathbf{A}_i^*(\beta)^{-1/2}\mathbf{A}_i(\beta)\mathbf{X}_i. \tag{2.1.16}$$

By taking the expectation on both sides of this inequality, we infer that

$$\mathbf{M}_n(\beta) \le \tau_n\mathbf{H}_n(\beta),$$

where $\mathbf{H}_n(\beta) = \mathrm{E}[\mathbf{H}_n^*(\beta)]$.

2.2 Existence and Consistency

In this section, we show that under certain conditions, equation (2.1.11) has a solution $\widehat{\beta}_n$ which is a consistent estimator of β. The proofs are very similar to the proofs of Theorems 1 and 2 of [12]. We include the details for the sake of completeness.

We consider the negative derivative of our estimating function $\mathbf{g}_n(\beta)$:

$$\mathcal{D}_n(\beta) = -\tfrac{\partial}{\partial\beta^T}\mathbf{g}_n(\beta)$$

Let $\mathbf{F}_n = \mathbf{H}_n \mathbf{M}_n^{-1} \mathbf{H}_n$ and $B_n(r) = \{\beta \ ; \ \|\mathbf{M}_n^{-1/2} \mathbf{H}_n(\beta - \beta_0)\| \leq r\}$. As in [12], we consider the following conditions:

(I_w) $\lambda_{\min}(F_n) \to \infty$

(L_w) There exists a constant $c_0 > 0$ such that for any $r > 0$
$P(\mathcal{D}_n^T(\beta)\mathbf{M}_n^{-1}\mathcal{D}_n(\beta) \geq c_o \mathbf{F}_n$ for all $\beta \in B_n(r)) \to 1$

(D_w) For any $r > 0, P(\mathcal{D}_n(\beta)$ is non $-$ singular, for any $\beta \in B_n(r)) \to 1$

We denote by $\lambda_{\min}(\mathbf{A})$ the minimum eigenvalue of a symmetric matrix $\mathbf{A}$ (see Appendix A).

Theorem 2.2.1. *Under conditions (I_w), (L_w) and (D_w) we have:*
a) $\lim_{n\to\infty} P(\text{there exist } r > 0 \text{ and } \widehat{\beta}_n \in B_n(r) \text{ such that } g_n(\widehat{\beta}_n) = 0) = 1$
b) $\widehat{\beta}_n \to \beta_0$.

The proof of this theorem relies on the following result:

Lemma 2.2.2. *[Lemma A of [4]] Let $H : \mathbb{R}^p \to \mathbb{R}^p$ be a one-to-one continuously differentiable function, and $x_0 \in \mathbb{R}^p$. If*

$$\|H(x_0)\| \leq \inf_{\|x-x_0\|=\delta} \|H(x) - H(x_0)\|,$$

then there exists $\widehat{x}$ with $\|\widehat{x} - x_0\| < \delta$ such that $H(\widehat{x}) = 0$.

Proof: Let $T_n(\beta) = \mathbf{M}_n^{-1/2} g_n(\beta)$. Let $\Omega_n(r)$ be the event where $\mathcal{D}_n^T(\beta)\mathbf{M}_n^{-1}\mathbf{D}_n(\beta) \geq c_o\mathbf{F}_n$ for all $\beta \in \mathbf{B}_n(r)$, and $\mathcal{D}_n(\beta)$ is non-singular for all $\beta \in \mathbf{B}_n(r)$. By conditions (L_w) and (D_w), $P(\Omega_n(r)) \to 1$ for any $r > 0$. On the event $\Omega_n(r)$ the function T_n is one-to-one since its derivative is non-singular. We let

$$E_n(r) = \left\{ \|T_n(\beta_0)\| \leq \inf_{\beta \in \partial B_n(r)} \|T_n(\beta) - T_n(\beta_0)\| \right\}$$

By Lemma 2.2.2,

$$E_n(r) \cap \Omega_n(r) \subset \overline{\Omega}_n(r), \tag{2.2.1}$$

where $\overline{\Omega}_n(r)$ is the event for which there exists $\widehat{\beta}_n \in B_n(r)$ such that $g_n(\widehat{\beta}_n) = 0$. (Note that on the event $\overline{\Omega}_n(r)$, $\hat{\beta}_n = \hat{\beta}_n(r)$ depends on r. But since g_n is one-to-one, if $r_1 < r_2$, then $\widehat{\beta}_n(r_1) = \widehat{\beta}_n(r_2)$. So the definition of $\hat{\beta}_n$ is correct.)

We will prove that for any $\varepsilon > 0$ there exists $r = r_\varepsilon > 0$ and an integer $N_\varepsilon \geq 1$ such that

$$P(E_n(r) \cap \Omega_n(r)) \geq 1 - \varepsilon \tag{2.2.2}$$

for all $n \geq N_\epsilon$. Using (2.2.1), it follows that $P(\bigcup_{r>0} \overline{\Omega}_n(r)) \geq P(\overline{\Omega}_n(r)) \geq 1 - \varepsilon$ for all $n \geq N_\epsilon$. This will conclude the proof of part a) since $\bigcup_{r>0} \overline{\Omega}_n(r)$ is exactly the event for which there exists $r > 0$ and $\hat{\beta}_n \in B_n(r)$ such that $g_n(\hat{\beta}_n) = 0$.

Now let $\varepsilon > 0$ be arbitrary and take $r = r_\varepsilon = \sqrt{\frac{2p}{c_0\epsilon}}$. For any $\beta \in \partial B_n(r)$, by Taylor's formula, there exists $\overline{\beta}_n \in B_n(r)$ such that

$$\begin{aligned}
\|T_n(\beta) - T_n(\beta_0)\|^2 &= \|\mathbf{M}_n^{-1/2}(g_n(\beta) - g_n(\beta_0))\|^2 \\
&= \|\mathbf{M}_n^{-1/2}\mathcal{D}_n(\bar{\beta}_n)(\beta - \beta_0)\|^2 \\
&= \|\mathbf{M}_n^{-1/2}\mathcal{D}_n(\overline{\beta}_n)\mathbf{H}_n^{-1}\mathbf{M}_n^{1/2}\mathbf{M}_n^{-1/2}\mathbf{H}_n(\beta - \beta_0)\|^2.
\end{aligned}$$

We use the fact that for any $p \times p$ matrix $\mathbf{A}$ and p-dimensional vector v,

$$\|\mathbf{A}v\|^2 = v^T\mathbf{A}^T\mathbf{A}v \geq \lambda_{\min}(\mathbf{A}^T\mathbf{A})v^T v, \tag{2.2.3}$$

(see Theorem A.1.5 in Appendix A). Applying this inequality with $\mathbf{A} = \mathbf{M}_n^{-1/2}\mathcal{D}_n(\bar{\beta}_n)H_n^{-1}\mathbf{M}_n^{1/2}$ and $v = \mathbf{M}_n^{-1/2}H_n(\beta - \beta_0)$, we obtain

$$\|T_n(\beta) - T_n(\beta_0)\|^2 \geq \lambda_{\min}(\mathbf{M}_n^{1/2}H_n^{-1}\mathcal{D}_n^T(\bar{\beta}_n)\mathbf{M}_n^{-1}\mathcal{D}_n(\bar{\beta}_n)H_n^{-1}\mathbf{M}_n^{1/2}) \cdot ||v||^2$$

By condition (L_w), on the event $\Omega_n(r)$,

$$\begin{aligned}
\lambda_{\min}(\mathbf{M}_n^{1/2}H_n^{-1}\mathcal{D}_n^T(\overline{\beta}_n)\mathbf{M}_n^{-1}\mathcal{D}_n(\overline{\beta}_n)H_n^{-1}\mathbf{M}_n^{1/2}) &\geq c_0\lambda_{\min}(\mathbf{M}_n^{1/2}H_n^{-1}F_nH_n^{-1}\mathbf{M}^{1/2}) \\
&= c_0\lambda_{\min}(I) = c_0.
\end{aligned}$$

Moreover, $\|v\| = r$ since $\beta \in \partial B_n(r)$. This shows that on the event $\Omega_n(r)$, $\|T_n(\beta) - T_n(\beta_0)\| \geq c_0 r^2$ for any $\beta \in \partial B_n(r)$. Hence

$$\Omega_n(r) \subseteq \left\{ \inf_{\beta \in \partial B_n(r)} \|T_n(\beta) - T_n(\beta_0)\| \geq c_0^{1/2} r \right\}.$$

On the other hand,

$$E_n(r) \supseteq \left\{ \|T_n(\beta_0)\| \leq c_0^{1/2} r \right\} \cap \left\{ c_0^{1/2} r \leq \inf_{\beta \in \partial B_n(r)} \|T_n(\beta) - T_n(\beta_0)\| \right\}$$

and hence

$$E_n(r) \cap \Omega_n(r) \supseteq \left\{ \|T_n(\beta_0)\| \leq c_0^{1/2} r \right\} \cap \Omega_n(r).$$

By Bonferroni's inequality,

$$P(E_n(r) \cap \Omega_n(r)) \geq P(\|T_n(\beta_0)\| \leq c_0^{1/2} r) + P(\Omega_n(r)) - 1 \tag{2.2.4}$$

By condition (L_w), $P(\Omega_n(r)) \to 1$ as $n \to \infty$. Hence there exists an $N_\varepsilon \geq 1$ such that

$$P(\Omega_n(r)) \geq 1 - \frac{\varepsilon}{2} \tag{2.2.5}$$

for all $n \geq N_\varepsilon$. By the Markov inequality,

$$\begin{aligned} P(\|T_n(\beta_0)\| \leq c_0^{1/2} r) &= 1 - P(\|T_n(\beta_0)\| > c_0^{1/2} r) \\ &\geq 1 - \frac{1}{c_0 r^2} \mathrm{E}(\|T_n(\beta_n)\|^2). \end{aligned}$$

We compute separately $\mathrm{E}(\|T_n(\beta_n)\|^2)$. Using the fact that $\|x\|^2 = \mathrm{tr}(xx^T)$ for any $x \in \mathbb{R}^p$, we have:

$$\begin{aligned} \mathrm{E}(\|T_n(\beta_n)\|^2) &= \mathrm{E}(\|\mathbf{M}_n^{-1/2} g_n(\beta_0)\|^2) = \mathrm{E}[\mathrm{tr}(\mathbf{M}_n^{-1/2} g_n(\beta_0) g_n(\beta_0)^T \mathbf{M}_n^{-1/2})] \\ &= \mathrm{tr}(\mathrm{E}(\mathbf{M}_n^{-1/2} g_n(\beta_0) g_n(\beta_0)^T \mathbf{M}_n^{-1/2})) \\ &= \mathrm{tr}(\mathbf{M}_n^{-1/2} \mathrm{E}[g_n(\beta_0) g_n(\beta_0)^T] \mathbf{M}_n^{-1/2}) \\ &= \mathrm{tr}(I) \\ &= p. \end{aligned}$$

Therefore,

$$P(\|T_n(\beta_0)\| \leq c_0^{1/2} r) \geq 1 - \frac{1}{c_0 r^2} p = \frac{\varepsilon}{2} \tag{2.2.6}$$

for all $n \geq 1$, where the last equality follows from the definition of r. Relation (2.2.2) follows from (2.2.4), (2.2.5) and (2.2.6).

b) Let $\delta > 0$ be arbitrary. We have to show that $P(\|\widehat{\beta}_n - \beta\| \leq \delta) \to 1$, i.e. for any $\varepsilon > 0$ there exists $N_{\varepsilon,\delta} \geq 1$ such that $P(\|\widehat{\beta}_n - \beta\| \leq \delta) \geq 1 - \varepsilon$ for all $n \geq N_{\varepsilon,\delta}$. Let $\varepsilon > 0$ be arbitrary and $r = r_\varepsilon$ from part a). On the event $\Omega_n(r) \cap E_n(r)$, we have:

$$\|\widehat{\beta}_n - \beta_0\| \leq \|\mathbf{H}_n^{-1} \mathbf{M}_n^{1/2}\| \cdot \|\mathbf{M}_n^{-1/2} \mathbf{H}_n(\widehat{\beta}_n - \beta_0)\| \leq (\tfrac{c_1 p}{\lambda_{\min} \mathbf{F}_n})^{1/2} r$$

since $\widehat{\beta}_n \in B_n(r)$ and

$$\|\mathbf{H}_n^{-1/2} \mathbf{M}_n^{1/2}\|^2 \leq c_1^2 p \|\mathbf{H}_n^{-1} \mathbf{M}_n^{1/2}\|_E^2 = c_1 \mathrm{tr}(\mathbf{F}_n^{-1}) \leq c_1 p \lambda_{\max}(\mathbf{F}_n^{-1}) = c_1 p \frac{1}{\lambda_{\min}(\mathbf{F}_n)}.$$

Here, we used the fact that $\|\mathbf{A}\|_E^2 = \mathrm{tr}(\mathbf{A}\mathbf{A}^T)$. By condition (I_w), there exists $N_\delta^* \geq 1$ such that $\lambda_{\min}(\mathbf{F}_n) \geq \frac{c_1 p r^2}{\delta^2}$ for all $n \geq N_\delta^*$. Hence, for any $n \geq N_\delta^*$,

$$\Omega_n(r) \cap E_n(r) \subset \{\|\widehat{\beta}_n - \beta_0\| \leq \delta\}.$$

Let $N_{\varepsilon,\delta} = \max(N_\varepsilon, N_\delta^*)$. By (2.2.2), it follows that for any $n \geq N_{\varepsilon,\delta}$,

$$P(\|\widehat{\beta}_n - \beta_0\| \leq \delta) \geq P(\Omega_n(r) \cap E_n(r)) \geq 1 - \varepsilon.$$

∎

Note that conditions (I_w) and (L_w) rely on the covariance matrix $\mathbf{M}_n$, which in turn depends on the unknown correlation matrix $\mathbf{R}_i^*$. In the next set of conditions, we eliminate the matrix $\mathbf{M}_n$, using the fact that $\mathbf{M}_n \leq \tau_n \mathbf{H}_n$ (see Remark 2.1.7). Note that τ_n also depends on $\mathbf{R}_i^*$, but by Lemma A.1.4 (Appendix A), we have

$$\tau_n \leq m\widetilde{\lambda}_n,$$

where $\widetilde{\lambda}_n = \max_{i \leq m} \lambda_{\max}(\mathbf{R}_i^{-1}(\alpha))$. Note that $\mathbf{R}_i^*$ is a correlation matrix, and hence all its entries are bounded in modulus by 1. Let $B_n^*(r) = \{\beta; \|\mathbf{H}_n^{1/2}(\beta - \beta_0)\| \leq \tau_n^{1/2} r\}$. The new conditions are:

(I_w^*) $\lambda_{\min}(\mathbf{H}_n)/\tau_n \to \infty$ as $n \to \infty$.
(L_w^*) There exists a constant $c_0 > 0$ such that for any $r > 0$,

$$P(\mathbf{x}^T \mathbf{H}_n^{-1/2} \mathcal{D}_n(\beta) \mathbf{H}_n^{-1/2} \mathbf{x} \geq c_0 \text{ for any } \beta \in B_n^*(r) \text{ and } \mathbf{x} \in \mathbb{R}^p \text{ with } \|\mathbf{x}\| = 1) \to 1.$$

(D_w^*) For any $r > 0$, $P(\mathcal{D}_n(\beta)$ is non-singular for any $\beta \in B_n^*(r)) \to 1$.

Under condition (D_w^*), with probability converging to 1, $\mathcal{D}_n(\beta)$ is non-singular. Observe that we cannot say that $\mathcal{D}_n(\beta) > 0$, since $\mathcal{D}_n(\beta)$ may not be a symmetric matrix.

Theorem 2.2.3. *Under conditions (I_w^*), (L_w^*) and (D_w^*), the conclusion of Theorem 2.2.1 remains valid, with $B_n(r)$ replaced by $B_n^*(r)$.*

Proof: We use an argument similar to the proof of Theorem 2 of [12]. Let $T_n^*(\beta) = \mathbf{H}_n^{-1/2} g_n(\beta)$ and $E_n^*(r) = \left\{ \|T_n^*(\beta_0)\| \leq \inf_{\beta \in \partial B_n^*(r)} \|T_n^*(\beta) - T_n^*(\beta_0)\| \right\}$. Let $\Omega_n^*(r)$ be the event where $\mathbf{x}^T \mathbf{H}_n^{-1/2} \mathcal{D}_n(\beta) \mathbf{H}_n^{-1/2} \mathbf{x} \geq c_0$ for any $\beta \in B_n^*(r)$ and for any $x \in \mathbb{R}^p$ with $\|\mathbf{x}\| = 1$, and $\mathcal{D}_n(\beta)$ is non-singular for any $\beta \in B_n^*(r)$. By conditions (L_w^*) and (D_w^*), $P(\Omega_n^*(r)) \to 1$ for any $r > 0$.
By Lemma 2.2.2 we know that

$$E_n^*(r) \cap \Omega_n^*(r) \subset \widetilde{\Omega}_n(r), \tag{2.2.7}$$

where $\widetilde{\Omega}_n(r)$ is the event that there exists $\widehat{\beta}_n \in B_n^*(r)$ such that $g_n(\widehat{\beta}_n) = 0$.
It suffices to prove that for any $\varepsilon > 0$, there exist $r = r_\varepsilon > 0$ and an integer $N_\varepsilon \geq 1$ such that for all $n \geq N_\varepsilon$,

$$P(\mathrm{E}_n^*(r) \cap \Omega_n^*(r)) \geq 1 - \varepsilon. \tag{2.2.8}$$

Let $\varepsilon > 0$ be arbitrary and $r = r_\varepsilon = \sqrt{\frac{2p}{c_0^2\varepsilon}}$. For any $\beta \in \partial B_n^*(r)$, by Taylor's formula, there exists $\overline{\beta}_n \in B_n^*(r)$ such that

$$\begin{aligned}\|T_n^*(\beta) - T_n^*(\beta_0)\|^2 &= \|\mathbf{H}_n^{-1/2}(g_n(\beta) - g_n(\beta_0)\|^2 = \|\mathbf{H}_n^{-1/2}\mathcal{D}_n(\overline{\beta}_n)(\beta-\beta_0)\|^2 \\ &= \|\mathbf{H}_n^{-1/2}\mathcal{D}_n(\overline{\beta}_n)\mathbf{H}_n^{-1/2}\mathbf{H}_n^{1/2}(\beta-\beta_0)||^2 \\ &\geq \lambda_{\min}(\mathbf{H}_n^{-1/2}\mathcal{D}_n(\overline{\beta}_n)\mathbf{H}_n^{-1}\mathcal{D}_n(\overline{\beta}_n)\mathbf{H}_n^{-1/2}) \cdot \|\mathbf{H}_n^{1/2}(\beta-\beta_0)\|^2,\end{aligned}$$

using (2.2.3).
Let $z_\lambda^* = \lambda_{min}(\mathbf{H}_n^{-1/2}\mathcal{D}_n(\overline{\beta}_n)\mathbf{H}_n^{-1}\mathcal{D}_n(\overline{\beta}_n)\mathbf{H}_n^{-1/2})$. We claim that $z_\lambda^* \geq c_0^2$ on the event $\Omega_n^*(r)$. To see this, let $\mathbf{A} = \mathbf{H}_n^{-1/2}\mathcal{D}_n(\overline{\beta}_n)\mathbf{H}_n^{-1/2}$. By Lemma 1 of [12], if x is a corresponding eigenvector of $\lambda_{\min}(\mathbf{A}^T\mathbf{A})$, then by (L_w^*)

$$\begin{aligned}z_\lambda^* &= \lambda_{\min}(\mathbf{A}^T\mathbf{A}) = x^T\mathbf{A}^T\mathbf{A}x \\ &\geq (x^T\mathbf{A}x)^2 = (x^T\mathbf{H}_n^{-1/2}\mathcal{D}_n(\overline{\beta}_n)\mathbf{H}_n^{-1/2}x)^2 \\ &\geq (c_0x^T\mathbf{H}_n^{-1/2}\mathbf{H}_n\mathbf{H}_n^{-1/2}x)^2 = (c_0x^Tx)^2 = c_0^2.\end{aligned}$$

Hence $\Omega_n^*(r) \subset \left\{\beta \in \partial\beta_n^*(r); \|T_n^*(\beta) - T_n^*(\beta_0)\| \geq c_0\tau_n^{1/2}r\right\}$.
On the other hand,

$$E_n^*(r) \supset \left\{\|T_n^*(\beta_0)\| \leq c_0\tau_n^{1/2}r\right\} \cap \left\{c_0\tau_n^{1/2}r \leq \inf_{\beta\in\partial B_n^*(r)} \|T_n^*(\beta) - T_n^*(\beta_0)\|\right\}$$

and therefore by

$$E_n^*(r) \cap \Omega_n^*(r) \supset \|T_n^*(\beta_0)\| \leq c_0\tau_n^{1/2}r \cap \Omega_n^*(r).$$

So, by Bonferroni's inequality,

$$P(E_n^*(r) \cap \Omega_n^*(r)) \geq P(||T_n^*(\beta_0)|| \leq c_0\tau_n^{1/2}r) + P(\Omega_n^*(r)) - 1. \tag{2.2.9}$$

By condition (L_w^*), there exists $N_\varepsilon^* \geq 1$ such that

$$P(\Omega_n^*(r)) \geq 1 - \frac{\varepsilon}{2} \tag{2.2.10}$$

for all n $\geq N_\varepsilon^*$. By Chebyshev's inequality,

$$P(||T_n^*(\beta_0)|| \leq c_0\tau_n^{1/2}r) = 1 - P(||T_n^*(\beta_0)|| > c_0\tau_n^{1/2}r) \geq 1 - \frac{1}{c_0^2\tau_n r^2}\mathrm{E}||T_n^*(\beta_0)||^2.$$

We compute $\mathrm{E}||T_n^*(\beta_0)||^2$ separately:

$$\mathrm{E}||T_n^*(\beta_0)||^2 = \mathrm{E}||\mathbf{H}_n^{-1/2}g_n||^2 = \mathrm{E}[\mathrm{tr}(\mathbf{H}_n^{-1/2}g_ng_n^T\mathbf{H}_n^{-1/2})]$$

$$= \text{trE}[\mathbf{H}_n^{-1/2} g_n g_n^T \mathbf{H}_n^{-1/2}] = \text{tr}(\mathbf{H}_n^{-1/2} \text{E}(g_n g_n^T) \mathbf{H}_n^{-1/2})$$
$$= \text{tr}(\mathbf{H}_n^{-1/2} \mathbf{M}_n \mathbf{H}_n^{-1/2}) \leq \tau_n tr(\mathbf{I}) = \tau_n p,$$

where for the last line we used the fact that $\mathbf{M}_n \leq \tau_n \mathbf{H}_n$. Therefore,

$$P(||T_n^*(\beta_0)|| \leq c_0 \tau_n^{1/2} r) \geq 1 - \frac{1}{c_0^2 \tau_n r^2} \tau_n p = 1 - \frac{\varepsilon}{2}. \tag{2.2.11}$$

From (2.2.9), (2.2.10), (2.2.11), it follows that

$$P(E_n^*(r) \cap \Omega_n^*(r)) \geq 1 - \varepsilon \tag{2.2.12}$$

for all $n \geq N_\varepsilon^*$.

b) Let $\delta > 0$ be arbitrary. We will prove that $P(\|\widehat{\beta}_n - \beta\| \leq \delta) \to 1$. For this, let $\varepsilon > 0$ be arbitrary, and $r = r_\varepsilon$ as in part a). On the event $\Omega_n^*(r_\varepsilon) \cap E_n^*(r_\varepsilon)$,

$$\|\widehat{\beta}_n - \beta_0\| \leq \|\mathbf{H}_n^{-1/2}\| \cdot \|\mathbf{H}_n^{1/2}(\widehat{\beta}_n - \beta_0)\| \leq \left[\frac{\tau_n}{\lambda_{\min}(\mathbf{H}_n)}\right]^{1/2} r$$

since $\|\mathbf{H}_n^{-1/2}\|^2 = \lambda_{\max}(\mathbf{H}_n^{-1/2}\mathbf{H}_n^{-1/2}) = \lambda_{\max}(\mathbf{H}_n^{-1}) = \frac{1}{\lambda_{\min}(\mathbf{H}_n)}$.

By condition (I_w^*), there exists $N_\delta^* \geq 1$ such that

$$\frac{\lambda_{\min}(\mathbf{H}_n)}{\tau_n} \geq \frac{\delta^2}{r}$$

for all $n \geq N_\delta^*$.

Hence, for all $n \geq N_\delta^*$, $\Omega_n^*(r_\varepsilon) \cap E_n^*(r_\epsilon) \subset \{\|\widehat{\beta}_n - \beta_0\| \leq \delta\}$. We take $N_{\delta,\varepsilon} = \max(N_\varepsilon, N_\delta^*)$, where N_ε is the same as in part a). Hence, by (2.2.12),

$$P(||\widehat{\beta}_n - \beta_0|| \leq \delta) \geq P(\Omega_n^*(r_\varepsilon) \cap E_n^*(r_\varepsilon)) \geq 1 - \varepsilon$$

for all $n \geq N_{\varepsilon,\delta}$. ∎

2.3 Asymptotic normality

In this section, we show that the estimator $\widehat{\beta}_n$ (defined in the previous section) is asymptotically normal, under some conditions. We consider the following condition:

(CC) For any $r > 0$ and $\delta > 0$,

$$P\Big(\sup_{||x||=1} \sup_{||y||=1} \sup_{\beta \in B_n^*(r)} |\mathbf{x}^T \mathbf{H}_n^{-1/2} \mathcal{D}_n(\beta) \mathbf{H}_n^{-1/2} \mathbf{y} - \mathbf{x}^T \mathbf{y}| \leq \delta\Big) \to 1.$$

Lemma 2.3.1. *Condition* (CC) *implies condition* (L_w^*).

Proof: We denote by $\Omega_n(\delta, r)$ the event in condition (CC). Choose $\delta \in (0,1)$ arbitrary. In particular, on the event $\Omega_n(\delta, r)$, for any $\beta \in B_n^*(r)$ and for any $\mathbf{x} \in \mathbb{R}^p$ with $\|\mathbf{x}\| = 1$,

$$|\mathbf{x}^T \mathbf{H}_n^{-1/2} \mathcal{D}_n(\beta) \mathbf{H}_n^{-1/2} \mathbf{x} - 1| \leq \delta,$$

and hence, $\mathbf{x}^T \mathbf{H}_n^{-1/2} \mathcal{D}_n(\beta) \mathbf{H}_n^{-1/2} \mathbf{x} \geq 1 - \delta =: c_0$. Therefore

$$\Omega_n(\delta, r) \subset \Omega_n^*(r), \tag{2.3.1}$$

where $\Omega_n^*(r) = \{\mathbf{x}^T \mathbf{H}_n^{-1/2} \mathcal{D}_n(\beta) \mathbf{H}_n^{-1/2} \mathbf{x} \geq c_0$ for all $\beta \in B_n^*(r)$ and $\mathbf{x} \in \mathbb{R}^p$ with $\|\mathbf{x}\| = 1\}$. Since $P(\Omega_n(\delta, r)) \to 1$, it follows that $P(\Omega_n^*(r)) \to 1$. ∎

Let $c_n = \lambda_{\max}(\mathbf{M}_n^{-1} \mathbf{H}_n)$. We consider the following boundedness condition:

(B) There exists $c > 0$ such that $\tau_n c_n \leq c$ for all n.

Theorem 2.3.2. *Under conditions* (I_w^*), (D_w^*), (CC) *and* (B),

$$\boldsymbol{M}_n^{-1/2} \boldsymbol{g}_n = \boldsymbol{M}_n^{-1/2} \boldsymbol{H}_n (\widehat{\beta}_n - \beta_0) + o_p(1).$$

Proof: We want to prove that $\mathbf{M}_n^{-1/2} \mathbf{g}_n - \mathbf{M}_n^{-1/2} \mathbf{H}_n (\widehat{\beta}_n - \beta_0) \xrightarrow{p} 0$, i.e. for any $\eta > 0$ and $\varepsilon > 0$, there exists $N_{\eta,\varepsilon} \geq 1$ such that for all $n \geq N_{\eta,\varepsilon}$,

$$P(||\mathbf{M}_n^{-1/2} \mathbf{g}_n - \mathbf{M}_n^{-1/2} \mathbf{H}_n (\widehat{\beta}_n - \beta_0)|| \leq \eta) \geq 1 - \varepsilon.$$

Let $\eta > 0$ and $\varepsilon > 0$ be arbitrary. Let $\delta = 1 - c_0$ and $r = \frac{1}{c_0}\sqrt{\frac{2p}{\varepsilon}}$, where $c_0 \in (0,1)$ is a constant which we will chose later. Let $E_n^*(r)$ and $\Omega_n^*(r)$ be the same events as in the proof of Theorem 2.2.3. By (2.3.1) and (2.2.7),

$$E_n^*(r) \cap \Omega_n^*(\delta, r) \subset E_n^*(r) \cap \Omega_n^*(r) \subset \widetilde{\Omega}_n(r).$$

Using Taylor's formula, on the event $E_n^*(r) \cap \Omega_n^*(\delta, r)$, there exists $\overline{\beta} \in B_n^*(r)$ such that

$$\mathbf{g}_n(\widehat{\beta}_n) - \mathbf{g}_n = \frac{\partial \mathbf{g}_n}{\partial \beta^T}(\overline{\beta}_n)(\widehat{\beta}_n - \beta_0).$$

Since $\mathbf{g}_n(\widehat{\beta}_n) = 0$, using the definition of $\mathcal{D}_n$, we obtain

$$\mathbf{g}_n = \mathcal{D}_n(\overline{\beta}_n)(\widehat{\beta}_n - \beta_0).$$

We multiply this identity by the matrix $\mathbf{M}_n^{-1/2}$. We denote $\mathbf{U}_n(\beta) = \mathbf{H}_n^{-1/2}\mathcal{D}_n(\beta)\mathbf{H}_n^{-1/2} - \mathbf{I}$. We obtain

$$\begin{aligned}\mathbf{M}_n^{-1/2}\mathbf{g}_n &= \mathbf{M}_n^{-1/2}\mathcal{D}_n(\overline{\beta}_n)(\widehat{\beta}_n - \beta_0) = \mathbf{M}_n^{-1/2}\mathbf{H}_n^{1/2}\mathbf{H}_n^{-1/2}\mathcal{D}_n(\overline{\beta}_n)\mathbf{H}_n^{-1/2}\mathbf{H}_n^{1/2}(\widehat{\beta}_n - \beta_0)\\ &= \mathbf{M}_n^{-1/2}\mathbf{H}_n^{1/2}\mathbf{H}_n^{1/2}\mathbf{U}_n(\beta)\mathbf{H}_n^{1/2}(\widehat{\beta}_n - \beta_0) + \mathbf{M}_n^{-1/2}\mathbf{H}_n(\widehat{\beta}_n - \beta_0).\end{aligned}$$

Note that

$$\begin{aligned}\|\mathbf{M}_n^{-1/2}\mathbf{H}_n^{1/2}\|^2 &= \lambda_{\max}(\mathbf{H}_n^{1/2}\mathbf{M}_n^{-1}\mathbf{H}_n^{1/2}) = \frac{1}{\lambda_{\min}(\mathbf{H}_n^{-1/2}\mathbf{M}_n\mathbf{H}_n^{-1/2})}\\ &= \frac{1}{\lambda_{\min}(\mathbf{H}_n^{-1}\mathbf{M}_n)} = \lambda_{\max}(\mathbf{M}_n^{-1}\mathbf{H}_n) = c_n.\end{aligned}$$

On the event $E_n^*(r) \cap \Omega_n(\delta, r)$, $\|\mathbf{U}_n(\beta)\| \leq c_1 p\delta$ for all $\beta \in B_n^*(r)$, and so by condition (B)

$$\begin{aligned}\|\mathbf{M}_n^{-1/2}\mathbf{g}_n - \mathbf{M}_n^{-1/2}\mathbf{H}_n(\widehat{\beta}_n - \beta)\| &\leq (c_n\tau_n)^{1/2}c_1 p\delta r \leq c^{1/2}c_1 p\delta r\\ &= c^{1/2}c_1 p\frac{1-c_0}{c_0}\sqrt{\frac{2p}{\varepsilon}} \leq \eta,\end{aligned}$$

if we choose $c_0 \in (0,1)$ (depending on η and ε) such that $\frac{1}{c_0} \leq 1 + \frac{\eta}{c^{1/2}c_1 p}\sqrt{\frac{\varepsilon}{2p}}$. This means that

$$E_n^*(r) \cap \Omega_n(\delta, r) \subset \{\|\mathbf{M}_n^{-1/2}\mathbf{g}_n - \mathbf{M}_n^{-1/2}\mathbf{H}_n(\widehat{\beta}_n - \beta_0)\| \leq \eta\}.$$

Similarly to (2.2.12), it can be shown that there exists $N_{\eta,\varepsilon} \geq 1$ such that

$$P(E_n^*(r) \cap \Omega_n(\delta, r)) \geq 1 - \varepsilon$$

for all $n \geq N_{\eta,\varepsilon}$. The conclusion follows.

∎

We define $\tilde{\mathbf{Y}}_i = (A_i^*)^{-1/2}\varepsilon_i^*$ and $\gamma_n^{(D)} = \max_{i \leq n} \gamma_{n,i}^{(D)}$, where

$$\gamma_{n,i}^{(D)} = \lambda_{\max}(\mathbf{H}_n^{-1/2}\mathbf{D}_i^T(\mathbf{V}_i^*)^{-1}\mathbf{D}_i\mathbf{H}_n^{-1/2}).$$

We assume that $\gamma_n^{(D)} \leq K_n^{(D)}$ where $K_n^{(D)}$ is not random. We consider the condition

(N_δ) There exist $\delta > 0$ and $K > 0$ such that $\mathrm{E}(\widetilde{Y}_{ij}^{2+2/\delta}|\mathbf{X}_i) \leq K$ for all $i = 1, \ldots, n,\ j = 1, \ldots, m,$ and $(c_n\widetilde{\lambda}_n)^{1+\delta}K_n^{(D)} \to 0$.

The following result is the analogue of Lemma 2 in [12] in our case.

Theorem 2.3.3. *Under condition* (N_δ)*, we have*

$$\boldsymbol{M}_n^{-1/2}\boldsymbol{g}_n \to N(0, \boldsymbol{I}).$$

Proof: By the Cramer-Wold theorem, it suffices to show that for any $\lambda \in \mathbb{R}^p$ with $\|\lambda\| = 1$, we have:

$$\lambda^T \mathbf{M}_n^{-1/2}\mathbf{g}_n \to N(0,1). \tag{2.3.2}$$

Note that $\lambda^T \mathbf{M}_n^{-1/2}\mathbf{g}_n = \sum_{i=1}^n Z_{n,i}$ where $Z_{n,i} = \lambda^T \mathbf{M}_n^{-1/2}\mathcal{D}_i^T(\mathbf{V}_i^*)^{-1}\varepsilon_i^*$. The variables $(Z_{n,i})_{i\leq n}$ are independent. By Lemma 2.1.4,

$$\mathrm{E}[Z_{n,i}] = \mathrm{E}[\lambda^T \mathbf{M}_n^{-1/2}(\mathbf{V}_i^*)^{-1}\mathrm{E}(\varepsilon_i^*|\mathbf{X}_i)] = 0.$$

Let $s_n^2 = \mathrm{Var}(\lambda^T \mathbf{M}_n^{-1/2}\mathbf{g}_n)$. Then, by Lemma 2.1.6,

$$\begin{aligned}
s_n^2 &= \sum_{i=1}^n \mathrm{E}(Z_{n,i}^2) = \sum_{i=1}^n \mathrm{E}[\mathrm{E}(Z_{n,i}^2|\mathbf{X}_i)] = \sum_{i=1}^n \mathrm{E}[\mathrm{E}(Z_{n,i}Z_{n,i}^T|\mathbf{X}_i)]\\
&= \sum_{i=1}^n \mathrm{E}[\lambda^T \mathbf{M}_n^{-1/2}\mathbf{D}_i^T(\mathbf{V}_i^*)^{-1}\mathrm{E}[\varepsilon_i^*(\varepsilon_i^*)^T](\mathbf{V}_i^*)^{-1}\mathbf{D}_i\mathbf{M}_n^{-1/2}\lambda]\\
&= \sum_{i=1}^n \mathrm{E}[\lambda^T \mathbf{M}_n^{-1/2}\mathbf{D}_i^T(\mathbf{V}_i^*)^{-1}\Sigma_i^*(\mathbf{V}_i^*)^{-1}\mathbf{D}_i\mathbf{M}_n^{-1/2}\lambda]\\
&= \mathrm{E}[\lambda^T \mathbf{M}_n^{-1/2}\Big(\sum_{i=1}^n \mathbf{D}_i^T(\mathbf{V}_i^*)^{-1}\Sigma_i^*(\mathbf{V}_i^*)^{-1}\mathbf{D}_i\Big)\mathbf{M}_n^{-1/2}\lambda]\\
&= \mathrm{E}[\lambda^T \mathbf{M}_n^{-1/2}\mathbf{M}_n^*\mathbf{M}_n^{-1/2}\lambda]\\
&= \lambda^T \mathbf{M}_n^{-1/2}\mathrm{E}[\mathbf{M}_n^*]\mathbf{M}_n^{-1/2}\lambda\\
&= \lambda^T \mathbf{M}_n^{-1/2}\mathbf{M}_n\mathbf{M}_n^{-1/2}\lambda\\
&= 1.
\end{aligned}$$

By the Central Limit Theorem for triangular arrays (Theorem 27.2 of [3]), relation (2.3.2) will follow, once we prove that the following Lindeberg condition holds:

$$\lim_{n\to\infty} \sum_{i=1}^n \mathrm{E}(Z_{n,i}^2 \mathrm{I}\|Z_{n,i}\| \geq \varepsilon) = 0 \tag{2.3.3}$$

for any $\varepsilon > 0$. We prove that (2.3.3) holds. Let $\varepsilon > 0$ be arbitrary. Using the Cauchy-Schwartz inequality $\mathbf{x}^T\mathbf{y} \leq \|\mathbf{x}\| \cdot \|\mathbf{y}\|$ for any p-dimensional vectors $\mathbf{x}$ and $\mathbf{y}$,

$$Z_{n,i} \leq \|\lambda^T \mathbf{M}_n^{-1/2}\mathbf{D}_i(\mathbf{V}_i^*)^{-1/2}\|^2 \cdot \|(\mathbf{V}_i^*)^{-1/2}\varepsilon_i^*\|^2$$

$$= (\lambda^T \mathbf{M}_n^{-1/2}\mathbf{D}_i^T(\mathbf{V}_i^*)^{-1}\mathbf{D}_i\mathbf{M}_n^{-1/2}\lambda)\cdot((\varepsilon_i^*)^T(\mathbf{V}_i^*)^{-1}\varepsilon_i^*)$$
$$= \overline{\gamma}_{n,i}(\varepsilon_i^*)^T(\mathbf{A}_i^*)^{-1/2}\mathbf{R}_i^{-1}(\alpha)(\mathbf{A}_i^*)^{-1/2}\varepsilon_i^*,$$

where $\overline{\gamma}_{n,i} = \lambda\mathbf{M}_n^{-1/2}\mathbf{D}_i^T(\mathbf{V}_i^*)^{-1}\mathbf{D}_i\mathbf{M}_n^{-1/2}\lambda$. Using Theorem A.1.5, (Appendix A.1), for any $i \le n$

$$(\varepsilon_i^*)^T(\mathbf{A}_i^*)^{-1/2}\mathbf{R}_i^{-1}(\alpha)(\mathbf{A}_i^*)^{-1/2}\varepsilon_i^* \le \lambda_{\max}(\mathbf{R}_i^{-1}(\alpha))\|\widetilde{\mathbf{Y}}_i\|^2 \le \widetilde{\lambda}_n\|\widetilde{\mathbf{Y}}_i\|^2.$$

We obtain

$$Z_{n,i}^2 \le \overline{\gamma}_{n,i}\widetilde{\lambda}_n\|\widetilde{\mathbf{Y}}_i\|^2, \tag{2.3.4}$$

for any $i \le n$. We also need another upper bound for $Z_{n,i}^2$, which is obtained as follows. Again, by the same Theorem A.1.5 (Appendix A.1),

$$\begin{aligned}\overline{\gamma}_{n,i} &= \lambda^T\mathbf{M}_n^{-1/2}\mathbf{H}_n^{1/2}(\mathbf{H}_n^{-1/2}\mathbf{D}_i^T(\mathbf{V}_i^*)^{-1}\mathbf{D}_i\mathbf{H}_n^{-1/2})\mathbf{H}_n^{1/2}\mathbf{M}_n^{-1/2}\lambda\\ &\le \lambda_{\max}(\mathbf{H}_n^{-1/2}\mathbf{D}_i^T(\mathbf{V}_i^*)^{-1}\mathbf{D}_i\mathbf{D}_i\mathbf{H}_n^{-1/2})\cdot\lambda^T\mathbf{M}_n^{-1/2}\mathbf{H}_n\mathbf{M}_n^{-1/2}\lambda\\ &\le \gamma_{n,i}^{(D)}\lambda_{\max}(\mathbf{M}_n^{-1/2}\mathbf{H}_n\mathbf{M}_n^{-1/2}).\end{aligned}$$

Recalling that $c_n = \lambda_{\max}(\mathbf{M}_n^{-1}\mathbf{H}_n) = \lambda_{\max}(\mathbf{M}_n^{-1/2}\mathbf{H}_n\mathbf{M}_n^{-1/2})$ (using Theorem A.1.3), we obtain that $\overline{\gamma}_{n,i} \le \gamma_{n,i}^{(D)}c_n$, and hence

$$Z_{n,i}^2 \le c_n\widetilde{\lambda}_n\gamma_n^{(D)}\|\widetilde{\mathbf{Y}}_i\|^2, \tag{2.3.5}$$

for any $i \le n$. Coming back to (2.3.3), and using (2.3.4) and (2.3.5), we obtain:

$$\begin{aligned}\sum_{i=1}^n \mathrm{E}[Z_{n,i}^2\mathrm{I}\{Z_{n,i}\ge\varepsilon\}] &\le \widetilde{\lambda}_n\sum_{i=1}^n\mathrm{E}\Big[\overline{\gamma}_{n,i}\ \mathrm{E}\Big[\|\widetilde{\mathbf{Y}}_i\|^2\mathrm{I}\left\{\|\widetilde{\mathbf{Y}}_i\|^2\ge\frac{\varepsilon^2}{c_n\widetilde{\lambda}_n\gamma_n^{(D)}}\right\}|\mathbf{X}_i\Big]\Big]\\ &\le \widetilde{\lambda}_n m\sum_{i=1}^n\mathrm{E}\Big[\overline{\gamma}_{n,i}\mathrm{E}\Big[\frac{\|\widetilde{\mathbf{Y}}_i\|^2}{m}\frac{(\|\widetilde{\mathbf{Y}}_i\|^2/m)^{1/\delta}}{(\varepsilon^2/(mc_n\widetilde{\lambda}_n\gamma_n^{(D)}))^{1/\delta}}I\left\{\|\widetilde{\mathbf{Y}}_i\|^2\ge\frac{\varepsilon^2}{c_n\widetilde{\lambda}_n\gamma_n^{(D)}}\right\}|\mathbf{X}_i\Big]\Big]\\ &\le \widetilde{\lambda}_n m\Big(\frac{mc_n\widetilde{\lambda}_nK_n^{(D)}}{\varepsilon^2}\Big)^{1/\delta}\sum_{i=1}^n\mathrm{E}\Big[\overline{\gamma}_{n,i}\mathrm{E}\Big[\Big(\frac{\|\widetilde{\mathbf{Y}}_i\|^2}{m}\Big)^{1+1/(\delta)}|\mathbf{X}_i\Big]\Big].\end{aligned}$$

Since the function $\phi(t) = t^{1+1/\delta}$ is convex,

$$\Big(\frac{1}{m}\|\widetilde{\mathbf{Y}}_i\|^2\Big)^{1+1/\delta} = \Big(\frac{1}{m}\sum_{j=1}^m\widetilde{Y}_{ij}^2\Big)^{1+1/\delta} \le \frac{1}{m}\sum_{j=1}^m\widetilde{Y}_{ij}^{2+2/\delta},$$

and hence, by condition (N_δ),

$$\mathrm{E}\Big[\Big(\frac{\|\widetilde{\mathbf{Y}}_i\|^2}{m}\Big)^{1+1/\delta}|\mathbf{X}_i\Big] \le \frac{1}{m}\sum_{j=1}^m\mathrm{E}(\widetilde{Y}_{ij}^{2+2/\delta}|\mathbf{X}_i) \le K.$$

Therefore

$$\sum_{i=1}^{n} \mathrm{E}[Z_{n,i}^2 I\{Z_{n,i} \geq \varepsilon\}] \leq \widetilde{\lambda}_n m \Big(\frac{m c_n \widetilde{\lambda}_n K_n^{(D)}}{\varepsilon^2}\Big)^{1/\delta} K \mathrm{E}\Big[\sum_{i=1}^{n} \overline{\gamma}_{n,i}\Big]. \tag{2.3.6}$$

Note that, by the definition of $\overline{\gamma}_{n,i}$,

$$\begin{aligned}\sum_{i=1}^{n} \overline{\gamma}_{n,i} &= \lambda^T \mathbf{M}_n^{-1/2} \Big(\sum_{i=1}^{n} \mathbf{D}_i^T (\mathbf{V}_i^*)^{-1} \mathbf{D}_i^T\Big) \mathbf{M}_n^{-1/2} \lambda \\ &= \lambda^T \mathbf{M}_n^{-1/2} \mathbf{H}_n^* \mathbf{M}_n^{-1/2} \lambda.\end{aligned}$$

Taking expectation on both sides of the previous equality, we obtain:

$$\begin{aligned}\mathrm{E}\Big[\sum_{i=1}^{n} \overline{\gamma}_{n,i}\Big] &= \lambda^T \mathbf{M}_n^{-1/2} \mathrm{E}(\mathbf{H}_n^*) \mathbf{M}_n^{-1/2} \lambda \\ &= \lambda^T \mathbf{M}_n^{-1/2} \mathbf{H}_n \mathbf{M}_n^{-1/2} \lambda \\ &\leq \lambda_{\max}(\mathbf{M}_n^{-1/2} \mathbf{H}_n \mathbf{M}_n^{-1/2}) \\ &= c_n.\end{aligned}$$

Introducing this in (2.3.6), we obtain:

$$\begin{aligned}\sum_{i=1}^{n} \mathrm{E}[Z_{n,i}^2 I Z_{n,i} \geq \varepsilon] &\leq \widetilde{\lambda}_n m \Big(\frac{m c_n \widetilde{\lambda}_n K_n^{(D)}}{\varepsilon^2}\Big)^{1/\delta} K c_n \\ &= m\Big(\frac{m}{\varepsilon^2}\Big)^{1/\delta} (c_n \widetilde{\lambda}_n)^{1+1/\delta} (K_n^{(D)})^{1/\delta} \to 0,\end{aligned}$$

by condition (N_δ). This finishes the proof. ∎

Corollary 2.3.4. *Under conditions* (I_w^*), (D_w^*), (CC), (C) *and* (N_δ),

$$\boldsymbol{M}_n^{-1/2} \boldsymbol{H}_n (\widehat{\beta}_n - \beta) \to \mathcal{N}(0, \boldsymbol{I}).$$

Remark 2.3.5. In practice, we replace the matrices $\mathbf{M}_n$ and $\mathbf{H}_n$ by

$$\begin{aligned}\widehat{\mathbf{M}}_n &= \sum_{i=1}^{n} \mathbf{D}_i(\widehat{\beta}_n)^T \mathbf{V}_i^*(\widehat{\beta}_n, \alpha)^{-1} \widehat{\Sigma}_i^*(\widehat{\beta}_n) \mathbf{V}_i^*(\widehat{\beta}_n, \alpha)^{-1} \mathbf{D}_i(\widehat{\beta}_n) \\ \widehat{\mathbf{H}}_n &= \sum_{i=1}^{n} \mathbf{D}_i(\widehat{\beta}_n)^T \mathbf{V}_i^*(\widehat{\beta}_n, \alpha)^{-1} \mathbf{D}_i(\widehat{\beta}_n),\end{aligned}$$

where $\widehat{\Sigma}_i^*(\beta) = (\mathbf{Y}_i^* - \mu_i(\beta))(\mathbf{Y}_i^* - \mu_i(\beta))^T$. Note that the weighted response Y_{ij}^* depends on the missingness probability π_{ij}, which is unknown. Moreover, the matrix $\mathbf{V}_i^*(\widehat{\beta}_n, \alpha)$ depends on $\mathbf{A}_i^*(\beta)$ (see (2.1.10)), which in turn depends on the probabilities $(\pi_{ij})_{1\leq j\leq m}$, which are unknown (see (2.1.9)). To circumvent this problem, we will use the method suggested by Yi, Ma, Carroll in [14]. This consists in fitting a logistic regression model to the data consisting of $(\mathbf{I}_i, \mathbf{X}_i)$ for $i = 1, \ldots, n$, with a new regression parameter γ. That is, we assume that I_{ij} is a Bernoulli random variable with mean

$$\pi_{ij} = \pi_{ij}(\gamma) = \frac{\exp(\mathbf{X}_{ij}^T\gamma)}{1 + \exp(\mathbf{X}_{ij}^T\gamma)} \tag{2.3.7}$$

(as in Example 2.1.3). To estimate γ, we solve the classical GEE with working independence matrices $\mathbf{R}_i(\alpha) = \mathbf{I}$ for all $i = 1, \ldots, n$:

$$\sum_{i=1}^{n}\sum_{j=1}^{m} \mathbf{X}_{ij}\Big(Y_{ij} - \frac{\exp(\mathbf{X}_{ij}^T\gamma)}{1 + \exp(\mathbf{X}_{ij}^T\gamma)}\Big) = 0 \tag{2.3.8}$$

(see equation (2.1.2)). In the calculation of the matrix $\mathbf{A}_i^*(\beta)$, we replace π_{ij} by $\widehat{\pi}_{ij}(\widehat{\gamma})$, where $\widehat{\gamma}$ is the solution of the equation (2.3.8).

2.4 Verification of condition (CC)

In this section, we give some sufficient conditions which ensure that condition (CC) holds. Proceeding as in Remark 1 of [12] (see also Appendix A of [12]), we write the derivative of $\mathbf{g}_n(\beta)$ as the sum of three terms:

$$-\mathcal{D}_n(\beta) = \frac{\partial}{\partial \beta^T}\mathbf{g}_n(\beta) = -\mathbf{H}_n^*(\beta) + \mathbf{B}_n(\beta) + \mathcal{E}_n(\beta),$$

where $\mathbf{H}_n^*(\beta)$ is given by (2.1.16), $\mathbf{B}_n(\beta) = \mathbf{B}_n^{(1)}(\beta) + \mathbf{B}_n^{(2)}(\beta)$, $\mathcal{E}_n(\beta) = \mathcal{E}_n^{(1)}(\beta) + \mathcal{E}_n^{(2)}(\beta)$ and

$$\begin{aligned}
\mathbf{B}_n^{(1)}(\beta) &= \sum_{i=1}^{n} \mathbf{X}_i^T \mathrm{diag}\big[\mathbf{R}_i(\alpha)^{-1}\mathbf{A}_i^*(\beta)^{-1/2}(\mu_i - \mu_i(\beta))\big]\mathbf{G}_i^{(1)}(\beta)\mathbf{X}_i \\
\mathbf{B}_n^{(2)}(\beta) &= \sum_{i=1}^{n} \mathbf{X}_i^T \mathbf{F}_i(\beta)\mathbf{R}_i(\alpha)^{-1}\mathrm{diag}\big[\mu_i - \mu_i(\beta)\big]\mathbf{G}_i^{(2)}(\beta)\mathbf{X}_i \\
\mathcal{E}_n^{(1)}(\beta) &= \sum_{i=1}^{n} \mathbf{X}_i^T \mathrm{diag}\big[\mathbf{R}_i(\alpha)^{-1}\mathbf{A}_i^*(\beta)^{-1/2}\varepsilon_i^*\big]\mathbf{G}_i^{(1)}(\beta)\mathbf{X}_i \\
\mathcal{E}_n^{(2)}(\beta) &= \sum_{i=1}^{n} \mathbf{X}_i^T \mathbf{F}_i(\beta)\mathbf{R}_i(\alpha)^{-1}\mathrm{diag}\big[\varepsilon_i^*\big]\mathbf{G}_i^{(2)}(\beta)\mathbf{X}_i.
\end{aligned}$$

Here $\mathbf{F}_i(\beta)$ is the diagonal matrix with entries $f_{ij}, 1 \leq j \leq m$, given by (2.1.13), and $\mathbf{G}_i^{(k)}(\beta) = \text{diag}\big(g_{i1}^{(k)}(\beta), \ldots, g_{im}^{(k)}(\beta)\big)$ for $k = 1, 2$, where

$$\frac{\partial}{\partial \beta^T} f_{ij}(\beta) = g_{ij}^{(1)}(\beta)\mathbf{X}_{ij}^T \quad \text{and} \quad \frac{\partial}{\partial \beta^T}[\sigma_{i,jj}^*(\beta)]^{-1/2} = g_{ij}^{(2)}(\beta)\mathbf{X}_{ij}^T,$$

with functions $g_{ij}^{(1)}(\beta)$ and $g_{ij}^{(2)}(\beta)$ given by:

$$g_{ij}^{(1)}(\beta) = \frac{\mu''(\mathbf{X}_{ij}^T\beta)}{[\sigma_{i,jj}^*(\beta)]^{1/2}} - \frac{2\left(\frac{1}{\pi_{ij}} - 1\right)\mu(\mathbf{X}_{ij}^T\beta)(\mu'(\mathbf{X}_{ij}^T\beta))^2 + \frac{1}{\pi_{ij}}\mu'(\mathbf{X}_{ij}^T\beta)\mu''(\mathbf{X}_{ij}^T\beta)}{2[\sigma_{i,jj}^*(\beta)]^{3/2}} \tag{2.4.1}$$

$$g_{ij}^{(2)}(\beta) = -\frac{2\left(\frac{1}{\pi_{ij}} - 1\right)\mu(\mathbf{X}_{ij}^T\beta)\mu'(\mathbf{X}_{ij}^T\beta) + \frac{1}{\pi_{ij}}\mu''(\mathbf{X}_{ij}^T\beta)}{2[\sigma_{i,jj}^*(\beta)]^{3/2}} \tag{2.4.2}$$

We treat separately the three terms. For this, we introduce the same constants and smoothness assumption as on pages 330-331 of [12]:

$$\gamma_n^{(0)} = \max_{i \leq n} \max_{j \leq m}(\mathbf{X}_{ij}^T \mathbf{H}_n^{-1} \mathbf{X}_{ij}), \qquad \gamma_n^* = \tau_n \gamma_n^{(0)}, \qquad \pi_n = \frac{\max_{1 \leq i \leq n} \lambda_{\max}(\mathbf{R}_i(\alpha)^{-1})}{\min_{1 \leq i \leq n} \lambda_{\min}(\mathbf{R}_i(\alpha)^{-1})}.$$

Assumption (AH). $k_n^{(i)} = O_p(1)$ for $i = 1, 2, 3$ where

$$k_n^{(0)} = \sup_{\beta \in B_n^*(r)} \max_{i,j} \frac{\mu'(\mathbf{X}_{ij}^T\beta)}{\mu(\mathbf{X}_{ij}^T\beta)} \qquad k_n^{(1)} = \sup_{\beta \in B_n^*(r)} \max_{i,j} \frac{\mu''(\mathbf{X}_{ij}^T\beta)}{\mu'(\mathbf{X}_{ij}^T\beta)}$$

We impose the following assumption on the missingness probabilities:

Assumption (M). $\rho_n = O_p(1)$, where

$$\rho_n = \max_{i \leq n} \max_{j \leq m} \frac{1}{\pi_{ij}}.$$

Assumption M says that for any $\varepsilon > 0$, there exists a constant $C_\varepsilon > 0$ and an integer $N_\varepsilon \geq 1$ such that for any $n \geq N_\varepsilon$, with probability greater than $1 - \varepsilon$, $\pi_{ij} \geq C_\varepsilon$ for all $i \leq n$ and $j \leq m$. Intuitively speaking, this means that the missingness probabilities π_{ij} are bounded away from 0. Note that the case when all probabilities π_{ij} are equal to 0 corresponds to the case when all the data is missing.

The following three lemmas are the counterparts of Lemmas A.1.(ii), A.2.(ii) and A.3.(ii) of [12], when the covariates are random and the responses are missing at random.

Lemma 2.4.1. *Suppose Assumptions (AH) and (M) hold. If* $\pi_n \gamma_n^* \xrightarrow{P} 0$ *then*

$$\sup_{\|\mathbf{x}\|=1} \sup_{\|\mathbf{y}\|=1} \sup_{\beta \in B_n^*(r)} |\mathbf{x}^T \mathbf{H}_n^{-1/2} \mathbf{H}_n^*(\beta) \mathbf{H}_n^{-1/2} \mathbf{y} - \mathbf{x}^T \mathbf{y}| \xrightarrow{P} 0.$$

Lemma 2.4.2. *Suppose Assumptions (AH) and (M) hold. If* $\pi_n^2 \gamma_n^* \xrightarrow{P} 0$ *then*

$$\sup_{\|\mathbf{x}\|=1} \sup_{\|\mathbf{y}\|=1} \sup_{\beta \in B_n^*(r)} |\mathbf{x}^T \mathbf{H}_n^{-1/2} \mathbf{B}_n(\beta) \mathbf{H}_n^{-1/2} \mathbf{y}| \xrightarrow{P} 0.$$

Lemma 2.4.3. *Suppose Assumptions (AH) and (M) hold. If* $\gamma_n^* \xrightarrow{P} 0$, $n\pi_n^2 \gamma_n^{(0)} = O_p(1)$ *and* $n\pi_n^2 \gamma_n^{(0)} E(\tau_n) \xrightarrow{P} 0$ *then*

$$\sup_{\|\mathbf{x}\|=1} \sup_{\|\mathbf{y}\|=1} \sup_{\beta \in B_n^*(r)} |\mathbf{x}^T \mathbf{H}_n^{-1/2} \mathcal{E}_n(\beta) \mathbf{H}_n^{-1/2} \mathbf{y}| \xrightarrow{P} 0.$$

Proof of Lemma 2.4.1: Writing $\mathbf{F}_i(\beta) = \mathbf{F}_i + (\mathbf{F}_i(\beta) - \mathbf{F}_i)$, we obtain that

$$\mathbf{x}^T \mathbf{H}_n^{-1/2} \mathbf{H}_n^*(\beta) \mathbf{H}_n^{-1/2} \mathbf{y} - \mathbf{x}^T \mathbf{y} = T_0(\mathbf{x}, \mathbf{y}) + \sum_{i=1}^{3} T_i(\beta, \mathbf{x}, \mathbf{y}),$$

where $T_0(\mathbf{x}, \mathbf{y}) = \mathbf{x}^T \mathbf{H}_n^{-1/2} \mathbf{H}_n^* \mathbf{H}_n^{-1/2} \mathbf{y} - \mathbf{x}^T \mathbf{y}$ and

$$T_1(\beta, \mathbf{x}, \mathbf{y}) = \sum_{i=1}^{n} \mathbf{x}^T \mathbf{H}_n^{-1/2} \mathbf{X}_i^T (\mathbf{F}_i(\beta) - \mathbf{F}_i) \mathbf{R}_i(\alpha)^{-1} (\mathbf{F}_i(\beta) - \mathbf{F}_i) \mathbf{X}_i \mathbf{H}_n^{-1/2} \mathbf{y} \quad (2.4.3)$$

$$T_2(\beta, \mathbf{x}, \mathbf{y}) = \sum_{i=1}^{n} \mathbf{x}^T \mathbf{H}_n^{-1/2} \mathbf{X}_i^T (\mathbf{F}_i(\beta) - \mathbf{F}_i) \mathbf{R}_i(\alpha)^{-1} \mathbf{F}_i \mathbf{X}_i \mathbf{H}_n^{-1/2} \mathbf{y}$$

$$T_3(\beta, \mathbf{x}, \mathbf{y}) = \sum_{i=1}^{n} \mathbf{x}^T \mathbf{H}_n^{-1/2} \mathbf{X}_i^T \mathbf{F}_i \mathbf{R}_i(\alpha)^{-1} (\mathbf{F}_i(\beta) - \mathbf{F}_i) \mathbf{X}_i \mathbf{H}_n^{-1/2} \mathbf{y}.$$

To treat $T_0(\mathbf{x}, \mathbf{y})$, note that $\mathbf{H}_n^* = \sum_{i=1}^n \mathbf{U}_i$, where $\mathbf{U}_i = \mathbf{D}_i^T (\mathbf{V}_i^*)^{-1} \mathbf{D}_i, i = 1, \ldots, n$ are i.i.d. random matrices. By the strong law of large numbers, $\frac{1}{n}\mathbf{H}_n^* \to E(\mathbf{U}_1)$ a.s. (component-wise), and hence $\|\frac{1}{n}\mathbf{H}_n^* - E(\mathbf{U}_1)\| \to 0$ a.s. Since $\mathbf{H}_n = nE(\mathbf{U}_1)$, we obtain:

$$\|\mathbf{H}_n^{-1/2} \mathbf{H}_n^* \mathbf{H}_n^{-1/2} - \mathbf{I}\| \to 0 \quad \text{a.s.} \quad (2.4.4)$$

Therefore, $\sup_{\mathbf{x},\mathbf{y}} |T_0(\mathbf{x}, \mathbf{y})| \to 0$ a.s. Using inequality (A.1.1), we have:

$$\begin{aligned} \sup_{\mathbf{x},\mathbf{y},\beta} |T_1(\beta, \mathbf{x}, \mathbf{y})| &\le \pi_n \sup_{\beta} \max_{i \le n} \lambda_{\max}^2 (\mathbf{F}_i^{-1} \mathbf{F}_i(\beta) - \mathbf{I}) \cdot \sup_{\mathbf{x},\mathbf{y}} |\mathbf{x}^T \mathbf{H}_n^{-1/2} \mathbf{H}_n^* \mathbf{H}_n^{-1/2} \mathbf{y}| \\ &= \pi_n O_p(\gamma_n^*) O_p(1) = \pi_n \gamma_n^* O_p(1) = o_p(1), \end{aligned} \quad (2.4.5)$$

where the first equality above is due to Lemma A.2.4 (Appendix A.2) and relation (2.4.4).

To treat $T_2(\beta, \mathbf{x}, \mathbf{y})$, we use Cauchy-Schwatz inequality: for any p-dimensional vectors $(\mathbf{a}_i)_{i=1,\ldots,n}$ and $(\mathbf{b}_i)_{i=1,\ldots,n}$,

$$\left|\sum_{i=1}^{n} \mathbf{a}_i^T \mathbf{b}_i\right| \le \Big(\sum_{i=1}^{n} \mathbf{a}_i^T \mathbf{a}_i\Big)^{1/2} \Big(\sum_{i=1}^{n} \mathbf{b}_i^T \mathbf{b}_i\Big)^{1/2}. \tag{2.4.6}$$

Letting $\mathbf{a}_i = \mathbf{x}^T \mathbf{H}_n^{-1/2} \mathbf{X}_i^T \mathbf{F}_i \mathbf{R}_i(\alpha)^{-1/2}$ and

$$\mathbf{b}_i^T = \mathbf{R}_i(\alpha)^{1/2} (\mathbf{F}_i^{-1} \mathbf{F}_i(\beta) - \mathbf{I}) \mathbf{R}_i(\alpha)^{-1} \mathbf{F}_i \mathbf{X}_i \mathbf{H}_n^{-1/2} \mathbf{y},$$

we obtain $|T_2(\beta, \mathbf{x}, \mathbf{y})| \le T_2'(\mathbf{x})^{1/2} T_2''(\beta, \mathbf{y})^{1/2}$, with $T_2'(\mathbf{x}) = \mathbf{x} \mathbf{H}_n^{-1/2} \mathbf{H}_n^* \mathbf{H}_n^{-1/2} \mathbf{x}$ and

$$T_2''(\beta, \mathbf{y}) \le \pi_n \max_{i \le n} \lambda_{\max}^2 (\mathbf{F}_i^{-1} \mathbf{F}_i(\beta) - \mathbf{I}) \mathbf{y} \mathbf{H}_n^{-1/2} \mathbf{H}_n^* \mathbf{H}_n^{-1/2} \mathbf{y}.$$

Arguing as above, we get $\sup_{\beta, \mathbf{x}, \mathbf{y}} |T_2(\beta, \mathbf{x}, \mathbf{y})| = o_p(1)$. The term $T_3(\beta, \mathbf{x}, \mathbf{y})$ is similar. $\square$

Proof of Lemma 2.4.2: We begin by treating $\mathbf{B}_n^{(1)}(\beta)$. Note that for any $p \times p$ diagonal matrix Δ and for any p-dimensional vectors $\mathbf{v}$ and $\mathbf{w}$,

$$\text{diag}(\mathbf{v}) \Delta \mathbf{w} = \Delta \text{diag}(\mathbf{w}) \mathbf{v}. \tag{2.4.7}$$

We use this with $\mathbf{v} = \mathbf{R}_i(\alpha)^{-1} \mathbf{A}_i^*(\beta)^{-1/2} (\mu_i - \mu_i(\beta))$, $\Delta = \mathbf{G}_i^{(1)}(\beta)$ and $\mathbf{w} = \mathbf{X}_i \mathbf{H}_n^{-1/2} \mathbf{y}$. We obtain that $\mathbf{x}^T \mathbf{H}_n^{-1/2} \mathbf{B}_n^{(1)}(\beta) \mathbf{H}_n^{-1/2} \mathbf{y}$ is equal to

$$\sum_{i=1}^{n} \mathbf{x}^T \mathbf{H}_n^{-1/2} \mathbf{X}_i^T \mathbf{G}_i^{(1)}(\beta) \text{diag}(\mathbf{X}_i \mathbf{H}_n^{-1/2} \mathbf{y}) \mathbf{R}_i(\alpha)^{-1} \mathbf{A}_i^*(\beta)^{-1/2} (\mu_i - \mu_i(\beta)).$$

Using Cauchy-Schwarz inequality (2.4.6), it follows that

$$|\mathbf{x}^T \mathbf{H}_n^{-1/2} \mathbf{B}_n^{(1)}(\beta) \mathbf{H}_n^{-1/2} \mathbf{y}| \le S_1(\beta, \mathbf{x}, \mathbf{y})^{1/2} S_2(\beta)^{1/2}, \tag{2.4.8}$$

where

$$S_1(\beta, \mathbf{x}, \mathbf{y}) = \sum_{i=1}^{n} \mathbf{x}^T \mathbf{H}_n^{-1/2} \mathbf{X}_i^T \mathbf{G}_i^{(1)}(\beta) \text{diag}(\mathbf{X}_i \mathbf{H}_n^{-1/2} \mathbf{y}) \mathbf{R}_i(\alpha)^{-1} \text{diag}(\mathbf{X}_i \mathbf{H}_n^{-1/2} \mathbf{y}) \tag{2.4.9}$$

$$\mathbf{G}_i^{(1)}(\beta) \mathbf{X}_i \mathbf{H}_n^{-1/2} \mathbf{x}$$

$$S_2(\beta) = \sum_{i=1}^{n} (\mu_i - \mu_i(\beta))^T \mathbf{A}_i^*(\beta)^{-1/2} \mathbf{R}_i(\alpha)^{-1} \mathbf{A}_i^*(\beta)^{-1/2} (\mu_i - \mu_i(\beta)). \tag{2.4.10}$$

Using (A.1.1) and the fact that

$$\lambda^2_{\max}(\text{diag}(\mathbf{X}_i\mathbf{H}_n^{-1/2}\mathbf{y})) \le \gamma_n^{(0)}, \tag{2.4.11}$$

it follows that

$$\begin{aligned} S_1(\beta,\mathbf{x},\mathbf{y}) &\le \tilde{\lambda}_n\gamma_n^{(0)}\sum_{i=1}^n \mathbf{x}^T\mathbf{H}_n^{-1/2}\mathbf{X}_i^T[\mathbf{G}_i^{(1)}(\beta)]^2\mathbf{X}_i\mathbf{H}_n^{-1/2}\mathbf{x} \qquad (2.4.12)\\ &\le \tilde{\lambda}_n\gamma_n^{(0)}\max_{i\le n}\lambda_{\max}\big(\mathbf{R}_i(\alpha)^{1/2}\mathbf{F}_i^{-1}[\mathbf{G}_i^{(1)}(\beta)]^2\mathbf{F}_i^{-1}\mathbf{R}_i(\alpha)^{1/2}\big)\cdot\mathbf{x}^T\mathbf{H}_n^{-1/2}\mathbf{H}_n^*\mathbf{H}_n^{-1/2}\mathbf{x}\\ &\le \pi_n\gamma_n^{(0)}\max_{i\le n}\lambda^2_{\max}\big(\mathbf{F}_i^{-1}\mathbf{G}_i^{(1)}(\beta)\big)\cdot\mathbf{x}\mathbf{H}_n^{-1/2}\mathbf{H}_n^*\mathbf{H}_n^{-1/2}\mathbf{x}. \end{aligned}$$

By relations (A.2.5) and (A.2.6) (given in Appendix A.2), $\max_{i\le n}\lambda^2_{\max}\big(\mathbf{F}_i^{-1}\mathbf{G}_i^{(1)}(\beta)\big) = O_p(1)$. By Lemma 2.4.1, $\sup_{\|\mathbf{x}\|=1}\big(\mathbf{x}\mathbf{H}_n^{-1/2}\mathbf{H}_n^*\mathbf{H}_n^{-1/2}\mathbf{x}\big) = O_p(1)$. From this, we infer that

$$\sup_{\|\mathbf{x}\|=1}\sup_{\|\mathbf{y}\|=1}\sup_{\beta\in B_n^*(r)} S_1(\beta,\mathbf{x},\mathbf{y}) \le \pi_n\gamma_n^{(0)}O_p(1). \tag{2.4.13}$$

We now treat $S_2(\beta)$. By Taylor's formula, for any $\beta\in B_n^*(r)$, there exists $\overline{\beta}_{ij}\in B_n^*(r)$ such that $\mu_{ij}(\beta)-\mu_{ij}(\beta_0) = \mu'(\mathbf{X}_{ij}^T\overline{\beta}_{ij})\mathbf{X}_{ij}^T(\beta-\beta_0)$. Then $\mu_i(\beta)-\mu_i = \overline{\mathbf{A}}_i\mathbf{X}_i(\beta-\beta_0)$, where $\overline{\mathbf{A}}_i$ is the diagonal matrix with entries $\mu'(\mathbf{X}_{ij}^T\overline{\beta}_{ij}), j=1,\ldots,m$. Note that $\overline{\mathbf{A}}_i\mathbf{A}_i^*(\beta)^{-1/2} = \mathbf{A}_i^*(\beta)^{-1/2}\overline{\mathbf{A}}_i$ since $\overline{\mathbf{A}}_i$ and $\mathbf{A}_i^*(\beta)^{-1/2}$ are diagonal matrices. Using inequality (A.1.1), we get:

$$\begin{aligned} S_2(\beta) &= \sum_{i=1}^n(\beta-\beta_0)^T\mathbf{X}_i^T\mathbf{A}_i^*(\beta)^{-1/2}\overline{\mathbf{A}}_i\mathbf{R}_i(\alpha)^{-1}\overline{\mathbf{A}}_i\mathbf{A}_i^*(\beta)^{-1/2}\mathbf{X}_i(\beta-\beta_0)\\ &= \sum_{i=1}^n(\beta-\beta_0)^T\mathbf{X}_i^T\mathbf{F}_i(\beta)\mathbf{A}_i(\beta)^{-1}\overline{\mathbf{A}}_i\mathbf{R}_i(\alpha)^{-1}\overline{\mathbf{A}}_i\mathbf{A}_i(\beta)^{-1}\mathbf{F}_i(\beta)\mathbf{X}_i(\beta-\beta_0)\\ &\le \widetilde{\lambda}_n\max_{i\le n}\lambda^2_{\max}(\overline{\mathbf{A}}_i\mathbf{A}_i(\beta)^{-1})\sum_{i=1}^n(\beta-\beta_0)^T\mathbf{X}_i^T[\mathbf{F}_i(\beta)]^2\mathbf{X}_i(\beta-\beta_0)\\ &\le \pi_n\max_{i\le n}\lambda^2_{\max}(\overline{\mathbf{A}}_i\mathbf{A}_i(\beta)^{-1})\cdot(\beta-\beta_0)^T\mathbf{H}_n^*(\beta)(\beta-\beta_0)\\ &\le \pi_n\max_{i\le n}\lambda^2_{\max}(\overline{\mathbf{A}}_i\mathbf{A}_i(\beta)^{-1})\cdot\|\mathbf{H}_n^{-1/2}\mathbf{H}_n^*(\beta)\mathbf{H}_n^{-1/2}\|\cdot\|\mathbf{H}_n^{-1/2}(\beta-\beta_0)\|^2\\ &\le \pi_n\max_{i\le n}\lambda^2_{\max}(\overline{\mathbf{A}}_i\mathbf{A}_i(\beta)^{-1})\cdot\|\mathbf{H}_n^{-1/2}\mathbf{H}_n^*(\beta)\mathbf{H}_n^{-1/2}\|\cdot\tau_n r^2 \end{aligned}$$

By Lemma 2.4.1, $\sup_{\beta\in B_n^*(r)}\|\mathbf{H}_n^{-1/2}\mathbf{H}_n^*(\beta)\mathbf{H}_n^{-1/2}\| = O_p(1)$. Note that $\overline{\mathbf{A}}_i\mathbf{A}_i(\beta)^{-1}$ is a diagonal matrix with entries $\mu'(\mathbf{X}_{ij}^T\beta_{ij})/\mu'(\mathbf{X}_{ij}^T\beta), j=1,\ldots,m$. By relation (A.2.5) (given in Appendix A.2), it follows that $\sup_{\beta \in B_n^*(r)} \max_{i \le n} \lambda^2_{\max}(\overline{\mathbf{A}}_i \mathbf{A}_i(\beta)^{-1}) = O_p(1)$. Hence,

$$\sup_{\beta \in B_n^*(r)} S_2(\beta) \le \pi_n \tau_n O_p(1). \tag{2.4.14}$$

Using relations (2.4.8), (2.4.13) and (2.4.14), we infer that

$$\sup_{\|x\|=1} \sup_{\|y\|=1} \sup_{\beta \in B_n^*(r)} |\mathbf{x}^T \mathbf{H}_n^{-1/2} \mathbf{B}_n^{(1)}(\beta) \mathbf{H}_n^{-1/2} \mathbf{y}| \le \pi_n (\gamma_n^*)^{1/2} O_p(1) = o_p(1).$$

We continue with the treatment of $\mathbf{B}_n^{(2)}(\beta)$. Using relation (2.4.7), we see that

$$\mathbf{x}^T \mathbf{H}_n^{-1/2} \mathbf{B}_n^{(2)}(\beta) \mathbf{H}_n^{-1/2} \mathbf{y} = \sum_{i=1}^n \mathbf{x}^T \mathbf{H}_n^{-1/2} \mathbf{X}_i^T \mathbf{F}_i(\beta) \mathbf{R}_i(\alpha)^{-1} \text{diag}(\mathbf{X}_i \mathbf{H}_n^{-1/2} \mathbf{y}) \mathbf{G}_i^{(2)}(\beta)(\mu_i - \mu_i(\beta)).$$

We use Cauchy-Schwarz inequality (2.4.6) with $\mathbf{a}_i^T = \mathbf{x}^T \mathbf{H}_n^{-1/2} \mathbf{X}_i^T \mathbf{F}_i(\beta) \mathbf{R}_i(\alpha)^{-1}$ $\text{diag}(\mathbf{X}_i \mathbf{H}_n^{-1/2} \mathbf{y}) \mathbf{G}_i^{(2)}(\beta) \mathbf{A}_i^*(\beta)^{1/2} \mathbf{R}_i(\alpha)^{1/2}$ and $\mathbf{b}_i = \mathbf{R}_i(\alpha)^{-1/2} \mathbf{A}_i^*(\beta)^{-1/2}(\mu_i - \mu_i(\beta))$. We obtain:

$$|\mathbf{x}^T \mathbf{H}_n^{-1/2} \mathbf{B}_n^{(2)}(\beta) \mathbf{H}_n^{-1/2} \mathbf{y}| \le S_3(\beta, \mathbf{x}, \mathbf{y})^{1/2} S_2(\beta)^{1/2}, \tag{2.4.15}$$

where $S_2(\beta)$ is given by (2.4.10) and

$$S_3(\beta, \mathbf{x}, \mathbf{y}) = \sum_{i=1}^n \mathbf{x}^T \mathbf{H}_n^{-1/2} \mathbf{X}_i^T \mathbf{F}_i(\beta) \mathbf{R}_i(\alpha)^{-1} \text{diag}(\mathbf{X}_i \mathbf{H}_n^{-1/2} \mathbf{y}) \mathbf{G}_i^{(2)}(\beta) \mathbf{A}_i^*(\beta)^{1/2} \mathbf{R}_i(\alpha) \tag{2.4.16}$$

$$\mathbf{A}_i^*(\beta)^{1/2} \mathbf{G}_i^{(2)}(\beta) \text{diag}(\mathbf{X}_i \mathbf{H}_n^{-1/2} \mathbf{y}) \mathbf{R}_i(\alpha)^{-1} \mathbf{F}_i(\beta) \mathbf{X}_i \mathbf{H}_n^{-1/2} \mathbf{x}.$$

Using inequalities (A.1.1) and (2.4.11), we obtain that:

$$S_3(\beta, \mathbf{x}, \mathbf{y}) \le \pi_n \gamma_n^{(0)} \max_{i \le n} \lambda^2_{\max}(\mathbf{A}_i^*(\beta)^{1/2} \mathbf{G}_i^{(2)}(\beta)) \cdot \mathbf{x}^T \mathbf{H}_n^{-1/2} \mathbf{H}_n^*(\beta) \mathbf{H}_n^{-1/2} \mathbf{x}.$$

Note that $\mathbf{A}_i^*(\beta)^{1/2} \mathbf{G}_i^{(2)}(\beta)$ is a diagonal matrix with elements $\sqrt{\sigma^*_{i,jj}(\beta)} g_{ij}^{(2)}(\beta), j = 1, \ldots, m$. By Lemma A.2.3 (Appendix A.2), $\sup_{\beta \in B_n^*(r)} \max_{i \le n} \lambda^2_{\max}(\mathbf{A}_i^*(\beta)^{1/2} \mathbf{G}_i^{(2)}(\beta)) = O_p(1)$. Using Lemma 2.4.1, we obtain:

$$\sup_{\|x\|=1} \sup_{\|y\|=1} \sup_{\beta \in B_n^*(r)} S_3(\beta, \mathbf{x}, \mathbf{y}) \le \pi_n \gamma_n^{(0)} O_p(1). \tag{2.4.17}$$

Using relations (2.4.15), (2.4.17) and (2.4.14), we infer that:

$$\sup_{\|x\|=1} \sup_{\|y\|=1} \sup_{\beta \in B_n^*(r)} |\mathbf{x}^T \mathbf{H}_n^{-1/2} \mathbf{B}_n^{(2)}(\beta) \mathbf{H}_n^{-1/2} \mathbf{y}| \le \pi_n (\gamma_n^*)^{1/2} O_p(1) = o_p(1).$$

□

Proof of Lemma 2.4.3: We first treat the term $\mathcal{E}_n^{(1)}(\beta)$. Using relation (2.4.7), we see that

$$\mathbf{x}^T\mathbf{H}_n^{-1/2}\mathcal{E}_n^{(1)}(\beta)\mathbf{H}_n^{-1/2}\mathbf{y} = \sum_{i=1}^n \mathbf{x}^T\mathbf{H}_n^{-1/2}\mathbf{X}_i^T\mathbf{G}_i^{(1)}(\beta)\mathrm{diag}(\mathbf{X}_i\mathbf{H}_n^{-1/2}\mathbf{y})\mathbf{R}_i(\alpha)^{-1}\mathbf{A}_i^*(\beta)^{-1/2}\varepsilon_i^*$$
$$= U_1(\mathbf{x},\mathbf{y}) + U_3(\beta,\mathbf{x},\mathbf{y}) + U_5(\beta,\mathbf{x},\mathbf{y}),$$

where

$$U_1(\mathbf{x},\mathbf{y}) = \sum_{i=1}^n \mathbf{x}^T\mathbf{H}_n^{-1/2}\mathbf{X}_i^T\mathbf{G}_i^{(1)}\mathrm{diag}(\mathbf{X}_i\mathbf{H}_n^{-1/2}\mathbf{y})\mathbf{R}_i(\alpha)^{-1}(\mathbf{A}_i^*)^{-1/2}\varepsilon_i^*$$
$$U_3(\beta,\mathbf{x},\mathbf{y}) = \sum_{i=1}^n \mathbf{x}^T\mathbf{H}_n^{-1/2}\mathbf{X}_i^T\mathbf{G}_i^{(1)}(\beta)\mathrm{diag}(\mathbf{X}_i\mathbf{H}_n^{-1/2}\mathbf{y})\mathbf{R}_i(\alpha)^{-1}\big(\mathbf{A}_i^*(\beta)^{-1/2} - (\mathbf{A}_i^*)^{-1/2}\big)\varepsilon_i^*$$
$$U_5(\beta,\mathbf{x},\mathbf{y}) = \sum_{i=1}^n \mathbf{x}^T\mathbf{H}_n^{-1/2}\mathbf{X}_i^T\big(\mathbf{G}_i^{(1)}(\beta) - \mathbf{G}_i^{(1)}\big)\mathrm{diag}(\mathbf{X}_i\mathbf{H}_n^{-1/2}\mathbf{y})\mathbf{R}_i(\alpha)^{-1}(\mathbf{A}_i^*)^{-1/2}\varepsilon_i^*.$$

We first treat $U_1(\mathbf{x},\mathbf{y})$. By the Cauchy-Schwarz inequality (2.4.6),

$$|U_1(\mathbf{x},\mathbf{y})| \le S_1(\beta_0,\mathbf{x},\mathbf{y})^{1/2}U^{1/2}, \tag{2.4.18}$$

where $S_1(\beta,\mathbf{x},\mathbf{y})$ is given by (2.4.9) and $U = \sum_{i=1}^n W_i$, with $W_i = (\varepsilon_i^*)^T(\mathbf{A}_i^*)^{-1/2}\mathbf{R}_i(\alpha)^{-1}$ $(\mathbf{A}_i^*)^{-1/2}\varepsilon_i^*$. Using the fact that $\mathbf{x}^T\mathbf{x} = \mathrm{tr}(\mathbf{x}\mathbf{x}^T)$ for any p-dimensional vector $\mathbf{x}$, we obtain:

$$E(W_i) = E[E[\mathrm{tr}\{\mathbf{R}_i(\alpha)^{-1/2}(\mathbf{A}_i^*)^{-1/2}\varepsilon_i^*(\varepsilon_i^*)^T(\mathbf{A}_i^*)^{-1/2}\mathbf{R}_i(\alpha)^{-1/2}\}|\mathbf{X}_i]$$
$$= E[\mathrm{tr}\{\mathbf{R}_i(\alpha)^{-1/2}(\mathbf{A}_i^*)^{-1/2}E[\varepsilon_i^*(\varepsilon_i^*)^T|\mathbf{X}_i](\mathbf{A}_i^*)^{-1/2}\mathbf{R}_i(\alpha)^{-1/2}\}]$$
$$= E[\mathrm{tr}\{\mathbf{R}_i(\alpha)^{-1/2}\mathbf{R}_i^*\mathbf{R}_i(\alpha)^{-1/2}\}] \le mE[\lambda_{\max}(\mathbf{R}_i(\alpha)^{-1/2}\mathbf{R}_i^*\mathbf{R}_i(\alpha)^{-1/2})] \le mE(\tau_n),$$

for any $i = 1,\dots,n$, using (2.1.15) for the last equality. Hence, $\sum_{i=1}^n E(W_i) \le mnE(\tau_n)$. Since $\{(\mathbf{Y}_i,\mathbf{X}_i,\mathbf{I}_i)\}_{i=1,\dots,n}$ are i.i.d., $(W_i)_{i=1,\dots,n}$ are independent. Therefore, by Chebyshev's weak law of large numbers, $\sum_{i=1}^n(W_i - E(W_i)) = o_p(n)$. Hence,

$$U \le o_p(n) + mnE(\tau_n). \tag{2.4.19}$$

Using (2.4.18), (2.4.13), (2.4.19) and the hypotheses of the lemma, it follows that

$$\sup_{\|\mathbf{x}\|=1}\sup_{\|\mathbf{y}\|=1}|U_1(\mathbf{x},\mathbf{y})| \le \{\big(\pi_n\gamma_n^{(0)}o_p(n)\big)^{1/2} + \big(\pi_n\gamma_n^{(0)}nE(\tau_n)\big)^{1/2}\}O_p(1) = o_p(1).$$

Next, we treat $U_3(\beta,\mathbf{x},\mathbf{y})$. By the Cauchy-Schwartz inequality (2.4.6), it follows that

$$|U_3(\beta,\mathbf{x},\mathbf{y})| \le U_3'(\beta,\mathbf{x},\mathbf{y})^{1/2}U^{1/2}, \tag{2.4.20}$$

where U is the same as above and

$$U_3'(\beta,\mathbf{x},\mathbf{y}) = \sum_{i=1}^{n} \mathbf{x}^T\mathbf{H}_n^{-1/2}\mathbf{G}_i^{(1)}(\beta)\text{diag}(\mathbf{X}_i\mathbf{H}_n^{-1/2}\mathbf{y})\mathbf{R}_i(\alpha)^{-1}\big(\mathbf{A}_i^*(\beta)^{-1/2}(\mathbf{A}_i^*)^{1/2}-\mathbf{I}\big)\mathbf{R}_i(\alpha)$$
$$\big(\mathbf{A}_i^*(\beta)^{-1/2}(\mathbf{A}_i^*)^{1/2}-\mathbf{I}\big)\mathbf{R}_i(\alpha)^{-1}\text{diag}(\mathbf{X}_i\mathbf{H}_n^{-1/2}\mathbf{y})\mathbf{G}_i^{(1)}(\beta)\mathbf{H}_n^{-1/2}\mathbf{x}.$$

Using inequalities (A.1.1) and (2.4.11), we see that

$$U_3'(\beta,\mathbf{x},\mathbf{y}) \le \pi_n\widetilde{\lambda}_n\gamma_n^{(0)}\max_{i\le n}\lambda_{\max}^2(\mathbf{A}_i^*(\beta)^{-1/2}(\mathbf{A}_i^*)^{1/2}-\mathbf{I})\cdot$$
$$\cdot\sum_{i=1}^{n}\mathbf{x}^T\mathbf{H}_n^{-1/2}\mathbf{X}_i^T[\mathbf{G}_i^{(1)}(\beta)]^2\mathbf{X}_i\mathbf{H}_n^{-1/2}\mathbf{x}.$$

Proceeding as in (2.4.12) and using Lemma A.2.2 (Appendix A.2), we get:

$$\sup_{\|x\|=1}\sup_{\|y\|=1}\sup_{\beta\in B_n^*(r)} U_3'(\beta,\mathbf{x},\mathbf{y}) \le \pi_n^2\gamma_n^{(0)}O_p(1). \tag{2.4.21}$$

Using (2.4.20), (2.4.21) and (2.4.19), we obtain by the hypotheses of the lemma that

$$\sup_{\|x\|=1}\sup_{\|y\|=1}\sup_{\beta\in B_n^*(r)}|U_3(\beta,\mathbf{x},\mathbf{y})| \le \{\big(\pi_n^2\gamma_n^{(0)}o_p(n)\big)^{1/2}+\big(\pi_n^2\gamma_n^{(0)}nE(\tau_n)\big)^{1/2}\}O_p(1)=o_p(1).$$

We now treat $U_5(\beta,\mathbf{x},\mathbf{y})$. By the Cauchy-Schwartz inequality (2.4.6), it follows that

$$|U_5(\beta,\mathbf{x},\mathbf{y})| \le U_5'(\beta,\mathbf{x},\mathbf{y})^{1/2}U^{1/2}, \tag{2.4.22}$$

where U is the same as above and

$$U_5'(\beta,\mathbf{x},\mathbf{y}) = \sum_{i=1}^{n}\mathbf{x}^T\mathbf{H}_n^{-1/2}\big(\mathbf{G}_i^{(1)}(\beta)-\mathbf{G}_i^{(1)}\big)\text{diag}(\mathbf{X}_i\mathbf{H}_n^{-1/2}\mathbf{y})\mathbf{R}_i(\alpha)^{-1}\text{diag}(\mathbf{X}_i\mathbf{H}_n^{-1/2}\mathbf{y})$$
$$\big(\mathbf{G}_i^{(1)}(\beta)-\mathbf{G}_i^{(1)}\big)\mathbf{X}_i\mathbf{H}_n^{-1/2}\mathbf{x}.$$

Using inequalities (A.1.1) and (2.4.11), it follows that

$$U_5'(\beta,\mathbf{x},\mathbf{y}) \le \pi_n\gamma_n^{(0)}\max_{i\le n}\lambda_{\max}^2(\mathbf{F}_i^{-1}(\mathbf{G}_i^{(1)}(\beta)-\mathbf{G}_i^{(1)}))\cdot\mathbf{x}^T\mathbf{H}_n^{-1/2}\mathbf{H}_n^*\mathbf{H}_n^{-1/2}\mathbf{x}.$$

The matrix $\mathbf{F}_i^{-1}(\mathbf{G}_i^{(1)}(\beta)-\mathbf{G}_i^{(1)})$ has j-th element given by

$$\frac{g_{ij}^{(1)}(\beta)-g_{ij}^{(1)}(\beta_0)}{f_{ij}(\beta_0)} = \frac{g_{ij}^{(1)}(\beta)}{f_{ij}(\beta)}\cdot\frac{f_{ij}(\beta)}{f_{ij}(\beta_0)} - \frac{g_{ij}^{(1)}(\beta_0)}{f_{ij}(\beta_0)}.$$

By relation (A.2.6) (Appendix A.2), $\max_{i\le n}\lambda_{\max}^2(\mathbf{F}_i^{-1}(\mathbf{G}_i^{(1)}(\beta)-\mathbf{G}_i^{(1)}))=O_p(1)$. By Lemma 2.4.1,

$$\sup_{\|x\|=1}\sup_{\|y\|=1}\sup_{\beta\in B_n^*(r)} U_5'(\beta,\mathbf{x},\mathbf{y}) \le \pi_n\gamma_n^{(0)}O_p(1). \tag{2.4.23}$$

Using (2.4.22), (2.4.23) and (2.4.19), we obtain by the hypotheses of the lemma that

$$\sup_{\|x\|=1} \sup_{\|y\|=1} \sup_{\beta \in B_n^*(r)} |U_5(\beta, \mathbf{x}, \mathbf{y})| \leq \{\left(\pi_n \gamma_n^{(0)} o_p(n)\right)^{1/2} + \left(\pi_n \gamma_n^{(0)} n E(\tau_n)\right)^{1/2}\} O_p(1) = o_p(1).$$

We now treat $\mathcal{E}_n^{(2)}(\beta)$. Using (2.4.7), we see that

$$\begin{aligned}\mathbf{x}^T \mathbf{H}_n^{-1/2} \mathcal{E}_n^{(2)}(\beta) \mathbf{H}_n^{-1/2} \mathbf{y} &= \sum_{i=1}^n \mathbf{x}^T \mathbf{H}_n^{-1/2} \mathbf{X}_i^T \mathbf{F}_i(\beta) \mathbf{R}_i(\alpha)^{-1} \mathrm{diag}(\mathbf{X}_i \mathbf{H}_n^{-1/2} \mathbf{y}) \mathbf{G}_i^{(2)}(\beta) \varepsilon_i^* \\ &= U_2(\mathbf{x}, \mathbf{y}) + U_4(\beta, \mathbf{x}, \mathbf{y}) + U_6(\beta, \mathbf{x}, \mathbf{y}),\end{aligned}$$

where

$$\begin{aligned}U_2(\mathbf{x}, \mathbf{y}) &= \sum_{i=1}^n \mathbf{x}^T \mathbf{H}_n^{-1/2} \mathbf{X}_i^T \mathbf{F}_i \mathbf{R}_i(\alpha)^{-1} \mathrm{diag}(\mathbf{X}_i \mathbf{H}_n^{-1/2} \mathbf{y}) \mathbf{G}_i^{(2)} \varepsilon_i^* \\ U_4(\beta, \mathbf{x}, \mathbf{y}) &= \sum_{i=1}^n \mathbf{x}^T \mathbf{H}_n^{-1/2} \mathbf{X}_i^T (\mathbf{F}_i(\beta) - \mathbf{F}_i) \mathbf{R}_i(\alpha)^{-1} \mathrm{diag}(\mathbf{X}_i \mathbf{H}_n^{-1/2} \mathbf{y}) \mathbf{G}_i^{(2)}(\beta) \varepsilon_i^* \\ U_6(\beta, \mathbf{x}, \mathbf{y}) &= \sum_{i=1}^n \mathbf{x}^T \mathbf{H}_n^{-1/2} \mathbf{X}_i^T \mathbf{F}_i \mathbf{R}_i(\alpha)^{-1} \mathrm{diag}(\mathbf{X}_i \mathbf{H}_n^{-1/2} \mathbf{y}) (\mathbf{G}_i^{(2)}(\beta) - \mathbf{G}_i^{(2)}) \varepsilon_i^*.\end{aligned}$$

By the Cauchy-Schwarz inequality (2.4.6), $|U_2(\mathbf{x}, \mathbf{y})| \leq S_3(\beta_0, \mathbf{x}, \mathbf{y})^{1/2} U^{1/2}$, where $S_3(\beta, \mathbf{x}, \mathbf{y})$ is given by (2.4.16) and U is the same as above. Using (2.4.17) and (2.4.19), it follows that

$$\sup_{\|x\|=1} \sup_{\|y\|=1} |U_2(\mathbf{x}, \mathbf{y})| = o_p(1).$$

Similarly, $|U_4(\beta, \mathbf{x}, \mathbf{y})| \leq U_4'(\beta_0, \mathbf{x}, \mathbf{y})^{1/2} U^{1/2}$, where

$$\begin{aligned}U_4'(\beta, \mathbf{x}, \mathbf{y}) = \sum_{i=1}^n & \mathbf{x}^T \mathbf{H}_n^{-1/2} \mathbf{X}_i^T (\mathbf{F}_i(\beta) - \mathbf{F}_i) \mathbf{R}_i(\alpha)^{-1} \mathrm{diag}(\mathbf{X}_i \mathbf{H}_n^{-1/2} \mathbf{y}) \mathbf{G}_i^{(2)}(\beta) (\mathbf{A}_i^*)^{1/2} \mathbf{R}_i(\alpha) \\ & (\mathbf{A}_i^*)^{1/2} \mathbf{G}_i^{(2)}(\beta) \mathrm{diag}(\mathbf{X}_i \mathbf{H}_n^{-1/2} \mathbf{y}) \mathbf{R}_i(\alpha)^{-1} (\mathbf{F}_i(\beta) - \mathbf{F}_i) \mathbf{X}_i \mathbf{H}_n^{-1/2} \mathbf{x}\end{aligned}$$

Using inequalities (A.1.1) and (2.4.11), it follows that

$$U_4'(\beta, \mathbf{x}, \mathbf{y}) \leq \pi_n \gamma_n^{(0)} \max_{i \leq n} \lambda_{\max}^2 (\mathbf{G}_i^{(2)}(\beta) (\mathbf{A}_i^*)^{1/2}) T_1(\beta, \mathbf{x}, \mathbf{x}),$$

where $T_1(\beta, \mathbf{x}, \mathbf{y})$ is given by (2.4.3). The matrix $\mathbf{G}_i^{(2)}(\beta) (\mathbf{A}_i^*)^{1/2}$ has j-th element given by:

$$\sqrt{\sigma_{i,jj}^*(\beta_0)}\, g_{ij}^{(2)}(\beta) = \frac{\sigma_{i,jj}^*(\beta_0)}{\sigma_{i,jj}^*(\beta)} \cdot \left(\sqrt{\sigma_{i,jj}^*(\beta)}\, g_{ij}^{(2)}(\beta) \right).$$

By Lemmas A.2.2 and A.2.3 (Appendix A.2),

$$\sup_{\beta \in B_n^*(r)} \max_{i \le n} \lambda_{\max}^2(\mathbf{G}_i^{(2)}(\beta)(\mathbf{A}_i^*)^{1/2}) = O_p(1). \tag{2.4.24}$$

Using (2.4.5) and the fact that $\gamma_n^* = o_p(1)$, it follows that

$$\sup_{\beta \in B_n^*(r)} \sup_{\|\mathbf{x}\|=1} \sup_{\|\mathbf{y}\|=1} U_4'(\beta, \mathbf{x}, \mathbf{y}) \le \pi_n^2 \gamma_n^{(0)} \gamma_n^* O_p(1) = \pi_n^2 \gamma_n^{(0)} o_p(1).$$

Using (2.4.19) and the hypotheses of the lemma it follows that

$$\sup_{\beta \in B_n^*(r)} \sup_{\|\mathbf{x}\|=1} \sup_{\|\mathbf{y}\|=1} |U_4(\beta, \mathbf{x}, \mathbf{y})| = o_p(1).$$

It remains to treat $U_6(\beta, \mathbf{x}, \mathbf{y})$. By Cauchy-Schwarz inequality (2.4.6),

$$|U_6(\beta, \mathbf{x}, \mathbf{y})| \le U_6'(\beta, \mathbf{x}, \mathbf{y})^{1/2} U^{1/2},$$

where U is the same as above and

$$U_6'(\beta, \mathbf{x}, \mathbf{y}) = \sum_{i=1}^{n} \mathbf{x}^T \mathbf{H}_n^{-1/2} \mathbf{X}_i^T \mathbf{F}_i \mathbf{R}_i(\alpha)^{-1} \text{diag}(\mathbf{X}_i \mathbf{H}_n^{-1/2} \mathbf{y})(\mathbf{G}_i^{(2)}(\beta) - \mathbf{G}_i^{(2)})(\mathbf{A}_i^*)^{1/2} \mathbf{R}_i(\alpha)$$
$$(\mathbf{A}_i^*)^{1/2}(\mathbf{G}_i^{(2)}(\beta) - \mathbf{G}_i^{(2)}) \text{diag}(\mathbf{X}_i \mathbf{H}_n^{-1/2} \mathbf{y}) \mathbf{R}_i(\alpha)^{-1} \mathbf{F}_i \mathbf{X}_i \mathbf{H}_n^{-1/2} \mathbf{x}.$$

Using inequalities (A.1.1) and (2.4.11), it follows that $U_6'(\beta, \mathbf{x}, \mathbf{y})$ is less than or equal to

$$\pi_n \gamma_n^{(0)} \max_{i \le n} \lambda_{\max}^2(\mathbf{G}_i^{(2)}(\mathbf{A}_i^*)^{1/2}) \max_{i \le n} \lambda_{\max}^2(\mathbf{G}_i^{(2)}(\beta)(\mathbf{G}_i^{(2)})^{-1} - \mathbf{I}) \cdot \mathbf{x}^T \mathbf{H}_n^{-1/2} \mathbf{H}_n^* \mathbf{H}_n^{-1/2} \mathbf{x}.$$

By (2.4.24), $\max_{i \le n} \lambda_{\max}^2(\mathbf{G}_i^{(2)}(\mathbf{A}_i^*)^{1/2}) = O_p(1)$. By Lemma 2.4.1,

$$\sup_x \mathbf{x}^T \mathbf{H}_n^{-1/2} \mathbf{H}_n^* \mathbf{H}_n^{-1/2} \mathbf{x} = O_p(1).$$

In can be shown that $\max_{i \le n} \max_{i \le n} \lambda_{\max}^2(\mathbf{G}_i^{(2)}(\beta)(\mathbf{G}_i^{(2)})^{-1} - \mathbf{I}) = O_p(1)$. Arguing as above, we infer that

$$\sup_{\beta \in B_n^*(r)} \sup_{\|\mathbf{x}\|=1} \sup_{\|\mathbf{y}\|=1} |U_6(\beta, \mathbf{x}, \mathbf{y})| = o_p(1).$$

□

Chapter 3

Real-life Example

In this section, we discuss an application of our method to a subset of the real-life dataset taken from [10]. This data consists of $n = 250$ preschool age rural Indonesian children which were examined every 3 months for 18 months for the presence of a respiratory disease. So each child was observed on $m = 6$ occasions. The response Y is a binary variable which takes value 1 if the respiratory disease is present and value 0 if the disease is absent. We consider a marginal logistic regression model (see Example 2.1.3) with intercept parameter β_0 and 3 covariates: $X^{(1)}$ is a binary variable with values 0 and 1 giving the gender, $X^{(2)}$ is another binary variable with values 0 and 1 giving the vitamin A deficiency, and $X^{(3)}$ is the child's age in years at the beginning of the study, with possible values $1, 2, \ldots, 7$. In this case, $\mu(x) = e^x/(1 + e^x)$. The model is

$$\text{logit}(Y_{ij}) = \beta_0 + \beta_1 X_{ij}^{(1)} + \beta_2 X_{ij}^{(2)} + \beta_3 X_{ij}^{(3)}, \quad i = 1, \ldots, n, \quad j = 1, \ldots, m.$$

Here $\beta = (\beta_0, \beta_1, \beta_2, \beta_3)$, so $p = 4$. We let $\mathbf{X}_{ij} = (X_{ij}^{(0)}, X_{ij}^{(1)}, X_{ij}^{(2)}, X_{ij}^{(3)})$ where $X_{ij}^{(0)} = 1$.

Since this data does not contain missing values, we generated missingness indicator variables I_{ij} using a Bernoulli distribution with probability 0.95 of success. This gives 3.33% missing responses.

We fit a logistic regression model with parameter γ to the complete data set consisting of $(\mathbf{I}_i, \mathbf{X}_i)$ for $i = 1, \ldots, 250$, and we solved equation (2.3.8). The root of this equation is $\widehat{\gamma} = (3.514, 0.025, 0.391, -0.076)$. The estimates $\widehat{\pi}_{ij}$ for the missingness probabilities π_{ij} are calcualted using the formula $\widehat{\pi}_{ij} = \pi_{ij}(\widehat{\gamma})$, where $\pi_{ij}(\gamma)$ is given by (2.3.7). We compute the inverse probability weighted responses

$$Y_{ij}^* = \frac{Y_{ij} I_{ij}}{\widehat{\pi_{ij}}} \quad i = 1, \ldots, n, \; j = 1, \ldots, m$$

and we solve the working independence GEE with weighted responses, which in this case is a system of 4 equations:

$$\sum_{i=1}^{n}\sum_{j=1}^{m}\mathbf{X}_{ij}^{(l)}\left(Y_{ij}^{*}-\frac{\exp(\mathbf{X}_{ij}^{T}\beta)}{1+\exp(\mathbf{X}_{ij}^{T}\beta)}\right)=0,\quad l=0,1,2,3$$

yielding the root $\beta^{\text{indep}}=(-0.444,-0.552,0.258,-0.066)$.

3.1 Our method

In this subsection, we apply our method to the dataset described above. We start by computing the standardized values $\widehat{Y}_{ij}=\widetilde{Y}_{ij}(\beta^{\text{indep}})$, where

$$\widetilde{Y}_{ij}(\beta)=\frac{Y_{ij}^{*}-\mu_{ij}(\beta)}{\sqrt{\sigma_{i,jj}^{*}(\beta)}},$$

and $\sigma_{i,jj}^{*}(\beta)$ was calculated using (2.1.9) with π_{ij} replaced by $\widehat{\pi}_{ij}$.

Recall that the conditional correlation matrix $\mathbf{R}_i^{*}$ of $\mathbf{Y}_i$ given $\mathbf{X}_i$ has elements:

$$r_{i,jk}^{*}=\frac{E[(Y_{ij}^{*}-\mu_{ij}(\beta))(Y_{ik}^{*}-\mu_{ik}(\beta))|\mathbf{X}_i]}{\sqrt{\sigma_{i,jj}^{*}(\beta)}\cdot\sqrt{\sigma_{i,kk}^{*}(\beta)}}=E[\widetilde{Y}_{ij}(\beta)\widetilde{Y}_{ik}(\beta)|\mathbf{X}_i].$$

To estimate the matrix $\mathbf{R}_i^{*}$, we use the same matrix $\mathbf{R}_i(\alpha)=\widehat{\mathbf{R}}=(\widehat{r}_{jk})_{j,k=1,\ldots,m}$ for all i, with $\widehat{r}_{jj}=1$ for all $j=1,\ldots,m$, and for $j\neq k$, $\widehat{r}_{jk}$ are as in Examples 2 and 3 of [7]:

Case 1: (1-dependent) $\widehat{r}_{jk}=0$ if $|j-k|\geq 2$ and $\widehat{r}_{j,j+1}=\widehat{r}_{j+1,j}=\widehat{\alpha}_j$ where

$$\widehat{\alpha}_j=\frac{1}{n-p}\sum_{i=1}^{n}\widehat{Y}_{ij}\widehat{Y}_{i,j+1},\quad j=1,\ldots,m-1$$

This produces the values $\widehat{\alpha}_1=0.534,\widehat{\alpha}_2=0.559,\widehat{\alpha}_3=0.562,\widehat{\alpha}_4=0.443,\widehat{\alpha}_5=0.521$.

Case 2: (exhangeable) $\widehat{r}_{jk}=\widehat{\alpha}$ for all $j,k=1,\ldots,m$ with $j\neq k$, where

$$\widehat{\alpha}=\frac{1}{N-p}\sum_{i=1}^{n}\sum_{k=2}^{m}\sum_{j=1}^{k-1}\widehat{Y}_{ij}\widehat{Y}_{ik}\quad\text{with}\quad N=n\frac{m(m-1)}{2}$$

This produces the value $\widehat{\alpha}=0.492$.

Using these two cases, we solve equation (2.1.11), taking into account that in the calculation of $\sigma_{i,jj}^{*}(\beta)$, π_{ij} is replaced by $\widehat{\pi}_{ij}$. This consists of a system of 4 equations:

$$\sum_{i=1}^{n}\sum_{j=1}^{m}X_{ij}^{(l)}\frac{\mu'(\mathbf{X}_{ij}^{T}\beta)}{\sqrt{\sigma_{i,jj}^{*}(\beta)}}\,w_{jk}\,\frac{Y_{ik}^{*}-\mu(\mathbf{X}_{ik}^{T}\beta)}{\sqrt{\sigma_{i,kk}^{*}(\beta)}}=0,\quad l=0,1,2,3,$$

1-*dependent*	estimate	s.e.	p-value	*exchangeable*	estimate	s.e.	p-value
intercept	-0.448	0.268	0.095	intercept	-0.447	0.272	0.101
gender	-0.487	0.219	0.026	gender	-0.550	0.220	0.012
vitamin A	0.243	0.224	0.279	vitamin A	0.256	0.226	0.256
age	-0.068	0.056	0.220	age	-0.065	0.056	0.241

Table 3.1: Parameter estimates, standard errors (s.e.) and p-values for Case 1 (1-dependent) and Case 2 (exchangeable) correlation matrices

where $\mathbf{W} = \widehat{\mathbf{R}}^{-1} = (w_{jk})_{j,k=1,\dots,m}$. We obtained the following estimates:

$$\widehat{\beta}^{(1)} = (-0.448, -0.487, 0.243, -0.068) \quad \text{for the 1-dependent case}$$
$$\widehat{\beta}^{(2)} = (-0.447, -0.550, 0.256, -0.065) \quad \text{for the exchangeable case.}$$

To evaluate the precision of these estimates, we compute the standard error of these estimates and the p-value of the two-sided test for $\beta = 0$, using the asymptotic normality of $\widehat{\beta}$ given by Corollary 2.3.4:

$$\mathbf{B}^{-1/2}(\widehat{\beta} - \beta) \approx N_p(\mathbf{0}, \mathbf{I}), \tag{3.1.1}$$

where $\mathbf{B} = \mathbf{H}_n^{-1}\mathbf{M}_n\mathbf{H}_n^{-1}$. We estimate the matrix $\mathbf{B}$ by $\widehat{\mathbf{B}} = \widehat{\mathbf{H}}_n^{-1}\widehat{\mathbf{M}}_n\widehat{\mathbf{H}}_n^{-1}$, with matrices $\widehat{\mathbf{M}}_n$ and $\widehat{\mathbf{H}}_n$ computed as in Remark 2.3.5. From (3.1.1), we deduce that $\widehat{\beta} - \beta$ has approximately a p-variate normal distribution with mean vector $\mathbf{0}$ and covariance matrix $\widehat{\mathbf{B}}$. Hence, for $l = 0, 1, 2, 3$, $\widehat{\beta}^{(l)} - \beta_l \approx N(0, b_l)$, where b_l is the l-th element on the diagonal of $\widehat{\mathbf{B}}$. It follows that the standard error (s.e.) of $\widehat{\beta}^{(l)}$ is $s\{\widehat{\beta}^{(l)}\} = \sqrt{b_l}$ and the p-value of the test of $H_0 : \beta_l = 0$ versus $H_1 : \beta_l \neq 0$ is $2P(Z > |\widehat{\beta}^{(l)}/\sqrt{b_l}|)$. In Table 1, we report the estimates, their standard errors and p-values for the two examples of correlation matrices considered above (1-dependent and exchangeable).

Finally, for the sake of comparison, we conclude this section with a quick analysis of the results obtained for the complete data. More precisely, we solve the Generalized Estimating Equation (2.1.1), which can be written as:

$$\sum_{i=1}^{n}\sum_{j=1}^{m} X_{ij}^{(l)} \sqrt{\mu'(\mathbf{X}_{ij}^T\beta)}\, w_{jk}\, \frac{Y_{ik} - \mu(\mathbf{X}_{ik}^T\beta)}{\sqrt{\mu'(\mathbf{X}_{ik}^T\beta)}} = 0, \quad l = 0, 1, 2, 3,$$

Solving this system of equations is very similar to the previous case with missing data. Once again we considered two cases (1-dependent and exchangeable) for the correlation matrix $\mathbf{R}_i(\alpha)$. The results are summarized in the following table:

1-*dependent*	estimate	s.e.	p-value	*exchangeable*	estimate	s.e.	p-value
intercept	-0.456	0.267	0.087	intercept	-0.430	0.270	0.112
gender	-0.467	0.215	0.030	gender	-0.557	0.216	0.010
vitamin A	0.282	0.221	0.202	vitamin A	0.223	0.226	0.198
age	-0.072	0.054	0.182	age	-0.071	0.054	0.188

Table 3.2: Classical GEE with no missing data: Parameter estimates, standard errors (s.e.) and p-values for Case 1 (1-dependent) and Case 2 (exchangeable) correlation matrices

We conclude that at a 5% significance level, we reject the hypothesis $\beta_1 = 0$, but we do not have enough evidence to reject the hypothesis $\beta_2 = 0$ or the hypothesis $\beta_3 = 0$. This means that the gender seems to have a significant effect on the presence of respiratory disease, but vitamin A deficiency and age do not influence the presence of this disease.

3.2 The generalized method of moments

In this subsection, we apply to the same dataset the generalized method of moments of [14], assuming that all covariates are observed exactly, i.e. without measurement error.

We recall briefly this method below. For each $i = 1, \ldots, n$ and $j = 1, \ldots, m$, let

$$\mathcal{G}_{ij}(Y_{ij}, X_{ij}, \beta) = \frac{\partial \mu_{ij}(\beta)}{\partial \beta^T}(\sigma_{ij}^2(\beta))^{-1}\big(Y_{ij} - \mu_{ij}(\beta)\big) = \mathbf{X}_{ij}\big(Y_{ij} - \mu_{ij}(\beta)\big)$$

and $\Phi_{ij}(\beta) = \mathcal{G}(Y_{ij}^*, \mathbf{X}_{ij}, \beta) = \mathbf{X}_{ij}\big(Y_{ij}^* - \mu_{ij}(\beta)\big)$. Note that $\frac{\partial}{\partial \beta^T}\Phi_{ij}(\beta) = -\mathbf{X}_{ij}\mathbf{X}_{ij}^T\mu'(\mathbf{X}_{ij}^T\beta)$ and $\mathrm{tr}(\frac{\partial}{\partial \beta^T}\Phi_{ij}(\beta)) = -\mathbf{X}_{ij}^T\mathbf{X}_{ij}\mu'(\mathbf{X}_{ij}^T\beta)$. Let $\widehat{\mathbf{D}} = \widehat{\mathbf{D}}(\beta^{\mathrm{indep}})$ where

$$\widehat{\mathbf{D}}(\beta) = \Big(-\frac{1}{n}\sum_{i=1}^n \mathbf{X}_{i1}^T\mathbf{X}_{i1}\mu'(\mathbf{X}_{i1}^T\beta), \ldots, -\frac{1}{n}\sum_{i=1}^n \mathbf{X}_{im}^T\mathbf{X}_{im}\mu'(\mathbf{X}_{im}^T\beta)\Big).$$

Let $\Phi_i(\beta) = \big(\Phi_{i1}^T(\beta), \ldots, \Phi_{im}^T(\beta)\big)^T$. Note that $\Phi_i(\beta)$ is an $m \times p$ matrix. Let $\mathbf{V}(\beta) = E[\Phi_i(\beta)\Phi_i(\beta)^T]$ which is estimated by $\widehat{\mathbf{V}} = \widehat{\mathbf{V}}(\beta^{\mathrm{indep}})$, where $\widehat{\mathbf{V}}(\beta) = \frac{1}{n}\sum_{i=1}^n \Phi_i(\beta)\Phi_i(\beta)^T$.

Let $\Psi_i(\beta) = \mathbf{K}\Phi_i(\beta) = \sum_{j=1}^m K_j\Phi_{ij}^T(\beta)$, where $\mathbf{K} = \widehat{\mathbf{D}}\widehat{\mathbf{V}}^{-1} = (K_1, \ldots, K_m)$. In Section 3.4 of [14], the authors suggested solving the equation

$$\sum_{i=1}^n \Psi_i(\beta) = 0,$$

	estimate	s.e.	p-value
intercept	1.345	0.271	0
gender	1.675	0.218	0
vitamin A	-0.833	0.224	0.0002
age	0.201	0.055	0.0002

Table 3.3: Parameter estimates, standard errors (s.e.) and p-values using the generalized method of moments of [14], in the absence of measurement error

which is equivalent to $\sum_{i=1}^{n}\sum_{j=1} K_j \Phi_{ij}(\beta) = 0$. In our case, this becomes a system of 4 equations:

$$\sum_{i=1}^{n}\sum_{j=1}^{m} K_j X_{ij}^{(l)}(Y_{ij}^{*} - \mu(\mathbf{X}_{ij}^{T}\beta)) = 0, \quad l = 0, 1, 2, 3$$

whose solution is $\widehat{\beta} = (1.345, 1.657, -0.833, 0.201)$. From Section 3.5 of [14], we know that

$$\sqrt{n}(\widehat{\beta} - \beta) \approx N_p(\mathbf{0}, \Sigma) \tag{3.2.1}$$

where $\Sigma = \Gamma_0^{-1}\Sigma_\Psi(\Gamma_0^{-1})^T$, $\Gamma_0 = E[\frac{\partial}{\partial \beta^T}\Psi_i^T(\beta)]$ and $\Sigma_\Psi = E[\Psi_i(\beta)^T\Psi_i(\beta)]$. We estimate the matrices Γ_0 and Σ_Ψ by $\widehat{\Gamma}_0 = \widehat{\Gamma}_0(\beta^{\text{indep}})$, respectively $\widehat{\Sigma}_\Psi = \widehat{\Sigma}_\Psi(\beta^{\text{indep}})$, where

$$\widehat{\Gamma}_0(\beta) = \frac{1}{n}\sum_{i=1}^{n}\frac{\partial}{\partial \beta^T}\Psi_i^T(\beta) = \frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{m} K_j \frac{\partial}{\partial \beta^T}\Phi_{ij}(\beta) = -\frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{m} K_j \mathbf{X}_{ij}\mathbf{X}_{ij}^T \mu'(\mathbf{X}_{ij}^T\beta),$$

and

$$\widehat{\Sigma}_\Psi(\beta) = \frac{1}{n}\sum_{i=1}^{n}\Psi_i(\beta)^T\Psi_i(\beta) = \frac{1}{n}\sum_{i=1}^{n}\Phi_i(\beta)^T\mathbf{K}^T\mathbf{K}\Phi_i(\beta).$$

From (3.2.1), we deduce that $\widehat{\beta} - \beta$ has approximately a p-variate normal distribution with mean vector $\mathbf{0}$ and covariance matrix $\widehat{\Sigma}/n$, where $\widehat{\Sigma} = \widehat{\Gamma}_0^{-1}\widehat{\Sigma}_\Psi(\widehat{\Gamma}_0^{-1})^T$. Hence, for $l = 0, 1, 2, 3$, $\widehat{\beta}_l - \beta_l \approx N(0, \sigma_l/n)$, where σ_l is the l-th element on the diagonal of $\widehat{\Sigma}$. Hence, the standard error (s.e.) of $\widehat{\beta}_l$ is $s\{\widehat{\beta}_l\} = \sqrt{\sigma_l/n}$ and the p-value of the test of $H_0 : \beta_l = 0$ versus $H_1 : \beta_l \neq 0$ is $2P(Z > |\widehat{\beta}_l/\sqrt{\sigma_l/n}|)$. These values are given in Table 2. Using this method, we conclude that all three covariates seem to have a significant effect on the presence of the respiratory disease.

Appendix A

Some Auxiliary results

A.1 Matrix analysis results

If $\mathbf{A}$ is a symmetric $p \times p$ matrix, all its eigenvalues are real and we denote by $\lambda_{\min}(\mathbf{A})$ and $\lambda_{\max}(\mathbf{A})$ its minimum, respectively maximum, eigenvalues.

For any $p \times p$ matrix $\mathbf{A} = (a_{ij})_{1 \le i,j \le p}$ we define the Euclidean norm of $\mathbf{A}$ by

$$\|\mathbf{A}\|_E = \Big(\sum_{i=1}^{p} \sum_{j=1}^{q} a_{ij}^2 \Big)^{1/2} = (\mathrm{tr}(\mathbf{A}^T\mathbf{A}))^{1/2} = (\mathrm{tr}(\mathbf{A}\mathbf{A}^T))^{1/2},$$

and its spectral norm by

$$\|\mathbf{A}\| = (\lambda_{\max}(\mathbf{A}^T\mathbf{A}))^{1/2} = \sup_{\|\mathbf{x}\|=1} \|\mathbf{A}\mathbf{x}\| = \sup_{\mathbf{x} \neq 0} \frac{\|\mathbf{A}\mathbf{x}\|}{\|\mathbf{x}\|}.$$

A $p \times p$ matrix $\mathbf{A}$ is non-negative definite if $\mathbf{x}^T\mathbf{A}\mathbf{x} \ge 0$ for any $\mathbf{x} \in \mathbb{R}^p$. In this case, we write $\mathbf{A} \ge 0$. We write $\mathbf{A} \le \mathbf{B}$ for $p \times p$ matrices $\mathbf{A}$ and $\mathbf{B}$ if $\mathbf{B} - \mathbf{A} \ge 0$.

Lemma A.1.1. *The Euclidean norm and spectral norm are equivalent:*

$$c_2(p)\|\boldsymbol{A}\|_E \le \|\boldsymbol{A}\| \le c_1(p)\|\boldsymbol{A}\|_E,$$

for some constants $c_1(p) > 0$, $c_2(p) > 0$ depending on p.

Theorem A.1.2. *Let $\boldsymbol{A}$ be a symmetric $p \times p$ matrix with eigenvalues $\lambda_1 \le \cdots \le \lambda_p$. Then*

$$\mathrm{tr}(\boldsymbol{A}) = \sum_{i=1}^{p} \lambda_i \text{ and } \det(\boldsymbol{A}) = \prod_{i=1}^{n} \lambda_i.$$

If $\boldsymbol{A} \ge 0$, then the eigenvalues of $\boldsymbol{A}^2$ are $\lambda_1^2 \le \cdots \le \lambda_p^2$, and in particular $\lambda_{\max}(\boldsymbol{A}^2) = [\lambda_{\max}(\boldsymbol{A})]^2$. If $\boldsymbol{A} > 0$, then $\lambda_1 > 0$, $\boldsymbol{A}$ is invertible and the eigenvalues of $\boldsymbol{A}^{-1}$ are $\lambda_1^{-1} \ge \cdots \ge \lambda_p^{-1}$; in particular, $\lambda_{\max}(\boldsymbol{A}^{-1}) = 1/\lambda_{\min}(\boldsymbol{A})$.

Theorem A.1.3. *Matrices* $\boldsymbol{A}^{-1}\boldsymbol{B}$ *and* $\boldsymbol{A}^{-1/2}\boldsymbol{B}\boldsymbol{A}^{-1/2}$ *have the same eigenvalues.*

Lemma A.1.4. *If* $\boldsymbol{A} = (a_{jk})_{1\leq j,k\leq m}$ *and* $|a_{jk}| \leq 1$ *for all* $j, k = 1, \ldots, m$, *then*

$$\lambda_{max}(\boldsymbol{B}\boldsymbol{A}) \leq m\lambda_{max}(\boldsymbol{B}).$$

Theorem A.1.5. *If* $\boldsymbol{A} \geq 0$, *then for any* $\boldsymbol{x} \in \mathbb{R}^p$

$$\lambda_{\min}(\boldsymbol{A})\boldsymbol{x}^T\boldsymbol{x} \leq \boldsymbol{x}^T\boldsymbol{A}\boldsymbol{x} \leq \lambda_{\max}(\boldsymbol{A})\boldsymbol{x}^T\boldsymbol{x}. \tag{A.1.1}$$

Lemma A.1.6. *If* $\boldsymbol{A}$ *and* $\boldsymbol{B}$ *are symmetric* $p \times p$ *matrices with respective eigenvalues* $\lambda_1(\boldsymbol{A}) \leq \cdots \leq \lambda_p(\boldsymbol{A})$ *and* $\lambda_1(\boldsymbol{B}) \leq \cdots \leq \lambda_p(\boldsymbol{B})$,*and* $\boldsymbol{A} \leq \boldsymbol{B}$, *then* $\lambda_i(\boldsymbol{A}) \leq \lambda_i(\boldsymbol{B})$ *for all* $i = 1, \ldots, p$. *In particular,* $\mathrm{tr}(\boldsymbol{A}) \leq \mathrm{tr}(\boldsymbol{B})$ *and* $\det(\boldsymbol{A}) \leq \det(\boldsymbol{B})$.

A.2 Asymptotic Results

In this appendix, we gather some auxiliary results which were used in the proofs of Lemmas 2.4.1, 2.4.2 and 2.4.3. Recall that Assumption (AH) is given on page 24.

Lemma A.2.1. *Suppose Assumption (AH) holds. If* $\gamma_n^* \xrightarrow{P} 0$, *then*

$$\sup_{\beta \in B_n^*(r)} \max_{i \le n} \max_{j \le m} \left| \frac{\mu'(\mathbf{X}_{ij}^T \beta)}{\mu'(\mathbf{X}_{ij}^T \beta_0)} - 1 \right| = O_p((\gamma_n^*)^{1/2}) \tag{A.2.1}$$

$$\sup_{\beta \in B_n^*(r)} \max_{i \le n} \max_{j \le m} \left| \frac{\mu^2(\mathbf{X}_{ij}^T \beta)}{\mu^2(\mathbf{X}_{ij}^T \beta_0)} - 1 \right| = O_p((\gamma_n^*)^{1/2}). \tag{A.2.2}$$

Proof: This follows by Taylor's formula, using the fact that $k_n^{(1)} = O_p(1)$, respectively $k_n^{(0)} = O_p(1)$. See also Lemma B.1 of [12]. □

Lemma A.2.2. *Suppose Assumption (AH) holds. If* $\gamma_n^* \xrightarrow{P} 0$ *then*

$$\sup_{\beta \in B_n^*(r)} \max_{i \le n} \max_{j \le m} \left| \frac{\sigma_{i,jj}^*(\beta)}{\sigma_{i,jj}^*(\beta_0)} - 1 \right| = O_p((\gamma_n^*)^{1/2}).$$

Proof: Note that $\frac{\sigma_{i,jj}^*(\beta)}{\sigma_{i,jj}^*(\beta_0)} - 1$ is equal to

$$\frac{\mu'(\mathbf{X}_{ij}^T \beta) - \mu'(\mathbf{X}_{ij}^T \beta_0) + \left(\frac{1}{\pi_{ij}} - 1\right)\left(\mu'(\mathbf{X}_{ij}^T \beta) - \mu'(\mathbf{X}_{ij}^T \beta_0) + \mu^2(\mathbf{X}_{ij}^T \beta) - \mu^2(\mathbf{X}_{ij}^T \beta_0)\right)}{\mu'(\mathbf{X}_{ij}^T \beta_0) + \left(\frac{1}{\pi_{ij}} - 1\right)\left(\mu'(\mathbf{X}_{ij}^T \beta_0) + \mu^2(\mathbf{X}_{ij}^T \beta_0)\right)}.$$

Since μ' is non-negative and $\pi_{ij} \le 1$,

$$\left| \frac{\sigma_{i,jj}^*(\beta)}{\sigma_{i,jj}^*(\beta_0)} - 1 \right| \le 2 \left| \frac{\mu'(\mathbf{X}_{ij}^T \beta) - \mu'(\mathbf{X}_{ij}^T \beta_0)}{\mu'(\mathbf{X}_{ij}^T \beta_0)} \right| + \left| \frac{\mu^2(\mathbf{X}_{ij}^T \beta) - \mu^2(\mathbf{X}_{ij}^T \beta_0)}{\mu^2(\mathbf{X}_{ij}^T \beta_0)} \right|.$$

The conclusion follows by Lemma A.2.1. □

Lemma A.2.3. *Suppose Assumptions (AH) and (M) hold. Then*

$$\sup_{\beta \in B_n^*(r)} \max_{i \le n} \max_{j \le m} \left(\sqrt{\sigma_{i,jj}^*(\beta)}\, |g_{ij}^{(2)}(\beta)| \right) = O_p(1).$$

Proof: Recalling definition (2.4.2) of $g_{ij}^{(2)}(\beta)$, we see that:

$$\sqrt{\sigma_{i,jj}^*(\beta)}\, g_{ij}^{(2)}(\beta) = -\frac{2\left(\frac{1}{\pi_{ij}} - 1\right)\mu(\mathbf{X}_{ij}^T\beta)\mu'(\mathbf{X}_{ij}^T\beta) + \frac{1}{\pi_{ij}}\mu''(\mathbf{X}_{ij}^T\beta)}{2\sigma_{i,jj}^*(\beta)}, \tag{A.2.3}$$

and hence

$$\sqrt{\sigma_{i,jj}^*(\beta)}\, |g_{ij}^{(2)}(\beta)| \le \left(\frac{1}{\pi_{ij}} - 1\right)\frac{|\mu(\mathbf{X}_{ij}^T\beta)\mu'(\mathbf{X}_{ij}^T\beta)|}{\sigma_{i,jj}^*(\beta)} + \frac{1}{\pi_{ij}} \cdot \frac{|\mu''(\mathbf{X}_{ij}^T\beta)|}{2\sigma_{i,jj}^*(\beta)}.$$

In the first term on the right-hand side of this inequality, we use the fact that $\sigma_{i,jj}^*(\beta) \ge \left(\frac{1}{\pi_{ij}} - 1\right)(\mu'(\mathbf{X}_{ij}^T\beta) + \mu^2(\mathbf{X}_{ij}^T\beta)) \ge \left(\frac{1}{\pi_{ij}} - 1\right)\mu^2(\mathbf{X}_{ij}^T\beta)$, whereas for the second term we use the fact that $\sigma_{i,jj}^*(\beta) \ge \mu'(\mathbf{X}_{ij}^T\beta)$. We obtain that

$$\sqrt{\sigma_{i,jj}^*(\beta)}\, |g_{ij}^{(2)}(\beta)| \le \left|\frac{\mu'(\mathbf{X}_{ij}^T\beta)}{\mu(\mathbf{X}_{ij}^T\beta)}\right| + \frac{1}{2\pi_{ij}}\left|\frac{\mu''(\mathbf{X}_{ij}^T\beta)}{\mu'(\mathbf{X}_{ij}^T\beta)}\right| \le k_n^{(0)} + \frac{1}{2}\rho_n k_n^{(1)}.$$

The conclusion follows by Assumptions (AH) and (M). □.

Lemma A.2.4. *Suppose Assumptions (AH) and (M) hold. If* $\gamma_n^* \xrightarrow{P} 0$ *then*

$$\sup_{\beta\in B_n^*(r)} \max_{i\le n} \max_{j\le m} \left|\frac{f_{ij}(\beta)}{f_{ij}(\beta_0)} - 1\right| = O_p((\gamma_n^*)^{1/2}).$$

Proof: Recall that $\frac{\partial}{\partial\beta^T} f_{ij}(\beta) = g_{ij}^{(1)}(\beta)\mathbf{X}_{ij}^T$. By Taylor's formula, for any $\beta \in B_n^*(r)$, there exists $\beta_{ij} \in B_n^*(r)$ such that $f_{ij}(\beta) - f_{ij}(\beta_0) = g_{ij}^{(1)}(\beta_{ij})\mathbf{X}_{ij}^T(\beta - \beta_0)$. Since

$$\|\mathbf{X}_{ij}^T(\beta - \beta_0)\| \le \|\mathbf{X}_{ij}^T\mathbf{H}_n^{-1/2}\| \cdot \|\mathbf{H}_n^{1/2}(\beta - \beta_0)\| \le (\gamma_n^{(0)})^{1/2}(\tau_n)^{1/2} r = (\gamma_n^*)^{1/2} r,$$

it follows that

$$\left|\frac{f_{ij}(\beta)}{f_{ij}(\beta_0)} - 1\right| \le \left|\frac{g_{ij}^{(1)}(\beta_{ij})}{f_{ij}(\beta_{ij})}\right| \cdot \left|\frac{f_{ij}(\beta_{ij})}{f_{ij}(\beta_0)}\right| (\gamma_n^*)^{1/2} r. \tag{A.2.4}$$

By definition (2.1.13) of $f_{ij}(\beta)$, we have $\frac{f_{ij}(\beta)}{f_{ij}(\beta_0)} = \frac{\mu'(\mathbf{X}_{ij}^T\beta)}{\mu'(\mathbf{X}_{ij}^T\beta_0)} \cdot \frac{\sqrt{\sigma_{i,jj}^*(\beta_0)}}{\sqrt{\sigma_{i,jj}^*(\beta)}}$. By Lemmas A.2.1 and A.2.2,

$$\sup_{\beta\in B_n^*(r)} \max_{i\le n} \max_{j\le m} \left|\frac{f_{ij}(\beta)}{f_{ij}(\beta_0)}\right| = O_p(1). \tag{A.2.5}$$

A direct calculation based on definition (2.4.1) of $g_{ij}^{(1)}(\beta)$ and relation (A.2.3) shows that

$$\frac{g_{ij}^{(1)}(\beta)}{f_{ij}(\beta)} = \frac{\mu''(\mathbf{X}_{ij}^T\beta)}{\mu'(\mathbf{X}_{ij}^T\beta)} + \sqrt{\sigma_{i,jj}^*(\beta)}\, g_{ij}^{(2)}(\beta).$$

By Assumption (AH) and Lemma A.2.3, it follows that

$$\sup_{\beta \in B_n^*(r)} \max_{i \leq n} \max_{j \leq m} \left| \frac{g_{ij}^{(1)}(\beta)}{f_{ij}(\beta)} \right| = O_p(1). \tag{A.2.6}$$

The conclusion follows from relations (A.2.4), (A.2.5) and (A.2.6). □

Appendix B

R Programs

```
#Installing the necessary packages
install.packages("DPpackage")
library(DPpackage)

install.packages("rootSolve")
library(rootSolve)

install.packages("tibble")
library(tibble)

#The dataset is located in the working directory
dataset = read.csv("ICHS.csv", header = TRUE)
attach(dataset)
head(dataset)

n = 250
m = 6
p = 4

cov2 = Gender
cov3 = VitaminA
cov4 = Age

y = Response

#We also need to introduce the intercept parameter beta_0
intercept = c(rep(1,len = n*m))
cov1 = intercept
```

```
dataset = add_column(dataset, cov1, .after = "Time")

set.seed(31417)
r = rbinom(n*m,1,0.95)

#Computing the missingness percentage
r1 = r[r<1]
missing = length(r1)/(n*m)
paste("missingness percentage: ",round(missing*100, digits=2),"%")

#Introduce missing responses
for (i in (1:n*m)){
if(r[i]<1) y[i] = NA
}

#Replace by 0 the missing responses
for (i in (1:n*m)){
if(r[i]<1) y[i] = 0
}

mu = function(x){
return(1/(1+exp(-x)))
}

mu_prime = function(x){
return(exp(x)/((1+exp(x))^2))
}

#Finding gamma_hat
logisticfunction_gamma = function(gamma){

prod = cov1*gamma[1] + cov2*gamma[2] + cov3*gamma[3] + cov4*gamma[4];
mean = 1/(1+exp(-prod));
residual = r-mean;
c(F1 = sum(cov1*residual), F2 = sum(cov2*residual),
F3 = sum(cov3*residual), F4 = sum(cov4*residual)) }

gamma_hat = multiroot(f = logisticfunction_gamma,
start = c(0,0,0,0))$root
```

```
#Computing estimated pi_{ij} using logit(pi_{ij})=gamma_hat^T X_{ij}
pi_hat = 1/(1+exp(-(gamma_hat[1]*cov1 +
gamma_hat[2]*cov2 + gamma_hat[3]*cov3 + gamma_hat[4]*cov4)))

#Computing y_{ij}^* (inverse probability weighted responses)
w = y*r/pi_hat

#Computing estimated beta in the case of working indepedence
logisticfunction0 = function(b){

prod = cov1*b[1] + cov2*b[2] + cov3*b[3] + cov4*b[4];
mean = mu(prod);
residual = w-mean;
c(F1 = sum(cov1*residual), F2 = sum(cov2*residual),
F3 = sum(cov3*residual), F4 = sum(cov4*residual)) }

beta_indep = multiroot(f = logisticfunction0, start = c(0,0,0,0))$root

logisticfunction00 = function(b){

prod = cov1*b[1] + cov2*b[2] + cov3*b[3] + cov4*b[4];
mean = mu(prod);
residual = y-mean;
c(F1 = sum(cov1*residual), F2 = sum(cov2*residual),
F3 = sum(cov3*residual), F4 = sum(cov4*residual)) }

beta_indep2 = multiroot(f = logisticfunction00, start = c(0,0,0,0))$root
beta_indep2
```

```
#Creating a list of 1500 matrices of dimension 6x3; only matrices
#List1[[k]] with k(mod6)=1 are defined
List1 = list()
for (k in 1:(n*m)) {
if ((k %% m) == 1) {
List1[[k]] = data.matrix(dataset[k:(k+m-1),4:7])
```

```
}
}

#Extracting matrices X_1,...,X_250 from List1 by mapping the indices
#from 1-250 to 1-1500
X = list()
for (i in 1:n){
X[[i]] = List1[[m*i-m+1]]
}

#Creating a list of 1500 6x1 matrices from y: only matrices y_list1[[k]]
with k(mod6)=1 are defined
y_list <- list()
for (k in 1:(n*m)){
if ((k %% m) == 1) {
y_list[[k]] = data.matrix(y[k:(k+m-1)])
}
}

#Creating a list of 1500 6x1 matrices from r: only matrices r_list1[[k]]
#with k(mod6)=1 are defined
r_list = list()
for (k in 1:(n*m)) {
if ((k %% m) == 1) {
r_list[[k]] = data.matrix(r[k:(k+m-1)])
}
}

#Creating a list of 1500 6x1 matrices from w: only matrices w_list1[[k]]
#with k(mod6)=1 are defined
w_list = list()
for (k in 1:(n*m)) {
if ((k %% m) == 1) {
w_list[[k]] = data.matrix(w[k:(k+m-1)])
}
}

#Extracting matrices y_star1,...,y_star250 from w_list by mapping the
#indices from 1-250 to 1-1500
#y_star[[i]] is a 6X1 matrix with values y_{ij}^*,j=1,...,6
y_star = list()
```

```
for (i in 1:n) {
y_star[[i]] = w_list[[m*i-m+1]]
}

y_new = list()
for (i in 1:n) {
y_new[[i]] = y_list[[m*i-m+1]]
}

#Creating a list of 1500 6x1 matrices from pi_hat: only matrices
#pi_list1[[k]] with k(mod6)=1 are defined
pi_list = list()
for (k in 1:(n*m)) {
if ((k %% m) == 1) {
pi_list[[k]] = data.matrix(pi_hat[k:(k+m-1)])
}
}

#Extracting matrices pi1,...,pi250 from pi_list by mapping the
#indices from 1-250 to 1-1500
#pi[i]] is a 6x1 matrix with values pi_{ij}, j=1,...,6
pi = list()
for (i in 1:n) {
pi[[i]] = pi_list[[6*i-5]]
}

#Computing inverse probabilities
#q[[i]] is a 6x1 matrix with values 1/pi_{ij}, j=1,...,6
q = list()
for (i in 1:n) {
q[[i]] = 1/pi[[i]]-1
}

#Creating a list of 250 6x1 matrices containing centered values when
#beta = beta_indep
#y_star_c[[i]] is a 6x1 matrix with values
#Y_{ij}^*-mu(X_{ij}^T beta_indep), j=1,...,6
y_star_c = list()
for (i in 1:n){
y_star_c[[i]] = matrix(rep(0,len=m),nrow=m)
}
```

```
for (i in 1:n){
for (j in 1:m){
y_star_c[[i]][j,] = y_star[[i]][j,] -
mu(t(X[[i]][j,]) %*% beta_indep)
}
}


#in case of no missingness, we will use the following:
y_c = list()
for (i in 1:n){
y_c[[i]] = matrix(rep(0,len=m),nrow=m)
}


for (i in 1:n){
for (j in 1:m){
y_c[[i]][j,] = y_new[[i]][j,] - mu(t(X[[i]][j,]) %*% beta_indep2)
}
}


#Creating a list of 250 6x1 matrices containing the variance of
#Y_{ij}^* when beta = beta_indep
#sigma[[i]]] is a 6x1 matrix with values
#sigma_{ij}^*(beta_indep), j=1,...,6
sigma = list()
for (i in 1:n){
sigma[[i]] = matrix(rep(0,len = m),nrow = m)
}


for (i in 1:n){
for (j in 1:m){
sigma[[i]][j,] = mu_prime(t(X[[i]][j,]) %*% beta_indep)  +
(1/pi_list[[6*i-5]][j,]-1)*(mu_prime(t(X[[i]][j,])
%*% beta_indep) + (mu(t(X[[i]][j,]) %*% beta_indep))^2)
}
}
```

```
#Creating a list of 250 6x1 matrices containing standardized
#Y_{ij}^* when beta = beta_indep
#y_tilde[[i]]] is a 6x1 matrix with values
#(Y_{ij}^*-mu(X_{ij}^T beta_indep))/sqrt{sigma_{ij}^*(beta_indep)},
j=1,...,6
y_tilde = list()
for (i in 1:n){
y_tilde[[i]] = matrix(rep(0,len=m),nrow=m)
}


for (i in 1:n){
for (j in 1:m){
y_tilde[[i]][j,] = y_star_c[[i]][j,] / sqrt( sigma[[i]][j,] )
}
}

####################################################
#in case of no missingness, we will use the following:
y_std = list()
for (i in 1:n){
y_std[[i]] = matrix(rep(0,len=m),nrow=m)
}


for (i in 1:n){
for (j in 1:m){
y_std[[i]][j,] = y_c[[i]][j,] / sqrt(mu_prime(t(X[[i]][j,]) %*%
beta_indep2))
}
}








#Computing the estimated correlation matrix
```

```
#Case 1: 1-dependent
R_hat1 = matrix(0,m,m)
R_hat1[m,m] = 1
for (j in 1:(m-1)){
R_hat1[j,j] = 1
for (i in 1:n){
R_hat1[j,j+1] = R_hat1[j,j+1] +
(1/(n-p))*y_tilde[[i]][j]*y_tilde[[i]][j+1]
R_hat1[j+1,j] =  R_hat1[j,j+1]
}
}

#Finding the inverse of R_hat1
R_hat1_inv = solve(R_hat1)


#Case 2: exchangeable
N = n*m*(m-1)/2

alpha = 0
for (i in 1:n){
for (k in 2:m){
for (j in 1:(k-1)){
alpha = alpha + (1/(N-p))*y_tilde[[i]][j]*y_tilde[[i]][k]
}
}
}


R_hat2 = matrix(0,m,m)
for (j in 1:m){
R_hat2[j,j] = 1
}
for (j in 1:(m-1)){
for (k in (j+1):m){
R_hat2[j,k] = alpha
}
}
for (j in 2:m){
for (k in 1:(j-1)){
R_hat2[j,k] = alpha
```

```
}
}

#Finding the inverse of R_hat2
R_hat2_inv = solve(R_hat2)



#### OUR METHOD #####
#Case 1: 1-dependent (R = R_hat1)
logisticfunction1 = function(b){
F = c(0,0,0,0)
b = c(b[1],b[2],b[3],b[4])

for (t in 1:p) {
for (i in 1:n){
for (k in 1:m){
for (j in 1:m){


F[t] = F[t] + X[[i]][j,t]*mu_prime(t(X[[i]][j,]) %*% b)/
sqrt(mu_prime(t(X[[i]][j,]) %*% b)  +
(1/pi_list[[m*i-m+1]][j,]-1)*(mu_prime(t(X[[i]][j,]) %*% b)
+(mu(t(X[[i]][j,]) %*% b))^2))*
R_hat1_inv[j,k]*(y_star[[i]][k]-mu(t(X[[i]][j,]) %*% b))/
sqrt(mu_prime(t(X[[i]][j,]) %*% b)  +
(1/pi_list[[m*i-m+1]][k,]-1)*(mu_prime(t(X[[i]][j,]) %*% b)
+(mu(t(X[[i]][j,]) %*% b))^2  ))

}
}
}
}

c(F[1],F[2],F[3],F[4])


} #end of logisticfunction1

bhat1 = multiroot(f = logisticfunction1, start =
c(beta_indep[1],beta_indep[2],beta_indep[3],beta_indep[4]))$root
```

```
#exchangeable case: R = R_hat2
logisticfunction2 = function(b){
F = c(0,0,0,0)
b = c(b[1],b[2],b[3],b[4])

for (t in 1:p) {
for (i in 1:n){
for (k in 1:m){
for (j in 1:m){


F[t] = F[t] + X[[i]][j,t]*mu_prime(t(X[[i]][j,]) %*% b)/
sqrt(mu_prime(t(X[[i]][j,]) %*% b)  +
(1/pi_list[[m*i-m+1]][j,]-1)*(mu_prime(t(X[[i]][j,]) %*% b) +
(mu(t(X[[i]][j,]) %*% b))^2))*
R_hat2_inv[j,k]*(y_star[[i]][k]-mu(t(X[[i]][j,]) %*% b))/
sqrt(mu_prime(t(X[[i]][j,]) %*% b)  +
(1/pi_list[[m*i-m+1]][k,]-1)*(mu_prime(t(X[[i]][j,]) %*% b)
+(mu(t(X[[i]][j,]) %*% b))^2  ))

}
}
}
}

c(F[1],F[2],F[3],F[4])


} #end of logisticfunction2
bhat2 = multiroot(f = logisticfunction2, start =
c(beta_indep[1],beta_indep[2],beta_indep[3],beta_indep[4]))$root


#solving the classical GEE without missing data

#finding the new correlation matrices R_hat12 and R_hat22

#case 1: 1-dependent
R_hat12 = matrix(0,m,m)
```

```
R_hat12[m,m] = 1
for (j in 1:(m-1)){
R_hat12[j,j] = 1
for (i in 1:n){
R_hat12[j,j+1] = R_hat12[j,j+1] + (1/(n-p))*y_std[[i]][j]*y_std[[i]][j+1]
R_hat12[j+1,j] =  R_hat12[j,j+1]
}
}

#Find the inverse of R_hat1
R_hat12_inv = solve(R_hat12)
R_hat12_inv


#Case 2: exhangeable
N = n*m*(m-1)/2

alpha = 0
for (i in 1:n){
for (k in 2:m){
for (j in 1:(k-1)){
alpha = alpha + (1/(N-p))*y_c[[i]][j]*y_c[[i]][k]
}
}
}


R_hat22 = matrix(0,m,m)
for (j in 1:m){
R_hat22[j,j] = 1
}
for (j in 1:(m-1)){
for (k in (j+1):m){
R_hat22[j,k] = alpha
}
}
for (j in 2:m){
for (k in 1:(j-1)){
R_hat22[j,k] = alpha
}
}
```

```
#Find the inverse of R_hat2
R_hat22_inv = solve(R_hat22)
R_hat22_inv






#Case 1: 1-dependent (R = R_hat12)
logisticfunction4 = function(b){
F = c(0,0,0,0)
b = c(b[1],b[2],b[3],b[4])

for (t in 1:p) {
for (i in 1:n){
for (k in 1:m){
for (j in 1:m){


F[t] = F[t] + X[[i]][j,t]*sqrt(mu_prime(t(X[[i]][j,]) %*% b))*
R_hat12_inv[j,k]*(y_new[[i]][k]-mu(t(X[[i]][k,]) %*% b))
/sqrt(mu_prime(t(X[[i]][k,]) %*% b))

}

}
}
}

c(F[1],F[2],F[3],F[4])


} #end of logisticfunction4
bhat4 = multiroot(f = logisticfunction4, start = c(beta_indep2[1],
beta_indep2[2], beta_indep2[3],beta_indep2[4]))$root
bhat4


#Case 2: exchangeable (R = R_hat22)
logisticfunction5 = function(b){
```

```
F = c(0,0,0,0)
b = c(b[1],b[2],b[3],b[4])

for (t in 1:p) {
for (i in 1:n){
for (k in 1:m){
for (j in 1:m){


F[t] = F[t] + X[[i]][j,t]*sqrt(mu_prime(t(X[[i]][j,]) %*% b))*
R_hat22_inv[j,k]*(y_new[[i]][k]-mu(t(X[[i]][k,]) %*% b))/
sqrt(mu_prime(t(X[[i]][k,]) %*% b))

}

}
}
}

c(F[1],F[2],F[3],F[4])


} #end of logisticfunction5
bhat5 = multiroot(f = logisticfunction5, start = c(beta_indep2[1],
beta_indep2[2], beta_indep2[3],beta_indep2[4]))$root
bhat5


A1_list = list()
for (i in 1:n){
A1_list[[i]] = matrix(rep(0,len=m*m),nrow=m)
for (j in 1:m){
A1_list[[i]][j,j] = mu_prime(t(X[[i]][j,]) %*%  bhat1 )
}
}

A2_list = list()
for (i in 1:n){
A2_list[[i]] = matrix(rep(0,len=m*m),nrow=m)
for (j in 1:m){
A2_list[[i]][j,j] = mu_prime(t(X[[i]][j,]) %*%  bhat2 )
```

```
}
}


D1_list = list()
for (i in 1:n){
for (j in 1:m){
D1_list[[i]] = A1_list[[i]]%*%X[[i]]
}
}

D2_list = list()
for (i in 1:n){
for (j in 1:m){
D2_list[[i]] = A2_list[[i]]%*%X[[i]]
}
}


prod1_list = list()
for (i in 1:n){
prod1_list[[i]] = matrix(rep(0,len=m),nrow=m)

for (j in 1:m){
prod1_list[[i]][j,] = 0
prod1_list[[i]][j,] = t(X[[i]][j,]) %*%  bhat1
}
}

prod2_list = list()
for (i in 1:n){
prod2_list[[i]] = matrix(rep(0,len=m),nrow=m)

for (j in 1:m){
prod2_list[[i]][j,] = 0
prod2_list[[i]][j,] = t(X[[i]][j,]) %*%  bhat2
}
}


mu1_list = list()
```

```
for (i in 1:n){
mu1_list[[i]] =  mu(prod1_list[[i]])
}

mu2_list = list()
for (i in 1:n){
mu2_list[[i]] =  mu(prod2_list[[i]])
}

mu1_prime_list = list()
for (i in 1:n){
mu1_prime_list[[i]] = matrix(rep(0,len=m*m),nrow=m)
mu1_prime_list[[i]] = mu_prime(prod1_list[[i]])
}

mu2_prime_list = list()
for (i in 1:n){
mu2_prime_list[[i]] = matrix(rep(0,len=m*m),nrow=m)
mu2_prime_list[[i]] = mu_prime(prod2_list[[i]])
}


A1_star_list = list()
for (i in 1:n){
A1_star_list[[i]] = matrix(rep(0,len=m*m),nrow=m)
for (j in 1:m){

A1_star_list[[i]][j,j] =  mu1_prime_list[[i]][j,]  +
q[[i]][j,]*( mu1_prime_list[[i]][j,] +
mu1_list[[i]][j,]^2  )
}
}

A2_star_list = list()
for (i in 1:n){
A2_star_list[[i]] = matrix(rep(0,len=m*m),nrow=m)
for (j in 1:m){

A2_star_list[[i]][j,j] =  mu2_prime_list[[i]][j,]  +
q[[i]][j,]*( mu2_prime_list[[i]][j,] +
mu2_list[[i]][j,]^2  )
```

```
}
}



V1_star_list = list()
for (i in 1:n){
V1_star_list[[i]] = A1_star_list[[i]]^(1/2) %*% R_hat1 %*%
A1_star_list[[i]]^(1/2)
}

V2_star_list = list()
for (i in 1:n){
V2_star_list[[i]] = A2_star_list[[i]]^(1/2) %*% R_hat2 %*%
A2_star_list[[i]]^(1/2)
}


Sigma1_star_list = list()
for (i in 1:n){
for (j in 1:m){
Sigma1_star_list[[i]] = (y_star[[i]]-mu1_list[[i]]) %*%
t(y_star[[i]]-mu1_list[[i]])
}
}

Sigma2_star_list = list()
for (i in 1:n){
for (j in 1:m){
Sigma2_star_list[[i]] = (y_star[[i]]-mu2_list[[i]]) %*%
t(y_star[[i]]-mu2_list[[i]])
}
}

M_hat1 = matrix(rep(0,len=p*p),nrow=p)
for (i in 1:n){
M_hat1 = M_hat1 + t(D1_list[[i]]) %*% solve(V1_star_list[[i]]) %*%
Sigma1_star_list[[i]] %*% solve(V1_star_list[[i]]) %*% D1_list[[i]]
}

M_hat2 = matrix(rep(0,len=p*p),nrow=p)
```

```
for (i in 1:n){
M_hat2 = M_hat2 + t(D2_list[[i]]) %*% solve(V2_star_list[[i]]) %*%
Sigma2_star_list[[i]] %*% solve(V2_star_list[[i]]) %*% D2_list[[i]]
}

H_hat1 = matrix(rep(0,len=p*p),nrow=p)
for (i in 1:n){
H_hat1 = H_hat1 + t(D1_list[[i]]) %*% solve(V1_star_list[[i]]) %*%
D1_list[[i]]
}

H_hat2 = matrix(rep(0,len=p*p),nrow=p)
for (i in 1:n){
H_hat2 = H_hat2 + t(D2_list[[i]]) %*% solve(V2_star_list[[i]]) %*%
D2_list[[i]]
}

B1 = solve(H_hat1) %*% M_hat1 %*% solve(H_hat1)
B2 = solve(H_hat2) %*% M_hat2 %*% solve(H_hat2)

### now all over again for bhat4 and bhat5
A1_list2 = list()
for (i in 1:n){
A1_list2[[i]] = matrix(rep(0,len=m*m),nrow=m)
for (j in 1:m){
A1_list2[[i]][j,j] = mu_prime(t(X[[i]][j,]) %*%  bhat4 )
}
}

A2_list2 = list()
for (i in 1:n){
A2_list2[[i]] = matrix(rep(0,len=m*m),nrow=m)
for (j in 1:m){
A2_list2[[i]][j,j] = mu_prime(t(X[[i]][j,]) %*%  bhat5 )
}
}


D1_list2 = list()
for (i in 1:n){
for (j in 1:m){
```

```
D1_list2[[i]] = A1_list2[[i]]%*%X[[i]]
}
}

D2_list2 = list()
for (i in 1:n){
for (j in 1:m){
D2_list2[[i]] = A2_list2[[i]]%*%X[[i]]
}
}


prod1_list2 = list()
for (i in 1:n){
prod1_list2[[i]] = matrix(rep(0,len=m),nrow=m)

for (j in 1:m){
prod1_list2[[i]][j,] = 0
prod1_list2[[i]][j,] = t(X[[i]][j,]) %*%  bhat4
}
}

prod2_list2 = list()
for (i in 1:n){
prod2_list2[[i]] = matrix(rep(0,len=m),nrow=m)

for (j in 1:m){
prod2_list2[[i]][j,] = 0
prod2_list2[[i]][j,] = t(X[[i]][j,]) %*%  bhat5
}
}


mu1_list2 = list()
for (i in 1:n){
mu1_list2[[i]] =  mu(prod1_list2[[i]])
}

mu2_list2 = list()
for (i in 1:n){
mu2_list2[[i]] =  mu(prod2_list2[[i]])
```

```
}

mu1_prime_list2 = list()
for (i in 1:n){
mu1_prime_list2[[i]] = matrix(rep(0,len=m*m),nrow=m)
mu1_prime_list2[[i]] = mu_prime(prod1_list2[[i]])
}

mu2_prime_list2 = list()
for (i in 1:n){
mu2_prime_list2[[i]] = matrix(rep(0,len=m*m),nrow=m)
mu2_prime_list2[[i]] = mu_prime(prod2_list2[[i]])
}


A1_star_list2 = list()
for (i in 1:n){
A1_star_list2[[i]] = matrix(rep(0,len=m*m),nrow=m)
for (j in 1:m){

A1_star_list2[[i]][j,j] =  mu1_prime_list2[[i]][j,]
}
}

A2_star_list2 = list()
for (i in 1:n){
A2_star_list2[[i]] = matrix(rep(0,len=m*m),nrow=m)
for (j in 1:m){

A2_star_list2[[i]][j,j] =  mu2_prime_list2[[i]][j,]
}
}



V1_star_list2 = list()
for (i in 1:n){
V1_star_list2[[i]] = A1_star_list2[[i]]^(1/2) %*% R_hat12 %*%
A1_star_list2[[i]]^(1/2)
}
```

```
V2_star_list2 = list()
for (i in 1:n){
V2_star_list2[[i]] = A2_star_list2[[i]]^(1/2) %*% R_hat22 %*%
A2_star_list2[[i]]^(1/2)
}


Sigma1_star_list2 = list()
for (i in 1:n){
for (j in 1:m){
Sigma1_star_list2[[i]] = (y_new[[i]]-mu1_list2[[i]]) %*%
t(y_new[[i]]-mu1_list2[[i]])
}
}

Sigma2_star_list2 = list()
for (i in 1:n){
for (j in 1:m){
Sigma2_star_list2[[i]] = (y_new[[i]]-mu2_list2[[i]]) %*%
t(y_new[[i]]-mu2_list2[[i]])
}
}

M_hat12 = matrix(rep(0,len=p*p),nrow=p)
for (i in 1:n){
M_hat12 = M_hat12 + t(D1_list2[[i]]) %*% solve(V1_star_list2[[i]]) %*%
Sigma1_star_list2[[i]] %*% solve(V1_star_list2[[i]]) %*% D1_list2[[i]]
}

M_hat22 = matrix(rep(0,len=p*p),nrow=p)
for (i in 1:n){
M_hat22 = M_hat22 + t(D2_list2[[i]]) %*% solve(V2_star_list2[[i]]) %*%
Sigma2_star_list2[[i]] %*% solve(V2_star_list2[[i]]) %*% D2_list2[[i]]
}

H_hat12 = matrix(rep(0,len=p*p),nrow=p)
for (i in 1:n){
H_hat12 = H_hat12 + t(D1_list2[[i]]) %*% solve(V1_star_list2[[i]]) %*%
D1_list2[[i]]
}
```

```
H_hat22 = matrix(rep(0,len=p*p),nrow=p)
for (i in 1:n){
H_hat22 = H_hat22 + t(D2_list2[[i]]) %*% solve(V2_star_list2[[i]]) %*%
D2_list2[[i]]
}

B12 = solve(H_hat12) %*% M_hat12 %*% solve(H_hat12)
B22 = solve(H_hat22) %*% M_hat22 %*% solve(H_hat22)


#Computing standard error of estimator and p-value of two-sided
#test for beta=0
#1-dependent case

SE1 = c()
for (i in 1:p){
SE1[i] = sqrt(B1[i,i])
}

test_stat1 = abs(bhat1/SE1)

p_value1 = c()
p_value1 = 2*(1-pnorm(test_stat1))

#exchangeable case

SE2 = c()
for (i in 1:p){
SE2[i] = sqrt(B2[i,i])
}

test_stat2 = abs(bhat2/SE2)

p_value2 = c()
p_value2 = 2*(1-pnorm(test_stat2))

#Compute standard error of estimator and p-value of
#two-sided test for beta=0
#bhat4 and bhat5
#1-dependent case
```

```
SE12 = c()
for (i in 1:p){
SE12[i] = sqrt(B12[i,i])
}
SE12

test_stat12 = abs(bhat4/SE12)

p_value12 = c()
p_value12 = 2*(1-pnorm(test_stat12))
p_value12

#exchangeable case

SE22 = c()
for (i in 1:p){
SE22[i] = sqrt(B22[i,i])
}
SE22

test_stat22 = abs(bhat5/SE22)

p_value22 = c()
p_value22 = 2*(1-pnorm(test_stat22))
p_value22


######## METHOD of Yi-Ma-Carroll ###########

#Creating the 6x1 matrix D
# D_j is -(1/n)sum_{i=1}^{n} X_{ij}^T X_{ij} \mu'(X_{ij}^T beta_indep)

D = matrix(rep(0,len=m),nrow=1)

for (j in 1:m){
for (i in 1:n){
D[1,j]=D[1,j] + (-1/n)*t(X[[i]][j,])%*%t(t(X[[i]][j,]))%*%
mu_prime(t(X[[i]][j,]) %*% beta_indep)
}
}
```

```
#Creating a list of 6x4 matrices Phi_1,...,Phi_250
#Phi[[i]]=Phi_i has j-th row X_{ij}^T (Y_{ij}^*- mu_{ij}(\beta_indep))
Phi = list()
PPhi = matrix(rep(0,len=m*p),nrow=m)

for (i in 1:n){
for (j in 1:m){

PPhi[j,] = y_star_c[[i]][j,]%*%t(X[[i]][j,])
}
Phi[[i]] = PPhi
}


#Computing the 6x6 matrix V
#V[j,k] is (1/n)sum_{i=1}^{250} \Phi_{ij}^T \Phi_{ik}
V = matrix(rep(0,len=m*m),nrow=m)

for (j in 1:m) {
for (k in 1:m){
for (i in 1:n){
V[j,k] = V[j,k] + (1/n)*y_star_c[[i]][j,]%*%y_star_c[[i]][k,]%*%
t(X[[i]][j,])%*%t(t(X[[i]][k,]))
}
}
}

#Computing the inverse of V
V_inv = solve(V)

#Computing the 1 x m vector K
#K=D V^{-1}
K = D%*%V_inv
K

#Defining new variables T
#T_{ij}^l=K_j X_{ij}^l, for l=1,2,3

ccov1 = c()
ccov2 = c()
ccov3 = c()
```

```
ccov4 = c()

for (i in 1:n){
for (j in 1:m){
ccov1[m*i-m+j] = K[j]*cov1[m*i-m+j]
ccov2[m*i-m+j] = K[j]*cov2[m*i-m+j]
ccov3[m*i-m+j] = K[j]*cov3[m*i-m+j]
ccov4[m*i-m+j] = K[j]*cov4[m*i-m+j]
}
}


#Computing the estimated beta using Yi-Ma-Carroll method
logisticfunction3 = function(b){

prod = ccov1*b[1] + ccov2*b[2] + ccov3*b[3] + ccov4*b[4];
mean = mu(prod);
residual = w-mean;
c(F1 = sum(ccov1*residual), F2 = sum(ccov2*residual),
F3 = sum(ccov3*residual), F4 = sum(ccov4*residual)) }

bhat3 = multiroot(f = logisticfunction3, start =
c(beta_indep[1],beta_indep[2],beta_indep[3],beta_indep[4]))$root

#Computing matrices needed for the asymptotic variance of estimator

#Computing the p x p matrix Gamma_0
Gamma0 = matrix(rep(0,len=p*p),nrow=p)

for (i in 1:n){
for (j in 1:m){

Gamma0 = Gamma0 + (-1/n)*(K[,j]*(t(t(X[[i]][j,]))
%*% t(X[[i]][j,])))*mu_prime(t(X[[i]][j,]) %*% beta_indep)[1,1]

}
}

#Computing the inverse of Gamma0
Gamma0_inv = solve(Gamma0)
```

```
#Computing the p x p matrix Sigma_psi
Sigma_psi = matrix(rep(0,len=p*p),nrow=p)
for (i in 1:n){
Sigma_psi = Sigma_psi + (1/n)* ( t(Phi[[i]])%*%t(K)%*%K%*%Phi[[i]])

}

#Computing the p x p matrix Sigma_beta
Sigma_beta = Gamma0_inv%*%Sigma_psi%*%t(Gamma0_inv)

#Computing standard error of estimator and p-value of two-sided
#test for beta=0

SE3 = c()
for (i in 1:p){
SE3[i] = sqrt(Sigma_beta[i,i]/n)
}

test_stat3 = abs(bhat3/SE3)

p_value3 = c()
p_value3 = 2*(1-pnorm(test_stat3))
```

Bibliography

[1] Balan, R.M. and Jankovic, D. (2018). Asymptotic theory for longitudinal data with missing responses adjusted by inverse probability weights. Preprint available on arXiv:1803.05836.

[2] Balan, R. M. and Schiopu-Kratina, I. (2005). Asymptotic Results with Generalized Estimating Equations for Longitudinal Data. *Ann. Statist.* **33**, 522-541.

[3] Billingsley, P. (1995). *Probability and Measure.* Third Edition. Wiley, New York.

[4] Chen, K., Hu, I. and Ying, Z. (1999). Strong consistency of maximum quasi-likelihood estimators in generalized linear models with fixed and adaptive designs. *Ann. Stat.* **27** 1155-1163.

[5] Chen, B., Yi, G.Y. and Cook, R.J. (2010). Weighted generalized estimating functions for longitudinal response and covariate data that are missing at random. . *J. Amer. Stat. Soc.* **105**, 336-353.

[6] Diggle, P.J., Liang, K.-Y. and Zeger, S.L. (1996). *Analysis of Longitudinal Data.* Clarendon Press, Oxford.

[7] Liang, K.-Y., and Zeger, S.L. (1986). Longitudinal data analysis using generalized linear models. *Biometrika*, **73** 13-22.

[8] McCullagh, P. and Nelder, J.A. (1989). *Generalized Linear Models.* Second edition. Chapman and Hall, Boca Raton.

[9] Schott, J.R. (1997). *Matrix Analysis for Statistics*, John Wiley, New York.

[10] Sommer, A., Katz, J. and Tarwotjo, I. (1984). Increased risk of respiratory disease and diarrhea in children with preexisting mild vitamin A deficiency. *Amer. J. Clinical Nutrition* **40**, 1090-1095.

[11] Wang, C.Y., Huang, Y., Chao, E.C. and Jeffcoat, M.K. (2008). Expected estimating equations for missing data, measurement error, and misclassification, with application to longitudinal nonignorable missing data. *Biometrics* **64**, 8595.

[12] Xie, M. and Yang, Y. (2003). Asymptotics for generalized estimating equations with large cluster sizes. *Ann. Stat.* **31** 310-347.

[13] Yi, G.Y., Liu, W. and Wu, L. (2011). Simultaneous inference and bias analysis for longitudinal data with covariate measurement error and missing responses. *Biometrics* **67**, 6775.

[14] Yi, G.Y., Ma, Y. and Carroll, R.J. (2012). A functional generalized method of moments approach for longitudinal studies with missing responses and covariate measurement error. *Biometrika* **99**, 151-165.

[15] Yi, G.Y., Tan, X. and Li, R. (2015). Variable selection and inference procedures for marginal analysis of longitudinal data with missing observations and covariate measurement error. *Canad. J. Stat.* **43**, 498-518.

[16] Yuan, K.-H. and Jennrich, R.I. (1998). Asymptotics of estimating equations under natural conditions. *J. Multiv. Anal.* **65**, 245-260.