Longitudinal data analysis using Generalized Linear Model with missing responses

by

Jeanseong Park

Thesis submitted to the
Faculty of Graduate and Postdoctoral Studies
In partial fulfillment of the requirements
For the Master of Science in Mathematics

Department of Mathematics and Statistics
Faculty of Sciences
University of Ottawa

Abstract

Longitudinal studies rely on data collected at several occasions from a set of selected individuals. The purpose of these studies is to use a regression-type model to express a response variable as a function of explanatory variables, or covariates. In this thesis, we use marginal models for the analysis of such data, which, coupled with the method of estimating equations, provide estimators of the main regression parameter. When some of the responses are missing or there is error in the recorded covariates, the original estimating equation may be biased. We use techniques available in the literature to modify it and regain the unbiasedness property. We prove the asymptotic normality of the regression estimator obtained under these more realistic circumstances, and provide theoretical and numerical examples to illustrate this approach.

Acknowledgements

I wish to take this opportunity to express my gratitude to my supervisors Raluca Balan and Ioana Schiopu-Kratina for patiently guiding me through the project with knowledge and passion.

I also want to thank my father, mother, and brother for their unconditional love and support.

I am especially grateful to my wife. It would not have been possible without her.

Finally, I give the deepest thanks to our Lord Jesus for His grace that allowed me to glorify His name.

Contents

Chapter 1

Introduction

Longitudinal models are widely used in the health, social, and behavioral sciences, as well as in biological and agricultural sciences, education, and economics. It is a type of observational study that involves repeated observations of the same subject over an extended period of time. As a simple illustration, consider a longitudinal study of n individuals, on which data is collected at m distinct occasions. The variable of interest (the response variable) is the systolic blood pressure. This response variable is continuous, if its actual value is recorded, or discrete, if it is recorded for example, in one of the three disjoint categories: low, normal, high. One could then study the evolution of the response variable over time as a function of covariates like height, weight, salt intake, amount of physical exercise, etc. This information is obtained on each subject at every occasion.

Liang and Zeger introduced in [3] the generalized estimating equations (GEE), which extend the use of generalized linear models (GLM) to longitudinal data. Generalized estimating equations are used in regression analysis of longitudinal data, where observations on the same subject are correlated. The solutions of these GEE provide a sequence of estimators $\widehat{\beta}_n$ of the regression parameter β_0. In [3], marginal models are used to express the mean and variance of the response variable in terms of covariates, for each individual and on every occasion. The correlation matrix within individual i, R_i, is assumed to exist but need not be modelled. The most important results of [3] are that the estimators $\widehat{\beta}_n$ are consistent, i.e., they converge to the true parameter β_0, and they are also asymptotically normal, even if the correlations R_i are misspecified. In this dissertation we consider the case of "working independence", where the covariance matrix of marginal responses within each subject is diagonal.

In [3], the covariates used in modelling are considered nonrandom. The extension to random covariates is important because, in many studies, one would like to include previous responses in a set of covariates for the current occasion. In the example above, one would like to include among covariates the values of the systolic blood pressure recorded on previous occasions. The extension of the GEE approach to include random covariates, which may vary over time, has proven difficult with the exception of the "working independence" case. Many papers in the field, including [8], used the generalized method of moments approach in [2] to represent the longitudinal correlations when marginal models are used on each occasion.

Despite efforts to obtain accurate and complete data on all variables and at all occasions, some data may still be missing in the response variable, and some covariates may be recorded with error. Disregarding the missing data in the analysis may lead to biased results, when the missing data depends on the response variable itself. A brief description of possible mechanisms underlining the pattern of missing data can be found in Chapter 6 of [5]. In our work we make the assumption that the probability of missingness can depend only on variables that are observed (the missing at random assumption, or MAR).

Some important methods for compensating for nonresponse are: likelihood methods, imputation methods, and inverse probability weighting methods. The first two methods have been studied in detail and are included in the classical monograph [4]. As in [8], we adopt the third method to compensate for nonresponse in the response variable. We also consider that the covariates could be recorded with error, and as in [8], we assume a classical additive model to link the observed covariates to their true values.

In this dissertation we study the work of Yi et al. in [8], which discusses marginal models with missing response data and random covariates that may have been recorded with error. We detail the results in [8] by providing proofs for most of our statements. For instance, in chapter 3, we give complete proofs of the asymptotic normality of the regression estimators, based on analytical and stochastic properties of the corresponding estimating functions. We illustrate our results with simulated data, assuming that the response variable follows one of the three distributions: normal linear regression, Poisson regression or logistic regression.

Our work is organized as follows. In Chapter 2, we discuss the GEE method introduced by Liang and Zeger in [3]. Using simulated data, we calculate the GEE estimator of the main regression parameter β in the "working independence" case, along with its bias (mean and absolute), standard error, and the mean squared error, for the three examples mentioned above. In Chapter 3, we consider random covariates in the specific case where they are recorded without error, and there is missingness among the response variables. We use a logit model for the probability of response, which is a function of covariates and a parameter α. This model is adapted to the situation when missingness follows an appropriate MAR assumption. The estimator $\widehat{\alpha}$ of α maximizes a likelihood-type function, and its asymptotic distribution is described in Lemma 3.4.4. Our main estimating functions, which give estimators $\widehat{\beta}$ for β, differ from those in [8] in that they are generated by the components of a "working independence" estimating function, which is simpler to use. These initial components are then "weighted up" to compensate for nonresponse and to ensure their unbiasedness. We illustrate our results with three theoretical examples and simulated data. In Theorem 3.4.5, we prove the asymptotic normality of $\widehat{\beta}$ in the general case when both α and β are unknown. In Chapter 4, we examine the model for longitudinal data introduced in [7] and [8], in which the covariates are measured with error. We first assume that there is no missingness in the response variables. Replacing the true (latent) covariates by the observed ones creates biasedness in the estimating equations. Following the ideas in [7], we analytically modify these equations to obtain unbiasedness in the three cases where the response variable has the exponential distributions mentioned above. Finally, we apply this technique to the general case when some covariates are measured with error, and there is missingness in the response variable.

Chapter 2

Generalized Estimating Equations for Longitudinal Data

The main objective of this chapter is to review the method of Generalized Estimating Equations (GEE) introduced by Liang and Zeger in [3]. The notations and the results introduced in this chapter will be employed in the next chapters. In this chapter, we assume that the covariates X_{ij} are not random.

2.1 Marginal Models for Longitudinal Data

We start by introducing some notation that will be used throughout the thesis. We assume that n independent subjects are measured repeatedly over time. We let Y_{ij} denote the response variable for the i^{th} individual on the j^{th} measurement occasion.

For each $1 \leq i \leq n$, we consider the m-dimensional vector of responses for the i^{th} individual:

$$Y_i = (Y_{i1}, \ldots, Y_{im})^T.$$

Associated with each response Y_{ij}, there is a $p \times 1$ vector of covariates

$$X_{ij} = \begin{bmatrix} X_{ij}^{(1)} \\ \vdots \\ X_{ij}^{(p)} \end{bmatrix}.$$

For each $1 \leq i \leq n$, we consider the $m \times p$ matrix of covariates for the i^{th} individual:

$$X_i = \begin{bmatrix} X_{i1}^T \\ \vdots \\ X_{im}^T \end{bmatrix}. \tag{2.1}$$

We let β be the p-dimensional regression parameter,

$$\beta = \begin{bmatrix} \beta_1 \\ \vdots \\ \beta_p \end{bmatrix}.$$

For a p-dimensional vector $x = (x_1, ..., x_p)^T$, we use the Euclidean norm

$$||x|| = \sqrt{x_1^2 + \cdots + x_p^2}.$$

We assume that there exists a continuously differentiable one-to-one function μ such that

$$\mu_{ij}(\beta) = E_\beta(Y_{ij}) = \mu(X_{ij}^T\beta) \text{ and } v_{ij}(\beta) = \text{Var}_\beta(Y_{ij}) = \phi_{ij}h(\mu(X_{ij}^T\beta)). \tag{2.2}$$

Here ϕ_{ij} is a parameter which has to be estimated separately. We will assume for simplicity that $\phi_{ij} = 1$.

We denote

$$\mu_i(\beta) = \begin{bmatrix} \mu_{i1}(\beta) \\ \vdots \\ \mu_{im}(\beta) \end{bmatrix}.$$

Note that $X_{ij}^T\beta = g(\mu_{ij}(\beta))$, where g is the inverse function of μ. The function g is sometimes called the *link function*.

We denote $\varepsilon_{ij}(\beta) = Y_{ij} - \mu_{ij}(\beta)$. This means that we consider the following model:

$$Y_{ij} = \mu(X_{ij}^T\beta) + \varepsilon_{ij}, \quad 1 \leq i \leq n, \quad 1 \leq j \leq m. \tag{2.3}$$

To simplify the notation, we omit writing the parameter β when no confusion may arise.

Note that this is a *marginal* model, i.e. it imposes a restriction only on the marginal distributions of Y_{ij}, but does not impose any assumption on the distribution of the vector $Y_i = (Y_{i1}, \ldots, Y_{im})^T$. We assume that the random vectors $Y_1, Y_2, \ldots, Y_n$ are independent.

In this section, we present the method of Generalized Estimating Equations (GEE) for estimating β, which was introduced by Liang and Zeger in [3].

This method consists in solving the following equation for β:

$$\sum_{i=1}^{n} D_i(\beta)^T V_i(\beta)^{-1}(Y_i - \mu_i(\beta)) = 0, \tag{2.4}$$

where

$$D_i(\beta) = \frac{\partial \mu_i}{\partial \beta^T}(\beta) = \begin{bmatrix} \frac{\partial \mu_{i1}}{\partial \beta_1}(\beta), \ldots, \frac{\partial \mu_{i1}}{\partial \beta_p}(\beta) \\ \vdots \\ \frac{\partial \mu_{im}}{\partial \beta_1}(\beta), \ldots, \frac{\partial \mu_{im}}{\partial \beta_p}(\beta) \end{bmatrix} \text{ is an } m \times p \text{ matrix, and} \tag{2.5}$$

$$\begin{aligned} V_i(\beta) &= \begin{bmatrix} \mathrm{Var}(Y_{i1}) & \mathrm{Cov}(Y_{i1}, Y_{i2}) & \ldots & \mathrm{Cov}(Y_{i1}, Y_{im}) \\ \mathrm{Cov}(Y_{i1}, Y_{i2}) & \mathrm{Var}(Y_{i2}) & \ldots & \mathrm{Cov}(Y_{i2}, Y_{im}) \\ \vdots & \vdots & \ddots & \vdots \\ \mathrm{Cov}(Y_{i1}, Y_{im}) & \mathrm{Cov}(Y_{i2}, Y_{im}) & \ldots & \mathrm{Var}(Y_{im}) \end{bmatrix} \\ &= E[(Y_i - \mu_i(\beta))(Y_i - \mu_i(\beta))^T] \\ &= \mathrm{Cov}(Y_i) \end{aligned}$$

is the covariance matrix of the vector Y_i. We will assume that $V_i(\beta)$ is invertible, and we denote by $V_i(\beta)^{-1}$ its inverse. We denote $v_{i,jk}(\beta) = \mathrm{Cov}(Y_{ij}, Y_{ik})$. Hence $v_{i,jj}(\beta) = \mathrm{Var}(Y_{ij}) := v_{ij}(\beta)$.

Note that for any $1 \leq i \leq n$ and for any $\beta \in \mathbb{R}^p$,

$$D_i(\beta) = A_i(\beta) X_i, \tag{2.6}$$

where X_i is the $m \times p$ matrix given by (2.1) and $A_i(\beta)$ is the diagonal matrix:

$$A_i(\beta) = \begin{bmatrix} \mu'(X_{i1}^T\beta) & 0 & \ldots & 0 \\ 0 & \mu'(X_{i2}^T\beta) & \ldots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \ldots & \mu'(X_{im}^T\beta) \end{bmatrix}.$$

Sometimes it is assumed that $h \circ \mu = \mu'$, where μ' denotes the derivative of the function μ. In this case, $A_i(\beta)$ is the diagonal matrix of variances:

$$A_i(\beta) = \begin{bmatrix} v_{i1}(\beta) & 0 & \ldots & 0 \\ 0 & v_{i2}(\beta) & \ldots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \ldots & v_{im}(\beta) \end{bmatrix}.$$

We emphasize that the components of the vector $Y_i = (Y_{i1}, \dots, Y_{im})$ are not independent, because they represent observations made on the same individual.

To examine the structure of the matrix $V_i(\beta)$, we look at the correlation between Y_{ij} and Y_{ik} for $1 \le j, k \le m$:

$$r_{i,jk} := \mathrm{Corr}(Y_{ij}, Y_{ik}) = \frac{\mathrm{Cov}(Y_{ij}, Y_{ik})}{\sqrt{\mathrm{Var}(Y_{ij})}\sqrt{\mathrm{Var}(Y_{ik})}} = \frac{v_{i,jk}(\beta)}{\sqrt{v_{ij}(\beta)}\sqrt{v_{ik}(\beta)}}.$$

Hence $v_{i,jk}(\beta) = r_{i,jk}\sqrt{v_{ij}(\beta)}\sqrt{v_{ik}(\beta)}$ for any $1 \le j,\ k \le m$.

In matrix notation,

$$V_i(\beta) = A_i(\beta)^{\frac{1}{2}} R_i A_i(\beta)^{\frac{1}{2}}, \tag{2.7}$$

where $R_i = (r_{i,jk})_{1 \le j,k \le m}$ is the correlation matrix for the i^{th} individual.

From (2.6) and (2.7), we infer that (2.4) can be written as

$$\sum_{i=1}^{n} X_i^T A_i(\beta) A_i(\beta)^{-\frac{1}{2}} R_i^{-1} A_i(\beta)^{-\frac{1}{2}} (Y_i - \mu_i(\beta)) = 0,$$

or equivalently,

$$\sum_{i=1}^{n} X_i^T A_i(\beta)^{\frac{1}{2}} R_i^{-1} A_i(\beta)^{-\frac{1}{2}} (Y_i - \mu_i(\beta)) = 0. \tag{2.8}$$

The case when, for each i, the variables $(Y_{ij})_{1 \le j \le m}$ are independent, is known in the literature under the name "working independence". In this case, $R_i = I_m$, the identity matrix in R^m,

$$V_i(\beta) = \begin{bmatrix} v_{i1}(\beta) & 0 & \dots & 0 \\ 0 & v_{i2}(\beta) & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & v_{im}(\beta) \end{bmatrix}, \quad \text{and} \tag{2.9}$$

$$V_i^{-1}(\beta) = \begin{bmatrix} v_{i1}^{-1}(\beta) & 0 & \dots & 0 \\ 0 & v_{i2}^{-1}(\beta) & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & v_{im}^{-1}(\beta) \end{bmatrix}.$$

Now,

$$
\begin{aligned}
& D_i^T(\beta)V_i(\beta)^{-1}(Y_i-\mu(\beta)) \\
&= \begin{bmatrix} \frac{\partial\mu_{i1}}{\partial\beta_1}(\beta),\ldots,\frac{\partial\mu_{im}}{\partial\beta_1}(\beta) \\ \vdots \\ \frac{\partial\mu_{i1}}{\partial\beta_p}(\beta),\ldots,\frac{\partial\mu_{im}}{\partial\beta_p}(\beta) \end{bmatrix} \begin{bmatrix} v_{i1}^{-1}(\beta) & 0 & \ldots & 0 \\ 0 & v_{i2}^{-1}(\beta) & \ldots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \ldots & v_{im}^{-1}(\beta) \end{bmatrix} \begin{bmatrix} Y_{i1}-\mu_{i1}(\beta) \\ \vdots \\ Y_{im}-\mu_{im}(\beta) \end{bmatrix} \\
&= \sum_{j=1}^{m} \frac{\partial\mu_{ij}}{\partial\beta^T}(\beta)v_{ij}^{-1}(\beta)(Y_{ij}-\mu_{ij}(\beta)).
\end{aligned}
$$

In the "working independence" case, equation (2.4) becomes

$$\sum_{i=1}^{n}\sum_{j=1}^{m} \frac{\partial\mu_{ij}}{\partial\beta^T}(\beta)v_{ij}^{-1}(\beta)(Y_{ij}-\mu_{ij}(\beta)) = 0. \tag{2.10}$$

Note that by the chain rule, $\frac{\partial\mu_{ij}}{\partial\beta^T}(\beta) = \mu'(X_{ij}\beta)X_{ij}$. Therefore, if we assume that $h\circ\mu = \mu'$ (i.e. $v_{ij}(\beta) = \mu'(X_{ij}^T\beta)$), then equation (2.10) becomes:

$$\sum_{i=1}^{n}\sum_{j=1}^{m} X_{ij}(Y_{ij}-\mu_{ij}(\beta)) = 0, \tag{2.11}$$

which is an expanded version of $\sum_{i=1}^{n} X_i^T(Y_i-\mu_i(\beta)) = 0$.

In the general case when, for each i, the random variables $(Y_{ij})_{1\leq j\leq m}$ are not independent, Liang and Zeger proposed to replace the unknown correlation matrix R_i by a correlation matrix $R_i(\alpha)$, which depends on an unknown parameter α. One possibility is to consider

$$R_i(\alpha) = \begin{bmatrix} 1 & \alpha & \ldots & \alpha \\ \alpha & 1 & \ldots & \alpha \\ \vdots & \vdots & \ddots & \vdots \\ \alpha & \alpha & \ldots & 1 \end{bmatrix} \text{ with } \alpha\in(-1,1).$$

The matrix $R_i(\alpha)$ is called the *working correlation matrix*.

Liang and Zeger suggested to use an algorithm, which iterates between the estimation of α and the estimation of β. The idea of this algorithm is the following. We start with the value $\alpha_0 = 0$, solve the GEE (2.8)(with $R_i(\alpha_0)$) for β, then use this value in the GEE to solve for α again, and iterate until convergence (i.e. until the values found for α and β stabilize around some values $\widehat{\alpha}$ and $\widehat{\beta}$).

In [6], Xie and Yang justified the fact that the GEE estimator $\widehat{\beta}$ is consistent and asymptotically normal.

Usually, observations Y_{ij} as in model (2.3), whose mean and variance satisfy hypothesis (2.2) are encountered when working with exponential families. More precisely, suppose that Y_{ij} has density:

$$f(y_{ij}|\theta_{ij}) = \exp[y_{ij}\theta_{ij} - a(\theta_{ij}) + b(y_{ij})], \tag{2.12}$$

where $\theta_{ij} = X_{ij}^T\beta$.

Theorem 2.1.1. *(Theorem 3.4.2 of [1]) Let Y be a random variable with probability density function (p.d.f.) or probability mass function (p.m.f.) of the form*

$$f(y|\theta) = h(y)c(\theta)\exp\left(\sum_{i=1}^{k} w_i(\theta)t_i(y)\right),$$

where $\theta \in \mathbb{R}^p$, $h(y)$ and $c(\theta)$ are positive-valued functions, $w_i(\theta)$ and $t_i(y)$ are real-valued functions, and $c(\theta)$ and $w_i(\theta)$ are twice differentiable, for $i = 1, ..., k$.

Then

$$E_\theta\left(\sum_{i=1}^{k}\frac{\partial w_i(\theta)}{\partial\theta_j}t_i(Y)\right) = -\frac{\partial}{\partial\theta_j}\log[c(\theta)] \quad \text{and}$$

$$\mathrm{Var}_\theta\left(\sum_{i=1}^{k}\frac{\partial w_i(\theta)}{\partial\theta_j}t_i(Y)\right) = -\frac{\partial^2}{\partial\theta_j^2}\log[c(\theta)] - E_\theta\left(\sum_{i=1}^{k}\frac{\partial^2 w_i(\theta)}{\partial\theta_j^2}t_i(Y)\right), \quad j = 1, ..., p.$$

Corollary 2.1.1.1. *When $p = 1$ and $f(y|\theta) = h(y)c(\theta)\exp(\theta y)$ (i.e., $k = 1$, $w_1(\theta) = \theta$, and $t_1(y) = y$), we have*

$$E_\theta(Y) = -\frac{d}{d\theta}\log[c(\theta)] \quad \text{and} \quad \mathrm{Var}_\theta(Y) = -\frac{d^2}{d\theta^2}\log[c(\theta)].$$

To fit (2.12) within the context of Theorem 2.1.1, we identify the functions in (2.12) as:

$$\exp[b(y_{ij})] = h(y),$$

$$\exp[-a(\theta_{ij})] = c(\theta),$$

$$\exp[y_{ij}\theta_{ij}] = \exp[w_1(\theta)t_1(y)].$$

By Theorem 2.1.1, it follows that

$$\mu_{ij}(\beta) = E_\theta(Y_{ij}) = \frac{d[a(\theta_{ij})]}{d\theta_{ij}} = a'(\theta_{ij}) = \mu(\theta_{ij}) \quad \text{and}$$

$$v_{ij}(\beta) = \mathrm{Var}_\theta(Y_{ij}) = \frac{d^2[a(\theta_{ij})]}{d\theta_{ij}^2} = a''(\theta_{ij}) = \mu'(\theta_{ij})$$

with $\mu = a'$. Therefore, in this case $h \circ \mu = \mu'$.

However, in the context of longitudinal measurements, we do not specify the full density of the vector $Y_i = (Y_{i1}, \ldots, Y_{im})^T$. We only specify the marginal density of each component Y_{ij}.

Therefore, the likelihood function of the data $Y_1, ..., Y_n$ cannot be computed. So equation (2.4) is not a likelihood equation.

We discuss three examples of exponential families. As noted above, in these examples, $h \circ \mu = \mu'$. For each of these examples, the roots of the corresponding GEE in the working independence case will be found with `R`, using the function `multiroot` in the library `rootSolve`. The `R` program is given in Appendix B. To simplify the writing in these examples, we do not indicate the dependence on β in μ_{ij} and v_{ij}.

Example 2.1.2. *(Normal Linear Regression for Quantitative Responses)*

Suppose that Y_{ij} has a normal distribution with a mean $\mu_{ij} = X_{ij}^T\beta$ and a known variance $v_{ij} = 1$. Then Y_{ij} has the density function

$$\begin{aligned} f(y_{ij}|\mu_{ij}) &= \frac{1}{\sqrt{2\pi}} \exp\left[-\frac{(y_{ij}-\mu_{ij})^2}{2}\right] \\ &= \frac{1}{\sqrt{2\pi}} \exp\left(y_{ij}\mu_{ij} - \frac{\mu_{ij}^2}{2} - \frac{y_{ij}^2}{2}\right), \quad -\infty < y_{ij} < \infty \end{aligned}$$

Let $\theta_{ij} = X_{ij}^T\beta$. Then

$$f(y_{ij}|\theta_{ij}) = \frac{1}{\sqrt{2\pi}} \exp\left(y_{ij}\theta_{ij} - \frac{\theta_{ij}^2}{2} - \frac{y_{ij}^2}{2}\right).$$

Therefore, $f(y_{ij}|\theta_{ij})$ is of the form (2.12) with

$$a(\theta_{ij}) = \theta_{ij}^2/2 \quad \text{and} \quad b(y_{ij}) = -(y_{ij}^2 + \log(2\pi))/2.$$

By Corollary 2.1.1.1, $\mu(x) = a'(x) = x$ and $\mu'(x) = 1$. The link function is $g(x) = x$. Note that $\mu_{ij} = E(Y_{ij}) = X_{ij}^T\beta$ and $v_{ij} = \text{Var}(Y_{ij}) = 1$.

Equation (2.11) becomes

$$\sum_{i=1}^{n}\sum_{j=1}^{m} X_{ij}(Y_{ij} - X_{ij}^T\beta) = 0. \tag{2.13}$$

Example 2.1.3. *(Log-Linear Regression for Count-type Responses)*

Suppose that Y_{ij} has a Poisson distribution with mean μ_{ij}, where $\log(\mu_{ij}) = X_{ij}^T\beta$. Hence $\mu_{ij} = \exp(X_{ij}^T\beta)$. Then Y_{ij} has the p.m.f.

$$f(y_{ij}|\mu_{ij}) = P(Y_{ij} = y_{ij}|\mu_{ij}) = \frac{e^{-\mu_{ij}}\mu_{ij}^{y_{ij}}}{y_{ij}!}, \quad y_{ij} = 0, 1, 2....$$

Let $\theta_{ij} = X_{ij}^T\beta$. Then

$$f(y_{ij}|\mu_{ij}) = \frac{\exp(-e^{\theta_{ij}} + \theta_{ij}y_{ij})}{y_{ij}!} = \exp(y_{ij}\theta_{ij} - e^{\theta_{ij}} - \log(y_{ij}!)).$$

Therefore, $f(y_{ij}|\theta_{ij})$ is of the form (2.12) with

$$a(\theta_{ij}) = e^{\theta_{ij}} \quad \text{and} \quad b(y_{ij}) = -\log(y_{ij}!).$$

In this case, $\mu(x) = a'(x) = e^x$ and $\mu'(x) = e^x$. Note that

$$\mu_{ij} = E(Y_{ij}) = \exp(X_{ij}^T\beta) \quad \text{and} \quad v_{ij} = \mathrm{Var}(Y_{ij}) = \exp(X_{ij}^T\beta).$$

The link function is $g(x) = \log x$.

Equation (2.11) becomes

$$\sum_{i=1}^{n}\sum_{j=1}^{m} X_{ij}(Y_{ij} - \exp(X_{ij}^T\beta)) = 0. \tag{2.14}$$

Example 2.1.4. *(Logistic Regression for Binary Responses)*

Suppose that each response Y_{ij} has a Bernoulli distribution with mean μ_{ij}, where $\mathrm{logit}(\mu_{ij}) = X_{ij}^T\beta$. Recall that the logit function is given by:

$$\mathrm{logit}(x) = \log\frac{x}{1-x}, \quad x \in (0, 1),$$

and its inverse function is $\mathrm{logit}^{-1}(y) = e^y/(1 + e^y)$.

Then $\mu_{ij}(\beta) = \exp(X_{ij}^T\beta)/(1+\exp(X_{ij}^T\beta))$, and Y_{ij} has the probability mass function

$$f(y_{ij}|\mu_{ij}) = P(Y_{ij} = y_{ij}|\mu_{ij}) = \mu_{ij}^{y_{ij}}(1 - \mu_{ij})^{1-y_{ij}}, \quad y_{ij} = 0, 1.$$

Let $\theta_{ij} = X_{ij}^T\beta$. Then

$$f(y_{ij}|\mu_{ij}) = \left(\frac{e^{\theta_{ij}}}{1 + e^{\theta_{ij}}}\right)^{y_{ij}} (1 + e^{\theta_{ij}})^{y_{ij}-1} = \exp(y_{ij}\theta_{ij} - \log(1 + e^{\theta_{ij}})).$$

Therefore, $f(y_{ij}|\mu_{ij})$ is of the form (2.12) with

$$a(\theta_{ij}) = \log(1 + e^{\theta_{ij}}) \quad \text{and} \quad b(y_{ij}) = 0.$$

In this case, $\mu(x) = a'(x) = e^x/(1+e^x)$ and $\mu'(x) = e^x/(1+e^x)^2$. Note that

$$\mu_{ij} = E(Y_{ij}) = \frac{\exp(X_{ij}^T\beta)}{1+\exp(X_{ij}^T\beta)} \quad \text{and} \quad v_{ij} = \text{Var}(Y_{ij}) = \frac{\exp(X_{ij}^T\beta)}{(1+\exp(X_{ij}^T\beta))^2}.$$

The link function is $g(x) = \text{logit}(x)$. Equation (2.11) becomes

$$\sum_{i=1}^{n}\sum_{j=1}^{m} X_{ij}\left(Y_{ij} - \frac{\exp(X_{ij}^T\beta)}{1+\exp(X_{ij}^T\beta)}\right) = 0. \tag{2.15}$$

2.2 Simulations

In this section, we consider separately the three examples mentioned above and we find the root of the corresponding GEE in the working independence case using the `multiroot` function in the library `rootSolve` of `R`. In these simulations, the covariates will be random, and we assume that $p = 2$ and $m = 3$. The true value of the regression parameter is denoted by β_0.

2.2.1 Normal Linear Regression

In this section, we assume that the conditional distribution of Y_{ij} given X_{ij} is normal with mean $\mu_{ij} = X_{ij}^T\beta$ and variance 1. In this case, we have seen that $\mu(x) = x$ and equation (2.11) becomes (2.13). This equation has a unique solution given by

$$\widehat{\beta} = (\sum_{i=1}^{n}\sum_{j=1}^{m} X_{ij}X_{ij}^T)^{-1}(\sum_{i=1}^{n}\sum_{j=1}^{m} X_{ij}Y_{ij}).$$

In the case when $p = 2$ and $m = 3$, equation (2.13) is a system of $p = 2$ equations given by:

$$\begin{cases} \sum_{i=1}^n X_{i1}^{(1)}(Y_{i1} - X_{i1}^{(1)}\beta_1 - X_{i1}^{(2)}\beta_2) \\ +\sum_{i=1}^n X_{i2}^{(1)}(Y_{i2} - X_{i2}^{(1)}\beta_1 - X_{i2}^{(2)}\beta_2) \\ +\sum_{i=1}^n X_{i3}^{(1)}(Y_{i3} - X_{i3}^{(1)}\beta_1 - X_{i3}^{(2)}\beta_2) = 0 \\ \sum_{i=1}^n X_{i1}^{(2)}(Y_{i1} - X_{i1}^{(1)}\beta_1 - X_{i1}^{(2)}\beta_2) \\ +\sum_{i=1}^n X_{i2}^{(2)}(Y_{i2} - X_{i2}^{(1)}\beta_1 - X_{i2}^{(2)}\beta_2) \\ +\sum_{i=1}^n X_{i3}^{(2)}(Y_{i3} - X_{i3}^{(1)}\beta_1 - X_{i3}^{(2)}\beta_2) = 0 \end{cases} \tag{2.16}$$

Recall that

$$\begin{cases} X_{i1} = (X_{i1}^{(1)}, X_{i1}^{(2)})^T & \text{are the covariates of individual } i \text{ at time } j = 1 \\ X_{i2} = (X_{i2}^{(1)}, X_{i2}^{(2)})^T & \text{are the covariates of individual } i \text{ at time } j = 2 \\ X_{i3} = (X_{i3}^{(1)}, X_{i3}^{(2)})^T & \text{are the covariates of individual } i \text{ at time } j = 3, \end{cases}$$

and Y_{ij} is the response of individual i at time j, for $j = 1, 2, 3$.

In our `R` program, the covariates are generated separately from the standard normal distribution. Below is the explanation of the `R` program given in Appendix A.1.

1. a) We generate the covariates for the n individuals at time $j = 1$:

$$\begin{cases} X_{i1}^{(1)}, i = 1, ..., n, & \text{are independent and identically distributed } N(0,1) \\ X_{i1}^{(2)}, i = 1, ..., n, & \text{are independent and identically distributed } N(0,1). \end{cases}$$

b) We generate the covariates for the n individuals at time $j = 2$:

$$\begin{cases} X_{i2}^{(1)}, i = 1, ..., n, & \text{are independent and identically distributed } N(0,1) \\ X_{i2}^{(2)}, i = 1, ..., n, & \text{are independent and identically distributed } N(0,1). \end{cases}$$

c) We generate the covariates for the n individuals at time $j = 3$:

$$\begin{cases} X_{i3}^{(1)}, i = 1, ..., n, & \text{are independent and identically distributed } N(0,1) \\ X_{i3}^{(2)}, i = 1, ..., n, & \text{are independent and identically distributed } N(0,1). \end{cases}$$

2. a) We generate the responses for the n individuals at time $j = 1$: we generate Y_{i1}, $i = 1, ..., n$, such that the distribution of Y_{i1} given $(X_{i1}^{(1)}, X_{i2}^{(2)})$ is $N(\mu_{i1}, 1)$ with

$$\mu_{i1} = X_{i1}^T \beta_0 = X_{i1}^{(1)} \beta_{01} + X_{i1}^{(2)} \beta_{02},$$

where $(X_{i1}^{(1)}, X_{i2}^{(2)})$ have been generated in Step 1. a) above and $\beta_0 = (\beta_{01}, \beta_{02})$ is the true parameter. We take $\beta_{01} = 0.2$ and $\beta_{02} = 0.5$.

b) Similarly to a), we generate the responses for the n individuals at time $j = 2$; we generate Y_{i2}, $i = 1, ..., n$, such that the distribution of Y_{i2} given $(X_{i2}^{(1)}, X_{i2}^{(2)})$ is $N(\mu_{i2}, 1)$ with

$$\mu_{i2} = X_{i2}^T \beta_0 = X_{i2}^{(1)} \beta_{01} + X_{i2}^{(2)} \beta_{02},$$

where $(X_{i2}^{(1)}, X_{i2}^{(2)})$ have been generated in Step 1. b) above and we keep $\beta_{01} = 0.2$ and $\beta_{02} = 0.5$ the same as above.

c) Similarly, we generate the responses for the n individuals at time $j = 3$; we generate Y_{i3}, $i = 1, ..., n$, such that the distribution of Y_{i3} given $(X_{i3}^{(1)}, X_{i3}^{(2)})$ is $N(\mu_{i3}, 1)$ with

$$\mu_{i3} = X_{i3}^T \beta_0 = X_{i3}^{(1)} \beta_{01} + X_{i3}^{(2)} \beta_{02},$$

where $(X_{i3}^{(1)}, X_{i3}^{(2)})$ have been generated in Step 1. c) above and we keep $\beta_{01} = 0.2$ and $\beta_{02} = 0.5$ as same as above.
3. We solve the system (2.16). The root $\widehat{\beta} = (\widehat{\beta}_1, \widehat{\beta}_2)$ is the GEE estimator of β in the working independence case.
4. We calculate the bias of $\widehat{\beta}$: $\text{Bias}(\widehat{\beta}_1) = \widehat{\beta}_1 - \beta_{01}$ and $\text{Bias}(\widehat{\beta}_2) = \widehat{\beta}_2 - \beta_{02}$.
5. We repeat Steps 1 - 4 above for $N = 1000$ times.

Now we compute the mean bias for the $N = 1000$ simulations:

$$\overline{\text{Bias}_1} = \frac{1}{N}\sum_{k=1}^{N} \text{Bias}_k(\widehat{\beta}_1)$$

$$\overline{\text{Bias}_2} = \frac{1}{N}\sum_{k=1}^{N} \text{Bias}_k(\widehat{\beta}_2),$$

where $\text{Bias}_k(\widehat{\beta}_1)$ and $\text{Bias}_k(\widehat{\beta}_2)$ denote the values determined at the k-th iteration. We can also compute the absolute Bias:

$$\text{AB}_1 = \frac{1}{N}\sum_{k=1}^{N} |\text{Bias}_k(\widehat{\beta}_1)|$$

$$\text{AB}_2 = \frac{1}{N}\sum_{k=1}^{N} |\text{Bias}_k(\widehat{\beta}_2)|.$$

We compute the standard error for the N simulations:

$$\text{SE}_1 = \sqrt{\frac{1}{N-1}\sum_{k=1}^{N}(\text{Bias}_k(\widehat{\beta}_1) - \overline{\text{Bias}_1})^2}$$

$$\text{SE}_2 = \sqrt{\frac{1}{N-1}\sum_{k=1}^{N}(\text{Bias}_k(\widehat{\beta}_2) - \overline{\text{Bias}_2})^2}.$$

Finally, we compute the mean squared error (MSE) for the N simulations:

$$\text{MSE}_1 = (\overline{\text{Bias}_1})^2 + (\text{SE}_1)^2$$

$$\text{MSE}_2 = (\overline{\text{Bias}_2})^2 + (\text{SE}_2)^2.$$

We report these values in the following table:

	Mean Bias	Absolute Bias	Standard Error	Mean Squared Error
Covariate1	0.0000788074	0.04636471	0.05862726	0.003437162
Covariate2	0.003399301	0.04646929	0.05773746	0.00334517

2.2.2 Poisson Regression

In this section, we assume that the conditional distribution of Y_{ij} given X_{ij} is Poisson with mean $\mu_{ij} = \exp(X_{ij}^T\beta)$, i.e. $\log(\mu_{ij}) = X_{ij}^T\beta$. In this case, we have seen that $\mu(x) = e^x$ and equation (2.11) becomes (2.14). There is no explicit formula for the root of this equation.

In the case when $p = 2$ and $m = 3$, equation (2.14) is a system of $p = 2$ equations given by:

$$\left\{ \begin{array}{l} \sum_{i=1}^n X_{i1}^{(1)}(Y_{i1} - \exp(X_{i1}^{(1)}\beta_1 + X_{i1}^{(2)}\beta_2)) \\ + \sum_{i=1}^n X_{i2}^{(1)}(Y_{i2} - \exp(X_{i2}^{(1)}\beta_1 + X_{i2}^{(2)}\beta_2)) \\ + \sum_{i=1}^n X_{i3}^{(1)}(Y_{i3} - \exp(X_{i3}^{(1)}\beta_1 + X_{i3}^{(2)}\beta_2)) = 0 \\ \sum_{i=1}^n X_{i1}^{(2)}(Y_{i1} - \exp(X_{i1}^{(1)}\beta_1 + X_{i1}^{(2)}\beta_2)) \\ + \sum_{i=1}^n X_{i2}^{(2)}(Y_{i2} - \exp(X_{i2}^{(1)}\beta_1 + X_{i2}^{(2)}\beta_2)) \\ + \sum_{i=1}^n X_{i3}^{(2)}(Y_{i3} - \exp(X_{i3}^{(1)}\beta_1 + X_{i3}^{(2)}\beta_2)) = 0 \end{array} \right. \tag{2.17}$$

Recall that

$$\left\{ \begin{array}{ll} X_{i1} = (X_{i1}^{(1)}, X_{i1}^{(2)})^T & \text{are the covariates of individual } i \text{ at time } j = 1 \\ X_{i2} = (X_{i2}^{(1)}, X_{i2}^{(2)})^T & \text{are the covariates of individual } i \text{ at time } j = 2 \\ X_{i3} = (X_{i3}^{(1)}, X_{i3}^{(2)})^T & \text{are the covariates of individual } i \text{ at time } j = 3, \end{array} \right.$$

and Y_{ij} is the response of individual i at time j, for $j = 1, 2, 3$.

In our `R` program, the covariates are generated separately from the standard normal distribution. Below is the explanation of the R program given in Appendix A.2.

1. a) We generate the covariates for the n individuals at time $j = 1$:

$$\left\{ \begin{array}{ll} X_{i1}^{(1)}, i = 1, ..., n, & \text{are independent and identically distributed } N(0, 1) \\ X_{i1}^{(2)}, i = 1, ..., n, & \text{are independent and identically distributed } N(0, 1). \end{array} \right.$$

b) We generate the covariates for the n individuals at time $j = 2$:

$$\left\{ \begin{array}{ll} X_{i2}^{(1)}, i = 1, ..., n, & \text{are independent and identically distributed } N(0, 1) \\ X_{i2}^{(2)}, i = 1, ..., n, & \text{are independent and identically distributed } N(0, 1). \end{array} \right.$$

c) We generate the covariates for the n individuals at time $j = 3$:

$$\left\{ \begin{array}{ll} X_{i3}^{(1)}, i = 1, ..., n, & \text{are independent and identically distributed } N(0, 1) \\ X_{i3}^{(2)}, i = 1, ..., n, & \text{are independent and identically distributed } N(0, 1). \end{array} \right.$$

2. a) We generate the responses for the n individuals at time $j = 1$: we generate Y_{i1}, $i = 1, ..., n$, such that the distribution of Y_{i1} given $(X_{i1}^{(1)}, X_{i2}^{(2)})$ is Poisson with mean

μ_{i1}, where

$$\mu_{i1} = \exp(X_{i1}^T \beta_0) = \exp(X_{i1}^{(1)} \beta_{01} + X_{i1}^{(2)} \beta_{02}),$$

$(X_{i1}^{(1)}, X_{i2}^{(2)})$ have been generated in Step 1. a) above, and $\beta_0 = (\beta_{01}, \beta_{02})$ is the true parameter. We take $\beta_{01} = 0.2$ and $\beta_{02} = 0.5$.

b) Similarly to a), we generate the responses for the n individuals at time $j = 2$; we generate Y_{i2}, $i = 1, ..., n$, such that the distribution of Y_{i2} given $(X_{i2}^{(1)}, X_{i2}^{(2)})$ is Poisson with mean μ_{i2}, where

$$\mu_{i2} = \exp(X_{i2}^T \beta_0) = \exp(X_{i2}^{(1)} \beta_{01} + X_{i2}^{(2)} \beta_{02}),$$

and $(X_{i2}^{(1)}, X_{i2}^{(2)})$ have been generated in Step 1. b) above. We keep $\beta_{01} = 0.2$ and $\beta_{02} = 0.5$ the same as above.

c) Similarly, we generate the responses for the n individuals at time $j = 3$; we generate Y_{i3}, $i = 1, ..., n$, such that the distribution of Y_{i3} given $(X_{i3}^{(1)}, X_{i3}^{(2)})$ is Poisson with mean μ_{i3}, where

$$\mu_{i3} = \exp(X_{i3}^T \beta_0) = \exp(X_{i3}^{(1)} \beta_{01} + X_{i3}^{(2)} \beta_{02}),$$

and $(X_{i3}^{(1)}, X_{i3}^{(2)})$ have been generated in Step 1. c) above. We keep $\beta_{01} = 0.2$ and $\beta_{02} = 0.5$ the same as above.

3. We solve the system (2.17). The root $\widehat{\beta} = (\widehat{\beta}_1, \widehat{\beta}_2)$ is the GEE estimator of β in the working independence case.

4. We calculate the bias of $\widehat{\beta}$: $\text{Bias}(\widehat{\beta}_1) = \widehat{\beta}_1 - \beta_{01}$ and $\text{Bias}(\widehat{\beta}_2) = \widehat{\beta}_2 - \beta_{02}$.

5. We repeat Steps 1 - 4 above for $N = 1000$ times.

Now we compute the mean bias for the $N = 1000$ simulations:

$$\overline{\text{Bias}}_1 = \frac{1}{N} \sum_{k=1}^{N} \text{Bias}_k(\widehat{\beta}_1)$$

$$\overline{\text{Bias}}_2 = \frac{1}{N} \sum_{k=1}^{N} \text{Bias}_k(\widehat{\beta}_2),$$

where $\text{Bias}_k(\widehat{\beta}_1)$ and $\text{Bias}_k(\widehat{\beta}_2)$ denote the values determined at the k-th iteration. We can also compute the absolute Bias:

$$\text{AB}_1 = \frac{1}{N} \sum_{k=1}^{N} |\text{Bias}_k(\widehat{\beta}_1)|$$

$$\text{AB}_2 = \frac{1}{N} \sum_{k=1}^{N} |\text{Bias}_k(\widehat{\beta}_2)|.$$

We compute the standard error for the N simulations:

$$\mathrm{SE}_1 = \sqrt{\frac{1}{N-1}\sum_{k=1}^{N}(\mathrm{Bias}_k(\widehat{\beta}_1) - \overline{\mathrm{Bias}}_1)^2}$$

$$\mathrm{SE}_2 = \sqrt{\frac{1}{N-1}\sum_{k=1}^{N}(\mathrm{Bias}_k(\widehat{\beta}_2) - \overline{\mathrm{Bias}}_2)^2}.$$

Finally, we compute the mean squared error (MSE) for the N simulations:

$$\mathrm{MSE}_1 = (\overline{\mathrm{Bias}}_1)^2 + (\mathrm{SE}_1)^2$$

$$\mathrm{MSE}_2 = (\overline{\mathrm{Bias}}_2)^2 + (\mathrm{SE}_2)^2.$$

We report these values in the following table:

	Mean Bias	Absolute Bias	Standard Error	Mean Squared Error
Covariate1	0.0004880474	0.04261351	0.05293236	0.002802073
Covariate2	-0.002170054	0.04012984	0.04998254	0.002502964

2.2.3 Logistic Regression

In this section, we assume that the conditional distribution of Y_{ij} given X_{ij} is Bernoulli with mean $\mu_{ij} = \exp(X_{ij}^T\beta)/(1+\exp(X_{ij}^T\beta))$, i.e. $\mathrm{logit}(\mu_{ij}) = X_{ij}^T\beta$. In this case, we have seen that $\mu(x) = e^x/(1+e)$ and equation (2.11) becomes (2.15). This equation cannot be solved explicitly.

In the case when $p = 2$ and $m = 3$, equation (2.14) is a system of $p = 2$ equations given by:

$$\begin{cases} \sum_{i=1}^{n} X_{i1}^{(1)}(Y_{i1} - \exp(X_{i1}^{(1)}\beta_1 + X_{i1}^{(2)}\beta_2)/(1+\exp(X_{i1}^{(1)}\beta_1 + X_{i1}^{(2)}\beta_2))) \\ +\sum_{i=1}^{n} X_{i2}^{(1)}(Y_{i2} - \exp(X_{i2}^{(1)}\beta_1 + X_{i2}^{(2)}\beta_2)/(1+\exp(X_{i2}^{(1)}\beta_1 + X_{i2}^{(2)}\beta_2))) \\ +\sum_{i=1}^{n} X_{i3}^{(1)}(Y_{i3} - \exp(X_{i3}^{(1)}\beta_1 + X_{i3}^{(2)}\beta_2)/(1+\exp(X_{i3}^{(1)}\beta_1 + X_{i3}^{(2)}\beta_2))) = 0 \\ \sum_{i=1}^{n} X_{i1}^{(2)}(Y_{i1} - \exp(X_{i1}^{(1)}\beta_1 + X_{i1}^{(2)}\beta_2)/(1+\exp(X_{i1}^{(1)}\beta_1 + X_{i1}^{(2)}\beta_2))) \\ +\sum_{i=1}^{n} X_{i2}^{(2)}(Y_{i2} - \exp(X_{i2}^{(1)}\beta_1 + X_{i2}^{(2)}\beta_2)/(1+\exp(X_{i2}^{(1)}\beta_1 + X_{i2}^{(2)}\beta_2))) \\ +\sum_{i=1}^{n} X_{i3}^{(2)}(Y_{i3} - \exp(X_{i3}^{(1)}\beta_1 + X_{i3}^{(2)}\beta_2)/(1+\exp(X_{i3}^{(1)}\beta_1 + X_{i3}^{(2)}\beta_2))) = 0 \end{cases} \tag{2.18}$$

Recall that

$$\begin{cases} X_{i1} = (X_{i1}^{(1)}, X_{i1}^{(2)})^T & \text{are the covariates of individual } i \text{ at time } j = 1 \\ X_{i2} = (X_{i2}^{(1)}, X_{i2}^{(2)})^T & \text{are the covariates of individual } i \text{ at time } j = 2 \\ X_{i3} = (X_{i3}^{(1)}, X_{i3}^{(2)})^T & \text{are the covariates of individual } i \text{ at time } j = 3, \end{cases}$$

and Y_{ij} is the response of individual i at time j, for $j = 1, 2, 3$.

In our R program, the covariates are generated separately from the standard normal distribution. Below is the explanation of the R program given in Appendix A.3.

1. a) We generate the covariates for the n individuals at time $j = 1$:

$$\begin{cases} X_{i1}^{(1)}, i = 1, ..., n, & \text{are independent and identically distributed } N(0,1) \\ X_{i1}^{(2)}, i = 1, ..., n, & \text{are independent and identically distributed } N(0,1). \end{cases}$$

b) We generate the covariates for the n individuals at time $j = 2$:

$$\begin{cases} X_{i2}^{(1)}, i = 1, ..., n, & \text{are independent and identically distributed } N(0,1) \\ X_{i2}^{(2)}, i = 1, ..., n, & \text{are independent and identically distributed } N(0,1). \end{cases}$$

c) We generate the covariates for the n individuals at time $j = 3$:

$$\begin{cases} X_{i3}^{(1)}, i = 1, ..., n, & \text{are independent and identically distributed } N(0,1) \\ X_{i3}^{(2)}, i = 1, ..., n, & \text{are independent and identically distributed } N(0,1). \end{cases}$$

2. a) We generate the responses for the n individuals at time $j = 1$: we generate Y_{i1}, $i = 1, ..., n$, such that the distribution of Y_{i1} given $(X_{i1}^{(1)}, X_{i2}^{(2)})$ is Bernoulli with probability of success μ_{i1}, where

$$\mu_{i1} = \exp(X_{i1}^T\beta_0)/(1+\exp(X_{i1}^T\beta_0)) = \exp(X_{i1}^{(1)}\beta_{01}+X_{i1}^{(2)}\beta_{02})/(1+\exp(X_{i1}^{(1)}\beta_{01}+X_{i1}^{(2)}\beta_{02})),$$

$(X_{i1}^{(1)}, X_{i2}^{(2)})$ have been generated in Step 1. a) above, and $\beta_0 = (\beta_{01}, \beta_{02})$ is the true parameter. We take $\beta_{01} = 0.2$ and $\beta_{02} = 0.5$.

b) Similarly to a), we generate the responses for the n individuals at time $j = 2$; we generate Y_{i2}, $i = 1, ..., n$, such that the distribution of Y_{i2} given $(X_{i2}^{(1)}, X_{i2}^{(2)})$ is Bernoulli with probability of success μ_{i2}, where

$$\mu_{i2} = \exp(X_{i2}^T\beta_0)/(1+\exp(X_{i2}^T\beta_0)) = \exp(X_{i2}^{(1)}\beta_{01}+X_{i2}^{(2)}\beta_{02})/(1+\exp(X_{i2}^{(1)}\beta_{01}+X_{i2}^{(2)}\beta_{02})),$$

and $(X_{i2}^{(1)}, X_{i2}^{(2)})$ have been generated in Step 1. b) above. We keep $\beta_{01} = 0.2$ and $\beta_{02} = 0.5$ the same as above.

c) Similarly, we generate the responses for the n individuals at time $j = 3$; we generate Y_{i3}, $i = 1, ..., n$, such that the distribution of Y_{i3} given $(X_{i3}^{(1)}, X_{i3}^{(2)})$ is Bernoulli with probability of success μ_{i3}, where

$$\mu_{i3} = \exp(X_{i3}^T\beta_0)/(1+\exp(X_{i3}^T\beta_0)) = \exp(X_{i3}^{(1)}\beta_{01}+X_{i3}^{(2)}\beta_{02})/(1+\exp(X_{i3}^{(1)}\beta_{01}+X_{i3}^{(2)}\beta_{02})),$$

and $(X_{i3}^{(1)}, X_{i3}^{(2)})$ have been generated in Step 1. c) above. We keep $\beta_{01} = 0.2$ and $\beta_{02} = 0.5$ the same as above.

3. We solve the system (2.18). The root $\widehat{\beta} = (\widehat{\beta}_1, \widehat{\beta}_2)$ is the GEE estimator of β in the working independence case.

4. We calculate the bias of $\widehat{\beta}$: $\text{Bias}(\widehat{\beta}_1) = \widehat{\beta}_1 - \beta_{01}$ and $\text{Bias}(\widehat{\beta}_2) = \widehat{\beta}_2 - \beta_{02}$.

5. We repeat the steps from 1 to 4 above for $N = 1000$ times.

Now we compute the mean bias for the $N = 1000$ simulations:

$$\overline{\text{Bias}_1} = \frac{1}{N} \sum_{k=1}^{N} \text{Bias}_k(\widehat{\beta}_1)$$

$$\overline{\text{Bias}_2} = \frac{1}{N} \sum_{k=1}^{N} \text{Bias}_k(\widehat{\beta}_2),$$

where $\text{Bias}_k(\widehat{\beta}_1)$ and $\text{Bias}_k(\widehat{\beta}_2)$ denote the values determined at the k-th iteration. We can also compute the absolute Bias:

$$\text{AB}_1 = \frac{1}{N} \sum_{k=1}^{N} |\text{Bias}_k(\widehat{\beta}_1)|$$

$$\text{AB}_2 = \frac{1}{N} \sum_{k=1}^{N} |\text{Bias}_k(\widehat{\beta}_2)|.$$

We compute the standard error for the N simulations:

$$\text{SE}_1 = \sqrt{\frac{1}{N-1} \sum_{k=1}^{N} (\text{Bias}_k(\widehat{\beta}_1) - \overline{\text{Bias}_1})^2}$$

$$\text{SE}_2 = \sqrt{\frac{1}{N-1} \sum_{k=1}^{N} (\text{Bias}_k(\widehat{\beta}_2) - \overline{\text{Bias}_2})^2}.$$

Finally, we compute the mean squared error (MSE) for the N simulations:

$$\text{MSE}_1 = (\overline{\text{Bias}_1})^2 + (\text{SE}_1)^2$$

$$\text{MSE}_2 = (\overline{\text{Bias}_2})^2 + (\text{SE}_2)^2.$$

We report these values in the following table:

	Mean Bias	Absolute Bias	Standard Error	Mean Squared Error
Covariate1	0.004727172	0.09930893	0.1257461	0.01583442
Covariate2	0.01805998	0.103723	0.1307979	0.01743426

Chapter 3

Longitudinal Studies with Missing Responses

A longitudinal study is an observational study in which the individuals are measured repeatedly throughout the duration of the study. During the study period, missing data can be encountered in many situations. For such a study, we can introduce a new random variable R_{ij}, which is the missingness indicator for the i^{th} individual at time j, i.e.

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

Note that we observe Y_{ij} only if $R_{ij} = 1$.

In this chapter, we assume that the covariates X_{ij} are p-dimensional random vectors. We also assume that all covariates are observed exactly without measurement error. The data consists of:

$$(R_{ij}Y_{ij}, X_{ij}, R_{ij}) \quad \text{for} \quad 1 \leq i \leq n \quad \text{and} \quad 1 \leq j \leq m.$$

Here i denotes the index of the individual and j is the time of the measurement. As in the previous chapter, we denote $R_i = (R_{i1}, ..., R_{im})$, $Y_i = (Y_{i1}, ..., Y_{im})^T$, and X_i is the $m \times p$ matrix of covariates given by (2.1). We assume that (Y_i, X_i, R_i), $i = 1, ..., n$ are independent and identically distributed.

It is important to emphasize the distinction between *full data, observed data*, and *complete data*, as these terminologies are often used with missing data. *Full data* are the data that we would want to have collected on all the individuals in the sample. *Observed data* are the data that are actually observed on the individuals in the study, some of which are missing. *Complete data* are the data from only the subset of individuals with no missing data.

As in the previous chapter, we consider the (conditional) marginal model:

$$\mu_{ij}(\beta) = E(Y_{ij}|X_{ij}) = \mu(X_{ij}^T\beta) \quad \text{and} \quad v_{ij}(\beta) = \text{Var}(Y_{ij}|X_{ij}) = h(\mu(X_{ij}^T\beta)).$$

Note that μ_{ij} and v_{ij} are random variables, which depend also on β. Recall that the model is given in (2.3).

3.1 Missing at Random for Classical Regression

In this section we review some standard techniques used for treating missing data in the case of classical regression (i.e. when $m = 1$). In this case, the observed data consists of:

$$(R_iY_i, X_i, R_i),\ i = 1, ..., n,$$

where Y_i is the response of the i^{th} individual, X_i is the covariate of this individual, and R_i is the missingness indicator:

$$R_i = \begin{cases} 1 & \text{if } Y_i \text{ is observed} \\ 0 & \text{if } Y_i \text{ is missing.} \end{cases}$$

In many practical situations, it is natural to assume that the reason for missingness of the i^{th} response depends on the covariate X_i, and Y_i has no additional effect on the probability of missingness. Formally, we have the following definition:

Definition 3.1.1. We say that the responses Y_i, $i = 1, ..., n$ are *missing at random* (MAR), if each Y_i is conditionally independent of the missingness indicator variable R_i, given the covariate X_i. In this case, we write

$$\text{(MAR)} \qquad R_i \perp\!\!\!\perp Y_i \,|\, X_i.$$

Under assumption (MAR), we have

$$E(Y_i|R_i, X_i) = E(Y_i|X_i). \tag{3.1}$$

3.1.1 Likelihood Methods

We review briefly the likelihood method under the MAR assumption as presented in Section 6.2 of [5]. This method will not be used in the present work.

We consider a parametric model of the form

$$f_{Y_i,X_i}(y_i,x_i) = f_{Y_i|X_i}(y_i|x_i,\gamma_1)f_{X_i}(x_i|\gamma_2), \tag{3.2}$$

where f_{Y_i,X_i} denotes the joint density of (Y_i,X_i), $f_{Y_i|X_i}$ is the conditional density of Y_i given X_i, and f_{X_i} is the density of X_i. We assume that γ_1 and γ_2 are unknown parameters.

The density function of (R_iY_i,X_i,R_i) is:

$$f_{R_iY_i,X_i,R_i}(r_iy_i,x_i,r_i) = \begin{cases} f_{Y_i,X_i,R_i}(y_i,x_i,1) & \text{if } r_i = 1 \\ f_{X_i,R_i}(x_i,0) & \text{if } r_i = 0, \end{cases}$$

where f_{Y_i,X_i,R_i} is the joint density of (Y_i,X_i,R_i) and f_{X_i,R_i} is the joint density of (X_i,R_i). Note that we can write

$$f_{R_iY_i,X_i,R_i}(r_iy_i,x_i,r_i) = f_{Y_i,X_i,R_i}(y_i,x_i,r_i)^{r_i}f_{X_i,R_i}(x_i,r_i)^{1-r_i} \tag{3.3}$$

for $r_i \in \{0,1\}$.

The joint density of (Y_i,X_i,R_i) can be calculated as:

$$\begin{aligned} f_{Y_i,X_i,R_i}(y_i,x_i,r_i) &= f_{Y_i|(X_i,R_i)}(y_i|x_i,r_i)f_{X_i,R_i}(x_i,r_i) \\ &= f_{Y_i|(X_i,R_i)}(y_i|x_i,r_i)f_{R_i|X_i}(r_i|x_i)f_{X_i}(x_i|\gamma_2), \end{aligned}$$

where $f_{Y_i|(X_i,R_i)}$ is the conditional density of Y_i given (X_i,R_i), f_{X_i,R_i} is the joint density of (X_i,R_i), and $f_{R_i|X_i}$ is the conditional density of R_i given X_i. For the second equality, we used:

$$f_{X_i,R_i}(x_i,r_i) = f_{R_i|X_i}(r_i|x_i)f_{X_i}(x_i|\gamma_2). \tag{3.4}$$

Under the MAR assumption and the parametric model (3.2),

$$f_{Y_i|X_i,R_i}(y_i|x_i,r_i) = f_{Y_i|X_i}(y_i|x_i,\gamma_1),$$

and therefore

$$f_{Y_i,X_i,R_i}(y_i,x_i,r_i) = f_{Y_i|X_i}(y_i|x_i,\gamma_1)f_{R_i|X_i}(r_i|x_i)f_{X_i}(x_i|\gamma_2). \tag{3.5}$$

Substituting (3.5) and (3.4) into (3.3), we obtain that the density function of (R_iY_i,X_i,R_i) is

$$\begin{aligned} f_{R_iY_i,X_i,R_i}(r_iy_i,x_i,r_i) &= [f_{Y_i|X_i}(y_i|x_i,\gamma_1)f_{R_i|X_i}(r_i|x_i)f_{X_i}(x_i|\gamma_2)]^{r_i} \cdot [f_{R_i|X_i}(r_i|x_i)f_{X_i}(x_i|\gamma_2)]^{1-r_i} \\ &= f_{Y_i|X_i}(y_i|x_i,\gamma_1)^{r_i}f_{R_i|X_i}(r_i|x_i)f_{X_i}(x_i|\gamma_2). \end{aligned}$$

We now impose an additional parametric model on the missingness mechanism:

$$\pi(x_i|\gamma_3) := P(R_i = 1|X_i = x_i) = f_{R_i|X_i}(1|x_i),$$

$$\text{and} \quad f_{R_i|X_i}(0|x_i) = P(R_i = 0|X_i = x_i) = 1 - \pi(x_i|\gamma_3).$$

In a compact notation, we have:

$$f_{R_i|X_i}(r_i|x_i) = \pi(x_i|\gamma_3)^{r_i}(1 - \pi(x_i|\gamma_3))^{1-r_i} \quad \text{for} \quad r_i \in \{0, 1\}.$$

Hence

$$f_{R_iY_i,X_i,R_i}(r_iy_i, x_i, r_i) = f_{Y_i|X_i}(y_i|x_i, \gamma_1)^{r_i} f_{X_i}(x_i|\gamma_2) \cdot \pi(x_i|\gamma_3)^{r_i} \cdot (1 - \pi(x_i|\gamma_3))^{1-r_i}.$$

Therefore, the likelihood function of the data (i.e. the joint density of (R_iY_i, X_i, R_i), $i = 1, ..., n$) is:

$$\begin{aligned} L(\gamma_1, \gamma_2, \gamma_3) &= \prod_{i=1}^{n} f_{R_iY_i,X_i,R_i}(r_iy_i, x_i, r_i) \\ &= \prod_{i=1}^{n} f_{Y_i|X_i}(y_i|x_i, \gamma_1)^{r_i} \cdot \prod_{i=1}^{n} f_{X_i}(x_i|\gamma_2) \cdot \prod_{i=1}^{n} [\pi(x_i|\gamma_3)^{r_i}(1 - \pi(x_i|\gamma_3))^{1-r_i}] \\ &= f_1(\gamma_1) f_2(\gamma_2) f_3(\gamma_3). \end{aligned} \tag{3.6}$$

The maximum likelihood estimators $\widehat{\gamma}_1$, $\widehat{\gamma}_2$, $\widehat{\gamma}_3$ of γ_1, γ_2, γ_3 respectively are found by maximizing separately the three functions f_1, f_2, f_3 above. Note that $\widehat{\gamma}_1$ uses only the complete cases (i.e. those y_i's for which $r_i = 1$), while $\widehat{\gamma}_2$ uses all the data.

3.1.2 Logit Model for the Missing Probability

We consider the random variable

$$\pi_i := P(R_i = 1|Y_i, X_i) = E(R_i|Y_i, X_i).$$

Note that $P(R_i = 0|Y_i, X_i) = 1 - \pi_i$.

Under assumption (MAR), π_i depends only on X_i:

$$\pi_i = P(R_i = 1|X_i) = E(R_i|X_i), \quad \text{and}$$

$$1 - \pi_i = P(R_i = 0|X_i).$$

To summarize, under assumption (MAR),

$$P(R_i = r_i|Y_i, X_i) = P(R_i = r_i|X_i) = \pi_i^{r_i}(1-\pi_i)^{1-r_i} \quad \text{for} \quad r_i \in \{0,1\}.$$

Under (MAR), conditionally on X_i, R_i is a Bernoulli random variable with probability of success π_i. A commonly used technique is to use a logit model for the distribution of R_i given X_i, namely

$$\text{logit}(\pi_i) = X_i^T\alpha, \tag{3.7}$$

where $X_i = X_{i1}$ in (2.1) and $\alpha = (\alpha_1, ..., \alpha_p)^T$ is a p-dimensional parameter. Hence

$$\pi_i = \exp(X_i^T\alpha)/(1+\exp(X_i^T\alpha)).$$

In this case, one possibility for estimating α is to use a "likelihood-type" method. We explain this method below.

We consider the following "likelihood-type" function:

$$L(\alpha) = \prod_{i=1}^{n} P(R_i = r_i|Y_i = y_i, X_i = x_i) = \prod_{i=1}^{n} \pi_i^{r_i}(1-\pi_i)^{1-r_i}.$$

The estimator $\widehat{\alpha}$ of α is obtained by maximizing the function $L(\alpha)$. Note that L is not the likelihood function of the data (R_iY_i, X_i, R_i), $i = 1, ..., n$. (See (3.6) above)

The logarithm of this function is:

$$\begin{aligned} l(\alpha) = \log L(\alpha) &= \sum_{i=1}^{n}(r_i \log \pi_i + (1-r_i)\log(1-\pi_i)) \\ &= \sum_{i=1}^{n}\left[r_i \log\frac{\pi_i}{1-\pi_i} + \log(1-\pi_i)\right] \\ &= \sum_{i=1}^{n}[r_i x_i^T\alpha - \log(1+\exp(x_i^T\alpha))]. \end{aligned}$$

The derivative of $l(\alpha)$ is:

$$\begin{aligned} \frac{\partial l}{\partial \alpha}(\alpha) &= \sum_{i=1}^{n}\left[r_i x_i^T - \frac{1}{1+\exp(x_i^T\alpha)} x_i^T \exp(x_i^T\alpha)\right] \\ &= \sum_{i=1}^{n}\left[r_i - \frac{\exp(x_i^T\alpha)}{1+\exp(x_i^T\alpha)}\right] x_i^T. \end{aligned}$$

The estimator $\widehat{\alpha}$ of α is found as root of the system of p-equations:

$$\sum_{i=1}^{n} X_i\left[R_i - \frac{\exp(X_i^T\alpha)}{1+\exp(X_i^T\alpha)}\right] = 0.$$

Recall that X_i is a p-dimensional vector when $m = 1$ in (2.1).

3.2 Missing at Random for Longitudinal Data

3.2.1 Various Missing Mechanisms

In this section, we introduce several missing mechanisms, which are used in the literature, in the case of longitudinal data. We write the response vector Y_i as $Y_i = (Y_i^{(o)}, Y_i^{(m)})$, where $Y_i^{(o)}$ contains the observed response measurements of Y_i and $Y_i^{(m)}$ represents the missing responses. For any $2 \leq j \leq m$, we let $\widetilde{R}_{ij} = (R_{ik})_{1 \leq k \leq j-1}$ be the history of the missing data indicator at time j. Similarly, we define $\widetilde{X}_{ij} = (X_{ik})_{1 \leq k \leq j-1}$, and we let $\widetilde{Y}_{ij}^{(o)}$ be the vector which contains all the observed responses Y_{ik} with $1 \leq k \leq j-1$.

In the longitudinal case, there are several models for the missing mechanism:

1) For any $i = 1, ..., n$, R_i is conditionally independent of Y_i given $(Y_i^{(o)}, X_i)$, i.e.

$$\text{(MAR1)} \qquad R_i \perp\!\!\!\perp Y_i \quad | \quad (Y_i^{(o)}, X_i).$$

2) For any $i = 1, ..., n$ and for any $j = 2, ..., m$, R_{ij} is conditionally independent of Y_i given $(\widetilde{R}_{ij}, Y_i^{(o)}, X_i)$, i.e.

$$\text{(MAR2)} \qquad R_{ij} \perp\!\!\!\perp Y_i \quad | \quad (\widetilde{R}_{ij}, Y_i^{(o)}, X_i).$$

3) For any $i = 1, ..., n$ and for any $j = 2, ..., m$, R_{ij} is conditionally independent of (Y_i, X_i) given $(\widetilde{R}_{ij}, \widetilde{Y}_{ij}^{(o)}, \widetilde{X}_{ij})$, i.e.

$$\text{(MAR3)} \qquad R_{ij} \perp\!\!\!\perp (Y_i, X_i) \quad | \quad (\widetilde{R}_{ij}, \widetilde{Y}_{ij}^{(o)}, \widetilde{X}_{ij}).$$

In this chapter, we assume that (MAR3) holds. Hence

$$\pi_{ij} := P(R_{ij} = 1 | \widetilde{R}_{ij}, Y_i, X_i) = P(R_{ij} = 1 | \widetilde{R}_{ij}, \widetilde{Y}_{ij}^{(o)} \widetilde{X}_{ij}),$$

$$\text{and} \quad 1 - \pi_{ij} = P(R_{ij} = 0 | \widetilde{R}_{ij}, Y_i, X_i) = P(R_{ij} = 0 | \widetilde{R}_{ij}, \widetilde{Y}_{ij}^{(o)} \widetilde{X}_{ij}).$$

To summarize,

$$P(R_{ij} = r_{ij} | \widetilde{R}_{ij}, Y_i, X_i) = \pi_{ij}^{r_{ij}} (1 - \pi_{ij})^{1 - r_{ij}}, \quad \text{for} \quad r_{ij} \in \{0, 1\}.$$

Note that

$$\pi_{ij} = E(R_{ij} | \widetilde{R}_{ij}, Y_i, X_i) = E(R_{ij} | \widetilde{R}_{ij}, \widetilde{Y}_{ij}^{(o)}, \widetilde{X}_{ij}).$$

In addition, we assume that $(R_{ij})_{j=1,\ldots,m}$ are conditionally independent given Y_i and X_i. Then the conditional density function of R_i given $Y_i = y_i$ and $X_i = x_i$ is:

$$\begin{aligned} f_{R_i|Y_i,X_i}(r_i|y_i,x_i) &= P(R_{i1} = r_{i1},\ldots,R_{im} = r_{im}|Y_i = y_i, X_i = x_i) \\ &= \prod_{j=1}^{m} P(R_{ij} = r_{ij}|Y_i = y_i, X_i = x_i) \\ &= \prod_{j=1}^{m} P(R_{ij} = r_{ij}|\widetilde{R}_{ij} = \widetilde{r}_{ij}, Y_i = y_i, X_i = x_i) \\ &= \prod_{j=1}^{m} \pi_{ij}^{r_{ij}}(1-\pi_{ij})^{1-r_{ij}}, \end{aligned}$$

where we used the conditional independence in the second and third equality.

Similarly to the case $m = 1$ (see Section 3.1.1), we consider a "likelihood-type" function:

$$L = \prod_{i=1}^{n} f_{R_i|Y_i,X_i}(r_i|y_i,x_i) = \prod_{i=1}^{n}\prod_{j=1}^{m} \pi_{ij}^{r_{ij}}(1-\pi_{ij})^{1-r_{ij}}.$$

Note that L is not the actual likelihood function of the data $(R_{ij}Y_{ij}, X_{ij}, R_{ij})$, $1 \le i \le n$, $1 \le j \le m$ (see relation (3.6)).

3.2.2 Logit Model for the Missingness Mechanism

Under (MAR3), a commonly used method for modelling the missingness mechanism is to consider a logistic regression model for the missing probability:

$$\text{logit}(\pi_{ij}) = U_{ij}^T\alpha, \quad 2 \le j \le m, \tag{3.8}$$

where $U_{ij}^T = (\widetilde{R}_{ij}, \widetilde{Y}_{ij}^{(o)}, \widetilde{X}_{ij})$ and $\alpha^T = (\alpha_1,\ldots,\alpha_q)$ is a q-dimensional parameter, whose dimension q coincides with the dimension of the vector U_{ij}. Given U_{ij}, the variable R_{ij} has a Bernoulli distribution with probability of success π_{ij} given by

$$\pi_{ij} = \frac{\exp(U_{ij}^T\alpha)}{1+\exp(U_{ij}^T\alpha)}, \quad \text{for any } 2 \le j \le m.$$

We assume that R_{i1} has a Bernoulli distribution with probability of success $\pi_{i1} = 0.5$. To estimate α, we compute the logarithm of the "likelihood-type" function, denoted now

$L(\alpha)$ since it depends on α:

$$\begin{aligned} l(\alpha) = \log L(\alpha) &= \sum_{i=1}^{n}\sum_{j=1}^{m}[r_{ij}\log(\pi_{ij}) + (1-r_{ij})\log(1-\pi_{ij})] \\ &= \sum_{i=1}^{n}\sum_{j=2}^{m}\left[r_{ij}\log\left(\frac{\pi_{ij}}{1-\pi_{ij}}\right) + \log(1-\pi_{ij})\right] + n\log(0.5) \\ &= \sum_{i=1}^{n}\sum_{j=2}^{m}[r_{ij}U_{ij}^{T}\alpha - \log(1+\exp(U_{ij}^{T}\alpha))] + n\log(0.5). \end{aligned}$$

The estimator of α is the value of $\widehat{\alpha}$ for which the function $L(\alpha)$ attains its maximum. Usually, $\widehat{\alpha}$ is the root of the equation

$$\frac{\partial l(\alpha)}{\partial \alpha} := \left(\frac{\partial l(\alpha)}{\partial \alpha^{T}}\right)^{T} = 0,$$

which can be written as the system of q-equations:

$$\sum_{i=1}^{n}\sum_{j=2}^{m}U_{ij}\left(R_{ij} - \frac{\exp(U_{ij}^{T}\alpha)}{1+\exp(U_{ij}^{T}\alpha)}\right) = 0. \tag{3.9}$$

We denote $S_i(\alpha) = \sum_{j=2}^{m}U_{ij}\left(R_{ij} - \frac{\exp(U_{ij}^{T}\alpha)}{1+\exp(U_{ij}^{T}\alpha)}\right)$. Hence equation (3.9) becomes

$$\sum_{i=1}^{n}S_i(\alpha) = 0. \tag{3.10}$$

3.3 Methodology

Now we let $\mathcal{G}_{ij}(Y_{ij}, X_{ij}, \beta)$ be estimating functions for β when there are no missing observations. Motivated by the "working independence" case (see equation (2.10)), for each $1 \le i \le n$ and $1 \le j \le m$, we consider the function

$$\mathcal{G}_{ij}(Y_{ij}, X_{ij}, \beta) = \frac{\partial \mu_{ij}}{\partial \beta^{T}}(\beta)v_{ij}^{-1}(\beta)(Y_{ij} - \mu_{ij}(\beta)).$$

By a double conditioning argument, since $\mu_{ij}(\beta)$ and $v_{ij}(\beta)$ depend only on X_{ij}, we have:

$$\begin{aligned} E[\mathcal{G}_{ij}(Y_{ij}, X_{ij}, \beta)] &= E[E[\mathcal{G}_{ij}(Y_{ij}, X_{ij}, \beta)|X_{ij}]] \\ &= E\left[\frac{\partial \mu_{ij}}{\partial \beta^{T}}(\beta)v_{ij}^{-1}(\beta)E(Y_{ij} - \mu_{ij}(\beta)|X_{ij})\right] \\ &= 0, \end{aligned} \tag{3.11}$$

since $\mu_{ij}(\beta) = E(Y_{ij}|X_{ij})$. Therefore, $\mathcal{G}_{ij}(Y_{ij}, X_{ij}, \beta)$ is an unbiased estimating function.

3.3.1 Inverse Probability Weights Adjusting for Missingness Effects

For each $i = 1, ..., n$ and $j = 1, ..., m$, we define

$$\begin{aligned}\Phi^*_{ij}(\beta) &= \frac{R_{ij}}{\pi_{ij}}\mathcal{G}_{ij}(Y_{ij}, X_{ij}, \beta)\\ &= \frac{R_{ij}}{\pi_{ij}}\frac{\partial\mu_{ij}}{\partial\beta^T}(\beta)v_{ij}^{-1}(\beta)(Y_{ij} - \mu_{ij}(\beta))\end{aligned}$$

Lemma 3.3.1. *For each $i = 1, ..., n$ and $j = 1, ..., m$,*

$$E[\Phi^*_{ij}(\beta)|Y_i, X_i] = \mathcal{G}_{ij}(Y_{ij}, X_{ij}, \beta) \tag{3.12}$$

and $\Phi^*_{ij}(\beta)$ is an unbiased estimating function of β, i.e. $E[\Phi^*_{ij}(\beta)] = 0$.

Proof: This is a simplified version of the argument given on page 157 of [8]. We fix $i = 1, ..., n$ and $j = 1, ..., m$. First note that by a conditioning argument,

$$\begin{aligned}E\left[\frac{R_{ij}}{\pi_{ij}}|Y_i, X_i\right] &= E\left[E\left[\frac{R_{ij}}{\pi_{ij}}|\widetilde{R}_{ij}, Y_i, X_i\right]|Y_i, X_i\right]\\ &= E\left[\frac{1}{\pi_{ij}}E[R_{ij}|\widetilde{R}_{ij}, Y_i, X_i]|Y_i, X_i\right]\\ &= E\left[\frac{1}{\pi_{ij}}\cdot\pi_{ij}|Y_i, X_i\right]\\ &= 1,\end{aligned} \tag{3.13}$$

where for the second equality we used the fact that $\pi_{ij} = P(R_{ij} = 1|\widetilde{R}_{ij}, Y_i, X_i)$ is measurable with respect to $(\widetilde{R}_{ij}, Y_i, X_i)$, and for the third equality we used the fact that $\pi_{ij} = E[R_{ij}|\widetilde{R}_{ij}, Y_i, X_i]$.

Therefore,

$$\begin{aligned}E[\Phi^*_{ij}(\beta)|Y_i, X_i] &= E\left[\frac{R_{ij}}{\pi_{ij}}\mathcal{G}_{ij}(Y_{ij}, X_{ij}, \beta)|Y_i, X_i\right]\\ &= \mathcal{G}_{ij}(Y_{ij}, X_{ij}, \beta)E\left[\frac{R_{ij}}{\pi_{ij}}|Y_i, X_i\right]\\ &= \mathcal{G}_{ij}(Y_{ij}, X_{ij}, \beta),\end{aligned}$$

where for the second equality above, we used the fact that $\mathcal{G}_{ij}(Y_{ij}, X_{ij}, \beta)$ is measurable with respect to (Y_i, X_i). This concludes the proof of (3.12).

Finally, to show that $\Phi^*_{ij}(\beta)$ is an unbiased estimating function, we observe that by a conditioning argument and relation (3.12),

$$\begin{aligned} E[\Phi^*_{ij}(\beta)] &= E[E[\Phi^*_{ij}(\beta)|Y_i, X_i]] \\ &= E[\mathcal{G}_{ij}(Y_{ij}, X_{ij}, \beta)] \\ &= 0, \end{aligned}$$

where the last equality is due to (3.11). □

We will use the following estimating function:

$$\begin{aligned} \Phi_{ij}(\beta) &= \mathcal{G}_{ij}\left(\frac{R_{ij}}{\pi_{ij}} Y_{ij}, X_{ij}, \beta\right) \\ &= \frac{\partial \mu_{ij}}{\partial \beta^T}(\beta) v_{ij}^{-1}(\beta) \left(\frac{R_{ij}}{\pi_{ij}} Y_{ij} - \mu_{ij}(\beta)\right) \end{aligned}$$

Lemma 3.3.2. *For any $i = 1, ..., n$ and $j = 1, ..., m$, $\Phi_{ij}(\beta)$ is an unbiased estimating function, i.e.*

$$E[\Phi_{ij}(\beta)] = 0.$$

Proof: We first calculate the difference between $\Phi_{ij}(\beta)$ and $\Phi^*_{ij}(\beta)$:

$$\begin{aligned} \Phi_{ij}(\beta) - \Phi^*_{ij}(\beta) &= \frac{\partial \mu_{ij}}{\partial \beta^T}(\beta) v_{ij}^{-1}(\beta) \frac{R_{ij}}{\pi_{ij}} Y_{ij} - \frac{\partial \mu_{ij}}{\partial \beta^T}(\beta) v_{ij}^{-1}(\beta) \mu_{ij}(\beta) \\ &\quad - \frac{R_{ij}}{\pi_{ij}} \frac{\partial \mu_{ij}}{\partial \beta^T}(\beta) v_{ij}^{-1}(\beta) Y_{ij} + \frac{R_{ij}}{\pi_{ij}} \frac{\partial \mu_{ij}}{\partial \beta^T}(\beta) v_{ij}^{-1}(\beta) \mu_{ij}(\beta) \\ &= \frac{\partial \mu_{ij}}{\partial \beta^T}(\beta) v_{ij}^{-1}(\beta) \mu_{ij}(\beta) \left(\frac{R_{ij}}{\pi_{ij}} - 1\right). \end{aligned}$$

Hence

$$\begin{aligned} E[\Phi_{ij}(\beta) - \Phi^*_{ij}(\beta)|\widetilde{R}_{ij}, Y_i, X_i] &= E\left[\frac{\partial \mu_{ij}}{\partial \beta^T}(\beta) v_{ij}^{-1}(\beta) \mu_{ij}(\beta) \left(\frac{R_{ij}}{\pi_{ij}} - 1\right) |\widetilde{R}_{ij}, Y_i, X_i\right] \\ &= \frac{\partial \mu_{ij}}{\partial \beta^T}(\beta) v_{ij}^{-1}(\beta) \mu_{ij}(\beta) \left\{E\left[\frac{R_{ij}}{\pi_{ij}} |\widetilde{R}_{ij}, Y_i, X_i\right] - 1\right\} \\ &= 0, \end{aligned}$$

where for the second equality we used the fact that $\mu_{ij}(\beta)$ and $v_{ij}(\beta)$ are functions of X_{ij}, hence they are measurable with respect to (Y_i, X_i), and for the last equality we used (3.13). From here, we obtain that

$$\begin{aligned} E[\Phi_{ij}(\beta) - \Phi^*_{ij}(\beta)] &= E[E[\Phi_{ij}(\beta) - \Phi^*_{ij}(\beta)|Y_i, X_i]] \\ &= 0. \end{aligned}$$

Hence, by Lemma 3.3.1

$$E[\Phi_{ij}(\beta)] = E[\Phi^*_{ij}(\beta)] = 0. \qquad \square$$

We will be interested in solving the following estimating equation:

$$g_n(\beta) := \sum_{i=1}^{n} \sum_{j=1}^{m} \Phi_{ij}(\beta) = 0. \tag{3.14}$$

Note that $g_n(\beta) = \sum_{i=1}^{n} \Psi_i(\beta)$, where

$$\begin{aligned}\Psi_i(\beta) \quad &:= \quad \sum_{j=1}^{m} \frac{\partial \mu_{ij}}{\partial \beta^T}(\beta) v_{ij}^{-1}(\beta) \left(\frac{R_{ij}}{\pi_{ij}} Y_{ij} - \mu_{ij}(\beta) \right) \\ &= \quad D_i(\beta)^T V_i(\beta)^{-1} (T_i(\beta) - \mu_i(\beta)),\end{aligned}$$

where $D_i(\beta)$ is given by (2.5), $V_i(\beta)$ is given by (2.9), and $T_i = (T_{i1}, ..., T_{im})^T$ with

$$T_{ij} = \frac{R_{ij}}{\pi_{ij}} Y_{ij}, \quad \text{for all } 1 \le j \le m.$$

Due to Lemma 3.3.2, $g_n(\beta)$ is an unbiased estimating function:

$$E[g_n(\beta)] = \sum_{i=1}^{n} \sum_{j=1}^{m} E[\Phi_{ij}(\beta)] = 0.$$

3.3.2 Examples

In this section, we consider again the three examples introduced in Section 2.1. In these examples,

$$\mu_{ij}(\beta) = \mu(X_{ij}^T \beta), \quad \text{and} \quad v_{ij} = \mu'(X_{ij}^T \beta).$$

Hence,

$$\frac{\partial \mu_{ij}}{\partial \beta^T}(\beta) v_{ij}^{-1}(\beta) = X_{ij} \mu'(X_{ij}^T \beta) \cdot \frac{1}{\mu'(X_{ij}^T \beta)} = X_{ij},$$

and equation (3.14) becomes:

$$\sum_{i=1}^{n} \sum_{j=1}^{m} X_{ij}(T_{ij} - \mu_{ij}(\beta)) = 0. \tag{3.15}$$

Note that equation (3.15) has the same form as (2.11) in which the response Y_{ij} is replaced by T_{ij}.

Example 3.3.3. *(Normal Linear Regression)*

We assume that the conditional distribution of Y_{ij} given X_{ij} is normal with mean $\mu_{ij}(\beta)$ and variance 1. In this case, equation (3.15) becomes:

$$\sum_{i=1}^{n}\sum_{j=1}^{m} X_{ij}(T_{ij} - X_{ij}^T\beta) = 0, \tag{3.16}$$

and the solution is

$$\widehat{\beta} = (\sum_{i=1}^{n}\sum_{j=1}^{m} X_{ij}X_{ij}^T)^{-1}(\sum_{i=1}^{n}\sum_{j=1}^{m} X_{ij}T_{ij}).$$

Example 3.3.4. *(Poisson Regression)*

We assume that the conditional distribution of Y_{ij} given X_{ij} is Poisson with mean $\mu_{ij}(\beta) = \exp(X_{ij}^T\beta)$. In this case, equation (3.15) becomes:

$$\sum_{i=1}^{n}\sum_{j=1}^{m} X_{ij}(T_{ij} - \exp(X_{ij}^T\beta)) = 0 \tag{3.17}$$

The solution of this equation cannot be written down explicitly. The root of equation (3.17) can be found using `R`.

Example 3.3.5. *(Logistic Regression)*

We assume that the conditional distribution of Y_{ij} given X_{ij} is Bernoulli with mean $\mu_{ij}(\beta) = \exp(X_{ij}^T\beta)/(1 + \exp(X_{ij}^T\beta))$. In this case, equation (3.15) becomes:

$$\sum_{i=1}^{n}\sum_{j=1}^{m} X_{ij}\left(T_{ij} - \frac{\exp(X_{ij}^T\beta)}{1 + \exp(X_{ij}^T\beta)}\right) = 0, \tag{3.18}$$

This equation can be solved with `R`.

3.4 Asymptotic Normality

3.4.1 The case when α is known

We consider the unbiased estimating function $g_n(\beta)$ given by

$$g_n(\beta) = \sum_{i=1}^{n}\sum_{j=1}^{m} \Phi_{ij}(\beta).$$

Then $g_n(\beta) = \sum_{i=1}^{n} \Psi_i(\beta)$, where $\Psi_i(\beta) = \sum_{j=1}^{m} \Phi_{ij}(\beta)$. Note that $\{\Psi_i(\beta)\}_{i\geq 1}$ are independent and identically distributed (i.i.d.) p-dimensional random vectors of zero mean.

We let β_0 be the true value of the parameter β. We consider the $p \times p$ matrix

$$\Gamma_0 = E\left[\frac{\partial \Psi_i}{\partial \beta^T}(\beta_0)\right],$$

and we assume that Γ_0 is invertible.

We denote Σ_Ψ by the covariance matrix of $\Psi_i(\beta_0)$, i.e.

$$\Sigma_\Psi = \mathrm{Cov}(\Psi(\beta_0)) = E[\Psi_i(\beta_0)\Psi_i(\beta_0)^T].$$

Definition 3.4.1. Consider a family of functions $f_\gamma : R^l \to R$, where $l > 0$ and the index γ lies in a set $\mathcal{F}$, and let $\eta_0 \in R^l$ be fixed. The family $\{f_\gamma\}_{\gamma\in\mathcal{F}}$ is *equicontinuous* at $\eta_0 \in R^l$ if, for any $\varepsilon > 0$, there exists $r_\varepsilon > 0$, such that for all $\gamma \in \mathcal{F}$:

$$(EQ_{\eta_0}) \quad |f_\gamma(\eta) - f_\gamma(\eta_0)| \leq \varepsilon, \quad \text{for all} \quad \eta, \quad ||\eta - \eta_0|| \leq r_\varepsilon.$$

Definition 3.4.2. The family $\left\{\frac{\partial \Phi_{ij}}{\partial \beta_k}(\beta); i \geq 1, 1 \leq j \leq m, 1 \leq k \leq p\right\}$ is *almost surely equicontinuous* at $\beta_0 \in R^P$, if, for $\varepsilon > 0$ fixed, there exists $r_\varepsilon > 0$, such that, for all $i \geq 1$, $1 \leq j \leq m$, $1 \leq k \leq p$, we have on a set of probability 1:

$$(EQ_{\beta_0}) \quad \left|\left|\frac{\partial \Phi_{ij}}{\partial \beta_k}(\beta) - \frac{\partial \Phi_{ij}}{\partial \beta_k}(\beta_0)\right|\right| \leq \varepsilon \quad \text{for all} \quad \beta, \quad ||\beta - \beta_0|| \leq r_\varepsilon.$$

Theorem 3.4.3. *Let $\widehat{\beta}_n$ be a solution of the equation*

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

Assume that $\widehat{\beta}_n \to \beta_0$ almost surely, and the family

$$\left\{\frac{\partial \Phi_{ij}}{\partial \beta_k}(\beta); i \geq 1, 1 \leq j \leq m, 1 \leq k \leq p\right\}$$

is almost surely equicontinuous at β_0. Then

$$\sqrt{n}(\widehat{\beta}_n - \beta_0) \xrightarrow{d} N_p(0, \Gamma_0^{-1}\Sigma_\Psi(\Gamma_0^{-1})^T).$$

Proof: We work on the event of probability 1 where the equicontinuous property given by definition 3.4.2 holds and $(\widehat{\beta}_n)_{n\geq 1}$ converges to β_0. Using Taylor's formula, there exists $\bar{\beta}_n$ with $||\bar{\beta}_n - \beta_0|| \leq ||\widehat{\beta}_n - \beta_0||$, such that

$$g_n(\widehat{\beta}_n) - g_n(\beta_0) = \frac{\partial g_n}{\partial \beta^T}(\bar{\beta}_n)(\widehat{\beta}_n - \beta_0).$$

Since $\widehat{\beta}_n$ is the root of the equation $g_n(\beta) = 0$, we have $g_n(\widehat{\beta}_n) = 0$, and the previous relation becomes

$$-g_n(\beta_0) = \frac{\partial g_n}{\partial \beta^T}(\bar{\beta}_n)(\widehat{\beta}_n - \beta_0). \tag{3.19}$$

Since $\left\{\frac{\partial \Psi_i}{\partial \beta^T}(\beta_0)\right\}_{i\geq 1}$, are i.i.d. random matrices, the strong law of large numbers can be applied, and we have the following:

$$\frac{1}{n}\cdot\frac{\partial g_n}{\partial \beta^T}(\beta_0) = \frac{1}{n}\sum_{i=1}^{n}\frac{\partial \Psi_i}{\partial \beta^T}(\beta_0) \to E\left[\frac{\partial \Psi_i}{\partial \beta^T}(\beta_0)\right] = \Gamma_0 \quad \text{almost surely,} \tag{3.20}$$

where $o(1)$ stands for a random variable χ_n that converges to 0 as $n \to \infty$.

We now prove that

$$\frac{1}{n}\cdot\frac{\partial g_n}{\partial \beta^T}(\bar{\beta}_n) - \frac{1}{n}\cdot\frac{\partial g_n}{\partial \beta^T}(\beta_0) = o(1), \quad \text{almost surely.} \tag{3.21}$$

To see this, let $\varepsilon > 0$ be arbitrary and $r_\varepsilon > 0$ be given by definition 3.4.2. Then there exists $N_\varepsilon > 0$ such that $||\widehat{\beta}_n - \beta_0|| \leq r_\varepsilon$ for all $n \geq N_\varepsilon$.

In particular, $||\bar{\beta}_n - \beta_0|| \leq r_\varepsilon$, for all $n \geq N_\varepsilon$. Using property (EQ_{β_0}) it follows that, for any $n \geq N_\varepsilon$ and for any $1 \leq k \leq p$, we have

$$\left|\left|\frac{1}{n}\cdot\frac{\partial g_n}{\partial \beta_k}(\bar{\beta}_n) - \frac{1}{n}\cdot\frac{\partial g_n}{\partial \beta_k}(\beta_0)\right|\right| \leq \frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{m}\left|\left|\frac{\partial \Phi_{ij}}{\partial \beta_k}(\bar{\beta}_n) - \frac{\partial \Phi_{ij}}{\partial \beta_k}(\beta_0)\right|\right| \leq \frac{1}{n}nm\varepsilon = m\varepsilon.$$

Since $\varepsilon > 0$ was arbitrary, this proves (3.21).

Now, from (3.19), combined with (3.20) and (3.21) we can write:

$$\begin{aligned} -g_n(\beta_0) &= n\left[\frac{1}{n}\cdot\frac{\partial g_n}{\partial \beta^T}(\bar{\beta}_n)\right](\widehat{\beta}_n - \beta_0) \\ &= n\left[\frac{1}{n}\cdot\frac{\partial g_n}{\partial \beta^T}(\bar{\beta}_n) - \frac{1}{n}\cdot\frac{\partial g_n}{\partial \beta^T}(\beta_0) + \frac{1}{n}\frac{\partial g_n}{\partial \beta^T}(\beta_0)\right](\widehat{\beta}_n - \beta_0) \\ &= n[o(1) + \Gamma_0](\widehat{\beta}_n - \beta_0) \quad \text{almost surely.} \end{aligned}$$

We multiply to the left both sides of this equality by $\Gamma_0^{-1} \cdot \frac{1}{\sqrt{n}}$. We obtain:

$$\begin{aligned} -\Gamma_0^{-1} \cdot \frac{1}{\sqrt{n}} g_n(\beta_0) &= \Gamma_0^{-1}[o(1) + \Gamma_0]\sqrt{n}(\widehat{\beta}_n - \beta_0) \\ &= [o(1) + I_p]\sqrt{n}(\widehat{\beta}_n - \beta_0) \quad \text{almost surely.} \end{aligned} \tag{3.22}$$

Since $\{\Psi_i(\beta_0)\}_{i\geq 1}$ are i.i.d. random vectors of mean 0 and covariance matrix Σ_Ψ, the central limit theorem holds and we infer that

$$\frac{1}{\sqrt{n}} \cdot g_n(\beta_0) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \Psi_i(\beta_0) \xrightarrow{d} N_p(0, \Sigma_\Psi).$$

Consequently,

$$-\Gamma_0^{-1} \cdot \frac{1}{\sqrt{n}} g_n(\beta_0) \xrightarrow{d} N_p(0, \Gamma_0^{-1}\Sigma_\Psi(\Gamma_0^{-1})^T). \tag{3.23}$$

The conclusion follows from (3.22) and (3.23), since $(o(1) + I_p)^{-1} = o(1) + I_p$, and the left hand side of (3.22) is bounded, by (3.23). □

3.4.2 The case when α is unknown

In this section, we let $\widehat{\alpha}_n$ be a solution of equation (3.10). Note that $\{S_i(\alpha)\}_{i\geq 1}$ are i.i.d. q-dimensional random vectors of mean 0. Let α_0 be the true value of the parameter α. We denote by Σ_S the covariance matrix of $S_i(\alpha_0)$, i.e.

$$\Sigma_S = \text{Cov}(S_i(\alpha_0)) = E[S_i(\alpha_0)S_i(\alpha_0)^T].$$

We denote

$$M = E\left[\frac{\partial S_i}{\partial \alpha^T}(\alpha_0)\right], \quad \text{and we write} \quad S_i(\alpha) = \sum_{j=2}^{m} S_{ij}(\alpha), \quad \text{where}$$

$$S_{ij}(\alpha) = U_{ij}\left(R_{ij} - \frac{\exp(U_{ij}^T\alpha)}{1 + \exp(U_{ij}^T\alpha)}\right).$$

We assume M is invertible.

Lemma 3.4.4. *Let $\widehat{\alpha}_n$ be a solution of equation* (3.10). *Assume that $\widehat{\alpha}_n \to \alpha_0$ almost surely, and the family $\left\{\left(\frac{\partial S_{ij}}{\partial \alpha_l}(\alpha)\right); i \geq 1,\, 2 \leq j \leq m,\, 1 \leq l \leq q\right\}$ is almost surely equicontinuous at α_0.*

Then

$$\sqrt{n}(\widehat{\alpha}_n - \alpha_0) \xrightarrow{d} N_q(0, M^{-1}\Sigma_S(M^{-1})^T).$$

Proof: We apply Taylor's formula to the function $h_n(\alpha) = \sum S_i(\alpha)$. The proof is identical to that of Theorem 3.4.3, where we replace β by α and Ψ_i by $S_i,\ i \geq 1$. □

The following result gives the asymptotic distribution of the estimator of the main parameter β_0, when α is unknown. In this case, g_n is a function of $\eta = (\beta, \alpha) \in \mathbb{R}^{p+q}$, since $\pi_{ij} = \pi_{ij}(\alpha),\ 1 \leq j \leq m,\ i \geq 1$.

We assume that the $p \times p$ matrix

$$\Gamma = E\left[\frac{\partial \Psi_i}{\partial \beta^T}(\beta_0, \alpha_0)\right] \quad \text{is invertible.}$$

Theorem 3.4.5. *Let $(\widehat{\beta}_n, \widehat{\alpha}_n)$ be a solution of the system*

$$\begin{cases} \sum_{i=1}^n \Psi_i(\beta, \alpha) = 0 \\ \sum_{i=1}^n S_i(\alpha) = 0. \end{cases}$$

Assume that the following conditions hold:

a) $\widehat{\alpha}_n \to \alpha_0$ almost surely, and $\widehat{\beta}_n \to \beta_0$ almost surely;

b) the family $\left\{\frac{\partial S_{ij}}{\partial \alpha_l}(\alpha); i \geq 1, 2 \leq j \leq m, 1 \leq l \leq q\right\}$ is almost surely equicontinuous at α_0;

c) the family $\left\{\frac{\partial \Psi_{ij}}{\partial \eta_k}(\eta); i \geq 1, 1 \leq j \leq m, 1 \leq k \leq p+q\right\}$ is almost surely equicontinuous at $\eta_0 = (\beta_0, \alpha_0)$.

Then

$$\sqrt{n}(\widehat{\beta}_n - \beta_0) \xrightarrow{d} N_p(0, \Gamma^{-1}\Omega(\Gamma^{-1})^T),$$

where

$$\Omega = E[Q_i(\beta_0, \alpha_0)Q_i^T(\beta_0, \alpha_0)], \quad Q_i(\beta, \alpha) := \Psi_i(\beta, \alpha) - KM^{-1}S_i(\alpha),$$

and

$$K := E\left[\frac{\partial \Psi_i}{\partial \alpha^T}(\beta_0, \alpha_0)\right].$$

Proof: The proof is similar to that of Theorem 3.4.3, and uses the results of Lemma 3.4.4.

We denote

$$G_n(\beta, \alpha) = \begin{bmatrix} g_n(\beta, \alpha) \\ h_n(\alpha) \end{bmatrix},$$

where $g_n(\beta, \alpha) = \sum_{i=1}^n \Psi_i(\beta, \alpha)$ and $h_n(\alpha) = \sum_{i=1}^n S_i(\alpha)$. We apply Taylor's formula to the $(p+q)$-dimensional function $G_n(\beta, \alpha)$. Continuing with the fact that $G_n(\widehat{\beta}_n, \widehat{\alpha}_n) = 0$, we obtain:

$$\begin{aligned}
-G_n(\beta_0,\alpha_0) &= G_n(\widehat{\beta}_n,\widehat{\alpha}_n) - G_n(\beta_0,\alpha_0)\\
&= \begin{bmatrix} \frac{\partial g_n}{\partial \beta^T}(\bar{\beta}_n,\bar{\alpha}_n) & \frac{\partial g_n}{\partial \alpha^T}(\bar{\beta}_n,\bar{\alpha}_n)\\ 0 & \frac{\partial h_n}{\partial \alpha^T}(\bar{\alpha}_n)\end{bmatrix}\begin{bmatrix}\widehat{\beta}_n-\beta_0\\ \widehat{\alpha}_n-\alpha_0\end{bmatrix}\\
&= \begin{bmatrix} \frac{\partial g_n}{\partial \beta^T}(\bar{\beta}_n,\bar{\alpha}_n)(\widehat{\beta}_n-\beta_0) + \frac{\partial g_n}{\partial \alpha^T}(\bar{\beta}_n,\bar{\alpha}_n)(\widehat{\alpha}_n-\alpha_0)\\ \frac{\partial h_n}{\partial \alpha^T}(\bar{\alpha}_n)(\widehat{\alpha}_n-\alpha_0)\end{bmatrix}.
\end{aligned}$$

This leads to the following two relations:

$$-g_n(\beta_0,\alpha_0) = \frac{\partial g_n}{\partial \beta^T}(\bar{\beta}_n,\bar{\alpha}_n)(\widehat{\beta}_n-\beta_0) + \frac{\partial g_n}{\partial \alpha^T}(\bar{\beta}_n,\bar{\alpha}_n)(\widehat{\alpha}_n-\alpha_0) \tag{3.24}$$

$$-h_n(\alpha_0) = \frac{\partial h_n}{\partial \alpha^T}(\bar{\alpha}_n)(\widehat{\alpha}_n-\alpha_0). \tag{3.25}$$

Relation (3.25) can be written as follows:

$$-h_n(\alpha_0) = n\left(\frac{1}{n}\cdot\frac{\partial h_n}{\partial \alpha^T}(\bar{\alpha}_n) - \frac{1}{n}\cdot\frac{\partial h_n}{\partial \alpha^T}(\alpha_0) + \frac{1}{n}\cdot\frac{\partial h_n}{\partial \alpha^T}(\alpha_0) - M + M\right)(\widehat{\alpha}_n-\alpha_0).$$

We multiply to the left both sides by $\frac{1}{\sqrt{n}}M^{-1}$. Then we have: $-\frac{1}{\sqrt{n}}M^{-1}h_n(\alpha_0) =$

$$M^{-1}\left(\frac{1}{n}\cdot\frac{\partial h_n}{\partial \alpha^T}(\bar{\alpha}_n) - \frac{1}{n}\cdot\frac{\partial h_n}{\partial \alpha^T}(\alpha_0) + \frac{1}{n}\cdot\frac{\partial h_n}{\partial \alpha^T}(\alpha_0) - M + M\right)\sqrt{n}(\widehat{\alpha}_n-\alpha_0)$$

$$= \left[M^{-1}\left(\frac{1}{n}\cdot\frac{\partial h_n}{\partial \alpha^T}(\bar{\alpha}_n) - \frac{1}{n}\cdot\frac{\partial h_n}{\partial \alpha^T}(\alpha_0)\right) + M^{-1}\left(\frac{1}{n}\cdot\frac{\partial h_n}{\partial \alpha^T}(\alpha_0) - M\right) + I_q\right]\sqrt{n}(\widehat{\alpha}_n-\alpha_0).$$

By the equicontinuity assumption b) and the fact that $\widehat{\alpha}_n \to \alpha_0$ almost surely, we have:

$$\frac{1}{n}\cdot\frac{\partial h_n}{\partial \alpha^T}(\bar{\alpha}_n) - \frac{1}{n}\cdot\frac{\partial h_n}{\partial \alpha^T}(\alpha_0) \to 0 \quad \text{almost surely.}$$

By the strong law of large numbers,

$$\frac{1}{n}\cdot\frac{\partial h_n}{\partial \alpha^T}(\alpha_0) = \frac{1}{n}\sum_{i=1}^{n}\frac{\partial S_i}{\partial \alpha^T}(\alpha_0) \to E\left[\frac{\partial S_i}{\partial \alpha^T}(\alpha_0)\right] = M \quad \text{almost surely.}$$

Hence

$$-\frac{1}{\sqrt{n}}M^{-1}h_n(\alpha_0) = (o(1)+I_q)\sqrt{n}(\widehat{\alpha}_n-\alpha_0). \tag{3.26}$$

We now rewrite relation (3.24) as follows:

$$
\begin{aligned}
-g_n(\beta_0,\alpha_0) &= n\left[\frac{1}{n}\cdot\frac{\partial g_n}{\partial\beta^T}(\bar\beta_n,\bar\alpha_n)\right](\widehat\beta_n-\beta_0)+n\left[\frac{1}{n}\cdot\frac{\partial g_n}{\partial\alpha^T}(\bar\beta_n,\bar\alpha_n)\right](\widehat\alpha_n-\alpha_0)\\
&= n\left[\frac{1}{n}\cdot\frac{\partial g_n}{\partial\beta^T}(\bar\beta_n,\bar\alpha_n)-\frac{1}{n}\cdot\frac{\partial g_n}{\partial\beta^T}(\beta_0,\alpha_0)+\frac{1}{n}\cdot\frac{\partial g_n}{\partial\beta^T}(\beta_0,\alpha_0)\right](\widehat\beta_n-\beta_0)\\
&\quad +n\left[\frac{1}{n}\cdot\frac{\partial g_n}{\partial\alpha^T}(\bar\beta_n,\bar\alpha_n)-\frac{1}{n}\cdot\frac{\partial g_n}{\partial\alpha^T}(\beta_0,\alpha_0)+\frac{1}{n}\cdot\frac{\partial g_n}{\partial\alpha^T}(\beta_0,\alpha_0)\right](\widehat\alpha_n-\alpha_0),
\end{aligned}
$$

where $||(\bar\beta_n,\bar\alpha_n)-(\beta_0,\alpha_0)||\le||(\widehat\beta_n,\widehat\alpha_n)-(\beta_0,\alpha_0)||$.

By the strong law of large numbers, we have:

$$\frac{1}{n}\cdot\frac{\partial g_n}{\partial\beta^T}(\beta_0,\alpha_0)=\frac{1}{n}\sum_{i=1}^{n}\frac{\partial\Psi_i}{\partial\beta^T}(\beta_0,\alpha_0)\to\Gamma\quad\text{almost surely.}\tag{3.27}$$

$$\frac{1}{n}\cdot\frac{\partial g_n}{\partial\alpha^T}(\beta_0,\alpha_0)=\frac{1}{n}\sum_{i=1}^{n}\frac{\partial\Psi_i}{\partial\alpha^T}(\beta_0,\alpha_0)\to K\quad\text{almost surely.}\tag{3.28}$$

As in the proof of Theorem 3.4.3, using the equicontinuity assumptions a) and c), we have:

$$
\begin{aligned}
&\frac{1}{n}\cdot\frac{\partial g_n}{\partial\beta^T}(\bar\beta_n,\bar\alpha_n)-\frac{1}{n}\cdot\frac{\partial g_n}{\partial\beta^T}(\beta_0,\alpha_0)=o(1)\quad\text{almost surely,}\\
&\frac{1}{n}\cdot\frac{\partial g_n}{\partial\alpha^T}(\bar\beta_n,\bar\alpha_n)-\frac{1}{n}\cdot\frac{\partial g_n}{\partial\alpha^T}(\beta_0,\alpha_0)=o(1)\quad\text{almost surely.}
\end{aligned}
$$

Combining these relations with (3.27) and (3.28), we obtain:

$$-g_n(\beta_0,\alpha_0)=n[o(1)+\Gamma](\widehat\beta_n-\beta_0)+n[o(1)+K](\widehat\alpha_n-\alpha_0).$$

Multiplying to the left both sides of the equation by $\frac{1}{\sqrt{n}}\Gamma^{-1}$, we obtain:

$$-\Gamma^{-1}\cdot\frac{1}{\sqrt{n}}g_n(\beta_0,\alpha_0)=[o(1)+I_p]\sqrt{n}(\widehat\beta_n-\beta_0)+[o(1)+\Gamma^{-1}K]\sqrt{n}(\widehat\alpha_n-\alpha_0)\quad\text{almost surely.}$$

In deducing the equation above, we used the fact that

$$\Gamma^{-1}o(1)=o(1)\quad\text{and}\quad o(1)+K\to K\quad\text{almost surely.}$$

From relation (3.26), we know that

$$\sqrt{n}(\widehat\alpha_n-\alpha_0)=-M^{-1}\cdot\frac{1}{\sqrt{n}}\sum_{i=1}^{n}S_i(\alpha_0)+o(1).\tag{3.29}$$

Therefore, we have:

$$\begin{aligned}(o(1)+I_p)\sqrt{n}(\widehat{\beta}_n-\beta_0) &= -\frac{1}{\sqrt{n}}\Gamma^{-1}g_n(\beta_0,\alpha_0)-(o(1)+\Gamma^{-1}K)\sqrt{n}(\widehat{\alpha}_n-\alpha_0)\\ &= -\frac{1}{\sqrt{n}}\Gamma^{-1}g_n(\beta_0,\alpha_0)-\Gamma^{-1}K\sqrt{n}(\widehat{\alpha}_n-\alpha_0)+o(1)\\ &= -\frac{1}{\sqrt{n}}\Gamma^{-1}\sum_{i=1}^{n}\Psi_i(\beta_0,\alpha_0)+\Gamma^{-1}KM^{-1}\frac{1}{\sqrt{n}}\sum_{i=1}^{n}S_i(\alpha_0)+o(1)\\ &= -\Gamma^{-1}\frac{1}{\sqrt{n}}\sum_{i=1}^{n}(\Psi_i(\beta_0,\alpha_0)-KM^{-1}S_i(\alpha_0))+o(1)\\ &= -\Gamma^{-1}\frac{1}{\sqrt{n}}\sum_{i=1}^{n}Q_i(\beta_0,\alpha_0)+o(1)\quad\text{almost surely,}\qquad(3.30)\end{aligned}$$

where we used Lemma 3.4.4 for the second equality and relation (3.29) for the third equality.

Since $\{Q_i(\beta_0,\alpha_0)\}_{i\geq 1}$ are i.i.d. p-dimensional random vectors of mean 0 and covariance matrix Ω, by the central limit theorem,

$$\frac{1}{\sqrt{n}}\sum_{i=1}^{n}Q_i(\beta_0,\alpha_0)\xrightarrow{d}N_p(0,\Omega).$$

Therefore, we have

$$\Gamma^{-1}\frac{1}{\sqrt{n}}\sum_{i=1}^{n}Q_i(\beta_0,\alpha_0)\xrightarrow{d}N_p(0,\Gamma^{-1}\Omega(\Gamma^{-1})^T).$$

The conclusion follows from (3.30) since $(o(1)+I_p)^{-1}\to I_p$ almost surely. □

3.5 Simulations

In this section, we consider again the three examples of Section 2.2. We assume that $p=2$, and $m=3$. To introduce the missingness mechanism, we use the logit model (3.8) with $U_{ij}=(Y_{i,j-1},X_{i,j-1})$. Given U_{ij}, R_{ij} has a Bernoulli distribution with probability of success π_{ij} given by:

$$\text{logit}(\pi_{ij})=\alpha_1Y_{i,j-1}+\alpha_2X_{i,j-1}^{(1)}+\alpha_3X_{i,j-1}^{(2)},\quad\text{for}\quad 2\leq j\leq 3.$$

3.5.1 Normal Linear Regression

In our R program, the covariates are generated separately from the standard normal distribution. Below is the explanation of the R program given in Appendix B.1.

1. For each $j = 1, 2, 3$, we generate variables $(X_{ij}^{(1)}, X_{ij}^{(2)}, Y_{ij})$ with $1 \leq i \leq n$ as in Steps 1 - 2 of the algorithm mentioned in Section 2.2.1.

2. We take $\alpha_{01} = 0.5$, $\alpha_{02} = 0.2$, and $\alpha_{03} = 0.3$.

a) For $j = 1$, we generate the missingness indicators R_{i1}, $i = 1, ..., n$, which are independent and identically distributed Bernoulli random variables with probability of success $\pi_{i1} = 0.5$.

b) For $j = 2$, we generate the missingness indicators R_{i2}, $i = 1, ..., n$ such that the distribution of R_{i2} given $U_{i2} = (Y_{i1}, X_{i1}^{(1)}, X_{i1}^{(2)})$ is Bernoulli with probability of success

$$\pi_{i2} = \frac{\exp(\alpha_{01} Y_{i1} + \alpha_{02} X_{i1}^{(1)} + \alpha_{03} X_{i1}^{(2)})}{1 + \exp(\alpha_{01} Y_{i1} + \alpha_{02} X_{i1}^{(1)} + \alpha_{03} X_{i1}^{(2)})}.$$

c) For $j = 3$, we generate the missingness indicators R_{i3}, $i = 1, ..., n$ such that the distribution of R_{i3} given $U_{i3} = (Y_{i2}, X_{i2}^{(1)}, X_{i2}^{(2)})$ is Bernoulli with probability of success

$$\pi_{i3} = \frac{\exp(\alpha_{01} Y_{i2} + \alpha_{02} X_{i2}^{(1)} + \alpha_{03} X_{i2}^{(2)})}{1 + \exp(\alpha_{01} Y_{i2} + \alpha_{02} X_{i2}^{(1)} + \alpha_{03} X_{i2}^{(2)})}.$$

3. We now solve for $\alpha = (\alpha_1, \alpha_2, \alpha_3)$ in the system (3.9), which in our case becomes a system of three equations:

$$\begin{cases} \sum_{i=1}^{n} \sum_{j=2}^{3} Y_{i,j-1} \left(R_{ij} - \frac{\exp(\alpha_1 Y_{i,j-1} + \alpha_2 X_{i,j-1}^{(1)} + \alpha_3 X_{i,j-1}^{(2)})}{1 + \exp(\alpha_1 Y_{i,j-1} + \alpha_2 X_{i,j-1}^{(1)} + \alpha_3 X_{i,j-1}^{(2)})} \right) = 0 \\ \sum_{i=1}^{n} \sum_{j=2}^{3} X_{i,j-1}^{(1)} \left(R_{ij} - \frac{\exp(\alpha_1 Y_{i,j-1} + \alpha_2 X_{i,j-1}^{(1)} + \alpha_3 X_{i,j-1}^{(2)})}{1 + \exp(\alpha_1 Y_{i,j-1} + \alpha_2 X_{i,j-1}^{(1)} + \alpha_3 X_{i,j-1}^{(2)})} \right) = 0 \\ \sum_{i=1}^{n} \sum_{j=2}^{3} X_{i,j-1}^{(2)} \left(R_{ij} - \frac{\exp(\alpha_1 Y_{i,j-1} + \alpha_2 X_{i,j-1}^{(1)} + \alpha_3 X_{i,j-1}^{(2)})}{1 + \exp(\alpha_1 Y_{i,j-1} + \alpha_2 X_{i,j-1}^{(1)} + \alpha_3 X_{i,j-1}^{(2)})} \right) = 0. \end{cases}$$

This system can be written as

$$\begin{cases} \sum_{i=1}^n Y_{i1}\left(R_{i2} - \frac{\exp(\alpha_1 Y_{i1}+\alpha_2 X_{i1}^{(1)}+\alpha_3 X_{i1}^{(2)})}{1+\exp(\alpha_1 Y_{i1}+\alpha_2 X_{i1}^{(1)}+\alpha_3 X_{i1}^{(2)})}\right) \\ \quad + \sum_{i=1}^n Y_{i2}\left(R_{i3} - \frac{\exp(\alpha_1 Y_{i2}+\alpha_2 X_{i2}^{(1)}+\alpha_3 X_{i2}^{(2)})}{1+\exp(\alpha_1 Y_{i2}+\alpha_2 X_{i2}^{(1)}+\alpha_3 X_{i2}^{(2)})}\right) = 0 \\ \sum_{i=1}^n X_{i1}^{(1)}\left(R_{i2} - \frac{\exp(\alpha_1 Y_{i1}+\alpha_2 X_{i1}^{(1)}+\alpha_3 X_{i1}^{(2)})}{1+\exp(\alpha_1 Y_{i1}+\alpha_2 X_{i1}^{(1)}+\alpha_3 X_{i1}^{(2)})}\right) \\ \quad + \sum_{i=1}^n X_{i2}^{(1)}\left(R_{i3} - \frac{\exp(\alpha_1 Y_{i2}+\alpha_2 X_{i2}^{(1)}+\alpha_3 X_{i2}^{(2)})}{1+\exp(\alpha_1 Y_{i2}+\alpha_2 X_{i2}^{(1)}+\alpha_3 X_{i2}^{(2)})}\right) = 0 \\ \sum_{i=1}^n X_{i1}^{(2)}\left(R_{i2} - \frac{\exp(\alpha_1 Y_{i1}+\alpha_2 X_{i1}^{(1)}+\alpha_3 X_{i1}^{(2)})}{1+\exp(\alpha_1 Y_{i1}+\alpha_2 X_{i1}^{(1)}+\alpha_3 X_{i1}^{(2)})}\right) \\ \quad + \sum_{i=1}^n X_{i2}^{(2)}\left(R_{i3} - \frac{\exp(\alpha_1 Y_{i2}+\alpha_2 X_{i2}^{(1)}+\alpha_3 X_{i2}^{(2)})}{1+\exp(\alpha_1 Y_{i2}+\alpha_2 X_{i2}^{(1)}+\alpha_3 X_{i2}^{(2)})}\right) = 0. \end{cases}$$

We solve this system in `R` using the multiroot function (the program is given in Appendix B.1).

We denote by $\widehat{\alpha} = (\widehat{\alpha}_1, \widehat{\alpha}_2, \widehat{\alpha}_3)$ the root of this system.

4. For $j = 1, 2, 3$, we compute

$$T_{ij} = \frac{R_{ij}}{\pi_{ij}} Y_{ij}, \quad i = 1, ..., n.$$

5. We solve equation (3.16), which in this case is a system of two equations:

$$\begin{cases} \sum_{i=1}^n X_{i1}^{(1)}(T_{i1} - X_{i1}^{(1)}\beta_1 - X_{i1}^{(2)}\beta_2) \\ + \sum_{i=1}^n X_{i2}^{(1)}(T_{i2} - X_{i2}^{(1)}\beta_1 - X_{i2}^{(2)}\beta_2) \\ + \sum_{i=1}^n X_{i3}^{(1)}(T_{i3} - X_{i3}^{(1)}\beta_1 - X_{i3}^{(2)}\beta_2) = 0 \\ \sum_{i=1}^n X_{i1}^{(2)}(T_{i1} - X_{i1}^{(1)}\beta_1 - X_{i1}^{(2)}\beta_2) \\ + \sum_{i=1}^n X_{i2}^{(2)}(T_{i2} - X_{i2}^{(1)}\beta_1 - X_{i2}^{(2)}\beta_2) \\ + \sum_{i=1}^n X_{i3}^{(2)}(T_{i3} - X_{i3}^{(1)}\beta_1 - X_{i3}^{(2)}\beta_2) = 0 \end{cases}$$

We denote the root by $\widehat{\beta} = (\widehat{\beta}_1, \widehat{\beta}_2)$.

6. We compute the bias of $\widehat{\beta}$:

$$\text{Bias}(\widehat{\beta}_1) = \widehat{\beta}_1 - \beta_{01}, \quad \text{Bias}(\widehat{\beta}_2) = \widehat{\beta}_2 - \beta_{02},$$

and we also compute the bias of $\widehat{\alpha}$:

$$\text{Bias}(\widehat{\alpha}_1) = \widehat{\alpha}_1 - \alpha_{01}, \quad \text{Bias}(\widehat{\alpha}_2) = \widehat{\alpha}_2 - \alpha_{02}, \quad \text{Bias}(\widehat{\alpha}_3) = \widehat{\alpha}_3 - \alpha_{03}.$$

7. We repeat Steps 1 to 6 above for $N = 1000$ times.

Now we compute the mean bias for the $N = 1000$ simulations for $\widehat{\alpha}_1$, $\widehat{\alpha}_2$, $\widehat{\alpha}_3$, and for $\widehat{\beta}_1$, $\widehat{\beta}_2$. We omit the calculations for $\widehat{\alpha}$, which is identical to the following calculations for $\widehat{\beta}$:

$$\overline{\text{Bias}_1} = \frac{1}{N}\sum_{k=1}^{N} \text{Bias}_k(\widehat{\beta}_1)$$

$$\overline{\text{Bias}_2} = \frac{1}{N}\sum_{k=1}^{N} \text{Bias}_k(\widehat{\beta}_2),$$

where $\text{Bias}_k(\widehat{\beta}_1)$ and $\text{Bias}_k(\widehat{\beta}_2)$ denote the values determined at the k-th iteration. We can also compute the absolute Bias:

$$\text{AB}_1 = \frac{1}{N}\sum_{k=1}^{N} |\text{Bias}_k(\widehat{\beta}_1)|$$

$$\text{AB}_2 = \frac{1}{N}\sum_{k=1}^{N} |\text{Bias}_k(\widehat{\beta}_2)|.$$

We compute the standard error for the N simulations:

$$\text{SE}_1 = \sqrt{\frac{1}{N-1}\sum_{k=1}^{N}(\text{Bias}_k(\widehat{\beta}_1) - \overline{\text{Bias}_1})^2}$$

$$\text{SE}_2 = \sqrt{\frac{1}{N-1}\sum_{k=1}^{N}(\text{Bias}_k(\widehat{\beta}_2) - \overline{\text{Bias}_2})^2}.$$

Finally, we compute the mean squared error (MSE) for the N simulations:

$$\text{MSE}_1 = (\overline{\text{Bias}_1})^2 + (\text{SE}_1)^2$$

$$\text{MSE}_2 = (\overline{\text{Bias}_2})^2 + (\text{SE}_2)^2.$$

We report these values in the following table:

	Mean Bias	Absolute Bias	Standard Error	Mean Squared Error
Covariate1	0.004721387	0.07449093	0.09496964	0.009041525
Covariate2	-0.0007333361	0.08371885	0.1072105	0.01149463

	Mean Bias	Absolute Bias	Standard Error	Mean Squared Error
α_1	0.003791433	0.1318648	0.1642152	0.02698099
α_2	0.006468194	0.1292527	0.1637577	0.02685842
α_3	0.01346578	0.1413862	0.1802378	0.03266698

3.5.2 Poisson Regression

In our `R` program, the covariates are generated separately from the standard normal distribution. Below is the explanation of the `R` program given in Appendix B.2.

1. For each $j = 1, 2, 3$, we generate variables $(X_{ij}^{(1)}, X_{ij}^{(2)}, Y_{ij})$ with $1 \leq i \leq n$ as in Steps 1 - 2 of the algorithm mentioned in Section 2.2.2.

2. We take $\alpha_{01} = 0.5$, $\alpha_{02} = 0.2$, and $\alpha_{03} = 0.3$.

a) For $j = 1$, we generate the missingness indicators R_{i1}, $i = 1, ..., n$, which are independent and identically distributed Bernoulli random variables with probability of success $\pi_{i1} = 0.5$.

b) For $j = 2$, we generate the missingness indicators R_{i2}, $i = 1, ..., n$ such that the distribution of R_{i2} given $U_{i2} = (Y_{i1}, X_{i1}^{(1)}, X_{i1}^{(2)})$ is Bernoulli with probability of success

$$\pi_{i2} = \frac{\exp(\alpha_{01}Y_{i1} + \alpha_{02}X_{i1}^{(1)} + \alpha_{03}X_{i1}^{(2)})}{1 + \exp(\alpha_{01}Y_{i1} + \alpha_{02}X_{i1}^{(1)} + \alpha_{03}X_{i1}^{(2)})}.$$

c) For $j = 3$, we generate the missingness indicators R_{i3}, $i = 1, ..., n$ such that the distribution of R_{i3} given $U_{i3} = (Y_{i2}, X_{i2}^{(1)}, X_{i2}^{(2)})$ is Bernoulli with probability of success

$$\pi_{i3} = \frac{\exp(\alpha_{01}Y_{i2} + \alpha_{02}X_{i2}^{(1)} + \alpha_{03}X_{i2}^{(2)})}{1 + \exp(\alpha_{01}Y_{i2} + \alpha_{02}X_{i2}^{(1)} + \alpha_{03}X_{i2}^{(2)})}.$$

3. We now solve for $\alpha = (\alpha_1, \alpha_2, \alpha_3)$ in the system (3.9), which in our case becomes a system of three equations:

$$\begin{cases} \sum_{i=1}^{n} \sum_{j=2}^{3} Y_{i,j-1} \left(R_{ij} - \frac{\exp(\alpha_1 Y_{i,j-1} + \alpha_2 X_{i,j-1}^{(1)} + \alpha_3 X_{i,j-1}^{(2)})}{1 + \exp(\alpha_1 Y_{i,j-1} + \alpha_2 X_{i,j-1}^{(1)} + \alpha_3 X_{i,j-1}^{(2)})} \right) = 0 \\ \sum_{i=1}^{n} \sum_{j=2}^{3} X_{i,j-1}^{(1)} \left(R_{ij} - \frac{\exp(\alpha_1 Y_{i,j-1} + \alpha_2 X_{i,j-1}^{(1)} + \alpha_3 X_{i,j-1}^{(2)})}{1 + \exp(\alpha_1 Y_{i,j-1} + \alpha_2 X_{i,j-1}^{(1)} + \alpha_3 X_{i,j-1}^{(2)})} \right) = 0 \\ \sum_{i=1}^{n} \sum_{j=2}^{3} X_{i,j-1}^{(2)} \left(R_{ij} - \frac{\exp(\alpha_1 Y_{i,j-1} + \alpha_2 X_{i,j-1}^{(1)} + \alpha_3 X_{i,j-1}^{(2)})}{1 + \exp(\alpha_1 Y_{i,j-1} + \alpha_2 X_{i,j-1}^{(1)} + \alpha_3 X_{i,j-1}^{(2)})} \right) = 0. \end{cases}$$

This system can be written as

$$\begin{cases} \sum_{i=1}^{n} Y_{i1}\left(R_{i2} - \frac{\exp(\alpha_1 Y_{i1}+\alpha_2 X_{i1}^{(1)}+\alpha_3 X_{i1}^{(2)})}{1+\exp(\alpha_1 Y_{i1}+\alpha_2 X_{i1}^{(1)}+\alpha_3 X_{i1}^{(2)})}\right) \\ \quad + \sum_{i=1}^{n} Y_{i2}\left(R_{i3} - \frac{\exp(\alpha_1 Y_{i2}+\alpha_2 X_{i2}^{(1)}+\alpha_3 X_{i2}^{(2)})}{1+\exp(\alpha_1 Y_{i2}+\alpha_2 X_{i2}^{(1)}+\alpha_3 X_{i2}^{(2)})}\right) = 0 \\ \sum_{i=1}^{n} X_{i1}^{(1)}\left(R_{i2} - \frac{\exp(\alpha_1 Y_{i1}+\alpha_2 X_{i1}^{(1)}+\alpha_3 X_{i1}^{(2)})}{1+\exp(\alpha_1 Y_{i1}+\alpha_2 X_{i1}^{(1)}+\alpha_3 X_{i1}^{(2)})}\right) \\ \quad + \sum_{i=1}^{n} X_{i2}^{(1)}\left(R_{i3} - \frac{\exp(\alpha_1 Y_{i2}+\alpha_2 X_{i2}^{(1)}+\alpha_3 X_{i2}^{(2)})}{1+\exp(\alpha_1 Y_{i2}+\alpha_2 X_{i2}^{(1)}+\alpha_3 X_{i2}^{(2)})}\right) = 0 \\ \sum_{i=1}^{n} X_{i1}^{(2)}\left(R_{i2} - \frac{\exp(\alpha_1 Y_{i1}+\alpha_2 X_{i1}^{(1)}+\alpha_3 X_{i1}^{(2)})}{1+\exp(\alpha_1 Y_{i1}+\alpha_2 X_{i1}^{(1)}+\alpha_3 X_{i1}^{(2)})}\right) \\ \quad + \sum_{i=1}^{n} X_{i2}^{(2)}\left(R_{i3} - \frac{\exp(\alpha_1 Y_{i2}+\alpha_2 X_{i2}^{(1)}+\alpha_3 X_{i2}^{(2)})}{1+\exp(\alpha_1 Y_{i2}+\alpha_2 X_{i2}^{(1)}+\alpha_3 X_{i2}^{(2)})}\right) = 0. \end{cases}$$

We solve this system in `R` using the multiroot function (the program is given in Appendix B.2).

We denote by $\widehat{\alpha} = (\widehat{\alpha}_1, \widehat{\alpha}_2, \widehat{\alpha}_3)$ the root of this system, and we compute the bias of $\widehat{\alpha}$:

$$\text{Bias}(\widehat{\alpha}_k) = \widehat{\alpha}_k - \alpha_{k0}, \quad \text{for} \quad k = 1, 2, 3.$$

4. For $j = 1, 2, 3$, we compute

$$T_{ij} = \frac{R_{ij}}{\pi_{ij}} Y_{ij}, \quad i = 1, ..., n.$$

5. We solve equation (3.17), which in this case is a system of two equations:

$$\begin{cases} \sum_{i=1}^{n} X_{i1}^{(1)}(T_{i1} - \exp(X_{i1}^{(1)}\beta_1 + X_{i1}^{(2)}\beta_2)) \\ + \sum_{i=1}^{n} X_{i2}^{(1)}(T_{i2} - \exp(X_{i2}^{(1)}\beta_1 + X_{i2}^{(2)}\beta_2)) \\ + \sum_{i=1}^{n} X_{i3}^{(1)}(T_{i3} - \exp(X_{i3}^{(1)}\beta_1 + X_{i3}^{(2)}\beta_2)) = 0 \\ \sum_{i=1}^{n} X_{i1}^{(2)}(T_{i1} - \exp(X_{i1}^{(1)}\beta_1 + X_{i1}^{(2)}\beta_2)) \\ + \sum_{i=1}^{n} X_{i2}^{(2)}(T_{i2} - \exp(X_{i2}^{(1)}\beta_1 + X_{i2}^{(2)}\beta_2)) \\ + \sum_{i=1}^{n} X_{i3}^{(2)}(T_{i3} - \exp(X_{i3}^{(1)}\beta_1 + X_{i3}^{(2)}\beta_2)) = 0 \end{cases}$$

We denote the root by $\widehat{\beta} = (\widehat{\beta}_1, \widehat{\beta}_2)$.

6. We compute the bias of $\widehat{\beta}$:

$$\text{Bias}(\widehat{\beta}_1) = \widehat{\beta}_1 - \beta_{01}, \quad \text{Bias}(\widehat{\beta}_2) = \widehat{\beta}_2 - \beta_{02},$$

and we also compute the bias of $\widehat{\alpha}$:

$$\text{Bias}(\widehat{\alpha}_1) = \widehat{\alpha}_1 - \alpha_{01}, \quad \text{Bias}(\widehat{\alpha}_2) = \widehat{\alpha}_2 - \alpha_{02}, \quad \text{Bias}(\widehat{\alpha}_3) = \widehat{\alpha}_3 - \alpha_{03}.$$

7. We repeat Steps 1 to 6 above for $N = 1000$ times.

Now we compute the mean bias for the $N = 1000$ simulations for $\widehat{\alpha}_1$, $\widehat{\alpha}_2$, $\widehat{\alpha}_3$, and for $\widehat{\beta}_1$, $\widehat{\beta}_2$. We omit the calculations for $\widehat{\alpha}$, which is identical to the following calculations for $\widehat{\beta}$:

$$\overline{\text{Bias}}_1 = \frac{1}{N}\sum_{k=1}^{N} \text{Bias}_k(\widehat{\beta}_1)$$

$$\overline{\text{Bias}}_2 = \frac{1}{N}\sum_{k=1}^{N} \text{Bias}_k(\widehat{\beta}_2),$$

where $\text{Bias}_k(\widehat{\beta}_1)$ and $\text{Bias}_k(\widehat{\beta}_2)$ denote the values determined at the k-th iteration. We can also compute the absolute Bias:

$$\text{AB}_1 = \frac{1}{N}\sum_{k=1}^{N} |\text{Bias}_k(\widehat{\beta}_1)|$$

$$\text{AB}_2 = \frac{1}{N}\sum_{k=1}^{N} |\text{Bias}_k(\widehat{\beta}_2)|.$$

We compute the standard error for the N simulations:

$$\text{SE}_1 = \sqrt{\frac{1}{N-1}\sum_{k=1}^{N}(\text{Bias}_k(\widehat{\beta}_1) - \overline{\text{Bias}}_1)^2}$$

$$\text{SE}_2 = \sqrt{\frac{1}{N-1}\sum_{k=1}^{N}(\text{Bias}_k(\widehat{\beta}_2) - \overline{\text{Bias}}_2)^2}.$$

Finally, we compute the mean squared error (MSE) for the N simulations:

$$\text{MSE}_1 = (\overline{\text{Bias}}_1)^2 + (\text{SE}_1)^2$$

$$\text{MSE}_2 = (\overline{\text{Bias}}_2)^2 + (\text{SE}_2)^2.$$

We report these values in the following table:

	Mean Bias	Absolute Bias	Standard Error	Mean Squared Error
Covariate1	-0.005004944	0.07176339	0.09033914	0.008186209
Covariate2	-0.002429365	0.07138262	0.08909096	0.007943101

	Mean Bias	Absolute Bias	Standard Error	Mean Squared Error
α_1	0.01670716	0.09973295	0.1253048	0.01598042
α_2	0.007270752	0.130634	0.1622003	0.02636181
α_3	0.007996393	0.1339124	0.171383	0.02943609

3.5.3 Logistic Regression

In our `R` program, the covariates are generated separately from the standard normal distribution. Below is the explanation of the `R` program given in Appendix B.3.

1. For each $j = 1, 2, 3$, we generate variables $(X_{ij}^{(1)}, X_{ij}^{(2)}, Y_{ij})$ with $1 \leq i \leq n$ as in Steps 1 - 2 of the algorithm mentioned in Section 2.2.3.

2. We take $\alpha_{01} = 0.5$, $\alpha_{02} = 0.2$, and $\alpha_{03} = 0.3$.

a) For $j = 1$, we generate the missingness indicators R_{i1}, $i = 1, ..., n$, which are independent and identically distributed Bernoulli random variables with probability of success $\pi_{i1} = 0.5$.

b) For $j = 2$, we generate the missingness indicators R_{i2}, $i = 1, ..., n$ such that the distribution of R_{i2} given $U_{i2} = (Y_{i1}, X_{i1}^{(1)}, X_{i1}^{(2)})$ is Bernoulli with probability of success

$$\pi_{i2} = \frac{\exp(\alpha_{01} Y_{i1} + \alpha_{02} X_{i1}^{(1)} + \alpha_{03} X_{i1}^{(2)})}{1 + \exp(\alpha_{01} Y_{i1} + \alpha_{02} X_{i1}^{(1)} + \alpha_{03} X_{i1}^{(2)})}.$$

c) For $j = 3$, we generate the missingness indicators R_{i3}, $i = 1, ..., n$ such that the distribution of R_{i3} given $U_{i3} = (Y_{i2}, X_{i2}^{(1)}, X_{i2}^{(2)})$ is Bernoulli with probability of success

$$\pi_{i3} = \frac{\exp(\alpha_{01} Y_{i2} + \alpha_{02} X_{i2}^{(1)} + \alpha_{03} X_{i2}^{(2)})}{1 + \exp(\alpha_{01} Y_{i2} + \alpha_{02} X_{i2}^{(1)} + \alpha_{03} X_{i2}^{(2)})}.$$

3. We now solve for $\alpha = (\alpha_1, \alpha_2, \alpha_3)$ in the system (3.9), which in our case becomes a system of three equations:

$$\begin{cases} \sum_{i=1}^{n}\sum_{j=2}^{3} Y_{i,j-1}\left(R_{ij} - \frac{\exp(\alpha_1 Y_{i,j-1}+\alpha_2 X_{i,j-1}^{(1)}+\alpha_3 X_{i,j-1}^{(2)})}{1+\exp(\alpha_1 Y_{i,j-1}+\alpha_2 X_{i,j-1}^{(1)}+\alpha_3 X_{i,j-1}^{(2)})}\right) = 0 \\ \sum_{i=1}^{n}\sum_{j=2}^{3} X_{i,j-1}^{(1)}\left(R_{ij} - \frac{\exp(\alpha_1 Y_{i,j-1}+\alpha_2 X_{i,j-1}^{(1)}+\alpha_3 X_{i,j-1}^{(2)})}{1+\exp(\alpha_1 Y_{i,j-1}+\alpha_2 X_{i,j-1}^{(1)}+\alpha_3 X_{i,j-1}^{(2)})}\right) = 0 \\ \sum_{i=1}^{n}\sum_{j=2}^{3} X_{i,j-1}^{(2)}\left(R_{ij} - \frac{\exp(\alpha_1 Y_{i,j-1}+\alpha_2 X_{i,j-1}^{(1)}+\alpha_3 X_{i,j-1}^{(2)})}{1+\exp(\alpha_1 Y_{i,j-1}+\alpha_2 X_{i,j-1}^{(1)}+\alpha_3 X_{i,j-1}^{(2)})}\right) = 0. \end{cases}$$

This system can be written as

$$\begin{cases} \sum_{i=1}^{n} Y_{i1}\left(R_{i2} - \frac{\exp(\alpha_1 Y_{i1}+\alpha_2 X_{i1}^{(1)}+\alpha_3 X_{i1}^{(2)})}{1+\exp(\alpha_1 Y_{i1}+\alpha_2 X_{i1}^{(1)}+\alpha_3 X_{i1}^{(2)})}\right) \\ \quad + \sum_{i=1}^{n} Y_{i2}\left(R_{i3} - \frac{\exp(\alpha_1 Y_{i2}+\alpha_2 X_{i2}^{(1)}+\alpha_3 X_{i2}^{(2)})}{1+\exp(\alpha_1 Y_{i2}+\alpha_2 X_{i2}^{(1)}+\alpha_3 X_{i2}^{(2)})}\right) = 0 \\ \sum_{i=1}^{n} X_{i1}^{(1)}\left(R_{i2} - \frac{\exp(\alpha_1 Y_{i1}+\alpha_2 X_{i1}^{(1)}+\alpha_3 X_{i1}^{(2)})}{1+\exp(\alpha_1 Y_{i1}+\alpha_2 X_{i1}^{(1)}+\alpha_3 X_{i1}^{(2)})}\right) \\ \quad + \sum_{i=1}^{n} X_{i2}^{(1)}\left(R_{i3} - \frac{\exp(\alpha_1 Y_{i2}+\alpha_2 X_{i2}^{(1)}+\alpha_3 X_{i2}^{(2)})}{1+\exp(\alpha_1 Y_{i2}+\alpha_2 X_{i2}^{(1)}+\alpha_3 X_{i2}^{(2)})}\right) = 0 \\ \sum_{i=1}^{n} X_{i1}^{(2)}\left(R_{i2} - \frac{\exp(\alpha_1 Y_{i1}+\alpha_2 X_{i1}^{(1)}+\alpha_3 X_{i1}^{(2)})}{1+\exp(\alpha_1 Y_{i1}+\alpha_2 X_{i1}^{(1)}+\alpha_3 X_{i1}^{(2)})}\right) \\ \quad + \sum_{i=1}^{n} X_{i2}^{(2)}\left(R_{i3} - \frac{\exp(\alpha_1 Y_{i2}+\alpha_2 X_{i2}^{(1)}+\alpha_3 X_{i2}^{(2)})}{1+\exp(\alpha_1 Y_{i2}+\alpha_2 X_{i2}^{(1)}+\alpha_3 X_{i2}^{(2)})}\right) = 0. \end{cases}$$

We solve this system in `R` using the multiroot function (the program is given in Appendix B.3).

We denote by $\widehat{\alpha} = (\widehat{\alpha}_1, \widehat{\alpha}_2, \widehat{\alpha}_3)$ the root of this system.

4. For $j = 1, 2, 3$, we compute

$$T_{ij} = \frac{R_{ij}}{\pi_{ij}} Y_{ij}, \quad i = 1, ..., n.$$

5. We solve equation (3.18), which in this case is a system of two equations:

$$\begin{cases} \sum_{i=1}^{n} X_{i1}^{(1)}(T_{i1} - \exp(X_{i1}^{(1)}\beta_1 + X_{i1}^{(2)}\beta_2)/(1+\exp(X_{i1}^{(1)}\beta_1 + X_{i1}^{(2)}\beta_2))) \\ + \sum_{i=1}^{n} X_{i2}^{(1)}(T_{i2} - \exp(X_{i2}^{(1)}\beta_1 + X_{i2}^{(2)}\beta_2)/(1+\exp(X_{i2}^{(1)}\beta_1 + X_{i2}^{(2)}\beta_2))) \\ + \sum_{i=1}^{n} X_{i3}^{(1)}(T_{i3} - \exp(X_{i3}^{(1)}\beta_1 + X_{i3}^{(2)}\beta_2)/(1+\exp(X_{i3}^{(1)}\beta_1 + X_{i3}^{(2)}\beta_2))) = 0 \\ \sum_{i=1}^{n} X_{i1}^{(2)}(T_{i1} - \exp(X_{i1}^{(1)}\beta_1 + X_{i1}^{(2)}\beta_2)/(1+\exp(X_{i1}^{(1)}\beta_1 + X_{i1}^{(2)}\beta_2))) \\ + \sum_{i=1}^{n} X_{i2}^{(2)}(T_{i2} - \exp(X_{i2}^{(1)}\beta_1 + X_{i2}^{(2)}\beta_2)/(1+\exp(X_{i2}^{(1)}\beta_1 + X_{i2}^{(2)}\beta_2))) \\ + \sum_{i=1}^{n} X_{i3}^{(2)}(T_{i3} - \exp(X_{i3}^{(1)}\beta_1 + X_{i3}^{(2)}\beta_2)/(1+\exp(X_{i3}^{(1)}\beta_1 + X_{i3}^{(2)}\beta_2))) = 0 \end{cases}$$

We denote the root by $\widehat{\beta} = (\widehat{\beta}_1, \widehat{\beta}_2)$.

6. We compute the bias of $\widehat{\beta}$:

$$\mathrm{Bias}(\widehat{\beta}_1) = \widehat{\beta}_1 - \beta_{01}, \quad \mathrm{Bias}(\widehat{\beta}_2) = \widehat{\beta}_2 - \beta_{02},$$

and we also compute the bias of $\widehat{\alpha}$:

$$\mathrm{Bias}(\widehat{\alpha}_1) = \widehat{\alpha}_1 - \alpha_{01}, \quad \mathrm{Bias}(\widehat{\alpha}_2) = \widehat{\alpha}_2 - \alpha_{02}, \quad \mathrm{Bias}(\widehat{\alpha}_3) = \widehat{\alpha}_3 - \alpha_{03}.$$

7. We repeat Steps 1 to 6 above for $N = 1000$ times.

Now we compute the mean bias for the $N = 1000$ simulations for $\widehat{\alpha}_1$, $\widehat{\alpha}_2$, $\widehat{\alpha}_3$, and for $\widehat{\beta}_1$, $\widehat{\beta}_2$. We omit the calculations for $\widehat{\alpha}$, which is identical to the following calculations for $\widehat{\beta}$:

$$\overline{\mathrm{Bias}}_1 = \frac{1}{N}\sum_{k=1}^{N} \mathrm{Bias}_k(\widehat{\beta}_1)$$

$$\overline{\mathrm{Bias}}_2 = \frac{1}{N}\sum_{k=1}^{N} \mathrm{Bias}_k(\widehat{\beta}_2),$$

where $\mathrm{Bias}_k(\widehat{\beta}_1)$ and $\mathrm{Bias}_k(\widehat{\beta}_2)$ denote the values determined at the k-th iteration. We can also compute the absolute Bias:

$$\mathrm{AB}_1 = \frac{1}{N}\sum_{k=1}^{N} |\mathrm{Bias}_k(\widehat{\beta}_1)|$$

$$\mathrm{AB}_2 = \frac{1}{N}\sum_{k=1}^{N} |\mathrm{Bias}_k(\widehat{\beta}_2)|.$$

We compute the standard error for the N simulations:

$$\mathrm{SE}_1 = \sqrt{\frac{1}{N-1}\sum_{k=1}^{N}(\mathrm{Bias}_k(\widehat{\beta}_1) - \overline{\mathrm{Bias}}_1)^2}$$

$$\mathrm{SE}_2 = \sqrt{\frac{1}{N-1}\sum_{k=1}^{N}(\mathrm{Bias}_k(\widehat{\beta}_2) - \overline{\mathrm{Bias}}_2)^2}.$$

Finally, we compute the mean squared error (MSE) for the N simulations:

$$\mathrm{MSE}_1 = (\overline{\mathrm{Bias}}_1)^2 + (\mathrm{SE}_1)^2$$

$$\mathrm{MSE}_2 = (\overline{\mathrm{Bias}}_2)^2 + (\mathrm{SE}_2)^2.$$

We report these values in the following table:

	Mean Bias	Absolute Bias	Standard Error	Mean Squared Error
Covariate1	0.01350436	0.1705404	0.2158289	0.0467645
Covariate2	0.01274661	0.1844335	0.2377458	0.05668553

	Mean Bias	Absolute Bias	Standard Error	Mean Squared Error
α_1	-0.0069865	0.1790168	0.2258149	0.05104117
α_2	0.003313798	0.1202162	0.149739	0.02243275
α_3	0.01888227	0.126332	0.1583981	0.0254465

Chapter 4

Measurement Error in the Covariates

In this chapter, we examine the regression model for longitudinal data introduced in [8], in which the covariates are measured with error. As before, the response variables follow the marginal model (2.3). We partition the covariates into two groups: covariates X_i, which are recorded with error, and covariates Z_i, which are observed and recorded without error. The covariates X_i are latent, in the sense that they are not observable, and they are represented in the analysis by surrogate variables W_i. As in [7] and [8], the connection between the latent and the surrogate variables is assumed to follow an additive error model. The unbiased estimating functions used so far were built on correctly specified covariates, and replacing some of these by surrogates will destroy the unbiasedness of these functions. The main purpose of this chapter is to "restore" unbiasedness, by building new unbiased estimating functions, which represent the old ones but use the observed, surrogate variables, in lieu of the unobserved, latent variables.

4.1 Measurement Error without Missing Response

In this section, we consider a regression model in which the covariates are measured with error and there are no missing data among the responses.

4.1.1 Methodology

As in Chapter 3, we denote by Y_{ij} the response of the individual i at time j. We let X_{ij} be the unobservable (or latent) covariate, and Z_{ij} the covariate observed exactly (without error). We assume that X_{ij} has dimension p_1 and Z_{ij} has dimention p_2, with $p_1+p_2=p$.

We let $Y_i=(Y_{i1},...,Y_{im})^T$, $X_i=(X_{i1},...,X_{im})^T$, and $Z_i=(Z_{i1},...,Z_{im})^T$.

We consider the model:

$$\begin{cases} Y_{ij}=\mu(X_{ij}^T\beta_x+Z_{ij}^T\beta_z)+\varepsilon_{ij} \\ W_{ij}=X_{ij}+e_{ij}, \end{cases} \tag{4.1}$$

where W_{ij} is the observed value of the covariate X_{ij} and e_{ij} is an error term. We define $\varepsilon_i=(\varepsilon_{i1},...,\varepsilon_{im})^T$, $W_i=(W_{i1},...,W_{im})^T$, $e_i=(e_{i1},...,e_{im})^T$, and $\beta=(\beta_x^T,\beta_z^T)^T$. Note that β is a parameter of dimension p. The first equality in (4.1) is the marginal model for the response variables, whereas the second is the link between the surrogates W_i, and the true, but unobserved covariates X_i.

We assume that:

$$\begin{cases} \mu_{ij}(\beta):=E(Y_{ij}|X_{ij},Z_{ij})=\mu(X_{ij}^T\beta_x+Z_{ij}^T\beta_z) \\ v_{ij}(\beta):=\mathrm{Var}(Y_{ij}|X_{ij},Z_{ij})=h(\mu_{ij}(\beta)) \end{cases}$$

for a certain function h. We impose the following conditions on the model:

a) $\{(X_i,Z_i,Y_i,e_i);1\le i\le n\}$ are independent and identically distributed;

b) e_{ij} is independent of (X_i,Z_i,Y_i) for any $1\le j\le m$ and $1\le i\le n$;

c) e_{ij} has a $N_{p_1}(0,\Sigma_j)$ distribution for any $1\le j\le m$ and $1\le i\le n$.

Note that by assumption b) and the fact that $W_{ij}=X_{ij}+e_{ij}$, it follows that W_{ij} is conditionally independent of (Y_{ij},Z_{ij}) given X_{ij}. Therefore, for any measurable function h,

$$E[h(W_{ij})|Y_{ij},X_{ij},Z_{ij}]=E[h(W_{ij})|X_{ij}]. \tag{4.2}$$

Relation (4.2) will be used frequently in what follows.

We remark that, since $W_{ij}=X_{ij}+e_{ij}$ and e_{ij} has a $N_{p_1}(0,\Sigma_j)$ distribution, W_{ij} has a $N_{p_1}(X_{ij},\Sigma_j)$ distribution given X_{ij}.

Lemma 4.1.1. *We have*

$$\begin{aligned} E[W_{ij}|Y_{ij},X_{ij},Z_{ij}] &= X_{ij} \\ E[W_{ij}W_{ij}^T|Y_{ij},X_{ij},Z_{ij}] &= X_{ij}X_{ij}^T+\Sigma_j. \end{aligned} \tag{4.3}$$

<u>Proof</u>: From the form of the error model (4.1), we have:

$$\begin{aligned} E[W_{ij}|Y_{ij},X_{ij},Z_{ij}] &= X_{ij}+E[e_{ij}|Y_{ij},X_{ij},Z_{ij}] \\ &= X_{ij}, \end{aligned}$$

since $E[e_{ij}|Y_{ij}, X_{ij}, Z_{ij}] = 0$, by assumptions b) and c).

Furthermore,

$$\begin{aligned} & E[W_{ij}W_{ij}^T|Y_{ij}, X_{ij}, Z_{ij}] \\ &= X_{ij}X_{ij}^T + E[e_{ij}|Y_{ij}, X_{ij}, Z_{ij}]X_{ij}^T + X_{ij}E[e_{ij}^T|Y_{ij}, X_{ij}, Z_{ij}] + E[e_{ij}e_{ij}^T|Y_{ij}, X_{ij}, Z_{ij}] \\ &= X_{ij}X_{ij}^T + \Sigma_j, \end{aligned}$$

where we used assumptions b) and c) for the last equality. □

We need the following result.

Lemma 4.1.2. *Let W be a p-dimensional random vector with $N_p(\mu, \Sigma)$ distribution, where $\mu \in \mathbb{R}^p$ and Σ is a symmetric and non-negative definite $p \times p$ matrix. Then, for any $a \in \mathbb{R}^p$, we have*

a) $E(e^{W^T a}) = \exp(\mu^T a + \frac{1}{2}a^T \Sigma a)$;

b) $E(We^{W^T a}) = (\mu + \Sigma a)\exp(\mu^T a + \frac{1}{2}a^T\Sigma a)$.

The goal of this section is to construct an unbiased estimating function, which depends only on the observed data $\{(Y_{ij}, W_{ij}, Z_{ij}); 1 \le i \le n,\ 1 \le j \le m\}$ and the parameter β. Similarly to Chapter 3, we consider the following estimating function:

$$\mathcal{G}_{ij}(Y_{ij}, X_{ij}, Z_{ij}, \beta) = \frac{\partial \mu_{ij}}{\partial \beta^T}(\beta)v_{ij}^{-1}(\beta)(Y_{ij} - \mu_{ij}(\beta)).$$

Example 4.1.3. *(Normal Linear Regression)*

Suppose that the conditional distribution of Y_{ij} given (X_{ij}, Z_{ij}) is normal with mean $\mu_{ij}(\beta) = X_{ij}^T\beta_x + Z_{ij}^T\beta_z$ and variance $v_{ij}(\beta) = 1$. In this case,

$$\mathcal{G}_{ij}(Y_{ij}, X_{ij}, Z_{ij}, \beta) = (X_{ij}^T, Z_{ij}^T)^T(Y_{ij} - X_{ij}^T\beta_x - Z_{ij}^T\beta_z). \tag{4.4}$$

Example 4.1.4. *(Poisson Regression)*

We assume that Y_{ij} is a count variable such that the conditional distribution of Y_{ij} given (X_{ij}, Z_{ij}) is Poisson with mean $\mu_{ij}(\beta) = \exp(X_{ij}^T\beta_x + Z_{ij}^T\beta_z)$. Hence, $v_{ij}(\beta) = \mu_{ij}(\beta)$. In this case,

$$\mathcal{G}_{ij}(Y_{ij}, X_{ij}, Z_{ij}, \beta) = (X_{ij}^T, Z_{ij}^T)^T\{Y_{ij} - \exp(X_{ij}^T\beta_x + Z_{ij}^T\beta_z)\}. \tag{4.5}$$

Example 4.1.5. *(Logistic Regression)*

We assume that Y_{ij} is a binary variable such that the conditional distribution of Y_{ij} given (X_{ij}, Z_{ij}) is Bernoulli with probability of success

$$\mu_{ij}(\beta) = \frac{\exp(X_{ij}^T\beta_x + Z_{ij}^T\beta_z)}{1 + \exp(X_{ij}^T\beta_x + Z_{ij}^T\beta_z)}.$$

In this case, $v_{ij}(\beta) = \mu_{ij}(\beta)(1 - \mu_{ij}(\beta))$, and

$$\mathcal{G}_{ij}(Y_{ij}, X_{ij}, Z_{ij}, \beta) = (X_{ij}^T, Z_{ij}^T)^T \left(Y_{ij} - \frac{\exp(X_{ij}^T\beta_x + Z_{ij}^T\beta_z)}{1 + \exp(X_{ij}^T\beta_x + Z_{ij}^T\beta_z)} \right). \tag{4.6}$$

We have the following preliminary result.

Lemma 4.1.6. *For any $1 \leq i \leq n$ and $1 \leq j \leq m$,*

$$E[\mathcal{G}_{ij}(Y_{ij}, X_{ij}, Z_{ij}, \beta) | X_{ij}, Z_{ij}] = 0. \tag{4.7}$$

Hence, $\mathcal{G}_{ij}(Y_{ij}, X_{ij}, Z_{ij}, \beta)$ is an unbiased estimating function, i.e.

$$E[\mathcal{G}_{ij}(Y_{ij}, X_{ij}, Z_{ij}, \beta)] = 0.$$

Proof: We have

$$E[\mathcal{G}_{ij}(Y_{ij}, X_{ij}, Z_{ij}, \beta) | X_{ij}, Z_{ij}] = \frac{\partial \mu_{ij}}{\partial \beta^T}(\beta) v_{ij}^{-1}(\beta)[E(Y_{ij}|X_{ij}, Z_{ij}) - \mu_{ij}(\beta)] = 0,$$

where we used the fact that $\mu_{ij}(\beta)$ and $v_{ij}(\beta)$ are functions of (X_{ij}, Z_{ij}). The fact that $\mathcal{G}_{ij}(Y_{ij}, X_{ij}, Z_{ij}, \beta)$ is unbiased follows since

$$E[\mathcal{G}_{ij}(Y_{ij}, X_{ij}, Z_{ij}, \beta)] = E[E[\mathcal{G}_{ij}(Y_{ij}, X_{ij}, Z_{ij}, \beta) | X_{ij}, Z_{ij}]] = 0. \qquad \square$$

Note that the estimating function $\mathcal{G}_{ij}(Y_{ij}, X_{ij}, Z_{ij}, \beta)$ depends on the unobservable variable X_{ij}. If we replace X_{ij} by its observed value W_{ij} in the expression of $\mathcal{G}_{ij}$, the resulting estimating function will no longer be unbiased. To avoid this problem, we proceed as in Section 3.2 of [8]. The idea is to construct another estimating function $\mathcal{G}_{ij}^*(Y_{ij}, W_{ij}, Z_{ij}, \beta)$ such that

$$E[\mathcal{G}_{ij}^*(Y_{ij}, W_{ij}, Z_{ij}, \beta) | Y_{ij}, X_{ij}, Z_{ij}] = \mathcal{G}_{ij}(Y_{ij}, X_{ij}, Z_{ij}, \beta). \tag{4.8}$$

Lemma 4.1.7. *If* (4.8) *holds, then $\mathcal{G}_{ij}^*(Y_{ij}, W_{ij}, Z_{ij}, \beta)$ is an unbiased estimating function, i.e.*

$$E[\mathcal{G}_{ij}^*(Y_{ij}, W_{ij}, Z_{ij}, \beta)] = 0.$$

Proof: By a conditioning argument,

$$\begin{aligned} E[\mathcal{G}^*_{ij}(Y_{ij}, W_{ij}, Z_{ij}, \beta)] &= E[E[\mathcal{G}^*_{ij}(Y_{ij}, W_{ij}, Z_{ij}, \beta)|Y_{ij}, X_{ij}, Z_{ij}]] \\ &= E[\mathcal{G}_{ij}(Y_{ij}, X_{ij}, Z_{ij}, \beta)] = 0, \end{aligned}$$

where we used relation (4.8) for the second equality, and Lemma 4.1.6 for the third equality. □

In the case of Examples 4.1.3 and 4.1.4, one can find an expression for $\mathcal{G}^*_{ij}$ such that (4.8) holds (see [7]. However, in the case of Example 4.1.5, there is no estimating function $\mathcal{G}^*_{ij}$ for which relation (4.8) holds. Therefore, the authors of [8] introduced a new method which works for the logistic regression. We explain this method below.

We define a *weighted* estimating function

$$\mathcal{G}_{wij}(Y_{ij}, X_{ij}, Z_{ij}, \beta) = \eta(X_{ij}, Z_{ij}, \beta)\mathcal{G}_{ij}(Y_{ij}, X_{ij}, Z_{ij}, \beta) \tag{4.9}$$

for a certain weight function $\eta(X_{ij}, Z_{ij}, \beta)$, which depends only on X_{ij}, Z_{ij}, and β. We construct another estimating function $\mathcal{G}^*_{wij}(Y_{ij}, W_{ij}, Z_{ij}, \beta)$ such that

$$E[\mathcal{G}^*_{wij}(Y_{ij}, W_{ij}, Z_{ij}, \beta)|Y_{ij}, X_{ij}, Z_{ij}] = \mathcal{G}_{wij}(Y_{ij}, X_{ij}, Z_{ij}, \beta). \tag{4.10}$$

Lemma 4.1.8. *If* (4.10) *holds, then* $\mathcal{G}^*_{wij}(Y_{ij}, W_{ij}, Z_{ij}, \beta)$ *is an unbiased estimating function i.e.*

$$E[\mathcal{G}^*_{wij}(Y_{ij}, W_{ij}, Z_{ij}, \beta)] = 0.$$

Proof: By a conditioning argument,

$$\begin{aligned} E[\mathcal{G}^*_{wij}(Y_{ij}, W_{ij}, Z_{ij}, \beta)] &= E[E[\mathcal{G}^*_{wij}(Y_{ij}, W_{ij}, Z_{ij}, \beta)|Y_{ij}, X_{ij}, Z_{ij}]] \\ &= E[\mathcal{G}_{wij}(Y_{ij}, X_{ij}, Z_{ij}, \beta)] \\ &= E[\eta(X_{ij}, Z_{ij}, \beta)\mathcal{G}_{ij}(Y_{ij}, Z_{ij}, \beta)] \\ &= E[\eta(X_{ij}, Z_{ij}, \beta)E[\mathcal{G}_{ij}(Y_{ij}, X_{ij}, Z_{ij}, \beta)|X_{ij}, Z_{ij}]] \\ &= 0 \end{aligned}$$

where we used relation (4.10) for the second equality, (4.9) for the third equality, and (4.7) for the last equality. □

4.1.2 Normal Linear Regression: Expression of $\mathcal{G}^*_{ij}$

For Example 4.1.3, we define:

$$\mathcal{G}^*_{ij}(Y_{ij}, W_{ij}, Z_{ij}, \beta) = (W_{ij}^T, Z_{ij}^T)^T [Y_{ij} - W_{ij}^T \beta_x - Z_{ij}^T \beta_z] + (\Sigma_j \beta_x, 0_{p_2}). \tag{4.11}$$

Proposition 4.1.9. *The estimating function defined by* (4.11) *satisfies* (4.7) *(hence it is unbiased, by Lemma 4.1.7).*

Proof: Note that $\mathcal{G}^*_{ij}(Y_{ij}, W_{ij}, Z_{ij}, \beta)$ is a vector of dimension $p = p_1 + p_2$, as we have

$$\mathcal{G}^*_{ij}(Y_{ij}, W_{ij}, Z_{ij}, \beta) = \begin{bmatrix} \mathcal{G}^*_{ij,1}(Y_{ij}, W_{ij}, Z_{ij}, \beta) \\ \mathcal{G}^*_{ij,2}(Y_{ij}, W_{ij}, Z_{ij}, \beta) \end{bmatrix},$$

where $\mathcal{G}^*_{ij,1}(Y_{ij}, W_{ij}, Z_{ij}, \beta)$ and $\mathcal{G}^*_{ij,2}(Y_{ij}, W_{ij}, Z_{ij}, \beta)$ are random vectors of dimensions p_1 and p_2, respectively, given by:

$$\begin{aligned} \mathcal{G}^*_{ij,1}(Y_{ij}, W_{ij}, Z_{ij}, \beta) &= W_{ij} Y_{ij} - W_{ij} W_{ij}^T \beta_x - W_{ij} Z_{ij}^T \beta_z + \Sigma_j \beta_x \\ \mathcal{G}^*_{ij,2}(Y_{ij}, W_{ij}, Z_{ij}, \beta) &= Z_{ij} Y_{ij} - Z_{ij} W_{ij}^T \beta_x - Z_{ij} Z_{ij}^T \beta_z. \end{aligned}$$

On the other hand, $\mathcal{G}_{ij}(Y_{ij}, X_{ij}, Z_{ij}, \beta)$ in (4.4) can also be written as

$$\mathcal{G}_{ij}(Y_{ij}, X_{ij}, Z_{ij}, \beta) = \begin{bmatrix} \mathcal{G}_{ij,1}(Y_{ij}, X_{ij}, Z_{ij}, \beta) \\ \mathcal{G}_{ij,2}(Y_{ij}, X_{ij}, Z_{ij}, \beta) \end{bmatrix},$$

where $\mathcal{G}_{ij,1}(Y_{ij}, X_{ij}, Z_{ij}, \beta)$ and $\mathcal{G}_{ij,2}(Y_{ij}, X_{ij}, Z_{ij}, \beta)$ are random vectors of dimensions p_1 and p_2, respectively, given by:

$$\begin{aligned} \mathcal{G}_{ij,1}(Y_{ij}, X_{ij}, Z_{ij}, \beta) &= X_{ij} Y_{ij} - X_{ij} X_{ij}^T \beta_x - X_{ij} Z_{ij}^T \beta_z \\ \mathcal{G}_{ij,2}(Y_{ij}, X_{ij}, Z_{ij}, \beta) &= Z_{ij} Y_{ij} - Z_{ij} X_{ij}^T \beta_x - Z_{ij} Z_{ij}^T \beta_z. \end{aligned}$$

We prove (4.8) on components, i.e.

$$E[\mathcal{G}^*_{ij,1}(Y_{ij}, W_{ij}, Z_{ij}, \beta) | Y_{ij}, X_{ij}, Z_{ij}] = \mathcal{G}_{ij,1}(Y_{ij}, X_{ij}, Z_{ij}, \beta) \tag{4.12}$$

$$E[\mathcal{G}^*_{ij,2}(Y_{ij}, W_{ij}, Z_{ij}, \beta) | Y_{ij}, X_{ij}, Z_{ij}] = \mathcal{G}_{ij,2}(Y_{ij}, X_{ij}, Z_{ij}, \beta). \tag{4.13}$$

We start with (4.12):

$$\begin{aligned} &E[\mathcal{G}^*_{ij,1}(Y_{ij}, W_{ij}, Z_{ij}, \beta) | Y_{ij}, X_{ij}, Z_{ij}] \\ &= Y_{ij} E[W_{ij} | Y_{ij}, X_{ij}, Z_{ij}] - E[W_{ij} W_{ij}^T | Y_{ij}, X_{ij}, Z_{ij}] \beta_x - E[W_{ij} | Y_{ij}, X_{ij}, Z_{ij}] Z_{ij}^T \beta_z + \Sigma_j \beta_x \\ &= X_{ij} Y_{ij} - X_{ij} X_{ij}^T \beta_x - \Sigma_j \beta_x + \Sigma_j \beta_x - X_{ij} Z_{ij}^T \beta_z \\ &= \mathcal{G}_{ij,1}(Y_{ij}, X_{ij}, Z_{ij}, \beta), \end{aligned}$$

where we used Lemma 4.1.1 in the second equality.

Similarly,

$$\begin{aligned}
&E[\mathcal{G}^*_{ij,2}(Y_{ij}, W_{ij}, Z_{ij}, \beta)|Y_{ij}, X_{ij}, Z_{ij}] \\
&\qquad = Z_{ij}Y_{ij} - Z_{ij}E[W_{ij}^T|Y_{ij}, X_{ij}, Z_{ij}]\beta_x - Z_{ij}Z_{ij}^T\beta_z \\
&\qquad = Z_{ij}Y_{ij} - Z_{ij}X_{ij}^T\beta_x - Z_{ij}Z_{ij}^T\beta_z \\
&\qquad = \mathcal{G}_{ij,2}(Y_{ij}, X_{ij}, Z_{ij}, \beta),
\end{aligned}$$

where we used Lemma 4.1.1 in the second equality. This completes the proof of Proposition 4.1.9. □

4.1.3 Poisson Regression: Expression of $\mathcal{G}^*_{ij}$

We use the notation $m_j(\beta_x) := \exp(\frac{1}{2}\beta_x^T\Sigma_j\beta_x)$ introduced in [7]. From Lemma 4.1.2 with $a = \beta_x$, $W = W_{ij}$, $\mu = X_{ij}$, and $\Sigma = \Sigma_j$, we have:

$$E(e^{W_{ij}^T\beta_x}|X_{ij}) = e^{X_{ij}^T\beta_x}m_j(\beta_x) \tag{4.14}$$

$$E(W_{ij}e^{W_{ij}\beta_x}|X_{ij}) = (X_{ij} + \Sigma_j\beta_x)e^{X_{ij}^T\beta_x}m_j(\beta_x). \tag{4.15}$$

For Example 4.1.4, we define:

$$\begin{aligned}
\mathcal{G}^*_{ij}(Y_{ij}, W_{ij}, Z_{ij}, \beta) = \; & (W_{ij}^T, Z_{ij}^T)^T[Y_{ij} - m_j^{-1}(\beta_x)\exp(W_{ij}^T\beta_x + Z_{ij}^T\beta_z)] \\
& +([m_j^{-1}(\beta_x)\Sigma_j\beta_x\exp(W_{ij}^T\beta_x + Z_{ij}^T\beta_z)]^T, 0_{p_2}^T)^T.
\end{aligned} \tag{4.16}$$

Note that $\mathcal{G}^*_{ij}(Y_{ij}, W_{ij}, Z_{ij}, \beta)$ is a vector of dimension $p_1 + p_2$ and we have:

$$\mathcal{G}^*_{ij}(Y_{ij}, W_{ij}, Z_{ij}, \beta) = \begin{bmatrix} \mathcal{G}^*_{ij,1}(Y_{ij}, W_{ij}, Z_{ij}, \beta) \\ \mathcal{G}^*_{ij,2}(Y_{ij}, W_{ij}, Z_{ij}, \beta) \end{bmatrix},$$

where $\mathcal{G}^*_{ij,1}(Y_{ij}, W_{ij}, Z_{ij}, \beta)$ and $\mathcal{G}^*_{ij,2}(Y_{ij}, W_{ij}, Z_{ij}, \beta)$ are random vectors of dimensions p_1 and p_2, respectively, given by:

$$\mathcal{G}^*_{ij,1}(Y_{ij}, W_{ij}, Z_{ij}, \beta) = W_{ij}Y_{ij} - m_j^{-1}(\beta_x)[W_{ij} - \Sigma_j\beta_x]\exp(W_{ij}^T\beta_x + Z_{ij}^T\beta_z) \tag{4.17}$$

$$\mathcal{G}^*_{ij,2}(Y_{ij}, W_{ij}, Z_{ij}, \beta) = Z_{ij}[Y_{ij} - m_j^{-1}(\beta_x)\exp(W_{ij}^T\beta_x + Z_{ij}^T\beta_z)]. \tag{4.18}$$

Proposition 4.1.10. *The estimating function defined by* (4.16) *satisfies* (4.8) *with* $\mathcal{G}_{ij}(Y_{ij}, X_{ij}, Z_{ij}, \beta)$ *given by* (4.5) *(and hence it is unbiased, by Lemma 4.1.8).*

Proof: We can write $\mathcal{G}_{ij}(Y_{ij}, X_{ij}, Z_{ij}, \beta)$ in (4.5) as the following:

$$\mathcal{G}_{ij}(Y_{ij}, X_{ij}, Z_{ij}, \beta) = \begin{cases} \mathcal{G}_{ij,1}(Y_{ij}, X_{ij}, Z_{ij}, \beta) \\ \mathcal{G}_{ij,2}(Y_{ij}, X_{ij}, Z_{ij}, \beta), \end{cases}$$

where $\mathcal{G}_{ij,1}(Y_{ij}, X_{ij}, Z_{ij}, \beta)$ and $\mathcal{G}_{ij,2}(Y_{ij}, X_{ij}, Z_{ij}, \beta)$ are random vectors of dimensions p_1 and p_2, respectively, given by:

$$\begin{aligned} \mathcal{G}_{ij,1}(Y_{ij}, X_{ij}, Z_{ij}, \beta) &= X_{ij}[Y_{ij} - \exp(X_{ij}^T\beta_x + Z_{ij}^T\beta_z)] \\ \mathcal{G}_{ij,2}(Y_{ij}, X_{ij}, Z_{ij}, \beta) &= Z_{ij}[Y_{ij} - \exp(X_{ij}^T\beta_x + Z_{ij}^T\beta_z)]. \end{aligned}$$

To prove (4.7), it suffices to prove that

$$E[\mathcal{G}^*_{ij,1}(Y_{ij}, W_{ij}, Z_{ij}, \beta)|Y_{ij}, X_{ij}, Z_{ij}] = X_{ij}[Y_{ij} - \exp(X_{ij}^T\beta_x + Z_{ij}^T\beta_z)] \quad (4.19)$$

$$E[\mathcal{G}^*_{ij,2}(Y_{ij}, W_{ij}, Z_{ij}, \beta)|Y_{ij}, X_{ij}, Z_{ij}] = Z_{ij}[Y_{ij} - \exp(X_{ij}^T\beta_x + Z_{ij}^T\beta_z)]. \quad (4.20)$$

We start by proving (4.19) from (4.17).

$$\begin{aligned} &E[\mathcal{G}^*_{ij,1}(Y_{ij}, W_{ij}, Z_{ij}, \beta)|Y_{ij}, X_{ij}, Z_{ij}] \\ &= X_{ij}Y_{ij} - m_j^{-1}(\beta_x)e^{Z_{ij}^T\beta_z}E[(W_{ij}e^{W_{ij}^T\beta_x} - \Sigma_j\beta_x e^{W_{ij}^T\beta_x})|X_{ij}] \\ &= X_{ij}Y_{ij} - m_j^{-1}(\beta_x)e^{Z_{ij}^T\beta_z}[(X_{ij} + \Sigma_j\beta_x)e^{X_{ij}^T\beta_x}m_j(\beta_x) - m_j^{-1}(\beta_x)\Sigma_j\beta_x e^{X_{ij}^T\beta_x}m_j(\beta_x)] \\ &= X_{ij}Y_{ij} - \exp(X_{ij}^T\beta_x + Z_{ij}^T\beta_z)[X_{ij} + \Sigma_j\beta_x - \Sigma_j\beta_x] \\ &= \mathcal{G}_{ij,1}(Y_{ij}, X_{ij}, Z_{ij}, \beta), \end{aligned}$$

where we used Lemma 4.1.1 for the first equality, and (4.14) and (4.15) for the second equality.

We now prove (4.20) from (4.18).

$$\begin{aligned} E[\mathcal{G}^*_{ij,2}(Y_{ij}, W_{ij}, Z_{ij}, \beta)|Y_{ij}, X_{ij}, Z_{ij}] &= Z_{ij}\{Y_{ij} - m_j^{-1}(\beta_x)e^{Z_{ij}^T\beta_z}E[e^{W_{ij}^T\beta_x}|X_{ij}]\} \\ &= Z_{ij}\{Y_{ij} - m_j^{-1}(\beta_x)e^{Z_{ij}^T\beta_z}e^{X_{ij}^T\beta_x}m_j(\beta_x)\} \\ &= \mathcal{G}_{ij}(Y_{ij}, X_{ij}, Z_{ij}, \beta), \end{aligned}$$

where we used (4.14) for the second equality. This completes the proof of Proposition 4.1.10. □

4.1.4 Logistic Regression: Expression of $\mathcal{G}^*_{wij}$

We now use Lemma 4.1.2 with $a = -\beta_x$ to obtain

$$E(e^{-W_{ij}^T\beta_x}|X_{ij}) = \exp\left(-X_{ij}^T\beta_x + \frac{1}{2}\beta_x^T\Sigma_j\beta_x\right), \quad \text{and} \tag{4.21}$$

$$E(W_{ij}e^{-W_{ij}^T\beta_x}|X_{ij}) = (X_{ij} - \Sigma_j\beta_x)\exp\left(-X_{ij}^T\beta_x + \frac{1}{2}\beta_x^T\Sigma_j\beta_x\right). \tag{4.22}$$

We define the weight function $\eta(X_{ij}, Z_{ij}, \beta)$ as follows:

$$\begin{aligned}\eta(X_{ij}, Z_{ij}, \beta) &= 1 + \exp[-(X_{ij}^T\beta_x + Z_{ij}^T\beta_z)] \\ &= 1 + \frac{1}{\exp(X_{ij}^T\beta_x + Z_{ij}^T\beta_z)} \\ &= \frac{1 + \exp(X_{ij}^T\beta_x + Z_{ij}^T\beta_z)}{\exp(X_{ij}^T\beta_x + Z_{ij}^T\beta_z)}\end{aligned}$$

and we observe that

$$\eta(X_{ij}, Z_{ij}, \beta) = \frac{1}{\mu(X_{ij}^T\beta_x + Z_{ij}^T\beta_z)},$$

where $\mu(x) = e^x/(1 + e^x)$.

Hence, using (4.9) and (4.6),

$$\begin{aligned}\mathcal{G}_{wij}(Y_{ij}, X_{ij}, Z_{ij}, \beta) &= \frac{1 + \exp(X_{ij}^T\beta_x + Z_{ij}^T\beta_z)}{\exp(X_{ij}^T\beta_x + Z_{ij}^T\beta_z)}(X_{ij}^T, Z_{ij}^T)^T\left(Y_{ij} - \frac{\exp(X_{ij}^T\beta_x + Z_{ij}^T\beta_z)}{1 + \exp(X_{ij}^T\beta_x + Z_{ij}^T\beta_z)}\right) \\ &= (X_{ij}^T, Z_{ij}^T)^T\left(\frac{1 + \exp(X_{ij}^T\beta_x + Z_{ij}^T\beta_z)}{\exp(X_{ij}^T\beta_x + Z_{ij}^T\beta_z)}Y_{ij} - 1\right) \\ &= (X_{ij}^T, Z_{ij}^T)^T\{Y_{ij}[1 + \exp[-(X_{ij}^T\beta_x + Z_{ij}^T\beta_z)]] - 1\}.\end{aligned} \tag{4.23}$$

Now we replace X_{ij} by W_{ij} in the previous expression, and we introduce a correction term for unbiasedness. More precisely, we define

$$\begin{aligned}\mathcal{G}^*_{wij}(Y_{ij}, W_{ij}, Z_{ij}, \beta) &= (W_{ij}^T, Z_{ij}^T)^T\left\{Y_{ij}\left[1 + \exp\left[-\left(W_{ij}^T\beta_x + Z_{ij}^T\beta_z + \frac{1}{2}\beta_x^T\Sigma_j\beta_x\right)\right]\right] - 1\right\} \\ &\quad + ((\Sigma_j\beta_x)^T, 0_{p_2}^T)^T Y_{ij}\left\{\exp\left[-\left(W_{ij}^T\beta_x + Z_{ij}^T\beta_z + \frac{1}{2}\beta_x^T\Sigma_j\beta_x\right)\right]\right\}.\end{aligned} \tag{4.24}$$

We have the following result.

Theorem 4.1.11. *The estimating functions $\mathcal{G}_{wij}$ and $\mathcal{G}^*_{wij}$ defined by* (4.23), (4.24), *respectively, satisfy* (4.10). *Consequently, $\mathcal{G}^*_{wij}$ is an unbiased estimating function.*

Proof: Note that $\mathcal{G}^*_{wij}(Y_{ij}, W_{ij}, Z_{ij}, \beta)$ is a vector of dimension p_1+p_2. More precisely, we write

$$\mathcal{G}^*_{wij}(Y_{ij}, W_{ij}, Z_{ij}, \beta) = \begin{bmatrix} \mathcal{G}^*_{wij,1}(Y_{ij}, W_{ij}, Z_{ij}, \beta) \\ \mathcal{G}^*_{wij,2}(Y_{ij}, W_{ij}, Z_{ij}, \beta) \end{bmatrix},$$

where $\mathcal{G}^*_{wij,1}(Y_{ij}, W_{ij}, Z_{ij}, \beta)$ and $\mathcal{G}^*_{wij,2}(Y_{ij}, W_{ij}, Z_{ij}, \beta)$ are random vectors of dimensions p_1 and p_2, respectively, given by:

$$\begin{aligned} \mathcal{G}^*_{wij,1}(Y_{ij}, W_{ij}, Z_{ij}, \beta) &= W_{ij}Y_{ij} + W_{ij}Y_{ij}\exp[-(W_{ij}^T\beta_x + Z_{ij}^T\beta_z + \frac{1}{2}\beta_x^T\Sigma_j\beta_x)] \\ &\quad -W_{ij} + \Sigma_j\beta_x Y_{ij}\exp\left[-\left(W_{ij}^T\beta_x + Z_{ij}^T\beta_z + \frac{1}{2}\beta_x^T\Sigma_j\beta_x\right)\right] \end{aligned} \tag{4.25}$$

$$\mathcal{G}^*_{wij,2}(Y_{ij}, W_{ij}, Z_{ij}, \beta) = Z_{ij}Y_{ij} + Z_{ij}Y_{ij}\exp\left[-\left(W_{ij}^T\beta_x + Z_{ij}^T\beta_z + \frac{1}{2}\beta_x^T\Sigma_j\beta_x\right)\right] - Z_{ij}. \tag{4.26}$$

On the other hand, $\mathcal{G}_{wij}(Y_{ij}, X_{ij}, Z_{ij}, \beta)$ is also a vector of dimension $p_1 + p_2$, which can be written as

$$\mathcal{G}_{wij}(Y_{ij}, X_{ij}, Z_{ij}, \beta) = \begin{bmatrix} \mathcal{G}_{wij,1}(Y_{ij}, X_{ij}, Z_{ij}, \beta) \\ \mathcal{G}_{wij,2}(Y_{ij}, X_{ij}, Z_{ij}, \beta) \end{bmatrix},$$

where $\mathcal{G}_{wij,1}(Y_{ij}, X_{ij}, Z_{ij}, \beta)$ and $\mathcal{G}_{wij,2}(Y_{ij}, X_{ij}, Z_{ij}, \beta)$ are random variables of dimension p_1 and p_2, respectively, given by:

$$\begin{aligned} \mathcal{G}_{wij,1}(Y_{ij}, X_{ij}, Z_{ij}, \beta) &= X_{ij}\{Y_{ij}[1 + \exp[-(X_{ij}^T\beta_x + Z_{ij}^T\beta_z)]] - 1\} \\ &= X_{ij}Y_{ij} + X_{ij}Y_{ij}\exp[-(X_{ij}^T\beta_x + Z_{ij}^T\beta_z)] - X_{ij}, \\ \mathcal{G}_{wij,2}(Y_{ij}, X_{ij}, Z_{ij}, \beta) &= Z_{ij}\{Y_{ij}[1 + \exp[-(X_{ij}^T\beta_x + Z_{ij}^T\beta_z)]] - 1\} \\ &= Z_{ij}Y_{ij} + Z_{ij}Y_{ij}\exp[-(X_{ij}^T\beta_x + Z_{ij}^T\beta_z)] - Z_{ij}. \end{aligned}$$

So to prove (4.10), it suffices to show that:

$$E[\mathcal{G}^*_{wij,1}(Y_{ij}, W_{ij}, Z_{ij}, \beta)|Y_{ij}, X_{ij}, Z_{ij}] = \mathcal{G}_{wij,1}(Y_{ij}, X_{ij}, Z_{ij}, \beta), \tag{4.27}$$

$$E[\mathcal{G}^*_{wij,2}(Y_{ij}, W_{ij}, Z_{ij}, \beta)|Y_{ij}, X_{ij}, Z_{ij}] = \mathcal{G}_{wij,2}(Y_{ij}, X_{ij}, Z_{ij}, \beta). \tag{4.28}$$

We prove (4.27) and (4.28) separately. We start with (4.27).

Using (4.25), we have:

$$
\begin{aligned}
&E[\mathcal{G}^*_{wij,1}(Y_{ij}, W_{ij}, Z_{ij}, \beta)|Y_{ij}, X_{ij}, Z_{ij}] \\
&= Y_{ij} E[W_{ij}|Y_{ij}, X_{ij}, Z_{ij}] - E(W_{ij}|Y_{ij}, X_{ij}, Z_{ij}) \\
&\quad + Y_{ij} \exp\left[-\left(Z_{ij}^T\beta_z + \frac{1}{2}\beta_x^T\Sigma_j\beta_x\right)\right] E(W_{ij} e^{-W_{ij}^T\beta_x}|Y_{ij}, X_{ij}, Z_{ij}) \\
&\quad + \Sigma_j\beta_x Y_{ij} \exp\left[-\left(Z_{ij}^T\beta_z + \frac{1}{2}\beta_x^T\Sigma_j\beta_x\right)\right] E(e^{-W_{ij}^T\beta_x}|Y_{ij}, X_{ij}, Z_{ij}) \\
&= Y_{ij}X_{ij} - X_{ij} \\
&\quad + Y_{ij} \exp\left[-\left(Z_{ij}^T\beta_z + \frac{1}{2}\beta_x^T\Sigma_j\beta_x\right)\right] (X_{ij} - \Sigma_j\beta_x) \exp\left[-\left(X_{ij}^T\beta_x - \frac{1}{2}\beta_x^T\Sigma_j\beta_x\right)\right] \\
&\quad + \Sigma_j\beta_x Y_{ij} \exp\left[-\left(Z_{ij}^T\beta_z + \frac{1}{2}\beta_x^T\Sigma_j\beta_x\right)\right] \exp\left[-\left(X_{ij}^T\beta_x - \frac{1}{2}\beta_x^T\Sigma_j\beta_x\right)\right] \\
&= Y_{ij}X_{ij} + Y_{ij}X_{ij} \exp[-(X_{ij}^T\beta_x + Z_{ij}^T\beta_z)] - Y_{ij}\Sigma_j\beta_x \exp[-(X_{ij}^T\beta_x + Z_{ij}^T\beta_z)] \\
&\quad - X_{ij} + \Sigma_j\beta_x Y_{ij} \exp[-(X_{ij}^T\beta_x + Z_{ij}^T\beta_z)] \\
&= \mathcal{G}_{wij,1}(Y_{ij}, X_{ij}, Z_{ij}),
\end{aligned}
$$

where we used (4.21), (4.22), and (4.2) for the second equality.

To prove (4.28), using (4.26), we have:

$$
\begin{aligned}
&E[\mathcal{G}^*_{wij,2}(Y_{ij}, W_{ij}, Z_{ij}, \beta)|Y_{ij}, X_{ij}, Z_{ij}] \\
&= Z_{ij}Y_{ij} + Z_{ij}Y_{ij} \exp\left[-\left(Z_{ij}^T\beta_z + \frac{1}{2}\beta_x^T\Sigma_j\beta_x\right)\right] E(e^{-W_{ij}^T\beta_x}|Y_{ij}, X_{ij}, Z_{ij}) - Z_{ij} \\
&= Z_{ij}Y_{ij} + Z_{ij}Y_{ij} \exp\left[-\left(Z_{ij}^T\beta_z + \frac{1}{2}\beta_x^T\Sigma_j\beta_x\right)\right] \exp\left[-\left(X_{ij}^T\beta_x - \frac{1}{2}\beta_x^T\Sigma_j\beta_x\right)\right] - Z_{ij} \\
&= Z_{ij}Y_{ij} + Z_{ij}Y_{ij} \exp[-(X_{ij}^T\beta_x + Z_{ij}^T\beta_z)] - Z_{ij} \\
&= \mathcal{G}_{wij,2}(Y_{ij}, X_{ij}, Z_{ij}),
\end{aligned}
$$

where we used (4.21), (4.22), and (4.2) for the second equality. This finishes the proof. □

4.2 Measurement Error with Missing Responses

In this case, the data consists of

$$(R_{ij}Y_{ij}, W_{ij}, Z_{ij}, R_{ij}) \quad \text{for} \quad 1 \le i \le n \quad \text{and} \quad 1 \le j \le m,$$

where Y_{ij} is the response of individual i at measurement time j, W_{ij} is the observed value of the covariate X_{ij}, Z_{ij} is the covariate observed exactly, and R_{ij} is the missingness indicator. We denote by $Y_i = (Y_{i1}, ..., Y_{im})^T$, $W_i = (W_{i1}, ..., W_{im})^T$, and $R_i = (R_{i1}, ..., R_{im})^T$.

We consider the following missing at random assumption: for each $1 \leq i \leq n$ and $2 \leq j \leq m$

$$\text{(MAR)} \qquad (Y_i, W_i, Z_i) \perp\!\!\!\perp R_{ij} | (\widetilde{R}_{ij}, \widetilde{Y}_{ij}^{(o)}, \widetilde{W}_{ij}, \widetilde{Z}_{ij}),$$

where the variables $\widetilde{R}_{ij}$, $\widetilde{Y}_{ij}^{(o)}$, $\widetilde{W}_{ij}$, and $\widetilde{Z}_{ij}$ are defined as in Chapter 3.

We define the missingness probability:

$$\begin{aligned} \pi_{ij} &= P(R_{ij} = 1 | \widetilde{R}_{ij}, Y_i, W_i, Z_i) \\ &= P(R_{ij} = 1 | \widetilde{R}_{ij}, \widetilde{Y}_{ij}^{(o)}, \widetilde{W}_{ij}, \widetilde{Z}_{ij}). \end{aligned}$$

Furthermore, $\pi_{ij} = E[R_{ij} | \widetilde{R}_{ij}, Y_i, W_i, Z_i]$.

We consider the estimating function

$$\Phi_{ij}^*(\beta) = \frac{R_{ij}}{\pi_{ij}} \mathcal{G}_{wij}^*(Y_{ij}, W_{ij}, Z_{ij}, \beta),$$

where $\mathcal{G}_{wij}^*(Y_{ij}, W_{ij}, Z_{ij}, \beta)$ is a weighted estimating function, which satisfies (4.10).

Similarly to Lemma 3.3.1, we have the following result:

Lemma 4.2.1. *For any $1 \leq i \leq n$ and $1 \leq j \leq m$,*

$$E[\Phi_{ij}^*(\beta) | Y_i, W_i, Z_i] = \mathcal{G}_{wij}^*(Y_{ij}, W_{ij}, Z_{ij}, \beta), \tag{4.29}$$

and $\Phi_{ij}^*(\beta)$ is an unbiased estimating function, i.e.

$$E[\Phi_{ij}^*(\beta)] = 0.$$

Proof: Fix $i = 1, ..., n$ and $j = 1, .., m$. First note that

$$\begin{aligned} E\left[\frac{R_{ij}}{\pi_{ij}} | Y_i, W_i, Z_i\right] &= E\left[E\left[\frac{R_{ij}}{\pi_{ij}} | \widetilde{R}_{ij}, Y_i, W_i, Z_i\right] | Y_i, W_i, Z_i\right] \\ &= E\left[\frac{1}{\pi_{ij}} E[R_{ij} | \widetilde{R}_{ij}, Y_i, W_i, Z_i] | Y_i, W_i, Z_i\right] \\ &= E\left[\frac{1}{\pi_{ij}} \pi_{ij} | Y_i, W_i, Z_i\right] \\ &= 1. \end{aligned} \tag{4.30}$$

Therefore,

$$\begin{aligned} E[\Phi_{ij}^*(\beta) | Y_i, W_i, Z_i] &= E\left[\frac{R_{ij}}{\pi_{ij}} \mathcal{G}_{wij}^*(Y_{ij}, W_{ij}, Z_{ij}, \beta) | Y_i, W_i, Z_i\right] \\ &= \mathcal{G}_{wij}^*(Y_{ij}, W_{ij}, Z_{ij}, \beta) E\left[\frac{R_{ij}}{\pi_{ij}} | Y_i, W_i, Z_i\right] \\ &= \mathcal{G}_{wij}^*(Y_{ij}, W_{ij}, Z_{ij}, \beta). \end{aligned}$$

This concludes the proof of (4.29).

Finally,

$$E[\Phi^*_{ij}(\beta)] = E[E[\Phi^*_{ij}(\beta)|Y_i, W_i, Z_i]] = E[\mathcal{G}^*_{wij}(Y_{ij}, W_{ij}, Z_{ij}, \beta)] = 0,$$

where we used (4.29) for the second equality and the fact that $\mathcal{G}^*_{wij}(Y_{ij}, W_{ij}, Z_{ij}, \beta)$ is an unbiased estimating function (by Lemma 4.1.8). □

We assume that $\mathcal{G}^*_{wij}(Y_{ij}, W_{ij}, Z_{ij}, \beta)$ is a linear function of Y_{ij}, i.e.

$$\mathcal{G}^*_{wij}(Y_{ij}, W_{ij}, Z_{ij}, \beta) = A(W_{ij}, Z_{ij}, \beta)Y_{ij} - B(W_{ij}, Z_{ij}, \beta), \tag{4.31}$$

for certain functions $A(W_{ij}, Z_{ij}, \beta)$ and $B(W_{ij}, Z_{ij}, \beta)$, which depend only on W_{ij}, Z_{ij}, and β. This assumption is satisfied in the case of the logistic regression, if we use the estimating function $\mathcal{G}^*_{wij}(Y_{ij}, W_{ij}, Z_{ij}, \beta)$ given by (4.24).

As in Chapter 3, our analysis will be based on the estimating function $\Phi_{ij}(\beta)$ given by:

$$\Phi_{ij}(\beta) = \mathcal{G}^*_{wij}\left(\frac{R_{ij}}{\pi_{ij}}Y_{ij}, W_{ij}, Z_{ij}, \beta\right). \tag{4.32}$$

Lemma 4.2.2. *If the estimating function $\mathcal{G}^*_{wij}(Y_{ij}, W_{ij}, Z_{ij}, \beta)$ satisfies* (4.10) *and* (4.31), *then $\Phi_{ij}(\beta)$ is unbiased.*

Proof: Note that

$$\begin{aligned}
\Phi_{ij}(\beta) &= \mathcal{G}^*_{wij}\left(\frac{R_{ij}}{\pi_{ij}}Y_{ij}, W_{ij}, Z_{ij}, \beta\right) \\
&= A(W_{ij}, Z_{ij}, \beta)\frac{R_{ij}}{\pi_{ij}}Y_{ij} - B(W_{ij}, Z_{ij}, \beta) \\
&= \frac{R_{ij}}{\pi_{ij}}[A(W_{ij}, Z_{ij}, \beta)Y_{ij} - B(W_{ij}, Z_{ij}, \beta)] + \frac{R_{ij}}{\pi_{ij}}B(W_{ij}, Z_{ij}, \beta) - B(W_{ij}, Z_{ij}, \beta) \\
&= \frac{R_{ij}}{\pi_{ij}}\mathcal{G}^*_{wij}(Y_{ij}, W_{ij}, Z_{ij}, \beta) + \left(\frac{R_{ij}}{\pi_{ij}} - 1\right) B(W_{ij}, Z_{ij}, \beta),
\end{aligned}$$

where we used (4.31) for the second equality and for the last equality. Hence, by the definition of $\Phi^*_{ij}(\beta)$, we see that

$$\Phi_{ij}(\beta) = \Phi^*_{ij}(\beta) + \left(\frac{R_{ij}}{\pi_{ij}} - 1\right) B(W_{ij}, Z_{ij}, \beta).$$

Hence,

$$\begin{aligned} E[\Phi_{ij}(\beta)] &= E[\Phi^*_{ij}(\beta)] + E\left[\left(\frac{R_{ij}}{\pi_{ij}} - 1\right) B(W_{ij}, Z_{ij}, \beta)\right] \\ &= E[\Phi^*_{ij}(\beta)] + E\left[B(W_{ij}, Z_{ij}, \beta) E\left(\frac{R_{ij}}{\pi_{ij}} - 1 | Y_i, W_i, Z_i\right)\right] \\ &= 0, \end{aligned}$$

where for the last equality we used (4.30) and the fact that $\Phi^*_{ij}(\beta)$ is unbiased (by Lemma 4.2.1). □

For the remaining part of this section, we deviate from the method of [8].

We are interested in solving the equation:

$$g_n(\beta) := \sum_{i=1}^{n} \sum_{j=1}^{m} \Phi_{ij}(\beta) = 0, \tag{4.33}$$

where $\Phi_{ij}(\beta)$ is given by (4.32).

By Lemma 4.2.2, $g_n(\beta)$ is an unbiased estimating function. As in Chapter 3, denoting $\Psi_i(\beta) = \sum_{j=1}^{m} \Phi_{ij}(\beta)$, we have

$$g_n(\beta) = \sum_{i=1}^{n} \Psi_i(\beta).$$

As in Chapter 3, we consider that conditionally on $U_{ij} = (\widetilde{R}_{ij}, Y_{ij}^{(o)}, W_{ij}, Z_{ij})$, R_{ij} is a Bernoulli random variable with probability of success π_{ij} given by (3.8).

The analogues of Theorem 3.4.3 and Theorem 3.4.5 continue to hold in the case when α is known, and α is unknown, respectively. We omit the proofs of these results since they are identical to the proofs in Chapter 3.

Appendix A

R programs for Chapter 2

A.1 Normal Linear Regression

Below is the R program, which is used for solving equation (2.13).

```
#Normal Linear Regression with p=2 and m=3

library(rootSolve)
N=1000
Bias1=rep(NA, N)
Bias2=rep(NA, N)
for (k in (1:N))
{
n=100; b01=0.2; b02=0.5; b0=c(b01,b02)

covariate11=rnorm(n,0,1); covariate21=rnorm(n,0,1)
covariate12=rnorm(n,0,1); covariate22=rnorm(n,0,1)
covariate13=rnorm(n,0,1); covariate23=rnorm(n,0,1)  #Generate covariates

response1=rep(NA,n); mu1=rep(NA,n); q1=rep(NA,n)
response2=rep(NA,n); mu2=rep(NA,n); q2=rep(NA,n)
response3=rep(NA,n); mu3=rep(NA,n); q3=rep(NA,n)

for (i in(1:n))
{
```

```
    q1[i]=covariate11[i]*b01+covariate21[i]*b02;
    mu1[i]=q1[i];
    response1[i]=rnorm(1,mu1[i],1);
    q2[i]=covariate12[i]*b01+covariate22[i]*b02;
    mu2[i]=q2[i];
    response2[i]=rnorm(1,mu2[i],1);
    q3[i]=covariate13[i]*b01+covariate23[i]*b02;
    mu3[i]=q3[i];
    response3[i]=rnorm(1,mu3[i],1)
}                                                    #Generate responses


normal.function23=function(x11,x12,x13,x21,x22,x23,y1,y2,y3,b)
{
    prod1=x11*b[1]+x21*b[2];
    prod2=x12*b[1]+x22*b[2];
    prod3=x13*b[1]+x23*b[2];
    mean1=prod1;
    mean2=prod2;
    mean3=prod3;
    residual1=y1-mean1;
    residual2=y2-mean2;
    residual3=y3-mean3
    c(F1=sum(x11*residual1)+sum(x12*residual2)+sum(x13*residual3),
      F2=sum(x21*residual1)+sum(x22*residual2)+sum(x23*residual3))
}                                             #Define estimating function

bhat=multiroot(f=normal.function23,x11=covariate11,x12=covariate12,
    x13=covariate13,x21=covariate21,x22=covariate22,x23=covariate23,
    y1=response1,y2=response2,y3=response3,start=c(0,0))$root
                                              #Find roots

bias=bhat-b0;                                 #Compute bias
Bias1[k]=bias[1]
Bias2[k]=bias[2]
```

```
}

mean.bias1=mean(Bias1)
mean.bias2=mean(Bias2)
AB1=mean(abs(Bias1))
AB2=mean(abs(Bias2))
se.bias1=sqrt(var(Bias1))
se.bias2=sqrt(var(Bias2))
MSE1=(mean.bias1)^2 + (se.bias1)^2
MSE2=(mean.bias2)^2 + (se.bias2)^2
print(mean.bias1)
print(mean.bias2)
print(AB1)
print(AB2)
print(se.bias1)
print(se.bias2)
print(MSE1)
print(MSE2)
```

A.2 Poisson Regression

Below is the `R` program, which is used for solving equation (2.14).

```
#Poisson Regression with p=2 and m=3

library(rootSolve)
N=1000
Bias1=rep(NA, N)
Bias2=rep(NA, N)
for (k in (1:N))
{
n=100; b01=0.2; b02=0.5; b0=c(b01,b02)

covariate11=rnorm(n,0,1); covariate21=rnorm(n,0,1)
covariate12=rnorm(n,0,1); covariate22=rnorm(n,0,1)
```

```
covariate13=rnorm(n,0,1); covariate23=rnorm(n,0,1)  #Generate Covariates

response1=rep(NA,n); mu1=rep(NA,n); q1=rep(NA,n)
response2=rep(NA,n); mu2=rep(NA,n); q2=rep(NA,n)
response3=rep(NA,n); mu3=rep(NA,n); q3=rep(NA,n)

for (i in(1:n))
{
    q1[i]=covariate11[i]*b01+covariate21[i]*b02;
    mu1[i]=exp(q1[i]);
    response1[i]=rpois(1,mu1[i]);
    q2[i]=covariate12[i]*b01+covariate22[i]*b02;
    mu2[i]=exp(q2[i]);
    response2[i]=rpois(1,mu2[i]);
    q3[i]=covariate13[i]*b01+covariate23[i]*b02;
    mu3[i]=exp(q3[i]);
    response3[i]=rpois(1,mu3[i])
}                    #Generate Responses

poisson.function23=function(x11,x12,x13,x21,x22,x23,y1,y2,y3,b)
{
    prod1=x11*b[1]+x21*b[2];
    prod2=x12*b[1]+x22*b[2];
    prod3=x13*b[1]+x23*b[2];
    mean1=exp(prod1);
    mean2=exp(prod2);
    mean3=exp(prod3);
    residual1=y1-mean1;
    residual2=y2-mean2;
    residual3=y3-mean3
    c(F1=sum(x11*residual1)+sum(x12*residual2)+sum(x13*residual3),
      F2=sum(x21*residual1)+sum(x22*residual2)+sum(x23*residual3))
}                                           #Define estimating function

bhat=multiroot(f=poisson.function23,x11=covariate11,x12=covariate12,
```

```
    x13=covariate13,x21=covariate21,x22=covariate22,x23=covariate23,
    y1=response1,y2=response2,y3=response3,start=c(0,5))$root
                                                 #Find roots
bias=bhat-b0;                                    #Compute bias
Bias1[k]=bias[1]
Bias2[k]=bias[2]
}

mean.bias1=mean(Bias1)
mean.bias2=mean(Bias2)
AB1=mean(abs(Bias1))
AB2=mean(abs(Bias2))
se.bias1=sqrt(var(Bias1))
se.bias2=sqrt(var(Bias2))
MSE1=(mean.bias1)^2 + (se.bias1)^2
MSE2=(mean.bias2)^2 + (se.bias2)^2
print(mean.bias1)
print(mean.bias2)
print(AB1)
print(AB2)
print(se.bias1)
print(se.bias2)
print(MSE1)
print(MSE2)
```

A.3 Logistic Regression

Below is the `R` program, which is used for solving equation (2.15).

```
#Logistic Regression with p=2 and m=3

library(rootSolve)
N=1000
Bias1=rep(NA, N)
Bias2=rep(NA, N)
```

```
for (k in (1:N))
{
n=100; b01=0.2; b02=0.5; b0=c(b01,b02)

covariate11=rnorm(n,0,1); covariate21=rnorm(n,0,1)
covariate12=rnorm(n,0,1); covariate22=rnorm(n,0,1)
covariate13=rnorm(n,0,1); covariate23=rnorm(n,0,1)   #Generate covariates

response1=rep(NA,n); mu1=rep(NA,n); q1=rep(NA,n)
response2=rep(NA,n); mu2=rep(NA,n); q2=rep(NA,n)
response3=rep(NA,n); mu3=rep(NA,n); q3=rep(NA,n)

for (i in(1:n))
{
    q1[i]=covariate11[i]*b01+covariate21[i]*b02;
    mu1[i]=exp(q1[i])/(1+exp(q1[i]));
    response1[i]=rbinom(1,1,mu1[i]);
    q2[i]=covariate12[i]*b01+covariate22[i]*b02;
    mu2[i]=exp(q2[i])/(1+exp(q2[i]));
    response2[i]=rbinom(1,1,mu2[i]);
    q3[i]=covariate13[i]*b01+covariate23[i]*b02;
    mu3[i]=exp(q3[i])/(1+exp(q3[i]));
    response3[i]=rbinom(1,1,mu3[i])
}                                                        #Generate responses

logistic.function23=function(x11,x12,x13,x21,x22,x23,y1,y2,y3,b)
{
    prod1=x11*b[1]+x21*b[2];
    prod2=x12*b[1]+x22*b[2];
    prod3=x13*b[1]+x23*b[2];
    mean1=exp(prod1)/(1+exp(prod1));
    mean2=exp(prod2)/(1+exp(prod2));
    mean3=exp(prod3)/(1+exp(prod3));
    residual1=y1-mean1;
    residual2=y2-mean2;
```

```
    residual3=y3-mean3
    c(F1=sum(x11*residual1)+sum(x12*residual2)+sum(x13*residual3),
      F2=sum(x21*residual1)+sum(x22*residual2)+sum(x23*residual3))
}                                            #Define estimating function

bhat=multiroot(f=logistic.function23,x11=covariate11,x12=covariate12,
    x13=covariate13,x21=covariate21,x22=covariate22,x23=covariate23,
    y1=response1,y2=response2,y3=response3,start=c(0,0))$root
                                             #Find roots

bias=bhat-b0;                                #Compute bias
Bias1[k]=bias[1]
Bias2[k]=bias[2]
}

mean.bias1=mean(Bias1)
mean.bias2=mean(Bias2)
AB1=mean(abs(Bias1))
AB2=mean(abs(Bias2))
se.bias1=sqrt(var(Bias1))
se.bias2=sqrt(var(Bias2))
MSE1=(mean.bias1)^2 + (se.bias1)^2
MSE2=(mean.bias2)^2 + (se.bias2)^2
print(mean.bias1)
print(mean.bias2)
print(AB1)
print(AB2)
print(se.bias1)
print(se.bias2)
print(MSE1)
print(MSE2)
```

Appendix B

R programs for Chapter 3

B.1 Normal Linear Regression

```
#Normal Linear Regression with p=2 and m=3

library(rootSolve)
N=1000
bBias1=rep(NA, N)
bBias2=rep(NA, N)
aBias1=rep(NA, N)
aBias2=rep(NA, N)
aBias3=rep(NA, N)
for (k in (1:N))
{
n=100; a01=0.5; a02=0.2; a03=0.3; a0=c(a01,a02,a03)
b01=0.2; b02=0.5; b0=c(b01,b02)

covariate11=rnorm(n,0,1); covariate21=rnorm(n,0,1)
covariate12=rnorm(n,0,1); covariate22=rnorm(n,0,1)
covariate13=rnorm(n,0,1); covariate23=rnorm(n,0,1)  #Generate covariates

response1=rep(NA,n); mu1=rep(NA,n); q1=rep(NA,n)
response2=rep(NA,n); mu2=rep(NA,n); q2=rep(NA,n)
response3=rep(NA,n); mu3=rep(NA,n); q3=rep(NA,n)
```

```
pi1=rep(NA,n); missing1=rep(NA,n); s1=rep(NA,n)
pi2=rep(NA,n); missing2=rep(NA,n); s2=rep(NA,n)
pi3=rep(NA,n); missing3=rep(NA,n); s3=rep(NA,n)
w1=rep(NA,n);
w2=rep(NA,n);
w3=rep(NA,n)

for (i in(1:n))
{
    q1[i]=covariate11[i]*b01+covariate21[i]*b02;
    mu1[i]=q1[i];
    response1[i]=rnorm(1,mu1[i],1);
    q2[i]=covariate12[i]*b01+covariate22[i]*b02;
    mu2[i]=q2[i];
    response2[i]=rnorm(1,mu2[i],1);
    q3[i]=covariate13[i]*b01+covariate23[i]*b02;
    mu3[i]=q3[i];
    response3[i]=rnorm(1,mu3[i],1)
}                                           #Generate responses

for (i in (1:n))
{
    pi1[i]=0.5;
    missing1[i]=rbinom(1,1,pi1[i]);

    s2[i]=response1[i]*a01+covariate11[i]*a02+covariate21[i]*a03;
    pi2[i]=exp(s2[i])/(1+exp(s2[i]));
    missing2[i]=rbinom(1,1,pi2[i]);

    s3[i]=response2[i]*a01+covariate12[i]*a02+covariate22[i]*a03;
    pi3[i]=exp(s3[i])/(1+exp(s3[i]));
    missing3[i]=rbinom(1,1,pi3[i]);
}                                #Generate missingness indicators

for(i in (1:n))
```

```
{
w1[i]=(missing1[i]/pi1[i])*response1[i];
w2[i]=(missing2[i]/pi2[i])*response2[i];
w3[i]=(missing3[i]/pi3[i])*response3[i];
} #Generate W variables

missing.function=function(x11,x12,x21,x22,y1,y2,r2,r3,a)
{
    m.prod1=y1*a[1]+x11*a[2]+x21*a[3];
    m.prod2=y2*a[1]+x12*a[2]+x22*a[3];
    m.mean1=exp(m.prod1)/(1+exp(m.prod1));
    m.mean2=exp(m.prod2)/(1+exp(m.prod2));
    m.residual1=r2-m.mean1;
    m.residual2=r3-m.mean2;
    c(F1=sum(y1*m.residual1)+sum(y2*m.residual2),
      F2=sum(x11*m.residual1)+sum(x12*m.residual2),
      F3=sum(x21*m.residual1)+sum(x22*m.residual2))
}                                    #Define estimating function

ahat=multiroot(f=missing.function,x11=covariate11,x12=covariate12,
    x21=covariate21,x22=covariate22,y1=response1,y2=response2,
    r2=missing2,r3=missing3,start=c(0,0,0))$root
                                                #Find roots
normal.function23=function(x11,x12,x13,x21,x22,x23,w1,w2,w3,b)
{
    prod1=x11*b[1]+x21*b[2]; prod2=x12*b[1]+x22*b[2]; prod3=x13*b[1]+x23*b[2];
    mean1=prod1; mean2=prod2; mean3=prod3;
    residual1=w1-mean1; residual2=w2-mean2; residual3=w3-mean3
    c(F1=sum(x11*residual1)+sum(x12*residual2)+sum(x13*residual3),
      F2=sum(x21*residual1)+sum(x22*residual2)+sum(x23*residual3))
}                                                  #Define estimating function

bhat=multiroot(f=normal.function23,x11=covariate11,x12=covariate12,
x13=covariate13,x21=covariate21,x22=covariate22,x23=covariate23,
w1=w1,w2=w2,w3=w3,start=c(0,0))$root
```

```
                                                  #Find roots

bbias=bhat-b0;                                   #Compute bias
abias=ahat-a0;
bBias1[k]=bbias[1]
bBias2[k]=bbias[2]
aBias1[k]=abias[1]
aBias2[k]=abias[2]
aBias3[k]=abias[3]
}

print(ahat)
print(bhat)
mean.bbias1=mean(bBias1)
mean.bbias2=mean(bBias2)
bAB1=mean(abs(bBias1))
bAB2=mean(abs(bBias2))
se.bbias1=sqrt(var(bBias1))
se.bbias2=sqrt(var(bBias2))
bMSE1=(mean.bbias1)^2 + (se.bbias1)^2
bMSE2=(mean.bbias2)^2 + (se.bbias2)^2
print(mean.bbias1)
print(mean.bbias2)
print(bAB1)
print(bAB2)
print(se.bbias1)
print(se.bbias2)
print(bMSE1)
print(bMSE2) #Calculation for Beta

mean.abias1=mean(aBias1)
mean.abias2=mean(aBias2)
mean.abias3=mean(aBias3)
aAB1=mean(abs(aBias1))
aAB2=mean(abs(aBias2))
```

```
aAB3=mean(abs(aBias3))
se.abias1=sqrt(var(aBias1))
se.abias2=sqrt(var(aBias2))
se.abias3=sqrt(var(aBias3))
aMSE1=(mean.abias1)^2 + (se.abias1)^2
aMSE2=(mean.abias2)^2 + (se.abias2)^2
aMSE3=(mean.abias3)^2 + (se.abias3)^2
print(mean.abias1)
print(mean.abias2)
print(mean.abias3)
print(aAB1)
print(aAB2)
print(aAB3)
print(se.abias1)
print(se.abias2)
print(se.abias3)
print(aMSE1)
print(aMSE2)
print(aMSE3) #Calculation for Alpha
```

B.2 Poisson Regression

```
#Poisson Regression with p=2 and m=3

library(rootSolve)
N=1000
bBias1=rep(NA, N)
bBias2=rep(NA, N)
aBias1=rep(NA, N)
aBias2=rep(NA, N)
aBias3=rep(NA, N)
for (k in (1:N))
{
```

```
n=100; a01=0.5; a02=0.2; a03=0.3; a0=c(a01,a02,a03)
b01=0.2; b02=0.5; b0=c(b01,b02)

covariate11=rnorm(n,0,1); covariate21=rnorm(n,0,1)
covariate12=rnorm(n,0,1); covariate22=rnorm(n,0,1)
covariate13=rnorm(n,0,1); covariate23=rnorm(n,0,1)  #Generate covariates

response1=rep(NA,n); mu1=rep(NA,n); q1=rep(NA,n)
response2=rep(NA,n); mu2=rep(NA,n); q2=rep(NA,n)
response3=rep(NA,n); mu3=rep(NA,n); q3=rep(NA,n)
pi1=rep(NA,n); missing1=rep(NA,n); s1=rep(NA,n)
pi2=rep(NA,n); missing2=rep(NA,n); s2=rep(NA,n)
pi3=rep(NA,n); missing3=rep(NA,n); s3=rep(NA,n)
w1=rep(NA,n);
w2=rep(NA,n);
w3=rep(NA,n)


for (i in(1:n))
{
    q1[i]=covariate11[i]*b01+covariate21[i]*b02;
    mu1[i]=exp(q1[i]);
    response1[i]=rpois(1,mu1[i]);
    q2[i]=covariate12[i]*b01+covariate22[i]*b02;
    mu2[i]=exp(q2[i]);
    response2[i]=rpois(1,mu2[i]);
    q3[i]=covariate13[i]*b01+covariate23[i]*b02;
    mu3[i]=exp(q3[i]);
    response3[i]=rpois(1,mu3[i])
}                 #Generate Responses

for (i in (1:n))
{
    pi1[i]=0.5;
    missing1[i]=rbinom(1,1,pi1[i]);
```

```
    s2[i]=response1[i]*a01+covariate11[i]*a02+covariate21[i]*a03;
    pi2[i]=exp(s2[i])/(1+exp(s2[i]));
    missing2[i]=rbinom(1,1,pi2[i]);

    s3[i]=response2[i]*a01+covariate12[i]*a02+covariate22[i]*a03;
    pi3[i]=exp(s3[i])/(1+exp(s3[i]));
    missing3[i]=rbinom(1,1,pi3[i]);
}                  #Generate missingness indicators

for(i in (1:n))
{
w1[i]=(missing1[i]/pi1[i])*response1[i];
w2[i]=(missing2[i]/pi2[i])*response2[i];
w3[i]=(missing3[i]/pi3[i])*response3[i];
} #Generate W variables


missing.function=function(x11,x12,x21,x22,y1,y2,r2,r3,a)
{
    m.prod1=y1*a[1]+x11*a[2]+x21*a[3];
    m.prod2=y2*a[1]+x12*a[2]+x22*a[3];
    m.mean1=exp(m.prod1)/(1+exp(m.prod1));
    m.mean2=exp(m.prod2)/(1+exp(m.prod2));
    m.residual1=r2-m.mean1;
    m.residual2=r3-m.mean2;
    c(F1=sum(y1*m.residual1)+sum(y2*m.residual2),
      F2=sum(x11*m.residual1)+sum(x12*m.residual2),
      F3=sum(x21*m.residual1)+sum(x22*m.residual2))
}                    #Define estimating function

ahat=multiroot(f=missing.function,x11=covariate11,x12=covariate12,
    x21=covariate21,x22=covariate22,y1=response1,y2=response2,
    r2=missing2,r3=missing3,start=c(0,0,0))$root
                       #Find roots
```

```
poisson.function23=function(x11,x12,x13,x21,x22,x23,w1,w2,w3,b)
{
prod1=x11*b[1]+x21*b[2];
prod2=x12*b[1]+x22*b[2];
prod3=x13*b[1]+x23*b[2];
mean1=exp(prod1);
mean2=exp(prod2);
mean3=exp(prod3);
residual1=w1-mean1;
residual2=w2-mean2;
residual3=w3-mean3
     c(F1=sum(x11*residual1)+sum(x12*residual2)+sum(x13*residual3),
        F2=sum(x21*residual1)+sum(x22*residual2)+sum(x23*residual3))
}       #Define estimating function

bhat=multiroot(f=poisson.function23,x11=covariate11,x12=covariate12,
x13=covariate13,x21=covariate21,x22=covariate22,x23=covariate23,
w1=w1,w2=w2,w3=w3,start=c(0,5))$root

bbias=bhat-b0;                                  #Compute bias
abias=ahat-a0;
bBias1[k]=bbias[1]
bBias2[k]=bbias[2]
aBias1[k]=abias[1]
aBias2[k]=abias[2]
aBias3[k]=abias[3]
}

print(ahat)
print(bhat)
mean.bbias1=mean(bBias1)
mean.bbias2=mean(bBias2)
bAB1=mean(abs(bBias1))
bAB2=mean(abs(bBias2))
se.bbias1=sqrt(var(bBias1))
```

```
se.bbias2=sqrt(var(bBias2))
bMSE1=(mean.bbias1)^2 + (se.bbias1)^2
bMSE2=(mean.bbias2)^2 + (se.bbias2)^2
print(mean.bbias1)
print(mean.bbias2)
print(bAB1)
print(bAB2)
print(se.bbias1)
print(se.bbias2)
print(bMSE1)
print(bMSE2) #Calculation for Beta

mean.abias1=mean(aBias1)
mean.abias2=mean(aBias2)
mean.abias3=mean(aBias3)
aAB1=mean(abs(aBias1))
aAB2=mean(abs(aBias2))
aAB3=mean(abs(aBias3))
se.abias1=sqrt(var(aBias1))
se.abias2=sqrt(var(aBias2))
se.abias3=sqrt(var(aBias3))
aMSE1=(mean.abias1)^2 + (se.abias1)^2
aMSE2=(mean.abias2)^2 + (se.abias2)^2
aMSE3=(mean.abias3)^2 + (se.abias3)^2
print(mean.abias1)
print(mean.abias2)
print(mean.abias3)
print(aAB1)
print(aAB2)
print(aAB3)
print(se.abias1)
print(se.abias2)
print(se.abias3)
print(aMSE1)
print(aMSE2)
```

```
print(aMSE3) #Calculation for Alpha
```

B.3 Logistic Regression

```
#Logistic Regression with p=2 and m=3

library(rootSolve)
N=1000
bBias1=rep(NA, N)
bBias2=rep(NA, N)
aBias1=rep(NA, N)
aBias2=rep(NA, N)
aBias3=rep(NA, N)
for (k in (1:N))
{
n=100; a01=0.5; a02=0.2; a03=0.3; a0=c(a01,a02,a03)
b01=0.2; b02=0.5; b0=c(b01,b02)

covariate11=rnorm(n,0,1); covariate21=rnorm(n,0,1)
covariate12=rnorm(n,0,1); covariate22=rnorm(n,0,1)
covariate13=rnorm(n,0,1); covariate23=rnorm(n,0,1)  #Generate covariates

response1=rep(NA,n); mu1=rep(NA,n); q1=rep(NA,n)
response2=rep(NA,n); mu2=rep(NA,n); q2=rep(NA,n)
response3=rep(NA,n); mu3=rep(NA,n); q3=rep(NA,n)
pi1=rep(NA,n); missing1=rep(NA,n); s1=rep(NA,n)
pi2=rep(NA,n); missing2=rep(NA,n); s2=rep(NA,n)
pi3=rep(NA,n); missing3=rep(NA,n); s3=rep(NA,n)
w1=rep(NA,n);
w2=rep(NA,n);
w3=rep(NA,n)
```

```
for (i in(1:n))
{
    q1[i]=covariate11[i]*b01+covariate21[i]*b02;
    mu1[i]=exp(q1[i])/(1+exp(q1[i]));
    response1[i]=rbinom(1,1,mu1[i]);
    q2[i]=covariate12[i]*b01+covariate22[i]*b02;
    mu2[i]=exp(q2[i])/(1+exp(q2[i]));
    response2[i]=rbinom(1,1,mu2[i]);
    q3[i]=covariate13[i]*b01+covariate23[i]*b02;
    mu3[i]=exp(q3[i])/(1+exp(q3[i]));
    response3[i]=rbinom(1,1,mu3[i])
}                  #Generate responses

for (i in (1:n))
{
    pi1[i]=0.5;
    missing1[i]=rbinom(1,1,pi1[i]);

    s2[i]=response1[i]*a01+covariate11[i]*a02+covariate21[i]*a03;
    pi2[i]=exp(s2[i])/(1+exp(s2[i]));
    missing2[i]=rbinom(1,1,pi2[i]);

    s3[i]=response2[i]*a01+covariate12[i]*a02+covariate22[i]*a03;
    pi3[i]=exp(s3[i])/(1+exp(s3[i]));
    missing3[i]=rbinom(1,1,pi3[i]);
}              #Generate missingness indicators

for(i in (1:n))
{
    w1[i]=(missing1[i]/pi1[i])*response1[i];
    w2[i]=(missing2[i]/pi2[i])*response2[i];
    w3[i]=(missing3[i]/pi3[i])*response3[i];
}                                       #Generate W variables

missing.function=function(x11,x12,x21,x22,y1,y2,r2,r3,a)
```

```
{
    m.prod1=y1*a[1]+x11*a[2]+x21*a[3];
    m.prod2=y2*a[1]+x12*a[2]+x22*a[3];
    m.mean1=exp(m.prod1)/(1+exp(m.prod1));
    m.mean2=exp(m.prod2)/(1+exp(m.prod2));
    m.residual1=r2-m.mean1;
    m.residual2=r3-m.mean2;
    c(F1=sum(y1*m.residual1)+sum(y2*m.residual2),
      F2=sum(x11*m.residual1)+sum(x12*m.residual2),
      F3=sum(x21*m.residual1)+sum(x22*m.residual2))
}                                      #Define estimating function

ahat=multiroot(f=missing.function,x11=covariate11,x12=covariate12,
    x21=covariate21,x22=covariate22,y1=response1,y2=response2,
    r2=missing2,r3=missing3,start=c(0,0,0))$root
                              #Find roots
logistic.function23=function(x11,x12,x13,x21,x22,x23,w1,w2,w3,b)
{
    prod1=x11*b[1]+x21*b[2];
    prod2=x12*b[1]+x22*b[2];
    prod3=x13*b[1]+x23*b[2];
    mean1=exp(prod1)/(1+exp(prod1));
    mean2=exp(prod2)/(1+exp(prod2));
    mean3=exp(prod3)/(1+exp(prod3));
    residual1=w1-mean1;
    residual2=w2-mean2;
    residual3=w3-mean3
    c(F1=sum(x11*residual1)+sum(x12*residual2)+sum(x13*residual3),
      F2=sum(x21*residual1)+sum(x22*residual2)+sum(x23*residual3))
}                                       #Define estimating function

bhat=multiroot(f=logistic.function23,x11=covariate11,x12=covariate12,
x13=covariate13,x21=covariate21,x22=covariate22,x23=covariate23,
w1=w1,w2=w2,w3=w3,start=c(0,0))$root
                                                #Find roots
```

```
bbias=bhat-b0;                                 #Compute bias
abias=ahat-a0;
bBias1[k]=bbias[1]
bBias2[k]=bbias[2]
aBias1[k]=abias[1]
aBias2[k]=abias[2]
aBias3[k]=abias[3]
}

print(ahat)
print(bhat)
mean.bbias1=mean(bBias1)
mean.bbias2=mean(bBias2)
bAB1=mean(abs(bBias1))
bAB2=mean(abs(bBias2))
se.bbias1=sqrt(var(bBias1))
se.bbias2=sqrt(var(bBias2))
bMSE1=(mean.bbias1)^2 + (se.bbias1)^2
bMSE2=(mean.bbias2)^2 + (se.bbias2)^2
print(mean.bbias1)
print(mean.bbias2)
print(bAB1)
print(bAB2)
print(se.bbias1)
print(se.bbias2)
print(bMSE1)
print(bMSE2) #Calculation for Beta

mean.abias1=mean(aBias1)
mean.abias2=mean(aBias2)
mean.abias3=mean(aBias3)
aAB1=mean(abs(aBias1))
aAB2=mean(abs(aBias2))
aAB3=mean(abs(aBias3))
```

```
se.abias1=sqrt(var(aBias1))
se.abias2=sqrt(var(aBias2))
se.abias3=sqrt(var(aBias3))
aMSE1=(mean.abias1)^2 + (se.abias1)^2
aMSE2=(mean.abias2)^2 + (se.abias2)^2
aMSE3=(mean.abias3)^2 + (se.abias3)^2
print(mean.abias1)
print(mean.abias2)
print(mean.abias3)
print(aAB1)
print(aAB2)
print(aAB3)
print(se.abias1)
print(se.abias2)
print(se.abias3)
print(aMSE1)
print(aMSE2)
print(aMSE3) #Calculation for Alpha
```

Bibliography

[1] Casella, G. and Berger, R. (2002). Statistical Inference. 2nd Edition.

[2] Lai, T.L. and Small, D. (2007). Marginal regression analysis of longitudinal data with time-dependent covariates: a generalized method-of-moments approach, *J. R. Statist. Soc. B*, 69, Part 1, pp. 79-99.

[3] Liang, K.-Y. and Zeger, S.L. (1986). Longitudinal data analysis using Generalized Linear Models, *Biometrika*, 73, pp. 13-22.

[4] Little, R. and Rubin, D. (2002). Statistical Analysis with Missing Data. 2nd Edition.

[5] Tsiatis, A.A. (2006). Semiparametric theory and missing data.

[6] Xie, M. and Yang, Y. (2003). Asymtotics for Generalized Estimating Equations with large cluster sizes, *The Annals of Statistics*, 31, pp. 310-347.

[7] Yi, G. (2005). Robust methods for incomplete longitudinal data with mismeasured covariates, *Far East J. Theo. Stat.* 16, pp. 205-234.

[8] 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, 1, pp. 151-165.