ANALYSIS OF PANEL DATA WITH INFORMATIVE MISSING RESPONSES

by

Jie Zhang

A dissertation submitted in partial fulfillment of

the requirements for the degree of

Doctor of Philosophy

(Statistics)

at the

UNIVERSITY OF WISCONSIN–MADISON

2013

Date of final oral examination: 08/16/2013

The dissertation is approved by the following Final Oral Committee members:

Jun Shao, Professor, Statistics

Kam-Wah Tsui, Professor, Statistics

Richard Chappell, Professor, Statistics & BMI

Menggang Yu, Associate Professor, Biostatistics & Medical Informatics

Jun Zhu, Professor, Statistics

Acknowledgments

First of all, I would like to express my deepest gratitude and appreciation toward my thesis advisor Professor Jun Shao. I learned from him not only the statistics knowledge, but also the passion toward research and life. His invaluable guidance and constant encouragement led me through the past five years of my graduate study, and made this thesis work possible. I feel so honored to have this opportunity to work closely and learn from him.

My sincere thanks also go to Professor Rick Chappell, Kam-Wah Tsui and Menggang Yu for serving on my preliminary exam and defense. Your helpful suggestions and resources helped me greatly in the completion of this thesis work. I would also like to thank Professor Jun Zhu for serving on my defense committee. I am grateful to all of them for their valuable time and kind help.

I would like to extend my thanks to all faculty members, staff and students in the Department of Statistics, University of Wisconsin - Madison. My life in Madison wouldn't have been so enjoyable without them.

Finally and most importantly, the endless love and support from my parents Zhiqiang Zhang and Linlin Guo, and my husband Fuqiang Gao, are essential to every of my successes. I owe them innumerable thanks.

Contents

List of Tables

List of Figures

Abstract

Missing responses is a common problem existing in panel data collection. In some studies, one can argue that the missingness does not depend on unobserved responses given all observed information. However, in many other studies, the missing pattern is more likely to be dependent on the unobserved responses. When this dependence is indirect through some panel level random effects, it is called "informative missingness". Some parametric and semi-parametric estimation methods have been proposed in the literature for informative missing responses, where the estimation processes are mainly likelihood-based.

In this thesis work, we proposed a semi-parametric method to solve this estimation problem. Because no specific distribution assumption is needed in our method, this method can be easily employed in practice without worrying about model mis-specification and struggling with maximization of complicated likelihoods. The estimation process contains two steps as we partition the parameters into two parts based on their relationship to the random effects. Parameters not related to random effects are estimated first, with the introduction of a linear transformation. In the second stage, these parameters are replaced by their estimates, and the remaining parameters can be estimated. Grouping of panels may be necessary if some panels do not have enough observation. The resulting estimators are proved

to be unbiased and asymptotically normal. Simulation studies comparing our estimator to some other existing estimators are conducted, which shows advantage of our methods under certain informative missingness setting. This proposed method is applied to a real data examples: the Health and Retirement Study.

Chapter 1

Introduction

1.1 Panel Data

Panel data are collected in a variety of studies. For example, cluster sampling in social studies and sample surveys results in panel data; repeated measurements in clinical trials and economic studies produce longitudinal data. Under panel data setting, observations are nested in their corresponding panels. Besides the observation level random error in measurements, there are usually panel level random effects that distinguish the panels.

A widely used modeling technique for panel data is *linear mixed-effect model* (Laird and Ware (1982)). Mixed-effect modeling takes into account both levels of randomness while calculating estimates for the fixed effects. Many standard statistical softwares include packages to fit linear mixed-effect model, such as package "*lme4*" in R and procedure"*Mixed*" in SAS.

1.2 Missing Data

Missing data is commonly encountered in the above data collecting processes. For example, some participants may refuse to answer income related questions in a survey and some patients may fail to show up in some visits due to poor health condition. There can also be times that the response collected is too good or too bad to be trusted. Instead of getting misleading result by analyzing this kind of records, we would rather treat them as missing.

Because the reasons for missingness vary in different studies, the way to handle the missing data should be adjusted accordingly to the missing mechanism. Data are said to be missing completely at random (MCAR) when the probability of missingness is independent of both the observed and unobserved data. In this case, standard analysis using only the portion of completely observed data will produce valid results. Missing at random (MAR) is a case where, conditional on the observed data, the probability of missingness does not depend on unobserved measurement. MCAR and MAR are both called "ignorable missingness", in which specifying the exact distribution of missingness given observed data is unnecessary in likelihood-based analysis. Likelihood-based analysis with ignorable missing data have been well explained in the book "Statistical Analysis with Missing Data" by Little and Rubin (2002). The most complicated situation arises when the missingness is not at random (MNAR), where the probability of missingness depends on both observed and unobserved responses. This mechanism of missingness is non-ignorable and the distribution of missingness must be explicitly specified in likelihood-based inferences.

The data resulting from missing observation are unbalanced panel data, that

is, panels have unequal number of measurements. For unbalanced panel data, R package *lme4* and SAS procedure *Mixed* can still be applied to fit a linear mixed-effect model, where the estimates are calculated by *restricted maximum likelihood* (**REML**) method. REML was first suggested by W. A. Thompson (1962), and was later formally described by Patterson and Thompson (1971). These tools are very useful in that panels with missing observations are still incorporated in the analysis, and we don't need to worry about loosing too much information and/or having a too small sample size. Under MCAR, one can easily argue that REML gives unbiased estimates. But it is not sure if the estimates are unbiased under MAR, and it is highly likely that the estimates will be biased under the non-ignorable missing mechanism.

1.3 Informative Missingness

When missingness is non-ignorable, the probability of missingness can depend on the unobserved responses either directly or indirectly through some parameters. When the missing probability is indirectly related to the missing observations through random effect(s), it was referred to as "informative" missingness by Wu and Carroll (1988) and Wu and Bailey (1988). Little (1995) named the same type of missingness as "nonignorable random-coefficient-based-missingness". This informative missingness was found to be a suitable assumption in many clinical and epidemiological studies (see, e.g., Albert and Follmann (2000), De Gruttola and Tu (1994), Follmann and Wu (1995), Pulkstenis et al. (1998), Ten Have et al. (1998), Wu and Bailey (1989) and Wu and Follmann (1999)). Moreover, in sample surveys with cluster sampling, the informative missingness is a natural assumption

when the nonresponse depends on a cluster-level random effect.

Wu and Carroll (1988) utilized likelihood-based approach to derive estimates with longitudinal data under informative censoring. Their method was completely parametric. In order to compute a closed form marginal joint likelihood, the distribution assumptions were extremely limited. Nevertheless, numerical methods, such as pseudo-maximize-likelihood estimation proposed by Gong and Samaniego (1981) and Bock and Petersen (1975) procedure, were still applied to find the maximum likelihood estimator. In addition, this parametric approach is very sensitive to distribution mis-specification. As an alternative, Wu and Bailey (1989) brought up the idea of conditional model. Follmann and Wu (1995) and Wu and Follmann (1999) further generalized this idea and proposed the *approximate conditional model* (ACM), which can be either parametric or semiparametric. The key step of this approach is to find an adequate ACM so that valid estimators for interested parameters can be computed. However, such an ACM may be difficult to find and some parameters may not be directly estimated by fitting the ACM (Albert and Follmann (2000)). A grouping approach was further proposed to modify the ACM by Park et al. (2002) and Xu and Shao (2009), where data were grouped according to a summary statistics $\boldsymbol{S}$. More discussions about the choices of such a sufficient statistic $\boldsymbol{S}$ can be found in their papers. A limitation of grouping method in ACM is that the summary statistic $\boldsymbol{S}$ has to be discrete with a small number of groups, which is often not the case.

1.4 Overview

Though the informative missing problem has been studied in the literature and some parametric or semi-parametric methods have been proposed, they are all likelihood-based where limitations naturally exist due to distribution assumptions. The purpose of my thesis is to investigate this estimation problem with informative missingness in a semi-parametric approach where we are not making any distribution assumptions. We assume the response model to be a ***linear mixed-effect model***, due to its popularity in analyzing panel data.

Because the impact of missing data depends on its type, separate modeling techniques are necessary for different types. Two common types of missing data are: *intermittent* missingness and *monotone* missingness. There has been quite some literature studying monotone missingness, which widely exists when censoring is inevitable. But in my thesis, we are focusing on ***intermittent*** missingness, which is a more general phenomenon in survey study and is also a common case in clinical trial, when participants miss certain visits but do not withdraw form the trial. Our method can be employed in monotone missingness settings as well.

Informative missingness is non-ignorable, which requires us to take into the account of missingness mechanism when estimating the parameters in the response model. Two general approaches have been proposed to jointly model the response and missingness process. They are *selection model* and *pattern mixture model*. Selection model approach assumes a model for the response which does not involve the missing process and makes the missingness depend on the response. Such a model can be naturally fitted using the method of maximum likelihood as in Schluchter (1992) and Ten Have et al. (1998). On the other hand, pattern mixture

model approach sets up a model for missingness first and makes the distribution of response depending on the missing pattern. This modeling approach does not sound natural, but it has the advantage that exact specification of missing model is not necessary and estimation within a pattern can be much simpler. This modeling approach was adopted in Mori et al. (1992) and Park and Lee (1997). More discussions about selection and pattern mixture models can be found in Little (1993) and Little (1995). Our approach in this thesis is the ***selection*** model approach, as a result of its naturalness. The result from selection model is also easier to understand and interpret.

The detail of our model and the basic assumptions are discussed in Chapter 2. We first consider studies where all variables are time-varying. We further divide the fixed time-varying parameters into two parts based on whether they are related to random effects. Estimators will be proposed part-by-part in Chapter 3 and Chapter 4. In Chapter 3, the parameters $\boldsymbol{\beta}$ we are looking at are not confounded by random effects. The novel idea we propose to estimate parameters $\boldsymbol{\beta}$ is a linear transformation $\boldsymbol{Q}$. After applying this transformation, the parameters $\boldsymbol{\alpha}$ related to random effect are all gone and the parameters in $\boldsymbol{\beta}$ are left in the model which can be estimated as if the missingness is ignorable. Similar transformation has been used in panel data with measurement error problems (Xiao et al. (2007), Xiao et al. (2010) and Shao et al. (2011)). In Chapter 4, our task is to provide estimators for the part, $\boldsymbol{\alpha}$, which was gone with Q-transformation. Two cases are separately considered. In the simple case where all panels have enough observations, under mild assumptions, we can construct an estimator for $\boldsymbol{\alpha}$. However, in the general case where some panels may have too few observed responses, we have to utilize a grouping technique to construct an estimator. After constructing estimators for

the slope of time-varying covariates, we extend our response model to include time-invariant covariates in Chapter 5. This extended work will be very helpful in studies like clinical trials, where researchers may also be interested in some baseline covariates. The key to estimate time-invariant parameters is the creation of a matrix $\boldsymbol{U}$, based on which the model is reformed to look like the model we proposed earlier. Simulation studies are conducted in Chapter 6 and real data examples are provided in Chapter 7.

Chapter 2

Model and Assumptions

2.1 Response Model

We perform analysis based on a linear mixed-effect model, where the K-dimensional panel response vector $\boldsymbol{y}$ is linearly related to $\boldsymbol{X\theta}$ and an unobserved q-dimensional random coefficient vector $\boldsymbol{b}$. $\boldsymbol{X}$ is a $K \times p$ covariate matrix, where columns of $\boldsymbol{X}$ are panel vectors of covariates, and $\boldsymbol{\theta}$ is an unknown p-dimensional fixed coefficient vector. K and p are fixed constants satisfying $K \geq p$. The linear mixed-effect model for response $\boldsymbol{y}$ is:

$$\boldsymbol{y} = \boldsymbol{X\theta} + \boldsymbol{Zb} + \boldsymbol{e},$$

where $\boldsymbol{Z}$ is a $K \times q$ submatrix of $\boldsymbol{X}$ including the intercept column $\mathbf{1}$. By assuming K is fixed, it is required that the same number of K responses were supposed to be collected from each panel.

With a total of n panels, let $(\boldsymbol{y}_i, \boldsymbol{X}_i)$, $i = 1, \ldots, n$ be independent samples, where

$(\boldsymbol{y}_i, \boldsymbol{X}_i)$ is the value of $(\boldsymbol{y}, \boldsymbol{X})$ for panel i. Then, our response model is:

$$\boldsymbol{y}_i = \boldsymbol{X}_i\boldsymbol{\theta} + \boldsymbol{Z}_i\boldsymbol{b}_i + \boldsymbol{e}_i, \quad i = 1, \ldots, n. \tag{2.1}$$

The elements in $\boldsymbol{X}_i$ can be either random or deterministic. If random, $\boldsymbol{X}_i$'s are independent and identically distributed. Let $\boldsymbol{b}_i$ and $\boldsymbol{e}_i$ be the random effects and random errors, respectively. As assumed widely in mixed-effect model, we assume that $(\boldsymbol{b}_i, \boldsymbol{e}_i)$ are independent and identically distributed for all panels, with $E(\boldsymbol{b}_i) = 0$, $var(\boldsymbol{b}_i) = \boldsymbol{\Sigma}_b$, and $E(\boldsymbol{e}_i) = 0$, $var(\boldsymbol{e}_i) = \boldsymbol{\Sigma}_e$. $\boldsymbol{b}_i$ and $\boldsymbol{e}_i$ are also independent within each panel. For the fixed effect $\boldsymbol{\theta}$, it remains the same across panels and is our interest to estimate.

The covariates in $\boldsymbol{X}_i$ can be partitioned into three parts as $\boldsymbol{X}_i = (\boldsymbol{Z}_i, \boldsymbol{T}_i, \mathbf{1} * \boldsymbol{u}_i')$, where:

(1) $\boldsymbol{Z}_i$ is a $K_i \times q$ matrix that contains the column of intercept and some time-variant covariates which have a random slope, such as age;

(2) $\boldsymbol{T}_i$ is a $K_i \times t$ matrix that contains all other time-variant covariates which do not have a random slope, such as weight;

(3) $\boldsymbol{u}_i$ is a s-dimensional ($s = p - q - t$) vector containing all time-invariant covariates, such as gender.

With this split, the model (2.1) can be reformed into

$$\boldsymbol{y}_i = \boldsymbol{Z}_i\boldsymbol{\alpha} + \boldsymbol{u}_i'\boldsymbol{\gamma}\mathbf{1} + \boldsymbol{T}_i\boldsymbol{\beta} + \boldsymbol{Z}_i\boldsymbol{b}_i + \boldsymbol{e}_i, \quad i = 1, \ldots, n. \tag{2.2}$$

Thus, the fixed parameter $\boldsymbol{\theta}$ is also split into three parts: $\boldsymbol{\alpha}$, $\boldsymbol{\beta}$ and $\boldsymbol{\gamma}$, and we are going to construct estimators for each part in the following three chapters.

Generally, in linear model, we require the covariates to be not linearly related as collinearity are problematic. This means that a covariate cannot be a linear combination of other covariates. Since $\boldsymbol{T}_i$ and $\boldsymbol{Z}_i$ both contain a subset of covariates and they are not overlapping based on the partition, they should not share any common covariate. Moreover, in the case of fixed covariate matrix, any covariate in $\boldsymbol{Z}_i$ should not be linear functions on covariates in $\boldsymbol{T}_i$ and vice versa. While in the case of random covariate matrix, we assume that the probability of $\boldsymbol{T}_i$ and $\boldsymbol{Z}_i$ being full rank is nonzero, and the probability for any column of $\boldsymbol{T}_i$ being linear combination of columns of $\boldsymbol{Z}_i$ or vice versa is less than 1. These assumptions help us to avoid issues with collinearity, which is not the focus in this work.

2.2 Missingness Model

We define a p-dimensional indicator vector $\boldsymbol{r}_i$, where the jth element $r_{ij} = 1$ if the jth response in $\boldsymbol{y}_i$ is observed and $r_{ij} = 0$ otherwise. The missingness we are focusing on here is informative missingness, where the missing probability depends on unobserved response $\boldsymbol{y}_{i,miss}$ indirectly through the unobserved random effects $\boldsymbol{b}_i$ as following:

$$p(\boldsymbol{r}_i|\boldsymbol{y}_i, \boldsymbol{X}_i) = p(\boldsymbol{r}_i|\boldsymbol{b}_i, \boldsymbol{X}_i). \tag{2.3}$$

With the i.i.d assumption on $\boldsymbol{X}_i$'s and $\boldsymbol{b}_i$'s, the missing indicator $\boldsymbol{r}_i$'s are also i.i.d. Let the number of observed responses in $\boldsymbol{y}_i$ be K_i. Naturally, K_i needs to be no less than one, since otherwise, that panel provides no information.

We define a $K \times K$ matrix $\boldsymbol{R}_i = diag(\boldsymbol{r}_i)$ for the purpose of calculation, whose

diagonal elements are $\boldsymbol{r}_i$ and off-diagonal elements are all zeros. If we multiply $\boldsymbol{X}_i$ on the left by $\boldsymbol{R}_i$, the rows corresponding to missing observations all become 0, and we have $\boldsymbol{X}_i'\boldsymbol{R}_i\boldsymbol{X}_i = \tilde{\boldsymbol{X}}_i{}'\tilde{\boldsymbol{X}}_i$, where $\tilde{\boldsymbol{X}}_i$ is the $K_i \times p$ observed submatrix of $\boldsymbol{X}_i$.

In the construction of the estimator for $\boldsymbol{\alpha}$, different cases are considered: (1) simple case, where all panels have enough observations and (2) general case, where some panels do not have enough observations. In the simple case, we need to assume $\boldsymbol{Z}_i'\boldsymbol{R}_i\boldsymbol{Z}_i = \tilde{\boldsymbol{Z}}_i{}'\tilde{\boldsymbol{Z}}_i$ is non-singular. In the general case, we need to utilize the grouping approach when informative missingness exists, so that we need some assumptions to guarantee the panels in the same group share some similarities. The assumption is that $p(\boldsymbol{r}_i|\boldsymbol{y}_i, \boldsymbol{X}_i)$ does not depend on $\boldsymbol{u}_i$ and $\boldsymbol{Z}_i$ and $E(\boldsymbol{b}_i|\boldsymbol{r}_i) = E(\boldsymbol{b}_i|k_i)$.

2.3 Summary of Assumptions

Finally, let's close this chapter with a summary of important assumptions we are making. The first four assumptions are always required. All the four assumptions are on the response model while the fourth assumption is also about missingness indicator $\boldsymbol{r}_i$.

(1) K is a fixed constant for all panels;

(2) Covariates in $\boldsymbol{Z}_i$ are not linear combinations of covariates in $\boldsymbol{T}_i$ and vice versa;

(3) $\boldsymbol{b}_i$ and $\boldsymbol{e}_i$ are independent within each panel and are independent of covariates;

(4) $(\boldsymbol{r}_i, \boldsymbol{b}_i, \boldsymbol{e}_i)$ are independent and identically distributed for all panels, with $E(\boldsymbol{b}_i) = 0$, $var(\boldsymbol{b}_i) = \boldsymbol{\Sigma}_b$, and $E(\boldsymbol{e}_i) = 0$, $var(\boldsymbol{e}_i) = \boldsymbol{\Sigma}_e$;

The last two assumptions are only needed in general case, where grouping idea is borrowed. These two assumptions are only necessary under informative missing mechanism. If missing mechanism if MAR, they are not necessary. The assumptions are:

(5) Within each panel, $P(r_{ij} = 1|\boldsymbol{b}_i, \boldsymbol{X}_i)$ does not depend on covariates in $\boldsymbol{u}_i$ and $\boldsymbol{Z}_i$;

(6) The conditional expectation $E(\boldsymbol{b}_i|\boldsymbol{r}_i) = E(\boldsymbol{b}_i|k_i)$.

Chapter 3

Estimation of Parameters β with A Linear Transformation

In this chapter and the next chapter, we consider response models without time-invariant covariate $\boldsymbol{u}_i$. That is,

$$\begin{aligned} \boldsymbol{y}_i &= \boldsymbol{Z}_i\boldsymbol{\alpha} + \boldsymbol{T}_i\boldsymbol{\beta} + \boldsymbol{Z}_i\boldsymbol{b}_i + \boldsymbol{e}_i, \\ &= \boldsymbol{Z}_i(\boldsymbol{\alpha} + \boldsymbol{b}_i) + \boldsymbol{T}_i\boldsymbol{\beta} + \boldsymbol{e}_i. \end{aligned} \tag{3.1}$$

Based on our previous literature review, many literature methods are likelihood-based, which usually try to estimate all parameters together by assuming some distributions and maximizing the joint likelihood. The maximizing process could be very complicated.

However, from the formulation above, we can see that parameters in $\boldsymbol{\beta}$ are not confounded by random effects, while parameters in $\boldsymbol{\alpha}$ do. This lead us to the thought that the estimation for $\boldsymbol{\beta}$ should have an easier solution. Based on model

(3.2), if we can get the random effects removed but keep the fixed effects β, the complexity for estimating β will be greatly reduced. With this motivation, we constructed a linear transformation, $\boldsymbol{Q}_i$, which can remove the random effects by eliminating $\boldsymbol{Z}_i$. After this step, the only parameters left in the model are those in β. With this linearly transformed model, we are able to construct an estimator $\hat{\beta}$ for parameters β, which carries a very simple form.

This transformation and the estimator for β will be presented in this chapter. The properties of the estimator will be discussed following its construction.

3.1 A Linear Transformation

The idea we propose to remove the random parts is to apply a linear transformation, $\boldsymbol{Q}_i$. In Chapter 2, we defined matrix $\boldsymbol{R}_i$ as :

$$\boldsymbol{R}_i = diag(\boldsymbol{r}_i), \;\; i = 1, \ldots, n.$$

Let $(\boldsymbol{Z}_i'\boldsymbol{R}_i\boldsymbol{Z}_i)^-$ be the generalized inverse of $\boldsymbol{Z}_i'\boldsymbol{R}_i\boldsymbol{Z}_i$, we define the linear transformation $\boldsymbol{Q}_i$ as

$$\boldsymbol{Q}_i = \boldsymbol{R}_i - \boldsymbol{R}_i\boldsymbol{Z}_i(\boldsymbol{Z}_i'\boldsymbol{R}_i\boldsymbol{Z}_i)^-\boldsymbol{Z}_i'\boldsymbol{R}_i, \;\; i = 1, \ldots, n. \tag{3.2}$$

With simple calculation, we can see that

$$\begin{aligned} \boldsymbol{Q}_i\boldsymbol{Z}_i &= (\boldsymbol{R}_i - \boldsymbol{R}_i\boldsymbol{Z}_i(\boldsymbol{Z}_i'\boldsymbol{R}_i\boldsymbol{Z}_i)^-\boldsymbol{Z}_i'\boldsymbol{R}_i)\boldsymbol{Z}_i \\ &= \boldsymbol{R}_i\boldsymbol{Z}_i(\boldsymbol{I} - (\boldsymbol{Z}_i'\boldsymbol{R}_i\boldsymbol{Z}_i)^-\boldsymbol{Z}_i'\boldsymbol{R}_i\boldsymbol{Z}_i) \\ &= \boldsymbol{0}. \end{aligned}$$

So, after applying $\boldsymbol{Q}_i$ to $\boldsymbol{Z}_i$, all columns of $\boldsymbol{Z}_i$ including intercept column become $\mathbf{0}$.

Apply this linear transformation to model (3.2), we get:

$$\boldsymbol{Q}_i\boldsymbol{y}_i = \boldsymbol{Q}_i\boldsymbol{T}_i\boldsymbol{\beta} + \boldsymbol{Q}_i\boldsymbol{e}_i, \;\; i = 1, \ldots, n. \tag{3.3}$$

The transformed model becomes a linear fixed-effect model, where the response $\boldsymbol{Q}_i\boldsymbol{y}_i$ is no longer affected by random effects. At the same time, the missingness (if any) in this transformed model becomes ignorable, in that the missing only depends on observed covariates. This step greatly reduced the complexity in estimating $\boldsymbol{\beta}$, because the trouble maker, random effects, are gone. The rows in $\boldsymbol{Q}_i$ corresponding to the missing responses are all zeros, so that there is no calculation issue.

3.2 Estimator for $\boldsymbol{\beta}$

For each panel i, we now have $\boldsymbol{Q}_i\boldsymbol{y}_i = \boldsymbol{Q}_i\boldsymbol{T}_i\boldsymbol{\beta} + \boldsymbol{Q}_i\boldsymbol{e}_i$. Applying $\boldsymbol{T}_i'$ on both sides, we have

$$\boldsymbol{T}_i'\boldsymbol{Q}_i\boldsymbol{y}_i = \boldsymbol{T}_i'\boldsymbol{Q}_i\boldsymbol{T}_i\boldsymbol{\beta} + \boldsymbol{T}_i'\boldsymbol{Q}_i\boldsymbol{e}_i.$$

Then, summing up across all panels on both sides of this equation, we get

$$\sum_{i=1}^{n}\boldsymbol{T}_i'\boldsymbol{Q}_i\boldsymbol{y}_i = \sum_{i=1}^{n}\boldsymbol{T}_i'\boldsymbol{Q}_i\boldsymbol{T}_i\boldsymbol{\beta} + \sum_{i=1}^{n}\boldsymbol{T}_i'\boldsymbol{Q}_i\boldsymbol{e}_i.$$

Let $\boldsymbol{S}_n = \sum_{i=1}^{n} \boldsymbol{T}_i' \boldsymbol{Q}_i \boldsymbol{T}_i$. If $\boldsymbol{S}_n$ is nonsingular, we construct the following estimator for $\boldsymbol{\beta}$:

$$\hat{\boldsymbol{\beta}} = \boldsymbol{S}_n^{-1} \sum_{i=1}^{n} \boldsymbol{T}_i' \boldsymbol{Q}_i \boldsymbol{y}_i = \boldsymbol{\beta} + \boldsymbol{S}_n^{-1} \sum_{i=1}^{n} \boldsymbol{T}_i' \boldsymbol{Q}_i \boldsymbol{e}_i. \tag{3.4}$$

Unbiasedness

Theorem 3.1. *(Unbiasedness of $\hat{\boldsymbol{\beta}}$) The estimator $\hat{\boldsymbol{\beta}}$ is unbiased.*

Proof. The difference between our estimator and the parameters to be estimated is a linear function of random errors. It is easy to show that the expectation is $\mathbf{0}$.

$$\begin{aligned} E(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}) &= E(\boldsymbol{S}_n^{-1} \sum_{i=1}^{n} \boldsymbol{T}_i' \boldsymbol{Q}_i \boldsymbol{e}_i) \\ &= E_{\boldsymbol{T},\boldsymbol{R}}\{\boldsymbol{S}_n^{-1} \sum_{i=1}^{n} \boldsymbol{T}_i' \boldsymbol{Q}_i E(\boldsymbol{e}_i | \boldsymbol{T}_i, \boldsymbol{R}_i)\} \\ &= E_{\boldsymbol{T},\boldsymbol{R}}\{\boldsymbol{S}_n^{-1} \sum_{i=1}^{n} \boldsymbol{T}_i' \boldsymbol{Q}_i E(\boldsymbol{e}_i)\} \\ &= \mathbf{0}. \end{aligned}$$

The third equality follows from that $\boldsymbol{e}_i$ is independent of $\boldsymbol{X}_i$ and $\boldsymbol{R}_i$, and the last equality follows from $E(\boldsymbol{e}_i) = 0$. This shows that our estimator $\hat{\boldsymbol{\beta}}$ is an unbiased estimator of $\boldsymbol{\beta}$.

This argument will go through for all linear functions of random errors only, i.e. with no random effects. In the rest of this thesis, we know the expectation of any linear function of random errors only is zero, without calculation. □

Asymptotic Property

Theorem 3.2. *(Asymptotic normality of $\hat{\beta}$) Under model* (2.2) *with informative missingness, assume that, as $n \to \infty$,*

$$\frac{\boldsymbol{S}_n}{n} = \frac{1}{n}\sum_{i=1}^{n} \boldsymbol{T}_i' \boldsymbol{Q}_i \boldsymbol{T}_i \to_p \boldsymbol{M}, \tag{3.5}$$

$$\frac{1}{n}\sum_{i=1}^{n} \boldsymbol{T}_i' \boldsymbol{Q}_i \boldsymbol{\Sigma}_e \boldsymbol{Q}_i \boldsymbol{T}_i \to_p \boldsymbol{L}. \tag{3.6}$$

According to the central limit theorem,

$$\sqrt{n}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}) \to N(\mathbf{0}, \boldsymbol{\Sigma}_\beta), \tag{3.7}$$

where $\boldsymbol{\Sigma}_\beta = \boldsymbol{M}^{-1}\boldsymbol{L}\boldsymbol{M}^{-1}$.

Proof. With our model and definition of $\boldsymbol{M}$, it follows that

$$\sqrt{n}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}) = \frac{1}{\sqrt{n}}\sum_{i=1}^{n} \boldsymbol{M}^{-1}\boldsymbol{T}_i'\boldsymbol{Q}_i\boldsymbol{e}_i + o_p(1). \tag{3.8}$$

Under the informative missingness assumption, conditioned on $\{\boldsymbol{R}_i, \boldsymbol{T}_i, i = 1, \ldots, n\}$, the right hand side of (3.8) is an average of independent random variables. The asymptotic normality follows directly from the central limit theorem.

The asymptotic covariance matrix $\boldsymbol{\Sigma}_\beta$ can be derived from a direct calculation of the average term.

$$
\begin{aligned}
& cov(\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\boldsymbol{M}^{-1}\boldsymbol{T}_i'\boldsymbol{Q}_i\boldsymbol{e}_i|\boldsymbol{R}_i,\boldsymbol{X}_i,i=1,\ldots,n) \\
= \quad & \frac{1}{n}\sum_{i=1}^{n}cov(\boldsymbol{M}^{-1}\boldsymbol{T}_i'\boldsymbol{Q}_i\boldsymbol{e}_i|\boldsymbol{R}_i,\boldsymbol{X}_i) \\
= \quad & \frac{1}{n}\sum_{i=1}^{n}\boldsymbol{M}^{-1}\boldsymbol{T}_i'\boldsymbol{Q}_i var(\boldsymbol{e}_i|\boldsymbol{R}_i,\boldsymbol{X}_i)\boldsymbol{Q}_i\boldsymbol{T}_i\boldsymbol{M}^{-1}\} \\
= \quad & \frac{1}{n}\sum_{i=1}^{n}E\{\boldsymbol{M}^{-1}\boldsymbol{T}_i'\boldsymbol{Q}_i\boldsymbol{\Sigma}_e\boldsymbol{Q}_i\boldsymbol{T}_i\boldsymbol{M}^{-1}\} \\
\rightarrow_p \quad & \boldsymbol{M}^{-1}\boldsymbol{L}\boldsymbol{M}^{-1}
\end{aligned}
$$

□

Consistent Estimator

In estimation problems, two characteristics are important to judge whether the estimates are good or not. One is bias and the other is variation. In Theorem 3.1, we showed that our estimator is unbiased. What about the variation? Can we have an estimate of the variance after we estimated the parameters?

From (3.4), we know that the variance of $\hat{\boldsymbol{\beta}}$ is the same as the variance of $\boldsymbol{S}_n^{-1}\sum_{i=1}^{n}\boldsymbol{T}_i'\boldsymbol{Q}_i\boldsymbol{e}_i$. In Theorem 3.2, we calculated the conditional limit of the variance of $\sqrt{n}\boldsymbol{S}_n^{-1}\sum_{i=1}^{n}\boldsymbol{T}_i'\boldsymbol{Q}_i\boldsymbol{e}_i$, and the limit is $\boldsymbol{\Sigma}_\beta = \boldsymbol{M}^{-1}\boldsymbol{L}\boldsymbol{M}^{-1}$. If we can construct a consistent estimator for this asymptotic covariance matrix $\boldsymbol{\Sigma}_\beta$, we can definitely use it as our estimate of the estimation variance.

To construct a consistent estimator for $\boldsymbol{\Sigma}_\beta$, we only need to find consistent estimators for $\boldsymbol{M}$ and $\boldsymbol{L}$. Based on the definition of $\boldsymbol{M}$, an estimator $\hat{\boldsymbol{M}}$ for $\boldsymbol{M}$

follows naturally.

$$\hat{\boldsymbol{M}} = \frac{1}{n}\sum_{i=1}^{n} \boldsymbol{T}_i'\boldsymbol{Q}_i\boldsymbol{T}_i = \boldsymbol{M} + o_p(1). \tag{3.9}$$

For matrix $\boldsymbol{L}$, we just need to estimate the random error covariance matrix $\boldsymbol{\Sigma}_e$. A good estimate can be constructed with squares of residuals. The consistent estimator $\hat{\boldsymbol{L}}$ for $\boldsymbol{L}$ we constructed is:

$$\begin{aligned}
\hat{\boldsymbol{L}} &= \frac{1}{n}\sum_{i=1}^{n} \boldsymbol{T}_i'\boldsymbol{Q}_i(\boldsymbol{y}_i - \boldsymbol{T}_i\hat{\boldsymbol{\beta}})(\boldsymbol{y}_i - \boldsymbol{T}_i\hat{\boldsymbol{\beta}})'\boldsymbol{Q}_i\boldsymbol{T}_i \qquad (3.10)\\
&= \frac{1}{n}\sum_{i=1}^{n} \boldsymbol{T}_i'\boldsymbol{Q}_i(\boldsymbol{Z}_i\boldsymbol{\alpha} + \boldsymbol{Z}_i\boldsymbol{b}_i + \boldsymbol{e}_i)(\boldsymbol{Z}_i\boldsymbol{\alpha} + \boldsymbol{Z}_i\boldsymbol{b}_i + \boldsymbol{e}_i)'\boldsymbol{Q}_i\boldsymbol{T}_i + o_p(1)\\
&= \frac{1}{n}\sum_{i=1}^{n} \boldsymbol{T}_i'\boldsymbol{Q}_i\boldsymbol{e}_i\boldsymbol{e}_i'\boldsymbol{Q}_i\boldsymbol{T}_i + o_p(1)\\
&= L + o_p(1).
\end{aligned}$$

Now a consistent estimator for $\boldsymbol{\Sigma}_\beta$ can be defined as $\hat{\boldsymbol{M}}^{-1}\hat{\boldsymbol{L}}\hat{\boldsymbol{M}}^{-1}$.

3.3 Discussion of Conditions

In the construction of estimator $\hat{\boldsymbol{\beta}}$, we need $\boldsymbol{S}_n$ to be invertible. Now, let's discuss the condition when this matrix is nonsingular.

(a) When $\boldsymbol{Z}_i = \boldsymbol{1}$ and $\boldsymbol{T}_i$ contains only 1 column.

Let $\boldsymbol{T}_i = (t_1, t_2, \ldots, t_K)'$. $\boldsymbol{Q}_i = \boldsymbol{R}_i - \frac{1}{K_i}\boldsymbol{r}_i\boldsymbol{r}_i'$. Then,

$$\begin{aligned}
\boldsymbol{T}_i'\boldsymbol{Q}_i\boldsymbol{T}_i &= (t_1, t_2, \ldots, t_K)\boldsymbol{R}_i(t_1, t_2, \ldots, t_K)' - \frac{1}{K_i}(t_1, t_2, \ldots, t_K)\boldsymbol{r}_i\boldsymbol{r}_i'(t_1, t_2, \ldots, t_K)'\\
&= \sum_{j\in\mathcal{O}} t_j^2 - \frac{1}{K_i}(\sum_{j\in\mathcal{O}} t_j)^2
\end{aligned}$$

where $\mathcal{O}$ is the set of observed entries. So, $\frac{1}{(K_i-1)}\boldsymbol{T}_i'\boldsymbol{Q}_i\boldsymbol{T}_i$ (when $K_i \geq 2$) is the sample variance based on the observed entries in $\boldsymbol{T}_i$. It only takes 0 when that panel has just one observation. But in practice, usually, there are panels with more than one observed responses. When $\boldsymbol{T}_i$ is fixed, we required that $\mathcal{L}(\boldsymbol{Z}_i) \cap \mathcal{L}(\boldsymbol{T}_i) = \emptyset$, which means $\boldsymbol{T}_i \neq a\mathbf{1}$ for any $a \in R$. There usually exists some i with $K_i > 1$, which implies $\boldsymbol{T}_i'\boldsymbol{Q}_i\boldsymbol{T}_i > 0$ for this i. So, the sum $\boldsymbol{S}_n = \sum_{i=1}^n \boldsymbol{T}_i'\boldsymbol{Q}_i\boldsymbol{T}_i > 0$. When $\boldsymbol{T}_i$ is a random vector, we required that $P(\boldsymbol{T}_i = a\mathbf{1}) \neq 1$. Then $E(\boldsymbol{T}_i'\boldsymbol{Q}_i\boldsymbol{T}_i) > 0$, and $E(\boldsymbol{S}_n) > 0$.

(b) When $\boldsymbol{Z}_i = \mathbf{1}$, but $\boldsymbol{T}_i$ has two or more columns.

Denote $\boldsymbol{T}_i$ as $\boldsymbol{T}_i = (\boldsymbol{T}_{i1}, \ldots, \boldsymbol{T}_{it})$, where $t = p - q - s$.

When $\boldsymbol{T}_i$ is fixed, we have

$$\boldsymbol{T}_i'\boldsymbol{Q}_i\boldsymbol{T}_i = \begin{bmatrix} \boldsymbol{T}_{i1}'\boldsymbol{Q}_i\boldsymbol{T}_{i1} & \boldsymbol{T}_{i1}'\boldsymbol{Q}_i\boldsymbol{T}_{i2} & \cdots & \boldsymbol{T}_{i1}'\boldsymbol{Q}_i\boldsymbol{T}_{it} \\ \boldsymbol{T}_{i2}'\boldsymbol{Q}_i\boldsymbol{T}_{i1} & \boldsymbol{T}_{i2}'\boldsymbol{Q}_i\boldsymbol{T}_{i2} & \cdots & \boldsymbol{T}_{i2}'\boldsymbol{Q}_i\boldsymbol{T}_{it} \\ \cdots & \cdots & \cdots & \cdots \\ \boldsymbol{T}_{it}'\boldsymbol{Q}_i\boldsymbol{T}_{i1} & \boldsymbol{T}_{it}'\boldsymbol{Q}_i\boldsymbol{T}_{i2} & \cdots & \boldsymbol{T}_{it}'\boldsymbol{Q}_i\boldsymbol{T}_{it}, \end{bmatrix}$$

which is the *covariance matrix*(up to a constant factor) based on the observed submatrix $\tilde{\boldsymbol{T}}_i$, a $K_i \times t$ matrix corresponding to the observed entries of subject i. We can conclude that, all $\boldsymbol{T}_i'\boldsymbol{Q}_i\boldsymbol{T}_i$ are positive semi-definite. In practice, there usually exist at least one panel i such that, $K_i \geq t$ and $\boldsymbol{T}_i$ is of full rank, which implies that $\boldsymbol{T}_i'\boldsymbol{Q}_i\boldsymbol{T}_i$ is positive definite. Therefore, the sum $\boldsymbol{S}_n = \sum_{i=1}^n \boldsymbol{T}_i'\boldsymbol{Q}_i\boldsymbol{T}_i$ is for sure to be positive definite.

When $\boldsymbol{T}_i$ is random, let's look at $E(\boldsymbol{S}_n)$ instead. If we further assume that $\pi_i = P(r_{ij} = 1|\boldsymbol{T}_i), \;\; j = 1, \ldots, K$ is constant for a given i, $E(\boldsymbol{Q}_i|\boldsymbol{T}_i)$ can be

calculated as follows.

From $\sum_{j=1}^{K} r_{ij} = K_i$, we have

$$1 = E\left(K_i^{-1} \sum_{j=1}^{K} r_{ij} \middle| \boldsymbol{T}_i \right) = \sum_{j=1}^{K} E(K_i^{-1} r_{ij} | \boldsymbol{T}_i) = K E(K_i^{-1} r_{i1} | \boldsymbol{T}_i),$$

and thus $E(K_i^{-1} r_{ij} | \boldsymbol{T}_i) = K^{-1}$. From

$$K_i^2 = \left(\sum_{j=1}^{K} r_{ij} \right)^2 = \sum_{j=1}^{K} r_{ij}^2 + \sum_{l \neq j} r_{ij} r_{il} = K_i + \sum_{l \neq j} r_{ij} r_{il},$$

we have

$$K_i = 1 + K_i^{-1} \sum_{l \neq j} r_{ij} r_{il},$$

then

$$K\pi_i = E(K_i | \boldsymbol{T}_i) = 1 + K(K-1) E(K_i^{-1} r_{ij} r_{il} | \boldsymbol{T}_i),\ j \neq l,$$

and thus

$$E(K_i^{-1} r_{ij} r_{il} | \boldsymbol{T}_i) = (\pi_i - K^{-1})(K-1),\ j \neq l.$$

From the definition, $\boldsymbol{Q}_i = \boldsymbol{R}_i - \frac{1}{K_i} \boldsymbol{r}_i \boldsymbol{r}_i'$, the diagonal elements of $\boldsymbol{Q}_i$ are $r_{ij} - \frac{1}{K_i} r_{ij}$, $j = 1, \ldots, K$, and the off-diagonal elements are $-\frac{1}{K_i} r_{ij} r_{il}, j \neq l$. Then

$$E(\boldsymbol{Q}_i | \boldsymbol{T}_i) = (\pi_i - K^{-1}) \begin{pmatrix} 1 & -(K-1)^{-1} & \cdots & -(K-1)^{-1} \\ -(K-1)^{-1} & 1 & \cdots & -(K-1)^{-1} \\ \cdots & \cdots & \cdots & \cdots \\ -(K-1)^{-1} & -(K-1)^{-1} & \cdots & 1 \end{pmatrix},$$

whose eigenvalues are 0 and $\frac{K\pi_i - 1}{(K-1)K}$. The only eigenvector for 0 is vector $\mathbf{1}$. Since we required that $P(T_{i,j} = a\mathbf{1}) \neq 1, a \in R$ and $P(\boldsymbol{a}'\boldsymbol{T}_i = 0) \neq 1, \boldsymbol{a} \in R^t$, with positive probability, columns of $\boldsymbol{T}_i$ are linearly independent and none of them is $a\mathbf{1}$. So, $E(\boldsymbol{T}_i'\boldsymbol{Q}_i\boldsymbol{T}_i) = E\{\boldsymbol{T}_i' E(\boldsymbol{Q}_i|\boldsymbol{T}_i)\boldsymbol{T}_i\}$ is positive definite, and thus, $E(\boldsymbol{S}_n)$ is positive definite.

(c) When $\boldsymbol{Z}_i$ has two or more columns.

The form of $\boldsymbol{T}_i'\boldsymbol{Q}_i\boldsymbol{T}_i$ or its expected value cannot be calculated explicitly. But as we are taking the sum across all panels, the sum is likely to be invertible. If they are random, the probability of columns of $\boldsymbol{T}_i$ being linear combination of columns of $\boldsymbol{Z}_i$ is less than 1, and the probability $\boldsymbol{T}_i$ being full rank is nonzero. Under mild condition, $E(\boldsymbol{S}_n)$ would be non-singular.

Chapter 4

Estimation of Parameters α Confounded by Random Effects

In the previous chapter, we proposed a linear transformation $\boldsymbol{Q}_i$ to the response model and removed the terms that are related to $\boldsymbol{\alpha}$ and $\boldsymbol{\gamma}$. In this chapter, we will introduce methods to construct estimator for $\boldsymbol{\alpha}$ with the same model as in Chapter 3,

$$\boldsymbol{y}_i = \boldsymbol{Z}_i\boldsymbol{\alpha} + \boldsymbol{T}_i\boldsymbol{\beta} + \boldsymbol{Z}_i\boldsymbol{b}_i + \boldsymbol{e}_i.$$

The estimator we constructed for $\boldsymbol{\beta}$ in Chapter 3 is

$$\hat{\boldsymbol{\beta}} = \boldsymbol{S}_n^{-1}\sum_{i=1}^{n}\boldsymbol{T}_i'\boldsymbol{Q}_i\boldsymbol{y}_i = \boldsymbol{\beta} + \boldsymbol{S}_n^{-1}\sum_{i=1}^{n}\boldsymbol{T}_i'\boldsymbol{Q}_i\boldsymbol{e}_i.$$

The difference between $\hat{\boldsymbol{\beta}}$ and $\boldsymbol{\beta}$ is $\boldsymbol{S}_n^{-1}\sum_{i=1}^{n}\boldsymbol{T}_i'\boldsymbol{Q}_i\boldsymbol{e}_i$, which is a linear function of random error $\boldsymbol{e}_i$'s. From the proof of unbiasedness of $\hat{\boldsymbol{\beta}}$, we know this linear sum has mean 0. If we subtract $\boldsymbol{T}_i\hat{\beta}$ on both sides of our response model above, we can

get:

$$\begin{aligned} \boldsymbol{y}_i - \boldsymbol{T}_i\hat{\boldsymbol{\beta}} &= \boldsymbol{Z}_i\boldsymbol{\alpha} + \boldsymbol{Z}_i\boldsymbol{b}_i + \boldsymbol{T}_i(\boldsymbol{\beta} - \hat{\boldsymbol{\beta}}) + \boldsymbol{e}_i \\ &= \boldsymbol{Z}_i\boldsymbol{\alpha} + \boldsymbol{Z}_i\boldsymbol{b}_i + (-\boldsymbol{S}_n^{-1}\sum_{i=1}^{n}\boldsymbol{T}_i'\boldsymbol{Q}_i\boldsymbol{e}_i + \boldsymbol{e}_i). \end{aligned} \tag{4.1}$$

Now, treating $\boldsymbol{\eta}_i = \boldsymbol{y}_i - \boldsymbol{T}_i\hat{\boldsymbol{\beta}}$ as the new response vector and $\boldsymbol{\epsilon}_i = -\boldsymbol{S}_n^{-1}\sum_{i=1}^{n}\boldsymbol{T}_i'\boldsymbol{Q}_i\boldsymbol{e}_i + \boldsymbol{e}_i$ as the new random error vector, the model has the following simple form:

$$\boldsymbol{\eta}_i = \boldsymbol{Z}_i(\boldsymbol{\alpha} + \boldsymbol{b}_i) + \boldsymbol{\epsilon}. \tag{4.2}$$

When all panels $i = 1, \ldots, n$ have enough observations, i.e. $K_i \geq q$, we can calculate an ordinary least square estimates separately for each panel. Otherwise, if some K_i is less than q, we need to utilize the information of these panels in a different way.

In the current chapter, we first consider this simple case when all panels have enough observed responses. Then, we generalize it to cases where some panels have $K_i < q$ observations.

4.1 Simple Case

In this section, we consider the estimator of $\boldsymbol{\alpha}$ under the simple case, where all panels have at least q observations.

The Estimator

In the simple case, all panels have $K_i \geq q$ observations. Under model (3.2), all covariates in $\boldsymbol{Z}_i$ are time-varying except the intercept term. Therefore, the observed submatrix $\tilde{\boldsymbol{Z}}_i$ usually is of full rank q. The naive OLS estimate for $\boldsymbol{\alpha}$ in panel i would be

$$\hat{\boldsymbol{\alpha}}_i = (\boldsymbol{Z}_i'\boldsymbol{R}_i\boldsymbol{Z}_i)^{-1}\boldsymbol{Z}_i'\boldsymbol{R}_i(\boldsymbol{y}_i - \boldsymbol{T}_i\hat{\boldsymbol{\beta}}).$$

In fact, this estimator is estimating $\boldsymbol{\alpha} + \boldsymbol{b}_i$ for panel i. So, it is not surprising that $\hat{\boldsymbol{\alpha}}_i$'s are different across panels. One commonly used method to deal with this issue is to calculate the average and use the average as the final estimator. By taking the average, we get

$$\begin{aligned}
\tilde{\boldsymbol{\alpha}} &= \frac{1}{n}\sum_{i=1}^{n}(\boldsymbol{Z}_i'\boldsymbol{R}_i\boldsymbol{Z}_i)^{-1}\boldsymbol{Z}_i'\boldsymbol{R}_i(\boldsymbol{y}_i - \boldsymbol{T}_i\hat{\boldsymbol{\beta}}) \\
&= \boldsymbol{\alpha} + \frac{1}{n}\sum_{i=1}^{n}\boldsymbol{b}_i + \frac{1}{n}\sum_{i=1}^{n}(\boldsymbol{Z}_i'\boldsymbol{R}_i\boldsymbol{Z}_i)^{-1}\boldsymbol{Z}_i'\boldsymbol{R}_i\boldsymbol{e}_i \\
&\quad -\frac{1}{n}\sum_{i=1}^{n}(\boldsymbol{Z}_i'\boldsymbol{R}_i\boldsymbol{Z}_i)^{-1}\boldsymbol{Z}_i'\boldsymbol{R}_i\boldsymbol{T}_i\boldsymbol{S}_n^{-1}(\sum_{j=1}^{n}\boldsymbol{T}_j'\boldsymbol{Q}_j\boldsymbol{e}_j) \qquad (4.3)
\end{aligned}$$

Properties of $\tilde{\boldsymbol{\alpha}}$

The following theorems state the properties of our estimator $\tilde{\boldsymbol{\alpha}}$. The proofs are provided in section 4.5.

Theorem 4.1. *The estimator $\tilde{\boldsymbol{\alpha}}$ is unbiased.*

Theorem 4.2. *Under model* (3.2) *and informative missing, assume that, as $n \to \infty$,* (3.5) *and* (3.6) *hold, and*

$$\frac{1}{n}\sum_{i=1}^{n}\boldsymbol{T}_i'\boldsymbol{R}_i\boldsymbol{Z}_i(\boldsymbol{Z}_i'\boldsymbol{R}_i\boldsymbol{Z}_i)^{-1} \to_p \boldsymbol{H}_n, \qquad (4.4)$$

$$\frac{1}{n}\sum_{i=1}^{n} \boldsymbol{T}_i'\boldsymbol{Q}_i\boldsymbol{\Sigma}_e\boldsymbol{R}_i\boldsymbol{Z}_i(\boldsymbol{Z}_i'\boldsymbol{R}_i\boldsymbol{Z}_i)^{-1} \to_p \boldsymbol{J}_n, \tag{4.5}$$

$$\frac{1}{n}\sum_{i=1}^{n}(\boldsymbol{Z}_i'\boldsymbol{R}_i\boldsymbol{Z}_i)^{-1}\boldsymbol{Z}_i'\boldsymbol{R}_i\boldsymbol{\Sigma}_e\boldsymbol{R}_i\boldsymbol{Z}_i(\boldsymbol{Z}_i'\boldsymbol{R}_i\boldsymbol{Z}_i)^{-1} \to_p \boldsymbol{K}_n. \tag{4.6}$$

According to the central limit theorem,

$$\sqrt{n}(\tilde{\boldsymbol{\alpha}} - \boldsymbol{\alpha}) \to_p N(0, \boldsymbol{\Sigma}_{\tilde{\boldsymbol{\alpha}}}), \tag{4.7}$$

where

$$\boldsymbol{\Sigma}_{\tilde{\boldsymbol{\alpha}}} = \boldsymbol{\Sigma}_b + \boldsymbol{K}_n + \boldsymbol{H}_n'\boldsymbol{M}^{-1}\boldsymbol{L}\boldsymbol{M}^{-1}\boldsymbol{H}_n - \boldsymbol{J}_n'\boldsymbol{M}^{-1}\boldsymbol{H}_n - \boldsymbol{H}_n'\boldsymbol{M}^{-1}\boldsymbol{J}_n.$$

A consistent estimator for the covariance matrix $\boldsymbol{\Sigma}_{\tilde{\boldsymbol{\alpha}}}$ is

$$\widehat{\boldsymbol{\Sigma}_b + \boldsymbol{K}}_n + \hat{\boldsymbol{H}}_n'\hat{\boldsymbol{M}}^{-1}\hat{\boldsymbol{L}}\hat{\boldsymbol{M}}^{-1}\hat{\boldsymbol{H}}_n - \hat{\boldsymbol{H}}_n'\hat{\boldsymbol{M}}^{-1}\hat{\boldsymbol{J}}_n - \hat{\boldsymbol{J}}_n'\hat{\boldsymbol{M}}^{-1}\hat{\boldsymbol{H}}_n, \tag{4.8}$$

where

$$\hat{\boldsymbol{H}}_n = \frac{1}{n}\sum_{i=1}^{n} \boldsymbol{T}_i'\boldsymbol{R}_i\boldsymbol{Z}_i(\boldsymbol{Z}_i'\boldsymbol{R}_i\boldsymbol{Z}_i)^{-1}, \tag{4.9}$$

$$\hat{\boldsymbol{J}}_n = \frac{1}{n}\sum_{i=1}^{n}\sum_{i=1}^{n} \boldsymbol{T}_i'\boldsymbol{Q}_i(\boldsymbol{y}_i - \boldsymbol{T}_i\hat{\boldsymbol{\beta}})(\boldsymbol{y}_i - \boldsymbol{T}_i\hat{\boldsymbol{\beta}})'\boldsymbol{R}_i\boldsymbol{Z}_i(\boldsymbol{Z}_i'\boldsymbol{R}_i\boldsymbol{Z}_i)^{-1}, \tag{4.10}$$

$$\widehat{\boldsymbol{\Sigma}_b + \boldsymbol{K}}_n = \frac{1}{n}\sum_{i=1}^{n}(\boldsymbol{Z}_i'\boldsymbol{R}_i\boldsymbol{Z}_i)^{-1}\boldsymbol{Z}_i'\boldsymbol{R}_i(\boldsymbol{y}_i - \boldsymbol{X}_i\hat{\boldsymbol{\theta}})(\boldsymbol{y}_i - \boldsymbol{X}_i\hat{\boldsymbol{\theta}})'\boldsymbol{R}_i\boldsymbol{Z}_i(\boldsymbol{Z}_i'\boldsymbol{R}_i\boldsymbol{Z}_i)^{-1}, \tag{4.11}$$

are consistent estimators for $\boldsymbol{H}_n$, $\boldsymbol{J}_n$ and $\boldsymbol{\Sigma}_b + \boldsymbol{K}_n$, respectively.

4.2 General Case with q-grouping

In many panel data studies with missing responses, there may be panels with very few observed responses. When the number of parameters q for $\boldsymbol{\alpha}$ is not very small, it is highly likely to have some panels with $K_i < q$. For the subgroup of panels satisfying $K_i \geq q$, we can use them to compute an estimator as in the simple case by treating this subgroup as if it is the whole sample. But is this estimator still unbiased? If not, how should we construct an unbiased estimator? In this section, we will answer these questions.

q-grouping Estimator $\hat{\boldsymbol{\alpha}}_q$

Let $I_{q+} = \{i : k_i \geq q\}$ denote the subset of panels with at least q observations. Let $M_q = |I_{q+}|$ be the number of panels in this subset. Using only observations in this subset, we can have the following estimator

$$\begin{aligned}
\hat{\boldsymbol{\alpha}}_{I_{q+}} &= \frac{1}{M_q} \sum_{i \in I_{q+}} (\boldsymbol{Z}_i' \boldsymbol{R}_i \boldsymbol{Z}_i)^{-1} \boldsymbol{Z}_i' \boldsymbol{R}_i (\boldsymbol{y}_i - \boldsymbol{T}_i \hat{\boldsymbol{\beta}}) \\
&= \boldsymbol{\alpha} + \frac{1}{M_q} \sum_{i \in I_{q+}} \boldsymbol{b}_i + \frac{1}{M_q} \sum_{i \in I_{q+}} (\boldsymbol{Z}_i' \boldsymbol{R}_i \boldsymbol{Z}_i)^{-1} \boldsymbol{Z}_i' \boldsymbol{R}_i \boldsymbol{e}_i \\
&\quad - \frac{1}{M_q} \sum_{i \in I_{q+}} (\boldsymbol{Z}_i' \boldsymbol{R}_i \boldsymbol{Z}_i)^{-1} \boldsymbol{Z}_i' \boldsymbol{R}_i \boldsymbol{T}_i \boldsymbol{S}_n^{-1} (\sum_{j=1}^{n} \boldsymbol{T}_j' \boldsymbol{Q}_j \boldsymbol{e}_j). \qquad (4.12)
\end{aligned}$$

The expectation of the last two terms are 0 because they are linear functions of $\boldsymbol{e}_i$'s. However, the expectation of the second term is

$$E(\frac{1}{M_q} \sum_{i \in I_{q+}} \boldsymbol{b}_i) = E_{\boldsymbol{\delta}_{q+}} \{ \frac{1}{M_q} \sum_{i \in I_{q+}} \delta_{i,q+} E(\boldsymbol{b}_i | \delta_{i,q+}) \},$$

where $\delta_{i,q+} = 1$ if panel i is in I_{q+}, and $\delta_{i,q+} = 0$ otherwise. This term is an expectation of weighted average of conditional expectations $E(\boldsymbol{b}_i|\delta_{i,q+}),\ i \in I_{q+}$. Because we are not assuming specific conditional distribution of missingness, we cannot conduct this calculation further. But in the studies by Follmann and Wu (1995) and Park et al. (2002), it is shown that $E(\boldsymbol{b}_i|k_i)$ is monotonic in k_i if the conditional distribution of $\boldsymbol{b}_i$ given k_i has density in exponential family (TP2). If this condition is satisfied, the average of $E(\boldsymbol{b}_i|\delta_{i,q+}),\ i \in I_{q+}$ is highly likely to be positive (negative) if $E(\boldsymbol{b}_i|k_i)$ is monotonic increasing (decreasing) with k_i.

At this point, we get the answer for the first question: the estimator using only panels with at least q observations may be biased. The possible linear trend between $E(\boldsymbol{b}_i|k_i)$ and k_i suggest that we need to include the remaining panels with $k_i < q$ to offset potential bias. The next question is how can we utilize the rest information and construct an unbiased estimator.

For those panels with $k_i < q$, $\boldsymbol{Z}_i'\boldsymbol{R}_i\boldsymbol{Z}_i$ is no longer invertible and we are no longer able to have the naive OLS estimator computed. To utilize these information, we will take a grouping approach to form groups and treat each group as a new individual, with which we are able to compute the naive OLS estimates.

We form the following groups for panels not in subset I_{q+} as:

$$I_k = \{i : k_i = k\}, \quad k = q, \ldots, q-1.$$

In group I_k, each individual i satisfies model (3.2). Applying $\boldsymbol{Z}_i'\boldsymbol{R}_i$ to both sides of the equation, we have

$$\boldsymbol{Z}_i'\boldsymbol{R}_i\boldsymbol{y}_i = \boldsymbol{Z}_i'\boldsymbol{R}_i\boldsymbol{Z}_i\boldsymbol{\alpha} + \boldsymbol{Z}_i'\boldsymbol{R}_i\boldsymbol{T}_i\boldsymbol{\beta} + \boldsymbol{Z}_i'\boldsymbol{R}_i\boldsymbol{Z}_i\boldsymbol{b}_i + \boldsymbol{Z}_i'\boldsymbol{R}_i\boldsymbol{e}_i.$$

Because $k_i < q$ for all individuals in group I_k, $\boldsymbol{Z}_i'\boldsymbol{R}_i\boldsymbol{Z}_i$ is not invertible. However, we can sum up all individuals in group I_k to get

$$\sum_{i\in I_k}\boldsymbol{Z}_i'\boldsymbol{R}_i\boldsymbol{y}_i = \sum_{i\in I_k}\boldsymbol{Z}_i'\boldsymbol{R}_i\boldsymbol{Z}_i\boldsymbol{\alpha} + \sum_{i\in I_k}\boldsymbol{Z}_i'\boldsymbol{R}_i\boldsymbol{T}_i\boldsymbol{\beta} + \sum_{i\in I_k}\boldsymbol{Z}_i'\boldsymbol{R}_i\boldsymbol{Z}_i\boldsymbol{b}_i + \sum_{i\in I_k}\boldsymbol{Z}_i'\boldsymbol{R}_i\boldsymbol{e}_i.$$

Each individual $\boldsymbol{Z}_i'\boldsymbol{R}_i\boldsymbol{Z}_i$ is singular, but the sum $\sum_{i\in I_k}\boldsymbol{Z}_i'\boldsymbol{R}_i\boldsymbol{Z}_i$ is very likely to be invertible. When this is true, we can get the following naive OLS estimator

$$\begin{aligned}
\hat{\boldsymbol{\alpha}}_{I_k} &= (\sum_{i\in I_k}\boldsymbol{Z}_i'\boldsymbol{R}_i\boldsymbol{Z}_i)^{-1}(\sum_{i\in I_k}\boldsymbol{Z}_i'\boldsymbol{R}_i(\boldsymbol{y}_i - \boldsymbol{T}_i\hat{\boldsymbol{\beta}})) \\
&= \alpha + (\sum_{i\in I_k}\boldsymbol{Z}_i'\boldsymbol{R}_i\boldsymbol{Z}_i)^{-1}(\sum_{i\in I_k}\boldsymbol{Z}_i'\boldsymbol{R}_i\boldsymbol{Z}_i\boldsymbol{b}_i) + (\sum_{i\in I_k}\boldsymbol{Z}_i'\boldsymbol{R}_i\boldsymbol{Z}_i)^{-1}(\sum_{i\in I_k}\boldsymbol{Z}_i'\boldsymbol{R}_i\boldsymbol{e}_i) \\
&\quad -(\sum_{i\in I_k}\boldsymbol{Z}_i'\boldsymbol{R}_i\boldsymbol{Z}_i)^{-1}(\sum_{i\in I_k}\boldsymbol{Z}_i'\boldsymbol{R}_i\boldsymbol{T}_i\boldsymbol{S}_n^{-1}(\sum_{j=1}^{n}\boldsymbol{T}_j'\boldsymbol{Q}_j\boldsymbol{e}_j)).
\end{aligned} \tag{4.13}$$

By applying this to all groups, we can get estimators $\hat{\boldsymbol{\alpha}}_{I_k},\ k = 1,\ldots,q-1$ besides estimator $\hat{\boldsymbol{\alpha}}_{I_{q+}}$. Let $m_k = |I_k|$ be the number of panels in group I_k, $k = 1,\ldots,q-1$, and $\delta_{i,k} = I[k_i == k]$ be the indicator taking 1 if subject i is in the kth group. We can compute the weighted average of all group-level estimators using the group proportions as the weights,

$$\hat{\boldsymbol{\alpha}}_q = \sum_{k=1}^{q-1}\frac{m_k}{n}\hat{\boldsymbol{\alpha}}_{I_k} + \frac{M_q}{n}\hat{\boldsymbol{\alpha}}_{I_{q+}}. \tag{4.14}$$

Properties of $\hat{\boldsymbol{\alpha}}_q$

With model (3.2) and informative missing, and the assumptions we are making in section 2, we can prove the following properties for estimator $\hat{\boldsymbol{\alpha}}_q$.

Theorem 4.3. *(Unbiasedness of $\hat{\boldsymbol{\alpha}}_q$) Estimator $\hat{\boldsymbol{\alpha}}_q$ with q-grouping is unbiased.*

We prove the unbiasedness of estimator $\hat{\boldsymbol{\alpha}}_q$, by showing: $E(\frac{m_k}{n}\hat{\boldsymbol{\alpha}_{I_k}}) = P(k_i = k)\cdot(\boldsymbol{\alpha}+E(\boldsymbol{b}_i|k_i = k))$ for $k = 1, \ldots, q-1$, and $E(\frac{M_q}{n}\hat{\boldsymbol{\alpha}_{I_{q+}}}) = P(k_i \geq q)\cdot(\boldsymbol{\alpha}+E(\boldsymbol{b}_i|k_i \geq q))$. Please see section 4.5 for detail.

Theorem 4.4. *(Asymptotic normality of $\hat{\boldsymbol{\alpha}}_q$) Under model* (3.2) *with informative missingness. As $n \to \infty$, assume the following limits exist besides limits* (3.9) *and* (3.11)*,*

$$m_k \to \infty, \quad \frac{m_k}{n} \to a_k \in [0,1], \quad k = 1, \ldots, q-1, \tag{4.15}$$

$$M_q \to \infty, \quad \frac{M_q}{n} \to a_q \in [0,1], \tag{4.16}$$

$$\frac{1}{m_k}\sum_{I_k} \boldsymbol{Z}_i'\boldsymbol{R}_i\boldsymbol{Z}_i \to_p \boldsymbol{A}_k, \quad k = 1, \ldots, q-1, \tag{4.17}$$

$$\frac{1}{m_k}\sum_{I_k} \boldsymbol{Z}_i'\boldsymbol{R}_i\boldsymbol{T}_i \to_p \boldsymbol{B}_k, \quad k = 1, \ldots, q-1, \tag{4.18}$$

$$\frac{1}{n}\sum_{I_k} \boldsymbol{Z}_i'\boldsymbol{R}_i\boldsymbol{Z}_i E(\boldsymbol{b}_i\boldsymbol{b}_i'|\delta_{i,k})\boldsymbol{Z}_i'\boldsymbol{R}_i\boldsymbol{Z}_i \to_p \boldsymbol{C}_k, \quad k = 1, \ldots, q-1, \tag{4.19}$$

$$\frac{1}{n}\sum_{I_k} \boldsymbol{Z}_i'\boldsymbol{R}_i\boldsymbol{Z}_i E(\boldsymbol{b}_i|\delta_{i,k})E(\boldsymbol{b}_i|\delta_{i,k})'\boldsymbol{Z}_i'\boldsymbol{R}_i\boldsymbol{Z}_i \to_p \boldsymbol{C}_{0k}, \quad k = 1, \ldots, q-1, \tag{4.20}$$

$$\frac{1}{n}\sum_{I_k} \boldsymbol{Z}_i'\boldsymbol{R}_i\boldsymbol{\Sigma}_e\boldsymbol{R}_i\boldsymbol{Z}_i \to_p \boldsymbol{D}_k, \quad k = 1, \ldots, q-1, \tag{4.21}$$

$$\frac{1}{n}\sum_{I_k} \boldsymbol{Z}_i'\boldsymbol{R}_i\boldsymbol{\Sigma}_e\boldsymbol{Q}_i\boldsymbol{T}_i \to_p \boldsymbol{F}_k, \quad k = 1, \ldots, q-1, \tag{4.22}$$

$$\frac{1}{n}\sum_{I_{q+}} E(\boldsymbol{b}_i\boldsymbol{b}_i'|\delta_{iq}^*) \to_p \boldsymbol{G}_q, \tag{4.23}$$

$$\frac{1}{n}\sum_{I_{q+}} E(\boldsymbol{b}_i|\delta_{i,q+})E(\boldsymbol{b}_i|\delta_{i,q+})' \to_p \boldsymbol{G}_{0q}, \tag{4.24}$$

$$\frac{1}{M_q}\sum_{I_{q+}} \boldsymbol{T}_i'\boldsymbol{R}_i\boldsymbol{Z}_i(\boldsymbol{Z}_i'\boldsymbol{R}_i\boldsymbol{Z}_i)^{-1} \to_p \boldsymbol{H}_q, \tag{4.25}$$

$$\frac{1}{M_q}\sum_{I_{q+}} \boldsymbol{T}_i'\boldsymbol{Q}_i\boldsymbol{\Sigma}_e\boldsymbol{R}_i\boldsymbol{Z}_i(\boldsymbol{Z}_i'\boldsymbol{R}_i\boldsymbol{Z}_i)^{-1} \to_p \boldsymbol{J}_q, \tag{4.26}$$

$$\frac{1}{n}\sum_{I_{q+}} (\boldsymbol{Z}_i'\boldsymbol{R}_i\boldsymbol{Z}_i)^{-1}\boldsymbol{Z}_i'\boldsymbol{R}_i\boldsymbol{\Sigma}_e\boldsymbol{R}_i\boldsymbol{Z}_i(\boldsymbol{Z}_i'\boldsymbol{R}_i\boldsymbol{Z}_i)^{-1} \to_p \boldsymbol{K}_q. \tag{4.27}$$

where $\to_p$ denotes convergence in probability.

Then, we have

$$\sqrt{n}(\hat{\boldsymbol{\alpha}}_q - \boldsymbol{\alpha}) \to_d N(\boldsymbol{0}, \boldsymbol{\Sigma}_{\hat{\boldsymbol{\alpha}}_q}), \tag{4.28}$$

where the covariance matrix

$$\begin{aligned}
\boldsymbol{\Sigma}_{\hat{\boldsymbol{\alpha}}_q} \;=\; & \sum_{k=1}^{q-1} \boldsymbol{A}_k^{-1}(\boldsymbol{C}_k - \boldsymbol{C}_{0k})\boldsymbol{A}_k^{-1} - \sum_{k=1}^{q-1} \boldsymbol{A}_k^{-1}\boldsymbol{D}_k\boldsymbol{A}_k^{-1} \\
& + \sum_{k=1}^{q-1}\sum_{l=1}^{q-1} a_k a_l \boldsymbol{A}_k^{-1}\boldsymbol{B}_k\boldsymbol{M}^{-1}\boldsymbol{L}\boldsymbol{M}^{-1}\boldsymbol{B}_l'\boldsymbol{A}_l^{-1} \\
& - \sum_{k=1}^{q-1}\sum_{l=1}^{q-1} a_k \boldsymbol{A}_k^{-1}\boldsymbol{B}_k\boldsymbol{M}^{-1}\boldsymbol{F}_l'\boldsymbol{A}_l \\
& - \sum_{k=1}^{q-1}\sum_{l=1}^{q-1} a_l \boldsymbol{A}_k^{-1}\boldsymbol{F}_k\boldsymbol{M}^{-1}\boldsymbol{B}_l'\boldsymbol{A}_l^{-1} \\
& + \boldsymbol{G}_q - \boldsymbol{G}_{0q} + \boldsymbol{K}_q + a_q^2\boldsymbol{H}_q'\boldsymbol{M}^{-1}\boldsymbol{L}\boldsymbol{M}^{-1}\boldsymbol{H}_q \\
& - a_q\boldsymbol{J}_q\boldsymbol{M}^{-1}\boldsymbol{H}_q - a_q\boldsymbol{H}_q'\boldsymbol{M}^{-1}\boldsymbol{J}_q \\
& - \sum_{k=1}^{q-1} a_k\boldsymbol{J}_q\boldsymbol{M}^{-1}\boldsymbol{B}_k'\boldsymbol{A}_k^{-1} - -\sum_{k=1}^{q-1} a_k\boldsymbol{A}_k^{-1}\boldsymbol{B}_k\boldsymbol{M}^{-1}\boldsymbol{J}_q \\
& - \sum_{k=1}^{q-1} a_q\boldsymbol{A}_k^{-1}\boldsymbol{F}_k\boldsymbol{M}^{-1}\boldsymbol{H}_q - \sum_{k=1}^{q-1} a_q\boldsymbol{H}_q'\boldsymbol{M}^{-1}\boldsymbol{F}_k\boldsymbol{A}_k^{-1} \\
& + \sum_{k=1}^{q-1} a_q a_k\boldsymbol{H}_q'\boldsymbol{M}^{-1}\boldsymbol{L}\boldsymbol{M}^{-1}\boldsymbol{B}_k'\boldsymbol{A}_k^{-1} \\
& + \sum_{k=1}^{q-1} a_q a_k\boldsymbol{A}_k^{-1}\boldsymbol{B}_k'\boldsymbol{M}^{-1}\boldsymbol{L}\boldsymbol{M}^{-1}\boldsymbol{H}_q'
\end{aligned}$$

Consistent Estimator

In the simple case, the asymptotic covariance matrix has a simple form, and we were able to create a consistent estimator by constructing consistent estimators for each single matrix involved in the covariance. Now, in the general case with q-grouping, we can see from Theorem 4.4 that the asymptotic covariance matrix $\boldsymbol{\Sigma}_{\hat{\boldsymbol{\alpha}}_q}$ is much more complicated than $\boldsymbol{\Sigma}_{\tilde{\boldsymbol{\alpha}}}$ in the simple case. In order to compute this asymptotic covariance matrix, we defined limits (4.15) to (4.27). Similar to the consistent estimator for $\boldsymbol{\Sigma}_{\tilde{\boldsymbol{\alpha}}}$, if we can get consistent estimators for all these limiting matrices, then a consistent estimator for $\boldsymbol{\Sigma}_{\hat{\boldsymbol{\alpha}}_q}$ follows naturally by replacing all the limiting matrices by their consistent estimators.

Consistent estimators for a_k, a_q, $\boldsymbol{A}_k$, $\boldsymbol{B}_k$ and $\boldsymbol{H}_q$ can simply be define as the left-hand side fraction or averages, because once the data and missingness information are available, they are known. The left-hand side averages for limits $\boldsymbol{D}_k$, $\boldsymbol{F}_k$, $\boldsymbol{J}_q$ and $\boldsymbol{K}_q$ all involved the covariance matrix $\boldsymbol{\Sigma}_e$. This covariance matrix is unknown even if given all observed data and missingness information. However, just as in consistent estimator $\hat{\boldsymbol{J}}_n$ (4.10) and $\widehat{\boldsymbol{\Sigma}_b + \boldsymbol{K}_n}$ (4.11), we can use residual to construct consistent estimators. So, we have the following consistent estimators,

$$\hat{\boldsymbol{F}}_k = \frac{1}{n}\sum_{I_k} \boldsymbol{Z}_i'\boldsymbol{R}_i(\boldsymbol{y}_i - \boldsymbol{T}_i'\hat{\boldsymbol{\beta}})(\boldsymbol{y}_i - \boldsymbol{T}_i'\hat{\boldsymbol{\beta}})'\boldsymbol{Q}_i\boldsymbol{T}_i, \quad k = 1, \ldots, q-1, \tag{4.29}$$

$$\widehat{\boldsymbol{C}_k + \boldsymbol{D}_k} = \frac{1}{n}\sum_{I_k} \boldsymbol{Z}_i'\boldsymbol{R}_i\boldsymbol{Z}_i(\boldsymbol{y}_i - \boldsymbol{X}_i'\hat{\boldsymbol{\theta}})(\boldsymbol{y}_i - \boldsymbol{X}_i'\hat{\boldsymbol{\theta}})'\boldsymbol{Z}_i'\boldsymbol{R}_i\boldsymbol{Z}_i, \quad k = 1, \ldots, q-1, \tag{4.30}$$

$$\hat{\boldsymbol{J}}_q = \frac{1}{M_q}\sum_{I_{q+}} \boldsymbol{T}_i'\boldsymbol{Q}_i(\boldsymbol{y}_i - \boldsymbol{T}_i'\hat{\boldsymbol{\beta}})(\boldsymbol{y}_i - \boldsymbol{T}_i'\hat{\boldsymbol{\beta}})'\boldsymbol{R}_i\boldsymbol{Z}_i(\boldsymbol{Z}_i'\boldsymbol{R}_i\boldsymbol{Z}_i)^{-1}, \tag{4.31}$$

$$\widehat{\boldsymbol{G}_k + \boldsymbol{K}_q} = \frac{1}{M_q}\sum_{I_{q+}} (\boldsymbol{Z}_i'\boldsymbol{R}_i\boldsymbol{Z}_i)^{-1}\boldsymbol{Z}_i'\boldsymbol{R}_i(\boldsymbol{y}_i - \boldsymbol{X}_i'\hat{\boldsymbol{\theta}})(\boldsymbol{y}_i - \boldsymbol{X}_i'\hat{\boldsymbol{\theta}})'\boldsymbol{R}_i\boldsymbol{Z}_i(\boldsymbol{Z}_i'\boldsymbol{R}_i\boldsymbol{Z}_i)^{-1}. \tag{4.32}$$

Note that $\widehat{\boldsymbol{C}_k + \boldsymbol{D}_k}$ is a consistent estimator for $\boldsymbol{C}_k + \boldsymbol{D}_k$, and $\widehat{\boldsymbol{G}_q + \boldsymbol{K}_q}$ is a consistent estimator for $\boldsymbol{G}_q + \boldsymbol{K}_q$.

Up to now, we constructed consistent estimator for all matrices involved in the limiting covariance matrix, except $\boldsymbol{C}_{0k},\ k = 1, ..., q-1$ and $\boldsymbol{G}_{0q}$. Consistent estimators for them cannot be easily constructed as a result of product of expectation terms $E(\boldsymbol{b}_i|\delta_i)E(\boldsymbol{b}_i|\delta_i)'$. Nevertheless, we can always use bootstrap resampling and get a consistent estimates in practice.

4.3 General Case with K-grouping

In the previous section, we grouped panels with less than q observations into groups based on their observed number of responses. But for panels with at least q observations, they are not further grouped, or they can be viewed a big group. Since we can do further partition to panels with less than q observations, we could do the same to panels with at least q observations too. In this section, we construct another estimator by partitioning panels into K groups as:

$$I_k = \{i : k_i = k\}, \quad k = q, \ldots, K.$$

K-grouping Estimator $\hat{\boldsymbol{\alpha}}_K$

Now, we have a total of K groups, similar to the first $q-1$ groups in q-grouping, we can calculate a naive OLS estimator for each of the K groups. Then, we have $\hat{\boldsymbol{\alpha}}_{I_k}$ defined exactly the same as (4.13), but k may be 1 to K instead of 1 to $q-1$.

$$\hat{\boldsymbol{\alpha}}_{I_k} = (\sum_{i \in I_k} \boldsymbol{Z}_i' \boldsymbol{R}_i \boldsymbol{Z}_i)^{-1} (\sum_{i \in I_k} \boldsymbol{Z}_i' \boldsymbol{R}_i (\boldsymbol{y}_i - \boldsymbol{T}_i \hat{\boldsymbol{\beta}}))$$

Our final estimator with K-grouping is a weighted average of these K local estimators, with the group proportion being the weights.

$$\hat{\boldsymbol{\alpha}}_K = \sum_{k=1}^{K} \frac{m_k}{n} \hat{\boldsymbol{\alpha}}_{I_k}. \tag{4.33}$$

Properties of $\hat{\boldsymbol{\alpha}}_K$

Similar to q-grouping estimator, K-grouping estimator $\hat{\boldsymbol{\alpha}}_K$ has both unbiasedness and asymptotic normality properties.

Theorem 4.5. *(Unbiasedness of $\hat{\boldsymbol{\alpha}}_K$) Estimator $\hat{\boldsymbol{\alpha}}_K$ with K-grouping is unbiased.*

Theorem 4.6. *(Asymptotic normality of $\hat{\boldsymbol{\alpha}}_K$) Under model* (3.2) *with informative missingness. As $n \to \infty$, assume the limits* (4.15),(4.17)-(4.22) *exists for $k = 1, \ldots, K$, we have*

$$\sqrt{n}(\hat{\boldsymbol{\alpha}}_K - \boldsymbol{\alpha}) \to_p N(\mathbf{0}, \boldsymbol{\Sigma}_{\hat{\boldsymbol{\alpha}}_K}), \tag{4.34}$$

where the covariance matrix

$$\begin{aligned}\boldsymbol{\Sigma}_{\hat{\alpha}_K} &= \sum_{k=1}^{K} \boldsymbol{A}_k^{-1}(\boldsymbol{C}_k - \boldsymbol{C}_{0k})\boldsymbol{A}_k^{-1} + \sum_{k=1}^{K} \boldsymbol{A}_k^{-1}\boldsymbol{D}_k\boldsymbol{A}_k^{-1} \\ &\quad + \sum_{k=1}^{K}\sum_{l=1}^{K} a_k a_l \boldsymbol{A}_k^{-1}\boldsymbol{B}_k\boldsymbol{M}^{-1}\boldsymbol{L}\boldsymbol{M}^{-1}\boldsymbol{B}_l'\boldsymbol{A}_l^{-1} \\ &\quad - \sum_{k=1}^{K}\sum_{l=1}^{K} a_k \boldsymbol{A}_k^{-1}\boldsymbol{B}_k\boldsymbol{M}^{-1}\boldsymbol{F}_l'\boldsymbol{A}_l \\ &\quad - \sum_{k=1}^{K}\sum_{l=1}^{K} a_l \boldsymbol{A}_k^{-1}\boldsymbol{F}_k\boldsymbol{M}^{-1}\boldsymbol{B}_l'\boldsymbol{A}_l^{-1}\end{aligned}$$

4.4 Discussion of Condition

In both q-grouping and K-grouping, for panels with $k_i \geq q$ observed responses, the matrix $\boldsymbol{Z}_i'\boldsymbol{R}_i\boldsymbol{Z}_i$ itself is usually invertible. But for groups $I_k,\ k = 1, \ldots, q-1$, this matrix is always singular for each individual panel. Instead, we calculated the sum of this matrix over all panels in the same group to get $\sum_{i\in I_k} \boldsymbol{Z}_i'\boldsymbol{R}_i\boldsymbol{Z}_i$, and constructed estimator assuming this sum is nonsingular. Now, let us look at the invertibility of this sum.

When $\boldsymbol{Z}_i$ is fixed, this sum cannot be shown to have positive definiteness theoretically. However, as long as the number of penels is not too small for that group, the sum $\sum_{i\in I_k} \boldsymbol{Z}_i'\boldsymbol{R}_i\boldsymbol{Z}_i$ is highly likely to be positive definite for real problem.

When $\boldsymbol{Z}_i$ is random, $E(\boldsymbol{Z}_i'\boldsymbol{R}_i\boldsymbol{Z}_i\delta_{i,k}) = E\{\boldsymbol{Z}_i'E(\boldsymbol{R}_i\delta_{i,k}|\boldsymbol{Z}_i)\boldsymbol{Z}_i\}$.

$$
\begin{aligned}
E(\boldsymbol{R}_i\delta_{i,k}|\boldsymbol{Z}_i) &= E_{\delta_{i,k}}\{E(\boldsymbol{R}_i\delta_{i,k}|\delta_{i,k},\boldsymbol{Z}_i)\} \\
&= E(\boldsymbol{R}_i|\delta_{i,k}=1,\boldsymbol{Z}_i) * P(\delta_{i,k}=1|\boldsymbol{Z}_i) \\
&= E(\boldsymbol{R}_i|K_i=k,\boldsymbol{Z}_i) * P(K_i=k|\boldsymbol{Z}_i) \\
&= diag(E(r_{ij}|\sum_{j=1}^{K} r_{ij}=k,\boldsymbol{Z}_i)) * P(K_i=k|\boldsymbol{Z}_i) \\
&= diag(\frac{1}{k},\ldots,\frac{1}{k}) * P(K_i=k|\boldsymbol{Z}_i)
\end{aligned}
$$

Then, $\boldsymbol{Z}_i' E(\boldsymbol{R}_i\delta_{i,k}|\boldsymbol{Z}_i)\boldsymbol{Z}_i = \frac{P(K_i=k|\boldsymbol{Z}_i)}{k}\boldsymbol{Z}_i'\boldsymbol{Z}_i$, which is positive definite if $\boldsymbol{Z}_i$ is of full rank. Then $E\{\boldsymbol{Z}_i' E(\boldsymbol{R}_i\delta_{i,k}|\boldsymbol{Z}_i)\boldsymbol{Z}_i\} = E\{\frac{P(K_i=k|\boldsymbol{Z}_i)}{k}\boldsymbol{Z}_i'\boldsymbol{Z}_i\}$ is positive definite if $\boldsymbol{Z}_i$ has nonzero probability of being full rank in the random covariate matrix case. As a result, we can conclude that $E(\sum_{i\in I_k}\boldsymbol{Z}_i'\boldsymbol{R}_i\boldsymbol{Z}_i)$ is positive definite.

4.5 Proof

Proof. of **Theorem 4.1**

First, $E(\frac{1}{n}\sum_{i=1}^{n}\boldsymbol{b}_i) = \frac{1}{n}\sum_{i=1}^{n}E(\boldsymbol{b}_i) = 0$ because $E(\boldsymbol{b}_i) = 0$ for all i. The second and last term are both linear functions of random error $\boldsymbol{e}_i$'s. Because the random error $\boldsymbol{e}_i$'s have mean 0 and they are independent among themselves and of everything else, the expectation of a linear combination of them can be shown to have mean 0 similarly to the proof for Theorem 3.1. □

Proof. of **Theorem 4.2**

With the definition of $\boldsymbol{M}$ and $\boldsymbol{L}$ in (3.9) and (3.11), we have

$$\sqrt{n}(\tilde{\boldsymbol{\alpha}}-\boldsymbol{\alpha}) = \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\boldsymbol{b}_i+\frac{1}{\sqrt{n}}\sum_{i=1}^{n}(\boldsymbol{Z}_i'\boldsymbol{R}_i\boldsymbol{Z}_i)^{-1}\boldsymbol{Z}_i'\boldsymbol{R}_i\boldsymbol{e}_i-\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\boldsymbol{H}_n'\boldsymbol{M}^{-1}\boldsymbol{T}_i'\boldsymbol{Q}_i\boldsymbol{e}_i+o_p(1) \tag{4.35}$$

Conditional on $\mathcal{C} = \{\boldsymbol{R}_i, \boldsymbol{Z}_i, \boldsymbol{T}_i, i = 1, \ldots, n\}$, the right hand side of (4.35) is a sum of averages of independent random variables. The asymptotic normality follows directly from central limit theorem. The asymptotic covariance matrix can be obtained by a direct calculation of the variances of $\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\boldsymbol{b}_i$, $\frac{1}{\sqrt{n}}\sum_{i=1}^{n}(\boldsymbol{Z}_i'\boldsymbol{R}_i\boldsymbol{Z}_i)^{-1}\boldsymbol{Z}_i'\boldsymbol{R}_i\boldsymbol{e}_i$ and $\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\boldsymbol{H}_n'\boldsymbol{M}^{-1}\boldsymbol{T}_i'\boldsymbol{Q}_i\boldsymbol{e}_i$, and the covariance between the last two terms plus its transpose.

(1)

$$Var(\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\boldsymbol{b}_i|\mathcal{C}) = \frac{1}{n}\sum_{i=1}^{n}Var(\boldsymbol{b}_i) = \frac{1}{n}\sum_{i=1}^{n}\boldsymbol{\Sigma}_b = \boldsymbol{\Sigma}_b,$$

(2)

$$\begin{aligned} & Var(\frac{1}{\sqrt{n}}\sum_{i=1}^{n}(\boldsymbol{Z}_i'\boldsymbol{R}_i\boldsymbol{Z}_i)^{-1}\boldsymbol{Z}_i'\boldsymbol{R}_i\boldsymbol{e}_i|\mathcal{C}) \\ = \quad & \frac{1}{n}\sum_{i=1}^{n}(\boldsymbol{Z}_i'\boldsymbol{R}_i\boldsymbol{Z}_i)^{-1}\boldsymbol{Z}_i'\boldsymbol{R}_iVar(\boldsymbol{e}_i|\boldsymbol{Z}_i,\boldsymbol{R}_i)\boldsymbol{R}_i\boldsymbol{Z}_i(\boldsymbol{Z}_i'\boldsymbol{R}_i\boldsymbol{Z}_i)^{-1} \\ = \quad & \frac{1}{n}\sum_{i=1}^{n}(\boldsymbol{Z}_i'\boldsymbol{R}_i\boldsymbol{Z}_i)^{-1}\boldsymbol{Z}_i'\boldsymbol{R}_i\boldsymbol{\Sigma}_e\boldsymbol{R}_i\boldsymbol{Z}_i(\boldsymbol{Z}_i'\boldsymbol{R}_i\boldsymbol{Z}_i)^{-1} \\ \to_p \quad & \boldsymbol{K}_n \end{aligned}$$

(3)

$$\begin{aligned} & Var(\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\boldsymbol{H}_n'\boldsymbol{M}^{-1}\boldsymbol{T}_i'\boldsymbol{Q}_i\boldsymbol{e}_i|\mathcal{C}) \\ = \quad & \frac{1}{n}\sum_{i=1}^{n}\boldsymbol{H}_n'\boldsymbol{M}^{-1}\boldsymbol{T}_i'\boldsymbol{Q}_iVar(\boldsymbol{e}_i|\boldsymbol{Z}_i,\boldsymbol{T}_i,\boldsymbol{R}_i)\boldsymbol{Q}_i\boldsymbol{T}_i\boldsymbol{M}^{-1}\boldsymbol{H}_n \end{aligned}$$

$$\begin{aligned} &= \frac{1}{n}\sum_{i=1}^{n} \boldsymbol{H}_n' \boldsymbol{M}^{-1} \boldsymbol{T}_i' \boldsymbol{Q}_i \boldsymbol{\Sigma}_e \boldsymbol{Q}_i \boldsymbol{T}_i \boldsymbol{M}^{-1} \boldsymbol{H}_n \\ \to_p \quad & \boldsymbol{H}_n' \boldsymbol{M}^{-1} \boldsymbol{L} \boldsymbol{M}^{-1} \boldsymbol{H}_n \end{aligned}$$

(4)

$$\begin{aligned} & Cov(\frac{1}{\sqrt{n}}\sum_{i=1}^{n}(\boldsymbol{Z}_i'\boldsymbol{R}_i\boldsymbol{Z}_i)^{-1}\boldsymbol{Z}_i'\boldsymbol{R}_i\boldsymbol{e}_i, \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\boldsymbol{H}_n'\boldsymbol{M}^{-1}\boldsymbol{T}_i'\boldsymbol{Q}_i\boldsymbol{e}_i|\mathcal{C}) \\ = \quad & \frac{1}{n}\sum_{i=1}^{n}(\boldsymbol{Z}_i'\boldsymbol{R}_i\boldsymbol{Z}_i)^{-1}\boldsymbol{Z}_i'\boldsymbol{R}_i Var(\boldsymbol{e}_i|\boldsymbol{Z}_i, \boldsymbol{T}_i, \boldsymbol{R}_i)\boldsymbol{Q}_i\boldsymbol{T}_i\boldsymbol{M}^{-1}\boldsymbol{H}_n \\ = \quad & \frac{1}{n}\sum_{i=1}^{n}(\boldsymbol{Z}_i'\boldsymbol{R}_i\boldsymbol{Z}_i)^{-1}\boldsymbol{Z}_i'\boldsymbol{R}_i\boldsymbol{\Sigma}_e\boldsymbol{Q}_i\boldsymbol{T}_i\boldsymbol{M}^{-1}\boldsymbol{H}_n \\ \to_p \quad & \boldsymbol{J}_n'\boldsymbol{M}^{-1}\boldsymbol{H}_n \end{aligned}$$

□

Proof. of **Theorem 4.3**

(1) For $\frac{m_k}{n}\hat{\boldsymbol{\alpha}}_{I_k}$:

$$\begin{aligned} E(\frac{m_k}{n}\hat{\boldsymbol{\alpha}}_{I_k}) \quad = \quad & E(\frac{m_k}{n})\boldsymbol{\alpha} + E(\frac{m_k}{n}(\sum_{i\in I_k}\boldsymbol{Z}_i'\boldsymbol{R}_i\boldsymbol{Z}_i)^{-1}(\sum_{i\in I_k}\boldsymbol{Z}_i'\boldsymbol{R}_i\boldsymbol{Z}_i\boldsymbol{b}_i)) \\ & + E(\frac{m_k}{n}(\sum_{i\in I_k}\boldsymbol{Z}_i'\boldsymbol{R}_i\boldsymbol{Z}_i)^{-1}(\sum_{i\in I_k}\boldsymbol{Z}_i'\boldsymbol{R}_i\boldsymbol{e}_i)) \\ & - E(\frac{m_k}{n}(\sum_{i\in I_k}\boldsymbol{Z}_i'\boldsymbol{R}_i\boldsymbol{Z}_i)^{-1}(\sum_{i\in I_k}\boldsymbol{Z}_i'\boldsymbol{R}_i\boldsymbol{T}_i\boldsymbol{S}_n^{-1}(\sum_{j=1}^{n}\boldsymbol{T}_j'\boldsymbol{Q}_j\boldsymbol{e}_j))) \end{aligned}$$

By the definition of $\delta_{i,k}$, we have $m_k = \sum_{i=1}^{n}\delta_{i,k}$. Then $E(\frac{m_k}{n}) = \sum_{i=1}^{n}\frac{E(\delta_{i,k})}{n} = \sum_{i=1}^{n}\frac{P(k_i=k)}{n}$. Since $k_i = \sum_{k=1}^{K} r_{ik}$ and $\boldsymbol{r}_i$'s are i.i.d, $E(\frac{m_k}{n}) = P(k_1 = k)$.

The second expectation term of $\frac{m_k}{n}\hat{\boldsymbol{\alpha}}_k$ is:

$$
\begin{aligned}
& E(\frac{m_k}{n}(\sum_{i\in I_k}\boldsymbol{Z}_i'\boldsymbol{R}_i\boldsymbol{Z}_i)^{-1}(\sum_{i\in I_k}\boldsymbol{Z}_i'\boldsymbol{R}_i\boldsymbol{Z}_i\boldsymbol{b}_i)) \\
= \;& E\{\frac{m_k}{n}(\sum_{i\in I_k}\boldsymbol{Z}_i'\boldsymbol{R}_i\boldsymbol{Z}_i)^{-1}(\sum_{i=1}^{n}\boldsymbol{Z}_i'\boldsymbol{R}_i\boldsymbol{Z}_iE(\boldsymbol{b}_i\delta_{i,k}|\boldsymbol{Z}_i,\boldsymbol{R}_i,\delta_{i,k}))\} \\
= \;& E\{\frac{m_k}{n}(\sum_{i\in I_k}\boldsymbol{Z}_i'\boldsymbol{R}_i\boldsymbol{Z}_i)^{-1}(\sum_{i=1}^{n}\boldsymbol{Z}_i'\boldsymbol{R}_i\boldsymbol{Z}_iE(\boldsymbol{b}_i\delta_{ik}|\delta_{i,k}))\} \\
= \;& E_{m_k}\{E_{\boldsymbol{Z},\boldsymbol{R},\boldsymbol{\delta}_k|m_k}[\frac{m_k}{n}(\sum_{i\in I_k}\boldsymbol{Z}_i'\boldsymbol{R}_i\boldsymbol{Z}_i)^{-1}(\sum_{i=1}^{n}\boldsymbol{Z}_i'\boldsymbol{R}_i\boldsymbol{Z}_iE(\boldsymbol{b}_i\delta_{i,k}|\delta_{i,k}))|m_k]\}
\end{aligned}
$$

The third equality follows from the assumption (4) that $\boldsymbol{b}_i$ is independent of $\boldsymbol{Z}_i$ and assumption (6) that the conditional distribution of $\boldsymbol{b}_i$ given $\boldsymbol{R}_i, \delta_{i,k}=1$ does not depend on the form of $\boldsymbol{R}_i$, as long as the diagonal elements of $\boldsymbol{R}_i$ sum up to k.

If $m_k = m$, WLOG, assume that $\delta_{i,k} = 1, i = 1,\ldots,m$ and $\delta_{i,k} = 0, i = m+1,\ldots,n$. Then,

$$
\begin{aligned}
& E_{\boldsymbol{Z},\boldsymbol{R},\boldsymbol{\delta}_k|m_k=m}[\frac{m}{n}(\sum_{i=1}^{m}\boldsymbol{Z}_i'\boldsymbol{R}_i\boldsymbol{Z}_i)^{-1}(\sum_{i=1}^{m}\boldsymbol{Z}_i'\boldsymbol{R}_i\boldsymbol{Z}_iE(\boldsymbol{b}_i|\delta_{i,k}=1))|m_k=m] \\
= \;& E_{\boldsymbol{Z},\boldsymbol{R},\boldsymbol{\delta}_k|m_k=m}[\frac{m}{n}(\sum_{i=1}^{m}\boldsymbol{Z}_i'\boldsymbol{R}_i\boldsymbol{Z}_i)^{-1}(\sum_{i=1}^{m}\boldsymbol{Z}_i'\boldsymbol{R}_i\boldsymbol{Z}_i)E(b_1|\delta_{k1}=1)] \\
= \;& \frac{m}{n}E(\boldsymbol{b}_1|\delta_{k1}=1)
\end{aligned}
$$

So, we have

$$
\begin{aligned}
& E(\frac{m_k}{n}(\sum_{i\in I_k}\boldsymbol{Z}_i'\boldsymbol{R}_i\boldsymbol{Z}_i)^{-1}(\sum_{i\in I_k}\boldsymbol{Z}_i'\boldsymbol{R}_i\boldsymbol{Z}_i\boldsymbol{b}_i)) \\
= \;& E(\frac{m_k}{n}E(\boldsymbol{b}_1|\delta_{k1}=1))
\end{aligned}
$$

$$
\begin{aligned}
&= E_{m_k}(\frac{m_k}{n})E(\boldsymbol{b}_1|\delta_{k1}=1) \\
&= P(k_i=k)E(\boldsymbol{b}_i|k_i=k)
\end{aligned}
$$

For the last two terms of $\hat{\boldsymbol{\alpha}}_k$, they are linear functions of random error $\boldsymbol{e}_i$'s and the means should be 0.

(2) For $\frac{M_q}{n}\hat{\boldsymbol{\alpha}}_{I_{q+}}$:

$$
\begin{aligned}
& E(\frac{M_q}{n}\hat{\boldsymbol{\alpha}}_{I_{q+}}) \\
= \; & E(\frac{M_q}{n})\boldsymbol{\alpha} + E(\frac{1}{n}\sum_{i\in I_{q+}}\boldsymbol{b}_i) + E(\frac{1}{n}\sum_{i\in I_{q+}}(\boldsymbol{Z}_i'\boldsymbol{R}_i\boldsymbol{Z}_i)^{-1}\boldsymbol{Z}_i'\boldsymbol{R}_i\boldsymbol{e}_i) \\
& -E(\frac{1}{n}\sum_{i\in I_{q+}}(\boldsymbol{Z}_i'\boldsymbol{R}_i\boldsymbol{Z}_i)^{-1}\boldsymbol{Z}_i'\boldsymbol{R}_i\boldsymbol{T}_i\boldsymbol{S}_n^{-1}(\sum_{j=1}^{n}\boldsymbol{T}_j'\boldsymbol{Q}_j\boldsymbol{e}_j)).
\end{aligned}
$$

Similarly to the proof in (1), we can show that $E(\frac{M_q}{n}) = P(k_i \geq q)$.

The second expectation is

$$
\begin{aligned}
E(\frac{1}{n}\sum_{i\in I_{q+}}\boldsymbol{b}_i) &= \frac{1}{n}E(\sum_{i=1}^{n}\delta_{i,q+}\boldsymbol{b}_i) \\
&= \frac{1}{n}\sum_{i=1}^{n}E_{\delta_{i,q+}}(\delta_{i,q+}E(\boldsymbol{b}_i|\delta_{i,q+})) \\
&= \frac{1}{n}\sum_{i=1}^{n}P(\delta_{i,q+}=1)E(\boldsymbol{b}_i|\delta_{i,q+}=1) \\
&= P(\delta_{1,q+}=1)E(\boldsymbol{b}_1|\delta_{1,q+}=1) \\
&= P(k_i\geq q)E(\boldsymbol{b}_i|k_i\geq q)
\end{aligned}
$$

Again, the last two terms for $\frac{M_q}{n}\hat{\boldsymbol{\alpha}}_{I_{q+}}$ are linear functions of $\boldsymbol{e}_i$'s, and their means are 0.

From (1) and (2), we can conclude that $E(\frac{m_k}{n}\hat{\boldsymbol{\alpha}}_{I_k}) = P(k_i = k)\cdot(\boldsymbol{\alpha}+E(\boldsymbol{b}_i|k_i = k))$ for $k = 1, \ldots, q-1$, and $E(\frac{M_q}{n}\hat{\boldsymbol{\alpha}}_{I_{q+}}) = P(k_i \geq q) \cdot (\boldsymbol{\alpha} + E(\boldsymbol{b}_i|k_i \geq q))$. Then, the weighted average is unbiased:

$$\begin{aligned} E(\hat{\boldsymbol{\alpha}}_q) &= \sum_{k=1}^{q-1} E(\frac{m_k}{n}\hat{\boldsymbol{\alpha}}_{I_k} + \frac{M_q}{n}\hat{\boldsymbol{\alpha}}_{I_{q+}}) \\ &= \sum_{k=1}^{q-1} P(k_i = k) \cdot (\boldsymbol{\alpha} + E(\boldsymbol{b}_i|k_i = k)) + P(k_i \geq q) \cdot (\boldsymbol{\alpha} + E(\boldsymbol{b}_i|k_i \geq q)) \\ &= \boldsymbol{\alpha} + E_{k_i}\{E(\boldsymbol{b}_i|k_i)\} \\ &= \boldsymbol{\alpha} + E(\boldsymbol{b}_i) \\ &= \boldsymbol{\alpha} \end{aligned}$$

□

Proof. of **Theorem 4.4**

With the definition of the limits $\boldsymbol{A}_k$, $\boldsymbol{B}_k$, $\boldsymbol{H}_q$ and $\boldsymbol{M}$, we have

$$\begin{aligned} & \sqrt{n}(\hat{\boldsymbol{\alpha}}_q - \boldsymbol{\alpha}) \\ = & \sum_{k=1}^{q-1} \frac{1}{\sqrt{n}} \boldsymbol{A}_k^{-1}(\sum_{i\in I_k} \boldsymbol{Z}_i'\boldsymbol{R}_i\boldsymbol{Z}_i\boldsymbol{b}_i) + \sum_{k=1}^{q-1} \frac{1}{\sqrt{n}} \boldsymbol{A}_k^{-1}(\sum_{i\in I_k} \boldsymbol{Z}_i'\boldsymbol{R}_i\boldsymbol{e}_i) \\ & - \sum_{k=1}^{q-1} \frac{m_k}{n^{3/2}} \boldsymbol{A}_k^{-1}\boldsymbol{B}_k\boldsymbol{M}^{-1}(\sum_{j=1}^{n} \boldsymbol{T}_j'\boldsymbol{Q}_j\boldsymbol{e}_j) + \frac{1}{\sqrt{n}} \sum_{i\in I_{q+}} \boldsymbol{b}_i + \frac{1}{\sqrt{n}} \sum_{i\in I_{q+}} (\boldsymbol{Z}_i'\boldsymbol{R}_i\boldsymbol{e}_i)^{-1}\boldsymbol{Z}_i'\boldsymbol{R}_i\boldsymbol{e}_i \\ & - \frac{M_q}{n^{3/2}} \boldsymbol{H}_q'\boldsymbol{M}^{-1}(\sum_{i=1}^{n} \boldsymbol{T}_j'\boldsymbol{Q}_j\boldsymbol{e}_j) + o_p(1). \end{aligned} \tag{4.36}$$

Under the informative missingness assumption, conditional on $\{\boldsymbol{R}_i, \boldsymbol{X}_i, i = 1, \ldots, n\}$ and grouping $\{\boldsymbol{I}_k, k = 1, \ldots, q-1, \boldsymbol{I}_{q+}\}$, the right hand side of (4.5) is a sum of averages of independent random variables. Based on the central limit theorem, we can prove the asymptotic normality.

The asymptotic covariance matrix can be derived from a direct calculation of all conditional variances of the terms on the right hand side and all pairwise conditional covariances between them. In the following calculation, $\mathcal{C} = \{\boldsymbol{R}_i, \boldsymbol{X}_i, \boldsymbol{I}_k, \boldsymbol{I}_{q+}, i = 1, \ldots, n, k = 1, \ldots, q-1\}$.

$$
\begin{aligned}
& Var\{\sum_{k=1}^{q-1} \frac{1}{\sqrt{n}} \boldsymbol{A}_k^{-1} (\sum_{i \in I_k} \boldsymbol{Z}_i' \boldsymbol{R}_i \boldsymbol{Z}_i \boldsymbol{b}_i) + \sum_{k=1}^{q-1} \frac{1}{\sqrt{n}} \boldsymbol{A}_k^{-1} (\sum_{i \in I_k} \boldsymbol{Z}_i' \boldsymbol{R}_i \boldsymbol{e}_i) + \frac{1}{\sqrt{n}} \sum_{i \in I_{q+}} \boldsymbol{b}_i \\
& + \frac{1}{\sqrt{n}} \sum_{i \in I_{q+}} (\boldsymbol{Z}_i' \boldsymbol{R}_i \boldsymbol{e}_i)^{-1} \boldsymbol{Z}_i' \boldsymbol{R}_i \boldsymbol{e}_i - \frac{M_q}{n^{3/2}} \boldsymbol{H}_q' \boldsymbol{M}^{-1} (\sum_{i=1}^{n} \boldsymbol{T}_j' \boldsymbol{Q}_j \boldsymbol{e}_j) \\
& - \sum_{k=1}^{q-1} \frac{m_k}{n^{3/2}} \boldsymbol{A}_k^{-1} \boldsymbol{B}_k \boldsymbol{M}^{-1} (\sum_{j=1}^{n} \boldsymbol{T}_j' \boldsymbol{Q}_j \boldsymbol{e}_j) | \mathcal{C}\} \\
= \; & \sum_{k=1}^{q-1} Var(\frac{1}{\sqrt{n}} \boldsymbol{A}_k^{-1} (\sum_{i \in I_k} \boldsymbol{Z}_i' \boldsymbol{R}_i \boldsymbol{Z}_i \boldsymbol{b}_i) | \mathcal{C}) + \sum_{k=1}^{q-1} Var(\frac{1}{\sqrt{n}} \boldsymbol{A}_k^{-1} (\sum_{i \in I_k} \boldsymbol{Z}_i' \boldsymbol{R}_i \boldsymbol{e}_i) | \mathcal{C}) \\
& + Var(\frac{1}{n} \sum_{i \in I_{q+}} \boldsymbol{b}_i | \mathcal{C}) + Var(\frac{1}{\sqrt{n}} \sum_{i \in I_{q+}} (\boldsymbol{Z}_i' \boldsymbol{R}_i \boldsymbol{Z}_i)^{-1} \boldsymbol{Z}_i' \boldsymbol{R}_i \boldsymbol{e}_i | \mathcal{C}) \\
& + Var(\frac{M_q}{n^{3/2}} \boldsymbol{H}_q' \boldsymbol{M}^{-1} (\sum_{i=1}^{n} \boldsymbol{T}_j' \boldsymbol{Q}_j \boldsymbol{e}_j) | \mathcal{C}) \\
& + \sum_{k=1}^{q-1} \sum_{l=1}^{q-1} n \cdot cov(-\frac{m_k}{n^2} \boldsymbol{A}_k^{-1} \boldsymbol{B}_k \boldsymbol{M}^{-1} (\sum_{j=1}^{n} \boldsymbol{T}_j' \boldsymbol{Q}_j \boldsymbol{e}_j), -\frac{m_l}{n^2} \boldsymbol{A}_l^{-1} \boldsymbol{B}_l \boldsymbol{M}^{-1} (\sum_{j=1}^{n} \boldsymbol{T}_j' \boldsymbol{Q}_j \boldsymbol{e}_j) | \mathcal{C}) \\
& + \sum_{k=1}^{q-1} \sum_{l=1}^{q-1} n \cdot cov(-\frac{m_k}{n^2} \boldsymbol{A}_k^{-1} \boldsymbol{B}_k \boldsymbol{M}^{-1} (\sum_{j=1}^{n} \boldsymbol{T}_j' \boldsymbol{Q}_j \boldsymbol{e}_j), \frac{1}{n} \boldsymbol{A}_l^{-1} (\sum_{i \in I_l} \boldsymbol{Z}_i' \boldsymbol{R}_i' \boldsymbol{e}_i') | \mathcal{C}) \\
& + \sum_{k=1}^{q-1} \sum_{l=1}^{q-1} n \cdot cov(\frac{1}{n} \boldsymbol{A}_k^{-1} (\sum_{i \in I_k} \boldsymbol{Z}_i' \boldsymbol{R}_i \boldsymbol{e}_i), -\frac{m_l}{n^2} \boldsymbol{A}_l^{-1} \boldsymbol{B}_l \boldsymbol{M}^{-1} (\sum_{j=1}^{n} \boldsymbol{T}_j' \boldsymbol{Q}_j \boldsymbol{e}_j) | \mathcal{C}) \\
& + cov(\frac{1}{\sqrt{n}} \sum_{i \in I_{q+}} (\boldsymbol{Z}_i' \boldsymbol{R}_i \boldsymbol{Z}_i)^{-1} \boldsymbol{Z}_i' \boldsymbol{R}_i \boldsymbol{e}_i, -\frac{M_q}{n^{3/2}} \boldsymbol{H}_q' \boldsymbol{M}^{-1} (\sum_{i=1}^{n} \boldsymbol{T}_j' \boldsymbol{Q}_j \boldsymbol{e}_j) | \mathcal{C}) \\
& + cov(-\frac{M_q}{n^{3/2}} \boldsymbol{H}_q' \boldsymbol{M}^{-1} (\sum_{i=1}^{n} \boldsymbol{T}_j' \boldsymbol{Q}_j \boldsymbol{e}_j), \frac{1}{\sqrt{n}} \sum_{i \in I_{q+}} (\boldsymbol{Z}_i' \boldsymbol{R}_i \boldsymbol{Z}_i)^{-1} \boldsymbol{Z}_i' \boldsymbol{R}_i \boldsymbol{e}_i | \mathcal{C})
\end{aligned}
$$

$$+cov(\frac{1}{\sqrt{n}}\sum_{i\in I_{q+}}(\boldsymbol{Z}_i'\boldsymbol{R}_i\boldsymbol{Z}_i)^{-1}\boldsymbol{Z}_i'\boldsymbol{R}_i\boldsymbol{e}_i, -\frac{m_k}{n^{3/2}}\boldsymbol{A}_k^{-1}\boldsymbol{B}_k\boldsymbol{M}^{-1}(\sum_{j=1}^{n}\boldsymbol{T}_j'\boldsymbol{Q}_j\boldsymbol{e}_j)|\mathcal{C})$$

$$+cov(-\frac{m_k}{n^{3/2}}\boldsymbol{A}_k^{-1}\boldsymbol{B}_k\boldsymbol{M}^{-1}(\sum_{j=1}^{n}\boldsymbol{T}_j'\boldsymbol{Q}_j\boldsymbol{e}_j), \frac{1}{\sqrt{n}}\sum_{i\in I_{q+}}(\boldsymbol{Z}_i'\boldsymbol{R}_i\boldsymbol{Z}_i)^{-1}\boldsymbol{Z}_i'\boldsymbol{R}_i\boldsymbol{e}_i|\mathcal{C})$$

$$+cov(\frac{1}{\sqrt{n}}\boldsymbol{A}_k^{-1}(\sum_{i\in I_k}\boldsymbol{Z}_i'\boldsymbol{R}_i\boldsymbol{e}_i), -\frac{M_q}{n^{3/2}}\boldsymbol{H}_q'\boldsymbol{M}^{-1}(\sum_{j=1}^{n}\boldsymbol{T}_j'\boldsymbol{Q}_j\boldsymbol{e}_j)|\mathcal{C})$$

$$+cov(-\frac{M_q}{n^{3/2}}\boldsymbol{H}_q'\boldsymbol{M}^{-1}(\sum_{j=1}^{n}\boldsymbol{T}_j'\boldsymbol{Q}_j\boldsymbol{e}_j), \frac{1}{\sqrt{n}}\boldsymbol{A}_k^{-1}(\sum_{i\in I_k}\boldsymbol{Z}_i'\boldsymbol{R}_i\boldsymbol{e}_i)|\mathcal{C})$$

$$+cov(-\frac{m_k}{n^{3/2}}\boldsymbol{A}_k^{-1}\boldsymbol{B}_k\boldsymbol{M}^{-1}(\sum_{j=1}^{n}\boldsymbol{T}_j'\boldsymbol{Q}_j\boldsymbol{e}_j), -\frac{M_q}{n^{3/2}}\boldsymbol{A}_k^{-1}\boldsymbol{B}_k\boldsymbol{M}^{-1}(\sum_{j=1}^{n}\boldsymbol{T}_j'\boldsymbol{Q}_j\boldsymbol{e}_j)|\mathcal{C})$$

$$+cov(-\frac{M_q}{n^{3/2}}\boldsymbol{A}_k^{-1}\boldsymbol{B}_k\boldsymbol{M}^{-1}(\sum_{j=1}^{n}\boldsymbol{T}_j'\boldsymbol{Q}_j\boldsymbol{e}_j), -\frac{m_k}{n^{3/2}}\boldsymbol{A}_k^{-1}\boldsymbol{B}_k\boldsymbol{M}^{-1}(\sum_{j=1}^{n}\boldsymbol{T}_j'\boldsymbol{Q}_j\boldsymbol{e}_j)|\mathcal{C})$$

(a) The first variance term is:

$$\begin{aligned}
& \sum_{k=1}^{q-1} var(\frac{1}{\sqrt{n}}\boldsymbol{A}_k^{-1}(\sum_{i\in I_k}\boldsymbol{Z}_i'\boldsymbol{R}_i\boldsymbol{Z}_i\boldsymbol{b}_i)|\mathcal{C}) \\
= & \sum_{k=1}^{q-1}\frac{1}{n}\boldsymbol{A}_k^{-1}\{\sum_{i\in I_k} var(\boldsymbol{Z}_i'\boldsymbol{R}_i\boldsymbol{Z}_i\boldsymbol{b}_i|\mathcal{C})\}\boldsymbol{A}_k^{-1} \\
= & \sum_{k=1}^{q-1}\frac{1}{n}\boldsymbol{A}_k^{-1}\{\sum_{i\in I_k} E(\boldsymbol{Z}_i'\boldsymbol{R}_i\boldsymbol{Z}_i\boldsymbol{b}_i\boldsymbol{b}_i'\boldsymbol{Z}_i'\boldsymbol{R}_i\boldsymbol{Z}_i|\mathcal{C}) - \sum_{i\in I_k} E(\boldsymbol{Z}_i'\boldsymbol{R}_i\boldsymbol{Z}_i\boldsymbol{b}_i|\mathcal{C})E(\boldsymbol{Z}_i'\boldsymbol{R}_i\boldsymbol{Z}_i\boldsymbol{b}_i|\mathcal{C})'\}\boldsymbol{A}_k^{-1} \\
= & \sum_{k=1}^{q-1}\frac{1}{n}\boldsymbol{A}_k^{-1}\{\sum_{i\in I_k}\boldsymbol{Z}_i'\boldsymbol{R}_i\boldsymbol{Z}_i E(\boldsymbol{b}_i\boldsymbol{b}_i'|\boldsymbol{R}_i,\boldsymbol{X}_i,\delta_{i,k})\boldsymbol{Z}_i'\boldsymbol{R}_i\boldsymbol{Z}_i \\
& -\sum_{i\in I_k}\boldsymbol{Z}_i'\boldsymbol{R}_i\boldsymbol{Z}_i E(\boldsymbol{b}_i|\boldsymbol{R}_i,\boldsymbol{X}_i,\boldsymbol{I}_k)E(\boldsymbol{b}_i|\boldsymbol{R}_i,\boldsymbol{X}_i,\boldsymbol{I}_k)'\boldsymbol{Z}_i'\boldsymbol{R}_i\boldsymbol{Z}_i\}\boldsymbol{A}_k^{-1} \\
\to_p & \sum_{k=1}^{q-1}\boldsymbol{A}_k^{-1}\boldsymbol{C}_k\boldsymbol{A}_k^{-1} - \sum_{k=1}^{q-1}\boldsymbol{A}_k^{-1}\boldsymbol{C}_{0k}\boldsymbol{A}_k^{-1}
\end{aligned}$$

The fourth equality follows from the assumption that r_{ij} are i.i.d for given subject i, so that given grouping $\boldsymbol{I}_k$, $\boldsymbol{R}_i$ are i.i.d for $i \in \boldsymbol{I}_k$, and thus random effect $\boldsymbol{b}_i$ is independent of $\boldsymbol{R}_i$ given $i \in \boldsymbol{I}_k$. Since missingness depends on $\boldsymbol{T}_i$ only, $\boldsymbol{X}_i$ can be reduced to $\boldsymbol{T}_i$.

(b) The second variance term is:

$$\begin{aligned}
& \sum_{k=1}^{q-1} var(\frac{1}{\sqrt{n}}\boldsymbol{A}_k^{-1}(\sum_{i\in I_k}\boldsymbol{Z}_i'\boldsymbol{R}_i\boldsymbol{e}_i)|\mathcal{C}) \\
= & \sum_{k=1}^{q-1}\frac{1}{n}\boldsymbol{A}_k^{-1}\{\sum_{i\in I_k} var(\boldsymbol{Z}_i'\boldsymbol{R}_i\boldsymbol{e}_i|\mathcal{C})\}\boldsymbol{A}_k^{-1} \\
= & \sum_{k=1}^{q-1}\frac{1}{n}\boldsymbol{A}_k^{-1}\{\sum_{i\in I_k} \boldsymbol{Z}_i'\boldsymbol{R}_i var(\boldsymbol{e}_i|\boldsymbol{Z}_i,\boldsymbol{R}_i,I_k)\boldsymbol{R}_i\boldsymbol{Z}_i\}\boldsymbol{A}_k^{-1} \\
= & \sum_{k=1}^{q-1}\frac{1}{n}\boldsymbol{A}_k^{-1}\{\sum_{i\in I_k} \boldsymbol{Z}_i'\boldsymbol{R}_i\boldsymbol{\Sigma}_e\boldsymbol{R}_i\boldsymbol{Z}_i\}\boldsymbol{A}_k^{-1} \\
\to_p & \sum_{k=1}^{K}\boldsymbol{A}_k^{-1}\boldsymbol{D}_k\boldsymbol{A}_k^{-1}
\end{aligned}$$

(c) The third variance term is:

$$\begin{aligned}
& Var(\frac{1}{\sqrt{n}}\sum_{i\in I_{q+}}\boldsymbol{b}_i|\mathcal{C}) \\
= & \frac{1}{n}\sum_{i\in I_{q+}} Var(\boldsymbol{b}_i|\mathcal{C}) \\
= & \frac{1}{n}\sum_{i\in I_{q+}}\{E(\boldsymbol{b}_i\boldsymbol{b}_i'|\delta_{i,q+}) - E(\boldsymbol{b}_i|\delta_{i,q+})E(\boldsymbol{b}_i|\delta_{i,q+})'\} \\
\to_p & \boldsymbol{G}_q - \boldsymbol{G}_{0q}
\end{aligned}$$

(d) The fourth variance term is:

$$\begin{aligned}
& Var(\frac{1}{\sqrt{n}}\sum_{i\in I_{q+}}(\boldsymbol{Z}_i'\boldsymbol{R}_i\boldsymbol{Z}_i)^{-1}\boldsymbol{Z}_i'\boldsymbol{R}_i\boldsymbol{e}_i|\mathcal{C}) \\
= & \frac{1}{n}\sum_{i\in I_{q+}}(\boldsymbol{Z}_i'\boldsymbol{R}_i\boldsymbol{Z}_i)^{-1}\boldsymbol{Z}_i'\boldsymbol{R}_i Var(\boldsymbol{e}_i)\boldsymbol{R}_i\boldsymbol{Z}_i(\boldsymbol{Z}_i'\boldsymbol{R}_i\boldsymbol{Z}_i)^{-1} \\
= & \frac{1}{n}\sum_{i\in I_{q+}}(\boldsymbol{Z}_i'\boldsymbol{R}_i\boldsymbol{Z}_i)^{-1}\boldsymbol{Z}_i'\boldsymbol{R}_i\boldsymbol{\Sigma}_e\boldsymbol{R}_i\boldsymbol{Z}_i(\boldsymbol{Z}_i'\boldsymbol{R}_i\boldsymbol{Z}_i)^{-1} \\
\to_p & \boldsymbol{K}_q
\end{aligned}$$

(e) The fifth variance term is:

$$\begin{aligned}
& Var(\frac{M_q}{n^{3/2}}\boldsymbol{H}_q'\boldsymbol{M}^{-1}(\sum_{i=1}^{n}\boldsymbol{T}_j'\boldsymbol{Q}_j\boldsymbol{e}_j)|\mathcal{C}) \\
= & \quad \frac{M_q}{n}^2\frac{1}{n}\boldsymbol{H}_q'\boldsymbol{M}^{-1}\sum_{i=1}^{n}\boldsymbol{T}_j'\boldsymbol{Q}_j Var(\boldsymbol{e}_j)\boldsymbol{Q}_j\boldsymbol{T}_j \\
= & \quad \frac{M_q}{n}^2\frac{1}{n}\boldsymbol{H}_q'\boldsymbol{M}^{-1}\sum_{i=1}^{n}\boldsymbol{T}_j'\boldsymbol{Q}_j\boldsymbol{\Sigma}_e\boldsymbol{Q}_j\boldsymbol{T}_j \\
\to_p & \quad a_q^2\boldsymbol{H}_q'\boldsymbol{M}^{-1}\boldsymbol{L}\boldsymbol{M}^{-1}\boldsymbol{H}_q'
\end{aligned}$$

(f) The first covariance term is:

$$\begin{aligned}
& n\cdot\sum_{k=1}^{q-1}\sum_{l=1}^{q-1}cov(-\frac{m_k}{\sqrt{n}}\boldsymbol{A}_k^{-1}\boldsymbol{B}_k\boldsymbol{M}^{-1}(\sum_{j=1}^{n}\boldsymbol{T}_j'\boldsymbol{Q}_j\boldsymbol{e}_j), -\frac{m_l}{n^{3/2}}\boldsymbol{A}_l^{-1}\boldsymbol{B}_l\boldsymbol{M}^{-1}(\sum_{j=1}^{n}\boldsymbol{T}_j'\boldsymbol{Q}_j\boldsymbol{e}_j)|\mathcal{C}) \\
= & \quad \sum_{k=1}^{q-1}\sum_{l=1}^{q-1}\frac{m_k m_l}{n^3}\boldsymbol{A}_k^{-1}\boldsymbol{B}_k\boldsymbol{M}^{-1}cov\{\sum_{j=1}^{n}\boldsymbol{T}_j'\boldsymbol{Q}_j\boldsymbol{e}_j, \sum_{j=1}^{n}\boldsymbol{T}_j'\boldsymbol{Q}_j\boldsymbol{e}_j|\mathcal{C}\}\boldsymbol{M}^{-1}\boldsymbol{B}_l'\boldsymbol{A}_l^{-1} \\
= & \quad \sum_{k=1}^{q-1}\sum_{l=1}^{q-1}\frac{m_k m_l}{n^3}\boldsymbol{A}_k^{-1}\boldsymbol{B}_k\boldsymbol{M}^{-1}\{\sum_{j=1}^{n}\boldsymbol{T}_j'\boldsymbol{Q}_j var(\boldsymbol{e}_j|\boldsymbol{Z}_i,\boldsymbol{R}_i)\}\boldsymbol{M}^{-1}\boldsymbol{B}_l'\boldsymbol{A}_l^{-1} \\
= & \quad \sum_{k=1}^{q-1}\sum_{l=1}^{q-1}\frac{m_k m_l}{n^3}\boldsymbol{A}_k^{-1}\boldsymbol{B}_k\boldsymbol{M}^{-1}\{\sum_{j=1}^{n}\boldsymbol{T}_j'\boldsymbol{Q}_j\boldsymbol{\Sigma}_e\boldsymbol{Q}_j\boldsymbol{T}_j\}\boldsymbol{M}^{-1}\boldsymbol{B}_l'\boldsymbol{A}_l^{-1} \\
\to_p & \quad \sum_{k=1}^{q-1}\sum_{l=1}^{q-1}a_k a_l\boldsymbol{A}_k^{-1}\boldsymbol{B}_k\boldsymbol{M}^{-1}\boldsymbol{L}\boldsymbol{M}^{-1}\boldsymbol{B}_l'\boldsymbol{A}_l^{-1}
\end{aligned}$$

(g) The second covariance term is (the third covariance term is a transpose of this covariance):

$$
\begin{aligned}
& \sum_{k=1}^{q-1}\sum_{l=1}^{q-1} cov(\frac{1}{\sqrt{n}}\boldsymbol{A}_k^{-1}(\sum_{i\in I_k}\boldsymbol{Z}_i'\boldsymbol{R}_i\boldsymbol{e}_i), -\frac{m_l}{n^{3/2}}\boldsymbol{A}_l^{-1}\boldsymbol{B}_l\boldsymbol{M}^{-1}(\sum_{j=1}^{2}\boldsymbol{T}_j'\boldsymbol{Q}_j\boldsymbol{e}_j)|\mathcal{C}) \\
= \quad & -\frac{m_l}{n^2}\boldsymbol{A}_k^{-1}cov\{\sum_{i\in I_k}\boldsymbol{Z}_i'\boldsymbol{R}_i\boldsymbol{e}_i, \sum_{i=1}^{n}\boldsymbol{T}_i'\boldsymbol{Q}_I\boldsymbol{e}_i|\mathcal{C}\}\boldsymbol{M}^{-1}\boldsymbol{B}_l'\boldsymbol{A}_l^{-1} \\
= \quad & -\frac{m_l}{n^2}\boldsymbol{A}_k^{-1}\{\sum_{i\in I_k}\boldsymbol{Z}_i'\boldsymbol{R}_iE(\boldsymbol{e}_i\boldsymbol{e}_i'|\boldsymbol{R}_i,\boldsymbol{Z}_i,I_k)\boldsymbol{Q}_i\boldsymbol{T}_i\}\boldsymbol{M}^{-1}\boldsymbol{B}_l'\boldsymbol{A}_l^{-1} \\
= \quad & -\frac{m_l}{n^2}\boldsymbol{A}_k^{-1}\{\sum_{i\in I_k}\boldsymbol{Z}_i'\boldsymbol{R}_i\boldsymbol{\Sigma}_e\boldsymbol{Q}_i\boldsymbol{T}_i)\}\boldsymbol{M}^{-1}\boldsymbol{B}_l'\boldsymbol{A}_l^{-1} \\
\to_p \quad & -a_l\boldsymbol{A}_k^{-1}\boldsymbol{F}_k\boldsymbol{M}^{-1}\boldsymbol{B}_l'\boldsymbol{A}_l^{-1}
\end{aligned}
$$

(h) The fourth covariance term is (the fifth covariance term is a transpose of this covariance):

$$
\begin{aligned}
& cov(\frac{1}{\sqrt{n}}\sum_{i\in I_{q+}}(\boldsymbol{Z}_i'\boldsymbol{R}_i\boldsymbol{Z}_i)^{-1}\boldsymbol{Z}_i'\boldsymbol{R}_i\boldsymbol{e}_i, -\frac{M_q}{n^{3/2}}\boldsymbol{H}_q'\boldsymbol{M}^{-1}(\sum_{j=1}^{n}\boldsymbol{T}_j'\boldsymbol{Q}_j\boldsymbol{e}_j)|\mathcal{C}) \\
= \quad & -\frac{M_q}{n^{3/2}}\sum_{i\in I_{q+}}cov((\boldsymbol{Z}_i'\boldsymbol{R}_i\boldsymbol{Z}_i)^{-1}\boldsymbol{Z}_i'\boldsymbol{R}_i\boldsymbol{e}_i, \boldsymbol{H}_q'\boldsymbol{M}^{-1}\boldsymbol{T}_i'\boldsymbol{Q}_i\boldsymbol{e}_i|\mathcal{C}) \\
= \quad & -\frac{M_q}{n^{3/2}}\sum_{i\in I_{q+}}(\boldsymbol{Z}_i'\boldsymbol{R}_i\boldsymbol{Z}_i)^{-1}\boldsymbol{Z}_i'\boldsymbol{R}_iVar(\boldsymbol{e}_i)\boldsymbol{Q}_i\boldsymbol{T}_i\boldsymbol{M}^{-1}\boldsymbol{H}_q \\
= \quad & -\frac{M_q}{n^{3/2}}\sum_{i\in I_{q+}}(\boldsymbol{Z}_i'\boldsymbol{R}_i\boldsymbol{Z}_i)^{-1}\boldsymbol{Z}_i'\boldsymbol{R}_i\boldsymbol{\Sigma}_e\boldsymbol{Q}_i\boldsymbol{T}_i\boldsymbol{M}^{-1}\boldsymbol{H}_q \\
\to_p \quad & -a_q\boldsymbol{J}_q\boldsymbol{M}^{-1}\boldsymbol{H}_q
\end{aligned}
$$

(i) The sixth covariance term is (the seventh covariance term is a transpose of this covariance):

$$
\begin{aligned}
& \sum_{k=1}^{q-1} cov(\frac{1}{\sqrt{n}} \sum_{i \in I_{q+}} (\boldsymbol{Z}_i' \boldsymbol{R}_i \boldsymbol{Z}_i)^{-1} \boldsymbol{Z}_i' \boldsymbol{R}_i \boldsymbol{e}_i, -\frac{m_k}{n^{3/2}} \boldsymbol{A}_k^{-1} \boldsymbol{B}_k \boldsymbol{M}^{-1} (\sum_{j=1}^{n} \boldsymbol{T}_j' \boldsymbol{Q}_j \boldsymbol{e}_j) | \mathcal{C}) \\
= & -\sum_{k=1}^{q-1} \sum_{i \in I_{q+}} \frac{m_k}{n^2} cov((\boldsymbol{Z}_i' \boldsymbol{R}_i \boldsymbol{Z}_i)^{-1} \boldsymbol{Z}_i' \boldsymbol{R}_i \boldsymbol{e}_i, \boldsymbol{A}_k^{-1} \boldsymbol{B}_k \boldsymbol{M}^{-1} \boldsymbol{T}_i' \boldsymbol{Q}_i \boldsymbol{e}_i) | \mathcal{C}) \\
= & -\sum_{k=1}^{q-1} \sum_{i \in I_{q+}} \frac{m_k}{n^2} (\boldsymbol{Z}_i' \boldsymbol{R}_i \boldsymbol{Z}_i)^{-1} \boldsymbol{Z}_i' \boldsymbol{R}_i Var(\boldsymbol{e}_i) \boldsymbol{Q}_i \boldsymbol{T}_i \boldsymbol{M}^{-1} \boldsymbol{B}_k' \boldsymbol{A}_k \\
= & -\sum_{k=1}^{q-1} \frac{m_k}{n} (\frac{1}{n} \sum_{i \in I_{q+}} (\boldsymbol{Z}_i' \boldsymbol{R}_i \boldsymbol{Z}_i)^{-1} \boldsymbol{Z}_i' \boldsymbol{R}_i \boldsymbol{\Sigma}_e \boldsymbol{Q}_i \boldsymbol{T}_i) \boldsymbol{M}^{-1} \boldsymbol{B}_k' \boldsymbol{A}_k \\
\rightarrow_p & \sum_{k=1}^{q-1} a_k \boldsymbol{J}_q \boldsymbol{M}^{-1} \boldsymbol{B}_k' \boldsymbol{A}_k^{-1}
\end{aligned}
$$

(j) The eighth covariance term is (the ninth covariance term is a transpose of this covariance):

$$
\begin{aligned}
& \sum_{k=1}^{q-1} cov(\frac{1}{\sqrt{n}} \boldsymbol{A}_k^{-1} (\sum_{i \in I_k} \boldsymbol{Z}_i' \boldsymbol{R}_i \boldsymbol{e}_i), -\frac{M_q}{n^{3/2}} \boldsymbol{H}_q' \boldsymbol{M}^{-1} (\sum_{j=1}^{n} \boldsymbol{T}_j' \boldsymbol{Q}_j \boldsymbol{e}_j) | \mathcal{C}) \\
= & -\sum_{k=1}^{q-1} \sum_{i \in I_k} \frac{M_q}{n^2} cov(\boldsymbol{A}_k^{-1} \boldsymbol{Z}_i' \boldsymbol{R}_i \boldsymbol{e}_i, \boldsymbol{H}_q' \boldsymbol{M}^{-1} \boldsymbol{T}_i' \boldsymbol{Q}_i \boldsymbol{e}_i) | \mathcal{C}) \\
= & -\sum_{k=1}^{q-1} \sum_{i \in I_k} \frac{M_q}{n^2} \boldsymbol{A}_k^{-1} \boldsymbol{Z}_i' \boldsymbol{R}_i Var(\boldsymbol{e}_i) \boldsymbol{Q}_i \boldsymbol{T}_i \boldsymbol{M}^{-1} \boldsymbol{H}_q \\
= & -\sum_{k=1}^{q-1} \frac{M_q}{n} \boldsymbol{A}_k^{-1} (\frac{1}{n} \sum_{i \in I_k} \boldsymbol{Z}_i' \boldsymbol{R}_i \boldsymbol{\Sigma}_e \boldsymbol{Q}_i \boldsymbol{T}_i) \boldsymbol{M}^{-1} \boldsymbol{H}_q \\
\rightarrow_p & -\sum_{k=1}^{q-1} a_q \boldsymbol{A}_k^{-1} \boldsymbol{F}_k \boldsymbol{M}^{-1} \boldsymbol{H}_q
\end{aligned}
$$

(k) The tenth covariance term is (the eleventh covariance term is a transpose of this covariance):

$$
\begin{aligned}
& \sum_{k=1}^{q-1} cov(-\frac{m_k}{n^{3/2}}\boldsymbol{A}_k^{-1}\boldsymbol{B}_k\boldsymbol{M}^{-1}(\sum_{i=1}^{n}\boldsymbol{T}_j'\boldsymbol{Q}_j\boldsymbol{e}_j), -\frac{M_q}{n^{3/2}}\boldsymbol{H}_q'\boldsymbol{M}^{-1}(\sum_{i=j}^{n}\boldsymbol{T}_j'\boldsymbol{Q}_j\boldsymbol{e}_j)|\mathcal{C}) \\
= & \sum_{k=1}^{q-1}\sum_{i=1}^{n}\frac{m_k M_q}{n^3} cov(\boldsymbol{A}_k^{-1}\boldsymbol{B}_k\boldsymbol{M}^{-1}\boldsymbol{T}_i'\boldsymbol{Q}_i\boldsymbol{e}_i, \boldsymbol{H}_q'\boldsymbol{M}^{-1}\boldsymbol{T}_i'\boldsymbol{Q}_i\boldsymbol{e}_i|\mathcal{C}) \\
= & \sum_{k=1}^{q-1}\sum_{i=1}^{n}\frac{m_k M_q}{n^3}\boldsymbol{A}_k^{-1}\boldsymbol{B}_k\boldsymbol{M}^{-1}\boldsymbol{T}_i'\boldsymbol{Q}_i Var(\boldsymbol{e}_i)\boldsymbol{Q}_i\boldsymbol{T}_i\boldsymbol{M}^{-1}\boldsymbol{H}_q \\
= & \sum_{k=1}^{q-1}\frac{m_k}{n}\frac{M_q}{n}\boldsymbol{A}_k^{-1}\boldsymbol{B}_k\boldsymbol{M}^{-1}(\frac{1}{n}\sum_{i=1}^{n}\boldsymbol{T}_i'\boldsymbol{Q}_i\boldsymbol{\Sigma}_e\boldsymbol{Q}_i\boldsymbol{T}_i)\boldsymbol{M}^{-1}\boldsymbol{H}_q \\
\to_p & \sum_{k=1}^{q-1} a_k a_q \boldsymbol{A}_k^{-1}\boldsymbol{B}_k\boldsymbol{M}^{-1}\boldsymbol{L}\boldsymbol{M}^{-1}\boldsymbol{H}_q
\end{aligned}
$$

□

Proof. of **Theorem 4.5**

From the proof of Theorem 4.3, we know that $E(\frac{m_k}{n}\hat{\boldsymbol{\alpha}}_{I_k}) = P(k_i = k)\cdot(\boldsymbol{\alpha} + E(\boldsymbol{b}_i|k_i = k))$ for group $k = 1, \ldots, q-1$. Exactly the same argument can prove that, for group $k = q, \ldots, K$, $E(\frac{m_k}{n}\hat{\boldsymbol{\alpha}}_{I_k}) = P(k_i = k)\cdot(\boldsymbol{\alpha} + E(\boldsymbol{b}_i|k_i = k))$. So,

$$
\begin{aligned}
E(\hat{\boldsymbol{\alpha}}_K) &= \sum_{k=1}^{K} E(\frac{m_k}{n}\hat{\boldsymbol{\alpha}}_{I_k}) \\
&= \sum_{k=1}^{K} P(k_i = k)\cdot(\boldsymbol{\alpha} + E(\boldsymbol{b}_i|k_i = k)) \\
&= \boldsymbol{\alpha} + E_{k_i}\{E(\boldsymbol{b}_i|k_i)\} \\
&= \boldsymbol{\alpha} + E(\boldsymbol{b}_i) \\
&= \boldsymbol{\alpha}
\end{aligned}
$$

is unbiased.

□

Proof. of **Theorem 4.6**

Again, with the definition of the limiting matrices, we can write the difference between estimator $\hat{\boldsymbol{\alpha}}_K$ and $\boldsymbol{\alpha}$ as

$$\begin{aligned}\sqrt{n}(\hat{\boldsymbol{\alpha}}_K - \boldsymbol{\alpha}) &= \sum_{k=1}^{K} \frac{1}{\sqrt{n}} \boldsymbol{A}_k^{-1} (\sum_{i \in I_k} \boldsymbol{Z}_i' \boldsymbol{R}_i \boldsymbol{Z}_i \boldsymbol{b}_i) + \sum_{k=1}^{K} \frac{1}{\sqrt{n}} \boldsymbol{A}_k^{-1} (\sum_{i \in I_k} \boldsymbol{Z}_i' \boldsymbol{R}_i \boldsymbol{e}_i) \\ &\quad - \sum_{k=1}^{K} \frac{m_k}{n^{3/2}} \boldsymbol{A}_k^{-1} \boldsymbol{B}_k \boldsymbol{M}^{-1} (\sum_{i=1}^{n} \boldsymbol{T}_i' \boldsymbol{Q}_i \boldsymbol{e}_i) + o_p(1), \end{aligned} \tag{4.37}$$

The right hand side is sum of averages over independent random variables, conditional on $\{\boldsymbol{R}_i, \boldsymbol{X}_i, i = 1, \ldots, n\}$ and grouping $\{\boldsymbol{I}_k, k = 1, \ldots, K\}$. The asymptotic normality follows from the central limit theorem, and the asymptotic covariance can be calculated exactly in the same way as in proof of Theorem 4.4.

□

Chapter 5

Including Time-invariant Covariates

In many studies there exist time-independent covariates, i.e., covariates whose values do not vary within each panel or subject. The intercept is a special case of time-invariant "covariate", but it has been included in model (3.2) and we already constructed estimator for the intercept. Beside this intercept term, there may be other variables which are time-invariant. In our model (2.2),

$$\boldsymbol{y}_i = \boldsymbol{Z}_i\boldsymbol{\alpha} + \boldsymbol{u}_i'\boldsymbol{\gamma}\mathbf{1} + \boldsymbol{T}_i\boldsymbol{\beta} + \boldsymbol{Z}_i\boldsymbol{b}_i + \boldsymbol{e}_i, \quad i = 1, \ldots, n,$$

we reserved the notation $\boldsymbol{u}_i$ for these time-invariant covariates, and $\boldsymbol{\gamma}$ is an s-dimensional vector of unknown parameters corresponding to $\boldsymbol{u}_i$.

Again, let us first do some modification to this model.

$$\begin{aligned} \boldsymbol{y}_i &= \boldsymbol{Z}_i\boldsymbol{\alpha} + \boldsymbol{u}_i'\boldsymbol{\gamma}\mathbf{1} + \boldsymbol{T}_i\boldsymbol{\beta} + \boldsymbol{Z}_i\boldsymbol{b}_i + \boldsymbol{e}_i, \\ &= \boldsymbol{Z}_i(\boldsymbol{\alpha} + \boldsymbol{b}_i) + \boldsymbol{u}_i'\boldsymbol{\gamma}\mathbf{1} + \boldsymbol{T}_i\boldsymbol{\beta} + \boldsymbol{e}_i. \end{aligned}$$

Since $\boldsymbol{Z}_i$ includes the intercept term, we can further split $\boldsymbol{Z}_i$ as $(\mathbf{1}, \boldsymbol{Z}_i^*)$. Similarly,

$\boldsymbol{\alpha} = (\alpha_0, \boldsymbol{\alpha}^*)$, $\boldsymbol{b}_i = (b_{i0}, \boldsymbol{b}_i^*)$. The above model becomes:

$$\begin{aligned} \boldsymbol{y}_i &= (\alpha_0 + b_{10})\mathbf{1} + \boldsymbol{Z}_i^*(\boldsymbol{\alpha}^* + \boldsymbol{b}_i^*) + \boldsymbol{u}_i'\boldsymbol{\gamma}\mathbf{1} + \boldsymbol{T}_i\boldsymbol{\beta} + \boldsymbol{e}_i, \\ &= (\alpha_0 + b_{10} + \boldsymbol{u}_i'\boldsymbol{\gamma})\mathbf{1} + \boldsymbol{Z}_i^*(\boldsymbol{\alpha}^* + \boldsymbol{b}_i^*) + \boldsymbol{T}_i\boldsymbol{\beta} + \boldsymbol{e}_i. \end{aligned} \tag{5.1}$$

With this reforming, we see that, parameter $\boldsymbol{\beta}$ can be separated from parameter $\boldsymbol{\alpha}$ and $\boldsymbol{\gamma}$, but $\boldsymbol{\alpha}$ and $\boldsymbol{\gamma}$ can no longer be divided due to the confounding intercept. So, similar to the previous two chapters, we estimate $\boldsymbol{\beta}$ and $(\boldsymbol{\alpha}, \boldsymbol{\gamma})$ separately.

5.1 Estimation of $\boldsymbol{\beta}$

In Chapter 3, we introduced a linear transformation matrix $\boldsymbol{Q}_i$, which projects $\boldsymbol{Z}_i$ to $\mathbf{0}$. Since $\mathbf{1}$ is a column included in $\boldsymbol{Z}_i$, we know that $\boldsymbol{Q}_i$ projects this vector to $\mathbf{0}$ as well. Applying $\boldsymbol{Q}_i$ to (5.1), we get $\boldsymbol{Q}_i\boldsymbol{y}_i = \boldsymbol{Q}_i\boldsymbol{T}_i\boldsymbol{\beta} + \boldsymbol{Q}_i\boldsymbol{e}_i$, which is exactly the same as (3.3) in Chapter 3.

This means the estimation for parameters in $\boldsymbol{\beta}$ stays the same as in Chapter 3. That is,

$$\hat{\boldsymbol{\beta}} = \boldsymbol{S}_n^{-1}\sum_{i=1}^{n}\boldsymbol{T}_i'\boldsymbol{Q}_i\boldsymbol{y}_i = \boldsymbol{\beta} + \boldsymbol{S}_n^{-1}\sum_{i=1}^{n}\boldsymbol{T}_i'\boldsymbol{Q}_i\boldsymbol{e}_i.$$

The unbiasedness and asymptotic normality of $\hat{\boldsymbol{\beta}}$ proved in Chapter 3 still holds, because the estimation for $\boldsymbol{\beta}$ is not affected by the introduction of time-invariant covariates $\boldsymbol{u}_i$.

5.2 Estimation of $\boldsymbol{\zeta}$

For parameters in $(\boldsymbol{\alpha}, \boldsymbol{\gamma})$, we would like to estimate them together in one process. However, if we follow the estimation process described in Chapter 4, the estimate for the intercept is actually for $\alpha_0 + \boldsymbol{u}_i'\boldsymbol{\gamma}$, from which we cannot give individual estimates for α_0 and $\boldsymbol{\gamma}$.

To solve this issue, we define a $q \times (q+s)$ matrix $\boldsymbol{U}_i$ as

$$\boldsymbol{U}_i = \begin{pmatrix} \boldsymbol{u}_i' & \\ \boldsymbol{0} & \boldsymbol{I}_q \end{pmatrix},$$

where $\boldsymbol{I}_q$ is the identity matrix of order q and $\boldsymbol{0}$ is the $(q-1) \times s$ matrix of zeros. Then, model (2.2) can be re-written as:

$$\boldsymbol{y}_i = \boldsymbol{Z}_i\boldsymbol{U}_i\boldsymbol{\zeta} + \boldsymbol{T}_i\boldsymbol{\beta} + \boldsymbol{Z}_i\boldsymbol{b}_i + \boldsymbol{e}_i, \quad i = 1, \ldots, n, \tag{5.2}$$

where $\boldsymbol{\zeta}' = (\boldsymbol{\gamma}', \boldsymbol{\alpha}')$. In this reformed model (5.2), $\boldsymbol{\gamma}$ is not nested in the intercept term anymore. What we need to do next is to build an estimator for $\boldsymbol{\zeta}$. Again, we try to construct estimators in the simple case first, and look at the more general case later on.

Simple Case

As in section 4.1, we start with the simple case, where all panels have at least q observed responses. When this is true, $\boldsymbol{Z}_i'\boldsymbol{R}_i\boldsymbol{Z}_i$ is invertible. But in model (5.2), we have an extra term $\boldsymbol{U}_i$, and the $(p+s) \times (p+s)$ matrix $(\boldsymbol{Z}_i\boldsymbol{U}_i)'\boldsymbol{R}_i(\boldsymbol{Z}_i\boldsymbol{U}_i)$ will never be invertible. So, we construct the estimator for $\boldsymbol{\zeta}'$ in a similar but slightly different

way.

First, multiplying $\boldsymbol{U}_i'(\boldsymbol{Z}_i'\boldsymbol{R}_i\boldsymbol{Z}_i)^{-1}\boldsymbol{Z}_i'\boldsymbol{R}_i$ to $\boldsymbol{y}_i - \boldsymbol{T}_i\hat{\boldsymbol{\beta}}$ we obtain that

$$\boldsymbol{U}_i'\boldsymbol{U}_i\boldsymbol{\zeta} + \boldsymbol{U}_i'\boldsymbol{b}_i + \boldsymbol{U}_i'(\boldsymbol{Z}_i'\boldsymbol{R}_i\boldsymbol{Z}_i)^{-1}\boldsymbol{Z}_i'\boldsymbol{R}_i\boldsymbol{e}_i + \boldsymbol{U}_i'(\boldsymbol{Z}_i'\boldsymbol{R}_i\boldsymbol{Z}_i)^{-1}\boldsymbol{Z}_i'\boldsymbol{R}_i\boldsymbol{T}_i\boldsymbol{S}_n^{-1}\sum_{j=1}^{n}\boldsymbol{T}_j'\boldsymbol{Q}_j\boldsymbol{e}_j.$$

The matrix

$$\boldsymbol{U}_i'\boldsymbol{U}_i = \begin{pmatrix} \boldsymbol{u}_i\boldsymbol{u}_i' & \boldsymbol{u}_i & \boldsymbol{0}' \\ \boldsymbol{u}_i' & 1 & 0 \\ \boldsymbol{0} & 0 & \boldsymbol{I}_{q-1} \end{pmatrix}$$

may not be invertable for any i, but usually $\boldsymbol{W}_n = \sum_{i=1}^{n}\boldsymbol{U}_i'\boldsymbol{U}_i$ is invertable for sufficiently large n. Then, an unbiased estimator of $\boldsymbol{\zeta}$ is

$$\begin{aligned} \tilde{\boldsymbol{\zeta}} &= \boldsymbol{W}_n^{-1}\sum_{i=1}^{n}\boldsymbol{U}_i'(\boldsymbol{Z}_i'\boldsymbol{R}_i\boldsymbol{Z}_i)^{-1}\boldsymbol{Z}_i'\boldsymbol{R}_i(\boldsymbol{y}_i - \boldsymbol{T}_i\hat{\boldsymbol{\beta}}) \\ &= \boldsymbol{\zeta} + \boldsymbol{W}_n^{-1}\sum_{i=1}^{n}\boldsymbol{U}_i'\boldsymbol{b}_i + \boldsymbol{W}_n^{-1}\sum_{i=1}^{n}\boldsymbol{U}_i'(\boldsymbol{Z}_i'\boldsymbol{R}_i\boldsymbol{Z}_i)^{-1}\boldsymbol{Z}_i'\boldsymbol{R}_i\boldsymbol{e}_i \\ &\quad -\boldsymbol{W}_n^{-1}\sum_{i=1}^{n}\boldsymbol{U}_i'(\boldsymbol{Z}_i'\boldsymbol{R}_i\boldsymbol{Z}_i)^{-1}\boldsymbol{Z}_i'\boldsymbol{R}_i\boldsymbol{T}_i\boldsymbol{S}_n^{-1}\sum_{j=1}^{n}\boldsymbol{T}_j'\boldsymbol{Q}_j\boldsymbol{e}_j. \end{aligned} \tag{5.3}$$

This estimator is very similar to the estimator $\tilde{\boldsymbol{\alpha}}$ for $\boldsymbol{\alpha}$. The only difference is that matrix $\boldsymbol{U}_i'$ is multiplied to the left of every term, and another matrix $\boldsymbol{W}_n$ is applied to cancel those extra $\boldsymbol{U}_i$'s.

Similarly, we can prove the unbiasedness and asymptotic normality of $\tilde{\boldsymbol{\zeta}}$.

Theorem 5.1. *Estimator $\tilde{\boldsymbol{\zeta}}$ is unbiased. As $n \to \infty$, assume the following besides* (3.9) *and* (3.11),

$$\sum_{i=1}^{n}\boldsymbol{T}_i'\boldsymbol{R}_i\boldsymbol{Z}_i(\boldsymbol{Z}_i'\boldsymbol{R}_i\boldsymbol{Z}_i)^{-1}\boldsymbol{U}_i\boldsymbol{W}_n^{-1} \to_p \boldsymbol{H}, \tag{5.4}$$

$$\sum_{i=1}^{n} \boldsymbol{T}_i' \boldsymbol{Q}_i \boldsymbol{\Sigma}_e \boldsymbol{R}_i \boldsymbol{Z}_i (\boldsymbol{Z}_i' \boldsymbol{R}_i \boldsymbol{Z}_i)^{-1} \boldsymbol{U}_i \boldsymbol{W}_n^{-1} \rightarrow_p \boldsymbol{J}, \tag{5.5}$$

$$n\boldsymbol{W}_n^{-1} \sum_{i=1}^{n} \boldsymbol{U}_i (\boldsymbol{Z}_i' \boldsymbol{R}_i \boldsymbol{Z}_i)^{-1} \boldsymbol{Z}_i' \boldsymbol{R}_i \boldsymbol{\Sigma}_e \boldsymbol{R}_i \boldsymbol{Z}_i (\boldsymbol{Z}_i' \boldsymbol{R}_i \boldsymbol{Z}_i)^{-1} \boldsymbol{U}_i \boldsymbol{W}_n^{-1} \rightarrow_p \boldsymbol{K}_n. \tag{5.6}$$

$$n\boldsymbol{W}_n^{-1} \sum_{i=1}^{n} \boldsymbol{U}_i' \boldsymbol{\Sigma}_b \boldsymbol{U}_i \boldsymbol{W}_n^{-1} \rightarrow_p \tilde{\boldsymbol{\Sigma}}_b \tag{5.7}$$

According to the central limit theorem,

$$\sqrt{n}(\tilde{\boldsymbol{\zeta}} - \boldsymbol{\zeta}) \rightarrow_p N(0, \boldsymbol{\Sigma}_{\tilde{\zeta}}), \tag{5.8}$$

where

$$\boldsymbol{\Sigma}_{\tilde{\zeta}} = \tilde{\boldsymbol{\Sigma}}_b + \boldsymbol{K}_n + \boldsymbol{H}_n' \boldsymbol{M}^{-1} \boldsymbol{L} \boldsymbol{M}^{-1} \boldsymbol{H}_n - \boldsymbol{H}_n' \boldsymbol{M}^{-1} \boldsymbol{J}_n - \boldsymbol{J}_n' \boldsymbol{M}^{-1} \boldsymbol{H}_n.$$

A consistent estimator for the covariance matrix $\boldsymbol{\Sigma}_{\tilde{\zeta}}$ is

$$\widehat{\boldsymbol{\Sigma}_b + \boldsymbol{K}}_n + \hat{\boldsymbol{H}}_n' \hat{\boldsymbol{M}}^{-1} \hat{\boldsymbol{L}} \hat{\boldsymbol{M}}^{-1} \hat{\boldsymbol{H}}_n - \hat{\boldsymbol{H}}_n' \hat{\boldsymbol{M}}^{-1} \hat{\boldsymbol{J}}_n - \hat{\boldsymbol{J}}_n' \hat{\boldsymbol{M}}^{-1} \hat{\boldsymbol{H}}_n, \tag{5.9}$$

where

$$\hat{\boldsymbol{H}}_n = \sum_{i=1}^{n} \boldsymbol{T}_i' \boldsymbol{R}_i \boldsymbol{Z}_i (\boldsymbol{Z}_i' \boldsymbol{R}_i \boldsymbol{Z}_i)^{-1} \boldsymbol{U}_i \boldsymbol{W}_n^{-1}, \tag{5.10}$$

$$\hat{\boldsymbol{J}}_n = \sum_{i=1}^{n} \boldsymbol{T}_i' \boldsymbol{Q}_i (\boldsymbol{y}_i - \boldsymbol{T}_i \hat{\boldsymbol{\beta}})(\boldsymbol{y}_i - \boldsymbol{T}_i \hat{\boldsymbol{\beta}})' \boldsymbol{R}_i \boldsymbol{Z}_i (\boldsymbol{Z}_i' \boldsymbol{R}_i \boldsymbol{Z}_i)^{-1} \boldsymbol{U}_i \boldsymbol{W}_n^{-1}, \tag{5.11}$$

$$\widehat{\boldsymbol{\Sigma}_b + \boldsymbol{K}}_n = n\boldsymbol{W}_n^{-1} \sum_{i=1}^{n} \boldsymbol{U}_i (\boldsymbol{Z}_i' \boldsymbol{R}_i \boldsymbol{Z}_i)^{-1} \boldsymbol{Z}_i' \boldsymbol{R}_i (\boldsymbol{y}_i - \boldsymbol{X}_i \hat{\boldsymbol{\theta}})(\boldsymbol{y}_i - \boldsymbol{X}_i \hat{\boldsymbol{\theta}})' \boldsymbol{R}_i \boldsymbol{Z}_i (\boldsymbol{Z}_i' \boldsymbol{R}_i \boldsymbol{Z}_i)^{-1} \boldsymbol{U}_i \boldsymbol{W}_n^{-1}, \tag{5.12}$$

are consistent estimators for $\boldsymbol{H}_n$, $\boldsymbol{J}_n$ and $\boldsymbol{\Sigma}_b + \boldsymbol{K}_n$, respectively.

General Case

In the general situation, where the observed number of responses k_i may be smaller than q for some panels, the non-invertibility problem shows up. So, we apply the grouping method again in this section to construct an estimator.

Let $\delta_{i,k}$ and $\delta_{i,q+}$, $i = 1, ..., n$ and $k = 1, ..., K$, be the same as in Chapter 4. In q-grouping, the groups are $I_k, k = 1, ..., q-1$ and I_{q+}; in K-grouping, the groups are $I_k, k = 1, ..., K$.

For group I_{q+}, estimator $\hat{\boldsymbol{\zeta}}_{I_{q+}}$ can be defined similarly as in the simple case:

$$\begin{aligned}
\hat{\boldsymbol{\zeta}}_{I_{q+}} &= \boldsymbol{W}_n^{-1} \sum_{i \in I_{q+}} \boldsymbol{U}_i' (\boldsymbol{Z}_i' \boldsymbol{R}_i \boldsymbol{Z}_i)^{-1} \boldsymbol{Z}_i' \boldsymbol{R}_i (\boldsymbol{y}_i - \boldsymbol{T}_i \hat{\boldsymbol{\beta}}) \\
&= \boldsymbol{\zeta} + \boldsymbol{W}_n^{-1} \sum_{i \in I_{q+}} \boldsymbol{U}_i' \boldsymbol{b}_i + \boldsymbol{W}_n^{-1} \sum_{i \in I_{q+}} \boldsymbol{U}_i' (\boldsymbol{Z}_i' \boldsymbol{R}_i \boldsymbol{Z}_i)^{-1} \boldsymbol{Z}_i' \boldsymbol{R}_i \boldsymbol{e}_i \\
&\quad - \boldsymbol{W}_n^{-1} \sum_{i \in I_{q+}} \boldsymbol{U}_i' (\boldsymbol{Z}_i' \boldsymbol{R}_i \boldsymbol{Z}_i)^{-1} \boldsymbol{Z}_i' \boldsymbol{R}_i \boldsymbol{T}_i \boldsymbol{S}_n^{-1} \sum_{j=1}^{n} \boldsymbol{T}_j' \boldsymbol{Q}_j \boldsymbol{e}_j,
\end{aligned}$$

where $\boldsymbol{W}_n = \sum_{i \in I_{q+}} \boldsymbol{U}_i' \boldsymbol{U}_i$.

For group I_k, we construct a group estimator in the same way as for $\boldsymbol{\alpha}$ in section 4. The only difference is that, we treat the product $\boldsymbol{Z}_i \boldsymbol{U}_i$ here as the $\boldsymbol{Z}_i$ in section 4.

$$\begin{aligned}
\hat{\boldsymbol{\zeta}}_{I_k} &= (\sum_{i \in \mathbf{1}_k} \boldsymbol{U}_i' \boldsymbol{Z}_i' \boldsymbol{R}_i \boldsymbol{Z}_i \boldsymbol{U}_i)^{-1} (\sum_{i \in \mathbf{1}_k} \boldsymbol{U}_i' \boldsymbol{Z}_i' \boldsymbol{R}_i (\boldsymbol{y}_i - \boldsymbol{T}_i' \hat{\boldsymbol{\beta}})) \\
&= \boldsymbol{\zeta} + (\sum_{i \in I_k} \boldsymbol{U}_i' \boldsymbol{Z}_i' \boldsymbol{R}_i \boldsymbol{Z}_i \boldsymbol{U}_i)^{-1} (\sum_{i \in I_k} \boldsymbol{U}_i' \boldsymbol{Z}_i' \boldsymbol{R}_i \boldsymbol{Z}_i \boldsymbol{b}_i) \\
&\quad + (\sum_{i \in I_k} \boldsymbol{U}_i' \boldsymbol{Z}_i' \boldsymbol{R}_i \boldsymbol{Z}_i \boldsymbol{U}_i)^{-1} (\sum_{i \in I_k} \boldsymbol{U}_i' \boldsymbol{Z}_i' \boldsymbol{R}_i \boldsymbol{T}_i (\beta - \hat{\beta})) \\
&\quad + (\sum_{i \in I_k} \boldsymbol{U}_i' \boldsymbol{Z}_i' \boldsymbol{R}_i \boldsymbol{Z}_i \boldsymbol{U}_i)^{-1} (\sum_{i \in I_k} \boldsymbol{U}_i' \boldsymbol{Z}_i' \boldsymbol{R}_i \boldsymbol{e}_i).
\end{aligned}$$

Then,

(1) the estimator under q-grouping is:

$$\hat{\boldsymbol{\zeta}}_q = \sum_{i=1}^{q-1} \frac{m_k}{n} \hat{\boldsymbol{\zeta}}_{I_k} + \frac{M_q}{n} \hat{\boldsymbol{\zeta}}_{I_{q+}}, \tag{5.13}$$

(2) the estimator under K-grouping is:

$$\hat{\boldsymbol{\zeta}}_K = \sum_{i=1}^{K} \frac{m_k}{n} \hat{\boldsymbol{\zeta}}_{I_k}. \tag{5.14}$$

Theorem 5.2. *The unbiasedness and asymptotic normality of $\hat{\boldsymbol{\zeta}}_q$ and $\hat{\boldsymbol{\zeta}}_K$ still hold, with the limits of* (4.17) *to* (4.27) *be re-defined as:*

$$\frac{1}{m_k} \sum_{I_k} \boldsymbol{U}_i' \boldsymbol{Z}_i' \boldsymbol{R}_i \boldsymbol{Z}_i \boldsymbol{U}_i \to_p \boldsymbol{A}_k, \quad k = 1, \ldots, q-1,$$

$$\frac{1}{m_k} \sum_{I_k} \boldsymbol{U}_i' \boldsymbol{Z}_i' \boldsymbol{R}_i \boldsymbol{T}_i \to_p \boldsymbol{B}_k, \quad k = 1, \ldots, q-1,$$

$$\frac{1}{n} \sum_{I_k} \boldsymbol{U}_i' \boldsymbol{Z}_i' \boldsymbol{R}_i \boldsymbol{Z}_i E(\boldsymbol{b}_i \boldsymbol{b}_i' | \delta_{i,k}) \boldsymbol{Z}_i' \boldsymbol{R}_i \boldsymbol{Z}_i \boldsymbol{U}_i \to_p \boldsymbol{C}_k, \quad k = 1, \ldots, q-1,$$

$$\frac{1}{n} \sum_{I_k} \boldsymbol{U}_i' \boldsymbol{Z}_i' \boldsymbol{R}_i \boldsymbol{Z}_i E(\boldsymbol{b}_i | \delta_{i,k}) E(\boldsymbol{b}_i | \delta_{i,k})' \boldsymbol{Z}_i' \boldsymbol{R}_i \boldsymbol{Z}_i \boldsymbol{U}_i \to_p \boldsymbol{C}_{0k}, \quad k = 1, \ldots, q-1,$$

$$\frac{1}{n} \sum_{I_k} \boldsymbol{U}_i' \boldsymbol{Z}_i' \boldsymbol{R}_i \boldsymbol{\Sigma}_e \boldsymbol{R}_i \boldsymbol{Z}_i \boldsymbol{U}_i \to_p \boldsymbol{D}_k, \quad k = 1, \ldots, q-1,$$

$$\frac{1}{n} \sum_{I_k} \boldsymbol{U}_i' \boldsymbol{Z}_i' \boldsymbol{R}_i \boldsymbol{\Sigma}_e \boldsymbol{Q}_i \boldsymbol{T}_i \to_p \boldsymbol{F}_k, \quad k = 1, \ldots, q-1,$$

$$\frac{1}{n} M_q \boldsymbol{W}_n^{-1} \sum_{I_{q+}} \boldsymbol{U}_i' E(\boldsymbol{b}_i \boldsymbol{b}_i' | \delta_{i,q+}) \boldsymbol{U}_i \boldsymbol{W}_n^{-1} M_q \to_p \boldsymbol{G}_q,$$

$$\frac{1}{n} M_q \boldsymbol{W}_n^{-1} \sum_{I_{q+}} \boldsymbol{U}_i' E(\boldsymbol{b}_i | \delta_{i,q+}) E(\boldsymbol{b}_i | \delta_{i,q+})' \boldsymbol{U}_i \boldsymbol{W}_n^{-1} M_q \to_p \boldsymbol{G}_{0q},$$

$$\sum_{I_{q+}} \boldsymbol{T}_i' \boldsymbol{R}_i \boldsymbol{Z}_i (\boldsymbol{Z}_i' \boldsymbol{R}_i \boldsymbol{Z}_i)^{-1} \boldsymbol{U}_i' \boldsymbol{W}_n^{-1} \to_p \boldsymbol{H}_q,$$

$$\sum_{I_{q+}} \boldsymbol{T}_i' \boldsymbol{Q}_i \boldsymbol{\Sigma}_e \boldsymbol{R}_i \boldsymbol{Z}_i (\boldsymbol{Z}_i' \boldsymbol{R}_i \boldsymbol{Z}_i)^{-1} \boldsymbol{U}_i' \boldsymbol{W}_n^{-1} \to_p \boldsymbol{J}_q,$$

$$\frac{1}{n} M_q \boldsymbol{W}_n^{-1} \sum_{I_{q+}} (\boldsymbol{Z}_i' \boldsymbol{R}_i \boldsymbol{Z}_i)^{-1} \boldsymbol{Z}_i' \boldsymbol{R}_i \boldsymbol{\Sigma}_e \boldsymbol{R}_i \boldsymbol{Z}_i (\boldsymbol{Z}_i' \boldsymbol{R}_i \boldsymbol{Z}_i)^{-1} \boldsymbol{W}_n^{-1} M_q \to_p \boldsymbol{K}_q.$$

Again, the form of the asymptotic covariance matrix is very complicated, and an easy consistent estimator could not be constructed due to the term $E(\boldsymbol{b}_i|\delta_i)E(\boldsymbol{b}_i|\delta_i)'$. We will apply bootstrap when we need a consistent estimator of the variances

5.3 Proof

Proof. of **Theorem 5.1**

To calculate the expectation of estimator $\boldsymbol{\zeta}$, we only need to do this for the second term $\boldsymbol{W}_n^{-1} \sum_{i=1}^n \boldsymbol{U}_i' \boldsymbol{b}_i$ in (5.13).

$$\begin{aligned} E(\boldsymbol{W}_n^{-1} \sum_{i=1}^n \boldsymbol{U}_i' \boldsymbol{b}_i) &= E_{\boldsymbol{U}}\{\boldsymbol{W}_n^{-1} \sum_{i=1}^n \boldsymbol{U}_i' E(\boldsymbol{b}_i|\boldsymbol{U}_i)\} \\ &= E_{\boldsymbol{U}}\{\boldsymbol{W}_n^{-1} \sum_{i=1}^n \boldsymbol{U}_i' E(\boldsymbol{b}_i)\} \\ &= E_{\boldsymbol{U}}\{\boldsymbol{W}_n^{-1} \sum_{i=1}^n \boldsymbol{U}_i' * \boldsymbol{0}\} \\ &= 0 \end{aligned}$$

Then, we can conclude that $\tilde{\boldsymbol{\zeta}}$ is an unbiased estimator of $\boldsymbol{\zeta}$.

With the definition of the limiting matrices, we can get

$$\sqrt{n}(\tilde{\boldsymbol{\zeta}} - \boldsymbol{\zeta}) = \sqrt{n}\boldsymbol{W}_n^{-1} \sum_{i=1}^n \boldsymbol{U}_i' \boldsymbol{b}_i + \sqrt{n}\boldsymbol{W}_n^{-1} \sum_{i=1}^n \boldsymbol{U}_i' (\boldsymbol{Z}_i' \boldsymbol{R}_i \boldsymbol{Z}_i)^{-1} \boldsymbol{Z}_i' \boldsymbol{R}_i \boldsymbol{e}_i - \frac{1}{\sqrt{n}} \sum_{i=1}^n \boldsymbol{H}_n' \boldsymbol{M}^{-1} \boldsymbol{T}_i' \boldsymbol{Q}_i \boldsymbol{e}_i + o_p(1). \tag{5.15}$$

Again, the right hand side is a sum of averages of independent random variables, conditional on $\{\boldsymbol{R}_i, \boldsymbol{X}_i, i = 1, \ldots, n\}$. The asymptotical normality follows directly from the central limit theorem. The asymptotic covariance matrix can be derived by a direct calculation of the conditional variances of the right hand side of (5.15), conditional on $\mathcal{C} = \{\boldsymbol{R}_i, \boldsymbol{X}_i, i = 1, \ldots, n\}$.

□

Proof. of **Theorem 5.2**

(1) Unbiasedness:

(1.1) Let $\boldsymbol{V}_i = \boldsymbol{Z}_i\boldsymbol{U}_i$ and $\boldsymbol{c}_i = \begin{pmatrix} \mathbf{0} \\ \boldsymbol{b}_i \end{pmatrix}$, where $\mathbf{0}$ is s-dimensional. Then,

$$\begin{aligned}
\frac{m_k}{n}\hat{\boldsymbol{\zeta}}_{I_k} &= \frac{m_k}{n}\boldsymbol{\zeta} + \frac{m_k}{n}(\sum_{i\in I_k}\boldsymbol{U}_i'\boldsymbol{Z}_i'\boldsymbol{R}_i\boldsymbol{Z}_i\boldsymbol{U}_i)^{-1}(\sum_{i\in I_k}\boldsymbol{U}_i'\boldsymbol{Z}_i'\boldsymbol{R}_i\boldsymbol{Z}_i\boldsymbol{b}_i) \\
&\quad -\frac{m_k}{n}(\sum_{i\in I_k}\boldsymbol{U}_i'\boldsymbol{Z}_i'\boldsymbol{R}_i\boldsymbol{Z}_i\boldsymbol{U}_i)^{-1}(\sum_{i\in I_k}\boldsymbol{U}_i'\boldsymbol{Z}_i'\boldsymbol{R}_i\boldsymbol{T}_i\boldsymbol{S}_n^{-1}\sum_{j=1}^{n}\boldsymbol{T}_j'\boldsymbol{Q}_j\boldsymbol{e}_j) \\
&\quad +\frac{m_k}{n}(\sum_{i\in I_k}\boldsymbol{U}_i'\boldsymbol{Z}_i'\boldsymbol{R}_i\boldsymbol{Z}_i\boldsymbol{U}_i)^{-1}(\sum_{i\in I_k}\boldsymbol{U}_i'\boldsymbol{Z}_i'\boldsymbol{R}_i\boldsymbol{e}_i) \\
&= \frac{m_k}{n}\boldsymbol{\zeta} + \frac{m_k}{n}(\sum_{i\in I_k}\boldsymbol{V}_i'\boldsymbol{R}_i\boldsymbol{V}_i)^{-1}(\sum_{i\in I_k}\boldsymbol{V}_i'\boldsymbol{R}_i\boldsymbol{Z}_i\boldsymbol{b}_i) \\
&\quad -\frac{m_k}{n}(\sum_{i\in I_k}\boldsymbol{V}_i'\boldsymbol{R}_i\boldsymbol{V}_i)^{-1}(\sum_{i\in I_k}\boldsymbol{V}_i'\boldsymbol{R}_i\boldsymbol{T}_i\boldsymbol{S}_n^{-1}\sum_{j=1}^{n}\boldsymbol{T}_j'\boldsymbol{Q}_j\boldsymbol{e}_j) \\
&\quad +\frac{m_k}{n}(\sum_{i\in I_k}\boldsymbol{V}_i'\boldsymbol{R}_i\boldsymbol{V}_i)^{-1}(\sum_{i\in I_k}\boldsymbol{V}_i'\boldsymbol{R}_i\boldsymbol{e}_i)
\end{aligned}$$

Treating $\boldsymbol{V}_i$ and $\boldsymbol{c}_i$ as $\boldsymbol{Z}_i$ and $\boldsymbol{b}_i$, respectively, in proof of Theorem 4.3, the same proof gives:

$$E(\frac{m_k}{n}\hat{\boldsymbol{\zeta}}_{I_k}) = p(k_i = k)\cdot(\boldsymbol{\zeta} + E(\boldsymbol{c}_i|k_i = k)).$$

(1.2)

$$\begin{aligned}\frac{M_q}{n}\hat{\boldsymbol{\zeta}}_{I_{q+}} &= \frac{M_q}{n}\boldsymbol{\zeta} + \frac{M_q}{n}\boldsymbol{W}_n^{-1}\sum_{i\in I_{q+}}\boldsymbol{U}_i'\boldsymbol{b}_i + \frac{M_q}{n}\boldsymbol{W}_n^{-1}\sum_{i\in I_{q+}}\boldsymbol{U}_i'(\boldsymbol{Z}_i'\boldsymbol{R}_i\boldsymbol{Z}_i)^{-1}\boldsymbol{Z}_i'\boldsymbol{R}_i\boldsymbol{e}_i \\ &\quad -\frac{M_q}{n}\boldsymbol{W}_n^{-1}\sum_{i\in I_{q+}}\boldsymbol{U}_i'(\boldsymbol{Z}_i'\boldsymbol{R}_i\boldsymbol{Z}_i)^{-1}\boldsymbol{Z}_i'\boldsymbol{R}_i\boldsymbol{T}_i\boldsymbol{S}_n^{-1}\sum_{j=1}^{n}\boldsymbol{T}_j'\boldsymbol{Q}_j\boldsymbol{e}_j\end{aligned}$$

Because the expectation of the first term is $p(k_i \geq q)\boldsymbol{\zeta}$ and the expectations of the last two terms are $\mathbf{0}$, we only calculate the expectation of the second term.

$$\begin{aligned}E(\frac{M_q}{n}\boldsymbol{W}_n^{-1}\sum_{i\in I_{q+}}\boldsymbol{U}_i'\boldsymbol{b}_i) &= E(\frac{M_q}{n}\boldsymbol{W}_n^{-1}\sum_{i\in I_{q+}}\boldsymbol{U}_i'\boldsymbol{U}_i\boldsymbol{c}_i) \\ &= E_{\boldsymbol{U},\boldsymbol{\delta}_{q+}}\{\frac{M_q}{n}\boldsymbol{W}_n^{-1}\sum_{i\in I_{q+}}\boldsymbol{U}_i'\boldsymbol{U}_i E(\boldsymbol{c}_i|\boldsymbol{U}_i,\delta_{i,q+})\} \\ &= E_{\boldsymbol{U},\boldsymbol{\delta}_{q+}}\{\frac{M_q}{n}\boldsymbol{W}_n^{-1}\sum_{i\in I_{q+}}\boldsymbol{U}_i'\boldsymbol{U}_i E(\boldsymbol{c}_i|\delta_{i,q+})\} \\ &= E_{\boldsymbol{U},\boldsymbol{\delta}_{q+}}\{\frac{M_q}{n}\boldsymbol{W}_n^{-1}(\sum_{i\in I_{q+}}\boldsymbol{U}_i'\boldsymbol{U}_i)E(\boldsymbol{c}_1|\delta_{1,q+})\} \\ &= E_{\boldsymbol{U},\boldsymbol{\delta}_{q+}}\{\frac{\sum_{i=1}^{n}\delta_{i,q+}}{n}E(\boldsymbol{c}_i|\delta_{i,q+})\} \\ &= \frac{1}{n}\sum_{i=1}^{n}E_{\delta_{i,q+}}\{\delta_{i,q+}E(\boldsymbol{c}_i|\delta_{i,q+})\} \\ &= \frac{1}{n}\sum_{i=1}^{n}P(\delta_{i,q+}=1)E(\boldsymbol{c}_i|\delta_{i,q+}=1) \\ &= P(k_i \geq q)E(\boldsymbol{c}_i|k_i \geq q),\end{aligned}$$

where the third equality follows from our assumption that missingness probability does not depend on $\boldsymbol{u}_i$.

So,

$$E(\frac{M_q}{n}\hat{\boldsymbol{\zeta}}_{I_{q+}}) = P(k_i \geq q) \cdot (\boldsymbol{\zeta} + E(\boldsymbol{c}_i|k_i \geq q)).$$

Based on the above results, it follows:

$$E(\hat{\boldsymbol{\zeta}}_q) = \sum_{k=1}^{q-1} P(k_i = k) \cdot (\boldsymbol{\zeta} + E(\boldsymbol{c}_i|k_i = k)) + P(k_i \geq q) \cdot (\boldsymbol{\zeta} + E(\boldsymbol{c}_i|k_i \geq q)) = \boldsymbol{\zeta},$$

$$E(\hat{\boldsymbol{\zeta}}_K) = \sum_{k=1}^{K} P(k_i = i) \cdot (\boldsymbol{\zeta} + E(\boldsymbol{c}_i|k_i = k)) = \boldsymbol{\zeta}.$$

(2) Asymptotic normality:

With the definition of the limiting matrices,

(2.1)

$$\begin{aligned}
\frac{m_k}{n}\hat{\boldsymbol{\zeta}}_{I_k} &= \frac{m_k}{n}\boldsymbol{\zeta} + \frac{m_k}{n}(\sum_{i\in I_k} \boldsymbol{V}_i'\boldsymbol{R}_i\boldsymbol{V}_i)^{-1}(\sum_{i\in I_k} \boldsymbol{V}_i'\boldsymbol{R}_i\boldsymbol{V}_i\boldsymbol{c}_i) \\
&\quad + \frac{m_k}{n}(\sum_{i\in I_k} \boldsymbol{V}_i'\boldsymbol{R}_i\boldsymbol{V}_i)^{-1}(\sum_{i\in I_k} \boldsymbol{V}_i'\boldsymbol{R}_i\boldsymbol{e}_i) \\
&\quad - \frac{m_k}{n}(\sum_{i\in I_k} \boldsymbol{V}_i'\boldsymbol{R}_i\boldsymbol{V}_i)^{-1}(\sum_{i\in I_k} \boldsymbol{V}_i'\boldsymbol{R}_i\boldsymbol{T}_i\boldsymbol{S}_n^{-1}(\sum_{j=1}^{n})\boldsymbol{T}_j'\boldsymbol{Q}_j\boldsymbol{e}_j) \\
&= \frac{m_k}{n}\boldsymbol{\zeta} + \frac{1}{n}\boldsymbol{A}_k^{-1}(\sum_{i\in I_k} \boldsymbol{V}_i'\boldsymbol{R}_i\boldsymbol{V}_i\boldsymbol{c}_i) + \frac{1}{n}\boldsymbol{A}_k^{-1}(\sum_{i\in I_k} \boldsymbol{V}_i'\boldsymbol{R}_i\boldsymbol{e}_i) \\
&\quad - \frac{m_k}{n^2}\boldsymbol{A}_k^{-1}\boldsymbol{B}_k\boldsymbol{M}^{-1}(\sum_{j=1}^{n})\boldsymbol{T}_j'\boldsymbol{Q}_j\boldsymbol{e}_j) + o_p(1),
\end{aligned}$$

(2.2)

$$\begin{aligned}
\frac{M_q}{n}\hat{\boldsymbol{\zeta}}_{I_{q+}} &= \frac{M_q}{n}\boldsymbol{\zeta} + \frac{M_q}{n}\boldsymbol{W}_n^{-1}\sum_{i\in I_{q+}}\boldsymbol{U}_i'\boldsymbol{b}_i \\
&\quad + \frac{M_q}{n}\boldsymbol{W}_n^{-1}\sum_{i\in I_{q+}}\boldsymbol{U}_i'(\boldsymbol{Z}_i'\boldsymbol{R}_i\boldsymbol{Z}_i)^{-1}\boldsymbol{Z}_i'\boldsymbol{R}_i\boldsymbol{e}_i \\
&\quad - \frac{M_q}{n}\boldsymbol{W}_n^{-1}\sum_{i\in I_{q+}}\boldsymbol{U}_i'(\boldsymbol{Z}_i'\boldsymbol{R}_i\boldsymbol{Z}_i)^{-1}\boldsymbol{Z}_i'\boldsymbol{R}_i\boldsymbol{T}_i\boldsymbol{S}_n^{-1}(\sum_{j=1}^{n}\boldsymbol{T}_j'\boldsymbol{Q}_j\boldsymbol{e}_j) \\
&= \frac{M_q}{n}\boldsymbol{\zeta} + \frac{M_q}{n}\boldsymbol{W}_n^{-1}\sum_{i\in I_{q+}}\boldsymbol{U}_i'\boldsymbol{b}_i \\
&\quad + \frac{M_q}{n}\boldsymbol{W}_n^{-1}\sum_{i\in I_{q+}}\boldsymbol{U}_i'(\boldsymbol{Z}_i'\boldsymbol{R}_i\boldsymbol{Z}_i)^{-1}\boldsymbol{Z}_i'\boldsymbol{R}_i\boldsymbol{e}_i \\
&\quad - \frac{M_q}{n^2}\boldsymbol{H}_q'\boldsymbol{W}_n^{-1}(\sum_{j=1}^{n}\boldsymbol{T}_j'\boldsymbol{Q}_j\boldsymbol{e}_j) + o_p(1),
\end{aligned}$$

So, the difference between q-grouping estimator and the true value $\boldsymbol{\zeta}$ multiplied by $\sqrt{n}$ is

$$\begin{aligned}
\sqrt{n}(\hat{\boldsymbol{\zeta}}_q - \boldsymbol{\zeta}) &= \sum_{k=1}^{q-1}\frac{1}{\sqrt{n}}\boldsymbol{A}_k^{-1}(\sum_{i\in I_k}\boldsymbol{V}_i'\boldsymbol{R}_i\boldsymbol{V}_i\boldsymbol{c}_i) + \sum_{k=1}^{q-1}\frac{1}{\sqrt{n}}\boldsymbol{A}_k^{-1}(\sum_{i\in I_k}\boldsymbol{V}_i'\boldsymbol{R}_i\boldsymbol{e}_i) \\
&\quad - \sum_{k=1}^{q-1}\frac{m_k}{n^{3/2}}\boldsymbol{A}_k^{-1}\boldsymbol{B}_k\boldsymbol{M}^{-1}(\sum_{j=1}^{n}\boldsymbol{T}_j'\boldsymbol{Q}_j\boldsymbol{e}_j) + \frac{M_q}{\sqrt{n}}\boldsymbol{W}_n^{-1}\sum_{i\in I_{q+}}\boldsymbol{U}_i'\boldsymbol{b}_i \\
&\quad \frac{M_q}{\sqrt{n}}\boldsymbol{W}_n^{-1}\sum_{i\in I_{q+}}\boldsymbol{U}_i'(\boldsymbol{Z}_i'\boldsymbol{R}_i\boldsymbol{Z}_i)^{-1}\boldsymbol{Z}_i'\boldsymbol{R}_i\boldsymbol{e}_i - \frac{M_q}{n^{3/2}}\boldsymbol{H}_q'\boldsymbol{M}^{-1}(\sum_{j=1}^{n}\boldsymbol{T}_j'\boldsymbol{Q}_j\boldsymbol{e}_j) + o_p(1),
\end{aligned}$$

and the difference between K-grouping estimator and the true value multiplied by $\sqrt{n}$ is

$$\begin{aligned}\sqrt{n}(\hat{\boldsymbol{\zeta}}_K - \boldsymbol{\zeta}) &= \sum_{k=1}^{K} \frac{1}{\sqrt{n}} \boldsymbol{A}_k^{-1} (\sum_{i \in I_k} \boldsymbol{V}_i' \boldsymbol{R}_i \boldsymbol{V}_i \boldsymbol{c}_i) + \sum_{k=1}^{K} \frac{1}{\sqrt{n}} \boldsymbol{A}_k^{-1} (\sum_{i \in I_k} \boldsymbol{V}_i' \boldsymbol{R}_i \boldsymbol{e}_i) \\ &\quad - \sum_{k=1}^{K} \frac{m_k}{n^{3/2}} \boldsymbol{A}_k^{-1} \boldsymbol{B}_k \boldsymbol{M}^{-1} (\sum_{j=1}^{n} \boldsymbol{T}_j' \boldsymbol{Q}_j \boldsymbol{e}_j) + o_p(1).\end{aligned}$$

Conditional on covariates and grouping indicators, the right hand side of the differences are sum of averages of independent random variables. So, the central limit theorems can be applied and the asymptotic normality follows. Similar to proof of Theorem 4.4, the asymptotic covariance matrix can be calculated directly.

□

Chapter 6

Simulation

In this chapter, we will test the performance of our method through simulation. When estimating the parameters in α and γ, if there are enough observations for each panel, a simple form of estimator was proposed (simple); while if some panels do not have enough observations, we utilized the grouping technique (general) to construct an unbiased estimator. In our simulation studies, we did simulations under both simple and general scenarios.

In the simulation studies, our response model is a linear mixed-effect model and our missingness model is a logistic model given covariates and random effects. Under each scenario, we also considered two situations: MAR and informative missingness. We let the missing probability depend on covariate only so that the missingness is MAR. On the other hand, for informative missingness, we allowed the missing probability to depend on random effect(s).

To see if our method is advantageous over other existing methods, we included OLS method, REML method and ACM method in the simulation for comparison. As mentioned in the introduction, REML is a standard method for analyzing panel

data even when the panels are unbalanced. This method is very popular because it is available in many statistics softwares and is very simple to use. However, the method was not designed for missing data. It may produce misleading result under some kind of missing mechanisms.

The missingness we are focusing on here is informative missing, where missing probability depends on panel level random effects. This missingness is not new and there already exist some parametric and semi-parametric methods in the literature. We choose the ACM method to compare because this method is relatively simple to implement and it can be used for intermittent missing data. More detailes about this method will be discussed in this chapter before we present our simulation findings.

In our simulation studies, we assume a total of $n = 1000$ panels with $K = 5$ measurements for each panel. Each simulation study contains $N = 500$ simulation runs. When bootstrap is needed, 100 bootstrap samples are generated. To summarize our simulation results, tables containing characteristics as: (1) bias, (2) standard deviation of estimates from simulation (SD_{est}), (3) standard error of estimate (SE), (4) standard deviation of SEs from simulation (SD_{se}) and (5) covering probability of 95% confidence interval are provided. Let $\boldsymbol{\theta}$ denote the fixed parameter vector we are trying to estimate. From 500 simulation runs, we can get 500 estimate $\hat{\boldsymbol{\theta}}^1, \ldots, \hat{\boldsymbol{\theta}}^{500}$ for the parameters, and we can also get estimates of the standard deviations $\hat{\boldsymbol{\sigma}}^1, \ldots, \hat{\boldsymbol{\sigma}}^{500}$ either by direct calculation or by bootstrap. In each simulation, a 95% confidence interval CI^i can be computed with $\hat{\boldsymbol{\theta}}^i$ and $\hat{\boldsymbol{\sigma}}^i$. The bias is calculated as $\frac{1}{500}\sum_{i=1}^{500}\hat{\boldsymbol{\theta}}^i - \boldsymbol{\theta}$, the SD_{est} is the standard deviation of $\hat{\boldsymbol{\theta}}^1, \ldots, \hat{\boldsymbol{\theta}}^{500}$, the SE is the square root of the average of $(\hat{\sigma}^1)^2, \ldots, (\hat{\sigma}^{500})^2$, and the SD_{se} is the standard deviation of $\hat{\boldsymbol{\sigma}}^1, \ldots, \hat{\boldsymbol{\sigma}}^{500}$. With each 95% confidence interval

CI^i, by comparing to the true parameter values $\boldsymbol{\theta}$, we can decide if the CI covers the true value or not. A covering probability is calculated as the percentage of simulated CIs that cover the truth.

Histograms of simulation bias as well as the trace plots of the simulation estimates and their 95% CIs are plotted for each parameter under each method. A few selected plots are provided in this chapter for a direct visual comparison. However, including all plots would be redundant and unnecessary.

Note that, *Monte Carlo errors* can be approximated by SD/$\sqrt{N}$ (Koehler et al. (2009), Ambegaokar and Troyer (2010)). So, by dividing the reported values for SD_{est} and SD_{se} by $\sqrt{500}$, we can get the estimates for Monte Carlo errors and can decide how many digits after decimal points are still meaningful for reported bias and SE values, respectively. However, in this thesis, all reported values are kept four decimal digits to make the tables look neat and clear.

6.1 Simulation Model

Response Model

The full response model is

$$\boldsymbol{y}_i = \boldsymbol{Z}_i\boldsymbol{\alpha} + \boldsymbol{u}_i'\boldsymbol{\gamma}\mathbf{1} + \boldsymbol{T}_i\boldsymbol{\beta} + \boldsymbol{Z}_i\boldsymbol{b}_i + \boldsymbol{e}_i, \quad i = 1, \ldots, n.$$

Specifically, in our simulation, our response model is:

$$y_{ij} = \alpha_0 + \alpha_1 * z_{ij} + \gamma * u_i + \beta * t_{ij} + b_{i0} + b_{i1} * z_{ij} + e_{ij}, \tag{6.1}$$

$$j = 1, \ldots, 5,\ i = 1, \ldots, 1000,$$

$$\alpha_0 = 2, \alpha_1 = -1, \gamma = -1, \beta = 1$$

where the random effects b_{i0} and b_{i1} are from $N(0, 1)$, and random error $\boldsymbol{e}_i$ is from $N(\mathbf{0}, 0.05^2\boldsymbol{I})$. Covariate z_{ij}'s are i.i.d from $Gamman(5, 2)$ where 5 is the shape parameter and 2 is the rate parameter. Covariate u_i's are i.i.d binary variables taking value 1 or 2 with equal probability. We make covariate t_{ij} related to index j through $t_{ij} \sim N(log(j + 1), 0.3)$. This is because in some longitudinal studies, there may be covariates which are changing with time.

Missingness Model

Once the complete data were generated, missingness need to be introduced. The target missing mechanism in this paper is informative missing, where missing probability depends on panel level random effect(s). Under the current simulation setting, we have the random intercept b_{i0} and one random slope b_{i1}. Four different logistic informative missingness models are set up as following. The performance of our method, OLS, REML and ACM will be compared under all missingness models.

(M1)

$$\text{simple: } P(r_{ij} = 0|\boldsymbol{X}_i, \boldsymbol{b}_i) = \frac{1}{1 + exp\{-2 * b_{i0}\}},$$

$$\text{general: } P(r_{ij} = 0|\boldsymbol{X}_i, \boldsymbol{b}_i) = \frac{1}{1 + exp\{-4 * b_{i0}\}};$$

(M2)

$$\text{simple: } P(r_{ij} = 0|\boldsymbol{X}_i, \boldsymbol{b}_i) = \frac{1}{1 + exp\{-2 * b_{i1}\}}$$

$$\text{general: } P(r_{ij}=0|\boldsymbol{X}_i,\boldsymbol{b}_i)=\frac{1}{1+exp\{-4*b_{i1}\}}$$

(M3)

$$\text{simple: } P(r_{ij}=0|\boldsymbol{X}_i,\boldsymbol{b}_i)=\frac{1}{1+exp(1-0.5*t_{ij}+b_{i0}*t_{ij})},$$

$$\text{general: } P(r_{ij}=0|\boldsymbol{X}_i,\boldsymbol{b}_i)=\frac{1}{1+exp(1-t_{ij}+b_{i0}*t_{ij})};$$

(M4)

$$\text{simple: } P(r_{ij}=0|\boldsymbol{X}_i,\boldsymbol{b}_i)=\frac{1}{1+exp(1-0.5*t_{ij}+b_{i1}*t_{ij})},$$

$$\text{general: } P(r_{ij}=0|\boldsymbol{X}_i,\boldsymbol{b}_i)=\frac{1}{1+exp(1-t_{ij}+b_{i1}*t_{ij})}.$$

As one can see in the above logistic missing probabilities, different values were used in simple case and general case. We slightly changed the parameter values in order to control the overall missing proportion to be no less than 30% in both cases.

Our method is proposed for informative missingness, but we would also like to know its performance when the true missing mechanism is MAR. Our proof for unbiasedness and asymptotic normality will still go through with assumption (1)-(4) in Chapter 2 when missing probability does not depend on random effects. So, in the simulation study, we also compare the performance of our method to OLS, REML and ACM under MAR. The missingness models for MAR setting are:

(M5)

$$\text{simple: } P(r_{ij}=0|\boldsymbol{X}_i,\boldsymbol{b}_i)=\frac{1}{1+exp\{1-0.5*t_{ij}\}},$$

$$\text{general: } P(r_{ij}=0|\boldsymbol{X}_i,\boldsymbol{b}_i)=\frac{1}{1+exp\{1-0.8*t_{ij}\}};$$

(M6)

$$\text{simple: } P(r_{ij}=0|\boldsymbol{X}_i,\boldsymbol{b}_i)=\frac{1}{1+exp\{1-0.5*u_i\}},$$

$$\text{general: } P(r_{ij}=0|\boldsymbol{X}_i,\boldsymbol{b}_i)=\frac{1}{1+exp\{1-0.8*u_i\}};$$

(M7)

$$\text{simple: } P(r_{ij}=0|\boldsymbol{X}_i,\boldsymbol{b}_i)=\frac{1}{1+exp\{1-0.3*z_{ij}\}},$$

$$\text{general: } P(r_{ij}=0|\boldsymbol{X}_i,\boldsymbol{b}_i)=\frac{1}{1+exp\{1-0.5*z_{ij}\}}.$$

6.2 The ACM Method

In the introduction section 1.3, we briefly reviewed the methods for analyzing informative missingness. They are all likelihood-based methods. One of the methods is ACM, where a second stage regression of $\boldsymbol{b}_i$ on some summary statistic $\boldsymbol{S}_i$ of missingness is involved.

With the original response model

$$\boldsymbol{y}_i=\boldsymbol{Z}_i\boldsymbol{\alpha}+\boldsymbol{u}_i'\boldsymbol{\gamma}\mathbf{1}+\boldsymbol{T}_i\boldsymbol{\beta}+\boldsymbol{Z}_i\boldsymbol{b}_i+\boldsymbol{e}_i,$$

if we take expectation on both sides conditional on some summary statistic $\boldsymbol{S}_i$, it becomes

$$E(\boldsymbol{y}_i|\boldsymbol{S}_i)=\boldsymbol{Z}_i\boldsymbol{\alpha}+\boldsymbol{u}_i'\boldsymbol{\gamma}\mathbf{1}+\boldsymbol{T}_i\boldsymbol{\beta}+\boldsymbol{Z}_iE(\boldsymbol{b}_i|\boldsymbol{S}_i)+\boldsymbol{e}_i.$$

The ACM approach is to approximate $E(\boldsymbol{b}_i|\boldsymbol{S}_i)$ by a function of $\boldsymbol{S}_i$. According to Follmann and Wu (1995), if the conditional distribution of $\boldsymbol{S}_i|\boldsymbol{b}_i$ is in the exponential family, then the conditional expectation $E(b_{ik}|\boldsymbol{S}_i)$ is monotone with respect to each

element in $\boldsymbol{S}_i$.

In our method, we used the number of observed responses k_i as our grouping criterion. Based on the discussion in Park et al. (2002) and Xu and Shao (2009), k_i can also serve as the summary statistics under some situations in the ACM model. So, we use k_i as $\boldsymbol{S}_i$ in all our simulation comparisons. The conditional model for our simulation study becomes

$$E(y_{ij}|k_i) = 2 - z_{ij} - u_i + t_{ij} + E(b_{i0}|k_i) + E(b_{i1}|k_i) * z_{ij}.$$

With the ACM approach, we can approximate $E(b_{i0}|k_i)$ and $E(b_{i1}|k_i)$ by functions of k_i. In our simulation studies, we considered both linear and quadratic regression on k_i. According to Follmann and Wu (1995), if the response model contains an intercept term, the approximation of $E(b_{ij}|k_i)$ on k_i does not need to include an intercept term. The approximations are

- ACM linear:

$$E(b_{i0}|k_i) \approx \eta_0 * k_i, \quad E(b_{i1}|k_i) \approx \eta_1 * k_i.$$

- ACM quadratic:

$$E(b_{i0}|k_i) \approx \eta_{01} * k_i + \eta_{02} * k_i^2, \quad E(b_{i1}|k_i) \approx \eta_{11} * k_i + \eta_{12} * k_i^2.$$

With these approximations, we have

- ACM linear:

$$E(y_{ij}|k_i) \approx 2 - z_{ij} - u_i + t_{ij} + \eta_0 * k_i + \eta_1 * k_i * z_{ij}.$$

- ACM quadratic:

$$E(y_{ij}|k_i) \approx 2 - z_{ij} - u_i + t_{ij} + \eta_{01} * k_i + \eta_{11} * k_i * z_{ij} + \eta_{02} * k_i^2 + \eta_{12} * k_i^2 * z_{ij}.$$

The model we fit with ACM becomes

- ACM linear:

$$y_{ij} = \alpha_0^* + \alpha_1^* z_{ij} + \eta_0 * k_i + \eta_1 * k_i * z_{ij} + \gamma u_i + \beta t_{ij} + e_{ij}.$$

- ACM quadratic:

$$y_{ij} = \alpha_0^* + \alpha_1^* z_{ij} + \eta_{01} * k_i + \eta_{11} * k_i * z_{ij} + \eta_{02} * k_i^2 + \eta_{12} * k_i^2 * z_{ij} + \gamma u_i + \beta t_{ij} + e_{ij}.$$

However, the estimated $\hat{\alpha}_0^*$ and $\hat{\alpha}_1^*$ are no longer for $\alpha_0 = 2$ and $\alpha_1 = -1$ in our original model. The details about reconstructing estimates for original parameters are not stated in Follmann and Wu (1995). We reconstructed estimate for α_0 by $\hat{\alpha}_0^* + \hat{\eta}_0 * \bar{k}_i$ in linear case and $\hat{\alpha}_0^* + \hat{\eta}_{01} * \bar{k}_i + \hat{\eta}_{02} * \bar{k}_i^2$ in quadratic case, where $\bar{k}_i$ and $\bar{k}_i^2$ are averages of k_i's and k_i^2's, respectively. The estimate for α_1 are calculated similarly. Since the estimates are not derived directly from the model fitting, the standard error estimates are not for the final estimators. To get standard errors, bootstrap resampling is utilized.

6.3 Simulation Under Simple Case

In the simple case, all panels have at least two observations. Parameter β is estimated with the Q-transformation in Chapter 3, while parameter $\boldsymbol{\alpha}$ and γ are estimated with the estimator proposed in Chapter 5, section 5.1. That is,

$$\hat{\boldsymbol{\beta}} = \boldsymbol{S}_n^{-1} \sum_{i=1}^{n} \boldsymbol{T}_i' \boldsymbol{Q}_i \boldsymbol{y}_i,$$

$$\tilde{\boldsymbol{\zeta}} = \boldsymbol{W}_n^{-1} \sum_{i=1}^{n} \boldsymbol{U}_i' (\boldsymbol{Z}_i' \boldsymbol{R}_i \boldsymbol{Z}_i)^{-1} \boldsymbol{Z}_i' \boldsymbol{R}_i (\boldsymbol{y}_i - \boldsymbol{T}_i \hat{\boldsymbol{\beta}}).$$

Note that $\boldsymbol{\zeta} = (\boldsymbol{\alpha}, \boldsymbol{\gamma})$.

Also, we are able to construct consistent estimators for the variances of our estimators in simple case. Therefore, in each simulation run, we can compute the standard error of our estimates using formula (5.9). To validate our formula, bootstrap standard errors are also computed. On the other hand, for ACM method, because the final estimates are not directly given by fitting the approximation model, so are the standard errors. Without existence of formulas for standard errors, we use bootstrap method to derive an estimate. As regards to OLS and REML, the standard errors are provided directly by the fitting the model in R.

To generate data satisfying the simple case, we first generate complete data based on our response model (6.1). Then, we introduce missingness to the complete data with the logistic missing probabilities (M1) - (M7). First, two indices within each panel are randomly sampled and the corresponding responses are forced to be observed. For the rest three indices, they are missing with the probability defined in (M1) - (M7).

Comparison Under Informative Missing

In missing models (M1) - (M4), missing probabilities depend on the random intercept and/or random slope. The missing mechanism is informative. The simulation results share similar patterns between missing models (M1) and (M2). The simulation results are summarized in Table 6.1, Table 6.2, Table 6.3 and Table 6.4, respectively. Q-tran represents our method.

From Table 6.1 to 6.4, we can clearly see that, OLS estimates are obviously biased when the corresponding covariate is involved in the missingness probability. OLS estimates for γ always have more than 1% bias. REML estimates are always slightly biased for parameter α_0 and α_1, but are almost unbiased for parameter γ and β. This finding suggests that REML still gives valid estimate for parameters not confounded by random effects, but may produce bias in the estimates of parameters confounded by random effects.

The biases for both ACM methods are: about $1.5\% \sim 2\%$ of the true values with missingness model (M1) and (M2) where missing probabilities only depend on random effect, about $1.5\% \sim 7.5\%$ with (M3) where missingness depends on random intercept and covariate $\boldsymbol{t}_i$, and can go up to $10\% \sim 18\%$ with (M4) where missingness depends on random slope and covariate $\boldsymbol{t}_i$. The estimates for intercept α_0 are always biased under all four missingness, the estimates for γ are slightly biased when missing probability depends on random intercept b_{i0} in (M1) and (M3), and the estimates for β become heavily biased when missing depends on covariate $\boldsymbol{t}_i$ in (M3) and (M4). This result is consistent with choice of summary statistics discussed in Wu and Follmann (1999), Park et al. (2002) and Xu and Shao (2009). Based on their papers, the choice of summary statistics highly depends on

	Parameter	OLS	REML	ACM (linear)	ACM (quadratic)	Q-tran
Bias	$\alpha_0 = 2$	-0.2861	-0.0410	0.0294	0.0273	0.0025
	$\alpha_1 = -1$	0.0012	0.0194	0.0005	0.0010	0.0028
	$\gamma = -1$	-0.0169	0.0004	-0.0178	-0.0174	-0.0036
	$\beta = 1$	0.0008	0.0000	0.0004	0.0003	0.0000
SD_{est}	$\alpha_0 = 2$	0.3144	0.1056	0.2989	0.2995	0.1709
	$\alpha_1 = -1$	0.0661	0.0319	0.0677	0.0675	0.0369
	$\gamma = -1$	0.1822	0.0674	0.1716	0.1719	0.1069
	$\beta = 1$	0.0771	0.0025	0.0726	0.0725	0.0026
SE	$\alpha_0 = 2$	0.2310	0.1018	0.3200	0.3184	0.1656
						0.1651
	$\alpha_1 = -1$	0.0439	0.0318	0.0707	0.0709	0.0364
						0.0362
	$\gamma = -1$	0.0980	0.0643	0.1772	0.1761	0.1036
						0.1033
	$\beta = 1$	0.0999	0.0024	0.0752	0.0754	0.0025
						0.0025
SD_{se}	$\alpha_0 = 2$	0.0066	0.0026	0.0268	0.0278	0.0554
	$\alpha_1 = -1$	0.0012	0.0007	0.0077	0.0076	0.0045
	$\gamma = -1$	0.0027	0.0015	0.0136	0.0135	0.0315
	$\beta = 1$	0.0029	0.0001	0.0066	0.0065	0.0002
CP	$\alpha_0 = 2$	0.712	0.926	0.968	0.968	0.948
	$\alpha_1 = -1$	0.794	0.916	0.962	0.954	0.944
	$\gamma = -1$	0.690	0.946	0.970	0.970	0.950
	$\beta = 1$	0.990	0.934	0.950	0.948	0.942

Table 6.1: Simulation results under missing pattern (M1) in simple case, $P(r_{ij} = 0|\boldsymbol{X}_i, \boldsymbol{b}_i) = \frac{1}{1+exp\{-2*b_{i0}\}}$. The SE in the first row for Q-tran is calculated by bootstrap, and the one in the second row is calculated by bootstrap.

	Parameter	OLS	REML	ACM (linear)	ACM (quadratic)	Q-tran
Bias	$\alpha_0 = 2$	0.0277	0.0202	0.0156	0.0159	0.0050
	$\alpha_1 = -1$	-0.3126	-0.0077	-0.0014	-0.0024	0.0013
	$\gamma = -1$	-0.1075	-0.0001	-0.0089	-0.0088	-0.0030
	$\beta = 1$	0.0018	0.0000	0.0037	0.0036	0.0000
SD_{est}	$\alpha_0 = 2$	0.3102	0.1086	0.2033	0.2033	0.1725
	$\alpha_1 = -1$	0.0644	0.0321	0.0497	0.0508	0.0379
	$\gamma = -1$	0.1687	0.0690	0.1162	0.1163	0.1109
	$\beta = 1$	0.0767	0.0024	0.0498	0.0497	0.0024
SE	$\alpha_0 = 2$	0.2223	0.1028	0.2067	0.2068	0.1637
						0.1645
	$\alpha_1 = -1$	0.0422	0.0317	0.0484	0.0494	0.0362
						0.0362
	$\gamma = -1$	0.0943	0.0649	0.1162	0.1167	0.1044
						0.1045
	$\beta = 1$	0.0961	0.0024	0.0649	0.0469	0.0025
						0.0024
SD_{se}	$\alpha_0 = 2$	0.0065	0.0025	0.0172	0.0177	0.0572
	$\alpha_1 = -1$	0.0012	0.0007	0.0049	0.0049	0.0053
	$\gamma = -1$	0.0026	0.0015	0.0092	0.0093	0.0363
	$\beta = 1$	0.0028	0.0001	0.0041	0.0042	0.0002
CP	$\alpha_0 = 2$	0.858	0.924	0.952	0.954	0.934
	$\alpha_1 = -1$	0.000	0.940	0.924	0.932	0.942
	$\gamma = -1$	0.700	0.934	0.950	0.956	0.934
	$\beta = 1$	0.992	0.940	0.930	0.920	0.940

Table 6.2: Simulation results under missing pattern (M2) in simple case, $P(r_{ij} = 0|\boldsymbol{X}_i, \boldsymbol{b}_i) = \frac{1}{1+exp\{-2*b_{i1}\}}$. The SE in the first row for Q-tran is calculated by bootstrap, and the one in the second row is calculated by bootstrap.

	Parameter	OLS	REML	ACM (linear)	ACM (quadratic)	Q-tran
Bias	$\alpha_0 = 2$	0.0641	0.0279	-0.0710	-0.0669	0.0043
	$\alpha_1 = -1$	0.0019	-0.0095	0.0003	0.0006	0.0002
	$\gamma = -1$	-0.0164	-0.0011	-0.0165	-0.0165	0.0003
	$\beta = 1$	0.1282	0.0002	0.0757	0.0743	0.0000
SD_{est}	$\alpha_0 = 2$	0.3224	0.1060	0.3161	0.3171	0.1463
	$\alpha_1 = -1$	0.0706	0.0318	0.0725	0.0727	0.0355
	$\gamma = -1$	0.1825	0.0676	0.1739	0.1740	0.0975
	$\beta = 1$	0.0771	0.0024	0.0744	0.0743	0.0025
SE	$\alpha_0 = 2$	0.2338	0.1018	0.3212	0.3171	0.1463
						0.1462
	$\alpha_1 = -1$	0.0449	0.0318	0.0711	0.0727	0.0355
						0.0348
	$\gamma = -1$	0.1005	0.0643	0.1773	0.1740	0.0975
						0.0934
	$\beta = 1$	0.1011	0.0025	0.0784	0.0743	0.0025
						0.0025
SD_{se}	$\alpha_0 = 2$	0.0068	0.0025	0.0277	0.0265	0.0513
	$\alpha_1 = -1$	0.0013	0.0007	0.0076	0.0071	0.0045
	$\gamma = -1$	0.0028	0.0015	0.0139	0.0136	0.0312
	$\beta = 1$	0.0030	0.0001	0.0067	0.0070	0.0002
CP	$\alpha_0 = 2$	0.834	0.926	0.952	0.948	0.934
	$\alpha_1 = -1$	0.784	0.940	0.944	0.946	0.944
	$\gamma = -1$	0.722	0.940	0.950	0.950	0.932
	$\beta = 1$	0.822	0.934	0.856	0.866	0.942

Table 6.3: Simulation results under missing pattern (M3) in simple case, $P(r_{ij} = 0|\boldsymbol{X}_i, \boldsymbol{b}_i) = \frac{1}{1+exp(1-0.5*t_{ij}+b_{i0}*t_{ij})}$. The SE in the first row for Q-tran is calculated by bootstrap, and the one in the second row is calculated by bootstrap.

	Parameter	OLS	REML	ACM (linear)	ACM (quadratic)	Q-tran
Bias	$\alpha_0 = 2$	-0.3852	-0.0086	-0.2209	-0.2169	0.0034
	$\alpha_1 = -1$	0.2051	0.0067	-0.0011	0.0010	0.0012
	$\gamma = -1$	-0.0128	-0.0006	-0.0087	-0.0091	-0.0007
	$\beta = 1$	0.3151	-0.0001	0.1849	0.1824	-0.0001
SD_{est}	$\alpha_0 = 2$	0.3164	0.1064	0.2551	0.2547	0.1562
	$\alpha_1 = -1$	0.0672	0.0317	0.0591	0.0594	0.0373
	$\gamma = -1$	0.1806	0.0678	0.1438	0.1435	0.0987
	$\beta = 1$	0.0787	0.0024	0.0631	0.0633	0.0025
SE	$\alpha_0 = 2$	0.2322	0.1023	0.2570	0.2560	0.1546
						0.1532
	$\alpha_1 = -1$	0.0446	0.0317	0.0587	0.0591	0.0354
						0.0352
	$\gamma = -1$	0.0998	0.0646	0.1425	0.1414	0.0970
						0.0962
	$\beta = 1$	0.1005	0.0025	0.0630	0.0629	0.0025
						0.0025
SD_{se}	$\alpha_0 = 2$	0.0069	0.0025	0.0211	0.0212	0.0560
	$\alpha_1 = -1$	0.0013	0.0007	0.0058	0.0058	0.0043
	$\gamma = -1$	0.0028	0.0015	0.0109	0.0108	0.0338
	$\beta = 1$	0.0032	0.0001	0.0053	0.0054	0.0002
CP	$\alpha_0 = 2$	0.562	0.930	0.864	0.860	0.950
	$\alpha_1 = -1$	0.042	0.944	0.948	0.942	0.946
	$\gamma = -1$	0.708	0.938	0.946	0.936	0.944
	$\beta = 1$	0.080	0.964	0.168	0.184	0.964

Table 6.4: Simulation results under missing pattern (M4) in simple case, $P(r_{ij} = 0|\boldsymbol{t}_i, b_{i1}) = \frac{1}{1+exp(1-0.5*t_{ij}+b_{i1}*t_{ij})}$. The SE in the first row for Q-tran is calculated by bootstrap, and the one in the second row is calculated by bootstrap.

the missingness probability, and under model (M3) and (M4), k_i is no longer a valid choice. However, in reality, we don't know what exactly the missing probability depends on and may end up with a bad choice of summary statistics just as in our simulation studies with model (M3) and (M4). Under missing model (M1) and (M2), k_i is a valid choice and we get estimates with very small bias.

Finally, for our transformation method, we consistently get estimates with negligible biases that are less that 0.5%. This strongly support our theoretical conclusion that our estimator is unbiased.

In terms of variation, OLS estimates have the largest SD and SE, REML estimates have the smallest SD and SE, ACM estimates have the second largest SD and SE, while our method is between REML and ACM and most of the times is closer to REML results. Even though OLS has largest SE, the CP can be as small as 0 in simulation with model (M2) because it produces sever bias. REML estimates are slightly biased and their SE are small, so the CPs are just about 90% and Type I error rate cannot be well controlled below 5%. Similar to OLS, ACM estimates always have large SE, but the CP can be as low as 20% in simulation with model (M4) due to its high bias. Our method gives negligible biases and have SEs slightly larger than REML. So, the CP for our method are always no less than 93% and are almost 95% in many cases, which indicates that Type I error rate can be controlled around 5%.

Comparison Under MAR

When the missing probabilities only depend on observed values, they are MAR. Among the seven missingness models, (M5) - (M7) satisfies MAR mechanism. We

did simulation under the three missingness models for our method, OLS, REML and ACM. The pattern in the results are very consistent across the three different missing models, so only the result from model (M5) is presented here in Table 6.5.

From this table, we can conclude the following:

(1) The biases for both REML and our method are almost zero, especially for parameters not confounded by random effects (i.e., β and γ). The bias from OLS and both ACM methods are also small, which are below 3%.

(2) Given a method, the standard deviation and standard errors are very close for all parameters. REML estimates possess the smallest SD and SE among all methods. Our transformation method has slightly higher SD and SE than REML. OLS and both ACM methods have very high SD and SE compared with REML and our method. For our method, by comparing the SE calculated with formula and from bootstrap, we are assured that our formula estimator for the covariance matrix are correct.

(3) Covering probability for all parameters and all methods except OLS are close to 95%, which is a sign that Type I error rates are well controlled. OLS still have low CPs even though its average bias is small, because it varies too much.

(4) For parameter β, all the four characteristics we are reporting here are almost identical between REML and our method.

The simulation results under missingness model (M6) and (M7) share similar pattern with the result under (M5). Based on these findings, REML might be the best method to use when one can argue that the missing mechanism is MAR,

	Parameter	OLS	REML	ACM (linear)	ACM (quadratic)	Q-tran
Bias	$\alpha_0 = 2$	0.0259	0.0012	0.0280	0.0302	0.0059
	$\alpha_1 = -1$	0.0027	0.0012	0.0019	0.0013	-0.0004
	$\gamma = -1$	-0.0204	0.0004	-0.0204	-0.0209	0.0000
	$\beta = 1$	0.0020	0.0000	0.0021	0.0021	0.0000
SD_{est}	$\alpha_0 = 2$	0.3095	0.1060	0.3126	0.3117	0.1336
	$\alpha_1 = -1$	0.0673	0.0320	0.0686	0.0681	0.0344
	$\gamma = -1$	0.1812	0.0677	0.1813	0.1812	0.0858
	$\beta = 1$	0.0743	0.0023	0.0739	0.0740	0.0024
SE	$\alpha_0 = 2$	0.2310	0.1017	0.3236	0.3220	0.1332
						0.1318
	$\alpha_1 = -1$	0.0443	0.0371	0.0699	0.0700	0.0337
						0.0336
	$\gamma = -1$	0.0991	0.0643	0.1797	0.1789	0.0839
						0.0835
	$\beta = 1$	0.1005	0.0024	0.0790	0.0793	0.0025
						0.0025
SD_{est}	$\alpha_0 = 2$	0.0067	0.0025	0.0273	0.0283	0.0440
	$\alpha_1 = -1$	0.0912	0.0007	0.0076	0.0074	0.0035
	$\gamma = -1$	0.0027	0.0015	0.0141	0.0138	0.0264
	$\beta = 1$	0.0029	0.0001	0.0067	0.0079	0.0001
CP	$\alpha_0 = 2$	0.848	0.932	0.952	0.960	0.946
	$\alpha_1 = -1$	0.802	0.950	0.952	0.954	0.946
	$\gamma = -1$	0.696	0.930	0.946	0.950	0.956
	$\beta = 1$	0.996	0.968	0.964	0.954	0.968

Table 6.5: Simulation results under missing pattern (M5) in simple case, $P(r_{ij} = 0|\boldsymbol{t}_i) = \frac{1}{1+exp\{1-0.5*t_{ij}\}}$. The SE in parenthesis for Q-tran is calculated by formula, and the one outside is calculated by bootstrap.

because it produces the smallest bias and smallest variation. Our transformation method and ACM are also good choices.

6.4 Simulation Under General Case

In the more general case, some panels can have less than $q = 2$ observations. Parameter β is estimated in the same way as in the simple case. However, $\boldsymbol{\alpha}$ and γ need to be estimated with the grouping approach proposed in Chapter 5, section 5.2. Two slightly different ways of grouping are proposed, they are:

(1) q-grouping:

$$\hat{\boldsymbol{\zeta}}_q = \sum_{i=1}^{q-1} \frac{m_k}{n} \hat{\boldsymbol{\zeta}}_{I_k} + \frac{M_q}{n} \hat{\boldsymbol{\zeta}}_{I_{q+}}, \tag{6.2}$$

(2) K-grouping :

$$\hat{\boldsymbol{\zeta}}_K = \sum_{i=1}^{K} \frac{m_k}{n} \hat{\boldsymbol{\zeta}}_{I_k}. \tag{6.3}$$

Under this general case, we are not able to construct a consistent estimator for the variances of our estimators, as we do not assume specific distributions on random effects $\boldsymbol{b}_i$. The standard errors of our estimators in general case are computed by bootstrap method. Bootstrap method is also used to compute standard errors with ACM method as in simple case.

Similar to the data generation under simple case, we first generate complete data based on our response model (6.1). Then, we randomly sample one index within each panel to force the corresponding response to be observed, and for the rest for measurements, they are missing with equal probability (M1) - (M7).

Comparison Under Informative Missing

The results under model (M1) to (M4) are summarized in Table 6.6, Table 6.7, Table 6.8 and Table 6.9, respectively. q-grouping is our method with q groups when estimating $\boldsymbol{\alpha}$ and $\boldsymbol{\gamma}$, while K-grouping is our method with K groups. Histograms of simulation estimates $\hat{\boldsymbol{\theta}}^1, \ldots, \hat{\boldsymbol{\theta}}^{500}$ and trace plots of these estimates and their CIs are shown in Figure 6.1 - 6.4 at the end of this section, for missingness model (M1) and (M4).

In this general case with the missingness probabilities (M1)-(M4), the results share similar pattern as in simple case. Even though we modified the parameters in missingness probabilities (M1)-(M7) between simple and general cases to control a minimum of 30% missingness in simulated data, the missing proportion is generally higher in general case ($33\% \sim 40\%$) than in simple case ($30\% \sim 34\%$). As a result, the difference among the compared methods are more significant.

OLS method still gives very high bias for parameters whose matching covariates are involved in the missingness probability. The SD and SEs are still the largest or the second largest in all simulations. The CPs are generally quite small. In the following, we will only compare the other methods without considering OLS.

REML starts to produce obviously biased estimates for α_0 and α_1, especially with model (M1) where the relative bias go up to 15%. However, the REML estimates are still almost unbiased for parameter β and γ. ACM methods give about 1% relative bias with model (M1) and (M2), $5\% \sim 9\%$ with model (M3), and go up to $15\% \sim 24\%$ with model (M4). The relative bias with our method are all below 0.5% no matter which criterion is used for grouping.

REML still gives smallest SE and SD. But due to its significant bias for parameter

	Parameter	OLS	REML	ACM (linear)	ACM (quadratic)	q-grouping	K-grouping
Bias	$\alpha_0 = 2$	-0.4236	-0.3059	-0.0179	-0.0215	0.0066	-0.0062
	$\alpha_1 = -1$	0.0063	0.1429	0.0076	0.0067	-0.0001	0.0060
	$\gamma = -1$	0.0054	-0.0004	0.0025	0.0031	-0.0014	-0.0019
	$\beta = 1$	0.0008	0.0001	0.0017	0.0019	0.0001	0.0001
SD_{est}	$\alpha_0 = 2$	0.3508	0.0958	0.3552	0.3554	0.2106	0.3301
	$\alpha_1 = -1$	0.0738	0.0346	0.0801	0.0813	0.0634	0.0811
	$\gamma = -1$	0.1923	0.0609	0.1869	0.1867	0.1117	0.1777
	$\beta = 1$	0.0893	0.0025	0.0877	0.0880	0.0026	0.0026
SE	$\alpha_0 = 2$	0.2483	0.0950	0.3428	0.3449	0.2106	0.3049
	$\alpha_1 = -1$	0.0471	0.0332	0.0799	0.0811	0.0634	0.0804
	$\gamma = -1$	0.1054	0.0601	0.1855	0.1854	0.1117	0.1697
	$\beta = 1$	0.1074	0.0027	0.0833	0.0831	0.0026	0.0027
SD_{se}	$\alpha_0 = 2$	0.0076	0.0028	0.0308	0.0299	0.0408	0.0294
	$\alpha_1 = -1$	0.0014	0.0007	0.0084	0.0096	0.0102	0.0097
	$\gamma = -1$	0.0030	0.0017	0.0154	0.0144	0.0259	0.0138
	$\beta = 1$	0.0032	0.0001	0.0075	0.0075	0.0002	0.0002
CP	$\alpha_0 = 2$	0.556	0.120	0.930	0.934	0.950	0.930
	$\alpha_1 = -1$	0.800	0.008	0.934	0.936	0.926	0.934
	$\gamma = -1$	0.714	0.948	0.948	0.950	0.950	0.932
	$\beta = 1$	0.984	0.956	0.932	0.922	0.950	0.948

Table 6.6: Simulation results under missing pattern (M1) in general case, $P(r_{ij} = 0|\boldsymbol{X}_i, \boldsymbol{b}_i) = \frac{1}{1+exp\{-4*b_{i0}\}}$.

	Parameter	OLS	REML	ACM (linear)	ACM (quadratic)	q-grouping	K-grouping
Bias	$\alpha_0 = 2$	-0.0114	0.1718	-0.0147	-0.0154	-0.0026	-0.0131
	$\alpha_1 = -1$	-0.3967	-0.0869	0.0055	0.0012	0.0027	0.0060
	$\gamma = -1$	0.0047	0.0029	0.0057	0.0057	0.0003	0.0021
	$\beta = 1$	-0.0025	0.0000	-0.0022	-0.0020	0.0000	0.0000
SD_{est}	$\alpha_0 = 2$	0.3234	0.1167	0.2225	0.2234	0.1671	0.2020
	$\alpha_1 = -1$	0.0662	0.0315	0.0519	0.0547	0.0477	0.0552
	$\gamma = -1$	0.1764	0.0732	0.1271	0.1271	0.0965	0.1134
	$\beta = 1$	0.0835	0.0026	0.0540	0.0539	0.0027	0.0027
SE	$\alpha_0 = 2$	0.2319	0.1160	0.2287	0.2313	0.1742	0.2049
	$\alpha_1 = -1$	0.0440	0.0305	0.0542	0.0569	0.0485	0.0565
	$\gamma = -1$	0.0985	0.0733	0.1266	0.1264	0.0999	0.1145
	$\beta = 1$	0.1003	0.0027	0.0541	0.0542	0.0027	0.0027
SD_{se}	$\alpha_0 = 2$	0.0072	0.0033	0.0187	0.0199	0.0369	0.0189
	$\alpha_1 = -1$	0.0014	0.0007	0.0060	0.0064	0.0070	0.0063
	$\gamma = -1$	0.0028	0.0019	0.0103	0.0098	0.0227	0.0087
	$\beta = 1$	0.0031	0.0001	0.0047	0.0048	0.0002	0.0002
CP	$\alpha_0 = 2$	0.844	0.668	0.950	0.956	0.958	0.954
	$\alpha_1 = -1$	0.000	0.198	0.952	0.958	0.948	0.946
	$\gamma = -1$	0.734	0.950	0.952	0.958	0.964	0.952
	$\beta = 1$	0.976	0.960	0.946	0.948	0.964	0.960

Table 6.7: Simulation results under missing pattern (M2) in general case, $P(r_{ij} = 0|\boldsymbol{X}_i, \boldsymbol{b}_i) = \frac{1}{1+exp\{-4*b_{i1}\}}$.

	Parameter	OLS	REML	ACM (linear)	ACM (quadratic)	q-grouping	K-grouping
Bias	$\alpha_0 = 2$	0.0472	0.1364	-0.1091	-0.1032	0.0017	0.0073
	$\alpha_1 = -1$	0.0004	-0.0614	0.0004	-0.0001	-0.0013	-0.0004
	$\gamma = -1$	-0.0008	0.0025	-0.0037	-0.0035	0.0041	-0.0021
	$\beta = 1$	0.1845	0.0004	0.0882	0.0881	0.0000	0.0000
SD_{est}	$\alpha_0 = 2$	0.3376	0.0962	0.3431	0.3476	0.1842	0.3117
	$\alpha_1 = -1$	0.0702	0.0338	0.0742	0.0765	0.0495	0.0758
	$\gamma = -1$	0.1890	0.0612	0.1879	0.1884	0.1001	0.1751
	$\beta = 1$	0.0843	0.0025	0.0805	0.0806	0.0025	0.0025
SE	$\alpha_0 = 2$	0.2356	0.0093	0.3267	0.3262	0.1766	0.2977
	$\alpha_1 = -1$	0.0452	0.0323	0.0745	0.0749	0.0476	0.0735
	$\gamma = -1$	0.1011	0.0628	0.1800	0.1793	0.1019	0.1701
	$\beta = 1$	0.1021	0.0025	0.0797	0.0792	0.0026	0.0026
SD_{se}	$\alpha_0 = 2$	0.0065	0.0027	0.0263	0.0271	0.0497	0.0252
	$\alpha_1 = -1$	0.0013	0.0007	0.0075	0.0081	0.0070	0.0075
	$\gamma = -1$	0.0027	0.0015	0.0129	0.0135	0.0287	0.0138
	$\beta = 1$	0.0029	0.0001	0.0070	0.0071	0.0002	0.0002
CP	$\alpha_0 = 2$	0.826	0.726	0.926	0.920	0.954	0.922
	$\alpha_1 = -1$	0.764	0.530	0.952	0.952	0.942	0.946
	$\gamma = -1$	0.714	0.960	0.928	0.940	0.956	0.932
	$\beta = 1$	0.588	0.952	0.788	0.794	0.940	0.942

Table 6.8: Simulation results under missing pattern (M3) in general case, $P(r_{ij} = 0|\boldsymbol{X}_i, \boldsymbol{b}_i) = \frac{1}{1+exp(1-t_{ij}+b_{i0}*t_{ij})}$.

	Parameter	OLS	REML	ACM (linear)	ACM (quadratic)	q-grouping	K-grouping
Bias	$\alpha_0 = 2$	-0.6124	-0.0696	-0.2996	-0.2998	-0.0032	-0.0008
	$\alpha_1 = -1$	0.2811	0.0362	-0.0033	0.0020	0.0016	0.0027
	$\gamma = -1$	-0.0007	0.0030	-0.0017	-0.0016	0.0034	-0.0005
	$\beta = 1$	0.4783	0.0001	0.2377	0.2376	0.0001	0.0001
SD_{est}	$\alpha_0 = 2$	0.3171	0.1071	0.2503	0.2502	0.1594	0.2305
	$\alpha_1 = -1$	0.0654	0.0324	0.0559	0.0565	0.0451	0.0569
	$\gamma = -1$	0.1753	0.0665	0.1426	0.1428	0.0938	0.1324
	$\beta = 1$	0.0823	0.0025	0.0625	0.0625	0.0025	0.0025
SE	$\alpha_0 = 2$	0.2257	0.1076	0.2501	0.2502	0.1602	0.2277
	$\alpha_1 = -1$	0.0432	0.0312	0.0578	0.0585	0.0416	0.0580
	$\gamma = -1$	0.0968	0.0032	0.1391	0.1385	0.0949	0.1306
	$\beta = 1$	0.0977	0.0001	0.0603	0.0604	0.0026	0.0026
SD_{se}	$\alpha_0 = 2$	0.0063	0.0028	0.0208	0.0200	0.0429	0.0192
	$\alpha_1 = -1$	0.0013	0.0007	0.0055	0.0060	0.0050	0.0057
	$\gamma = -1$	0.0026	0.0016	0.0108	0.0104	0.0248	0.0099
	$\beta = 1$	0.0028	0.0001	0.0053	0.0055	0.0002	0.0002
CP	$\alpha_0 = 2$	0.298	0.912	0.768	0.770	0.964	0.950
	$\alpha_1 = -1$	0.004	0.766	0.946	0.964	0.934	0.958
	$\gamma = -1$	0.714	0.960	0.934	0.934	0.954	0.938
	$\beta = 1$	0.000	0.958	0.026	0.032	0.960	0.962

Table 6.9: Simulation results under missing pattern (M4) in general case, $P(r_{ij} = 0|\boldsymbol{t}_i, b_{i1}) = \frac{1}{1+exp(1-t_{ij}+b_{i1}*t_{ij})}$.

α_0 and α_1, the CPs are much smaller than in the simple case. They are only 0.8% for α_1 and 12% for α_0 with model (M1) and 19.8% for α_1 with model (M2). All other CPs for α_0 and α_1 are well below 90% except for α_0 with model (m4). Both ACM methods again give the highest SE and SD. The CP can be about 95% when missingness only depends on random effects, but not on covariate. However, once the missingness probability depends on covariate $\boldsymbol{t}_i$ as with model (M3) and (M4), the CPs are well below 90%. In the general case, two different grouping criteria were adopted with our method. The q-grouping estimator is exactly the same as the estimator in simple case, if the simple case holds. So, the simulation result with q-grouping in general case share the same feature as the result in simple case. For example, the estimates are unbiased and the SEs are between REML and ACM. When we adopted the K-grouping criterion, the results about bias and CP are similar to that of q-grouping. But the SE and SD of K-grouping are at a similar level as ACM. This finding is reasonable, because it is likely to have more variation when more groups are formed.

As shown in Figure 6.1, the estimates for α_0 and α_1 by REML are severely biased. All estimates for α_0 are biased downwards, while all estimates for α_1 are biased upwards. For parameter γ and β, REML estimates are symmetrically and closely distributed around 0. The estimates for all parameters with ACM and our method are symmetrically distributed around 0, which is a sign of unbiasedness. We can also see that the range of estimates for a fixed parameter is smallest with REML and largest with ACM. From this phenomenon, we can tell that REML estimates have the smallest variation, ACM estimates have the largest variation, while our method gives variation in between. From Figure 6.2, by comparing the estimates(black) and their CIs(red for upper bound and blue for lower bound) to the true value(green),

we see that the CIs from REML method rarely cover the truth for α_0 and α_1. For all the other plots, the true values is mostly covered by their corresponding CIs. There findings are consistent with the summary characteristics in Table 6.6. It is reasonable to conclude that, when missing probability is a logistic regression on a linear function of random intercept, both ACM and our method produce almost unbiased estimates, but REML gives biased estimates to parameters confounded by random effects.

Figure 6.3 and Figure 6.4 are plots for missingness model (M4), where missing probability depends on both random slope and covariate $\boldsymbol{t}_i$. REML estimates for α_0 and α_1 are still not centered around 0, while its estimates for β and γ are symmetrically distributed around 0. Our method still produces estimates symmetrically distributed around 0. However, compared to Figure 6.1, ACM estimates for α_0 and β do not center around 0 anymore. In the trace plots for α_0 and α_1 with REML and trace plots for α_0 and β with ACM, we see that the green lines intersect heavily with either the red curve or the blue curve, which indicate that the covering probability is low. Again, these findings are consistent with the results presented in Table 6.9. We can conclude that, our method still gives unbiased estimates in this setting, while ACM and REML both cause bias.

Comparison Under MAR

Similar to the simple case, the patterns in the results are very consistent across the three different missing models. Simulation result from model (M5) is presented here in Table 6.10.

From this table, we can get the same conclusions as from Table 6.5, except the

	Parameter	OLS	REML	ACM (linear)	ACM (quadratic)	q-grouping	K-grouping
Bias	$\alpha_0 = 2$	0.0093	-0.0017	0.0061	0.0053	-0.0070	-0.0005
	$\alpha_1 = -1$	0.0013	0.0007	0.0023	0.0022	0.0015	0.0023
	$\gamma = -1$	-0.0016	0.0030	-0.0016	-0.0014	0.0055	-0.0004
	$\beta = 1$	-0.0038	0.0000	-0.0040	-0.0038	0.0001	0.0001
SD_{est}	$\alpha_0 = 2$	0.3363	0.1042	0.3387	0.3404	0.1609	0.3152
	$\alpha_1 = -1$	0.0687	0.0329	0.0713	0.0712	0.0399	0.0722
	$\gamma = -1$	0.1830	0.0660	0.1831	0.1833	0.0959	0.1782
	$\beta = 1$	0.0825	0.0025	0.0826	0.0827	0.0026	0.0026
SE	$\alpha_0 = 2$	0.2354	0.1033	0.3217	0.3279	0.1524	0.3036
	$\alpha_1 = -1$	0.0451	0.0318	0.0721	0.0725	0.0386	0.0715
	$\gamma = -1$	0.1009	0.0653	0.1814	0.1812	0.0932	0.1761
	$\beta = 1$	0.1024	0.0025	0.0820	0.0812	0.0026	0.0026
SD_{se}	$\alpha_0 = 2$	0.0064	0.0026	0.0253	0.0278	0.0477	0.0246
	$\alpha_1 = -1$	0.0013	0.0007	0.0068	0.0073	0.0055	0.0067
	$\gamma = -1$	0.0027	0.0015	0.0141	0.0138	0.0298	0.0135
	$\beta = 1$	0.0029	0.0001	0.0073	0.0073	0.0002	0.0002
CP	$\alpha_0 = 2$	0.840	0.946	0.954	0.954	0.936	0.940
	$\alpha_1 = -1$	0.794	0.944	0.944	0.952	0.936	0.944
	$\gamma = -1$	0.718	0.944	0.948	0.944	0.950	0.940
	$\beta = 1$	0.990	0.964	0.942	0.942	0.962	0.956

Table 6.10: Simulation results under missing pattern (M5) in general case, $P(r_{ij} = 0|\boldsymbol{t}_i) = \frac{1}{1+exp\{1-0.8*t_{ij}\}}$.

following:

(1) The biases for all methods are well below 1%.

(2) Our transformation method with q-grouping has slightly higher SD and SE

than REML. Our method with K-grouping, OLS and both ACM methods have much higher SD and SE. This finding is not surprising as there are more groups in K-grouping, which may increase the variation.

Histograms of estimates and trace plots of estimates and CIs are provided in Figure 6.5 and Figure 6.6. All histograms are centered around zero, which indicates that all methods produce unbiased estimates under this setting. In the trace plots, except for OLS, the estimates represented by the black curves slightly fluctuate around the true values represented by the green lines, and the CIs represented by the red curves and blue curves mostly enclose the green lines. In the trace plots for OLS, we can still see that the black curve fluctuates around the green line, but the fluctuation is much larger which results in small CP.

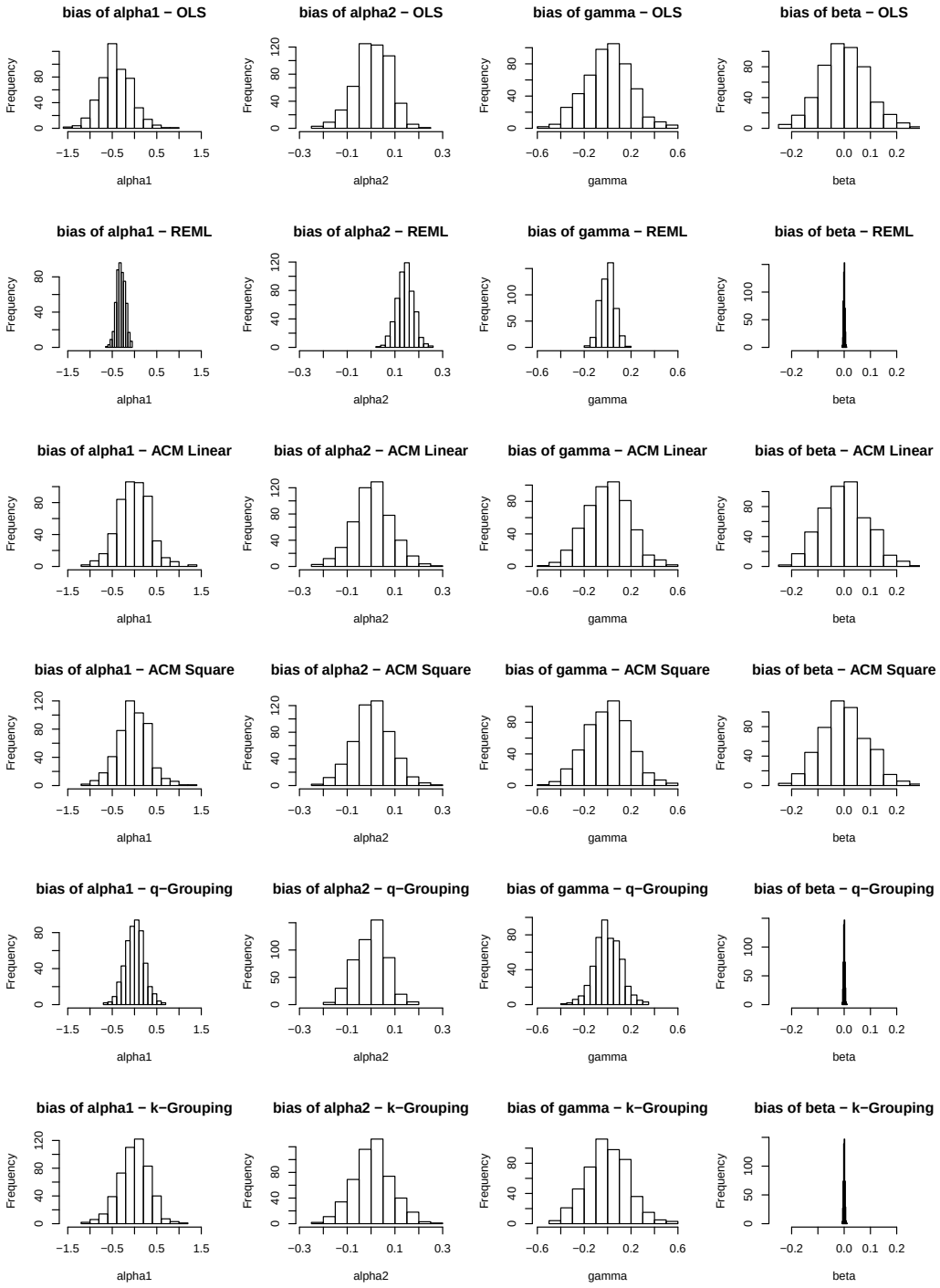

Figure 6.1: Histograms of simulated estimates under missing pattern (M1) in general case, $P(r_{ij} = 0|\boldsymbol{X}_i, \boldsymbol{b}_i) = \frac{1}{1+exp\{-4*b_{i0}\}}$.

The histogram of simulated estimates in all 500 simulations are plotted. All histograms are plotted in the same range.

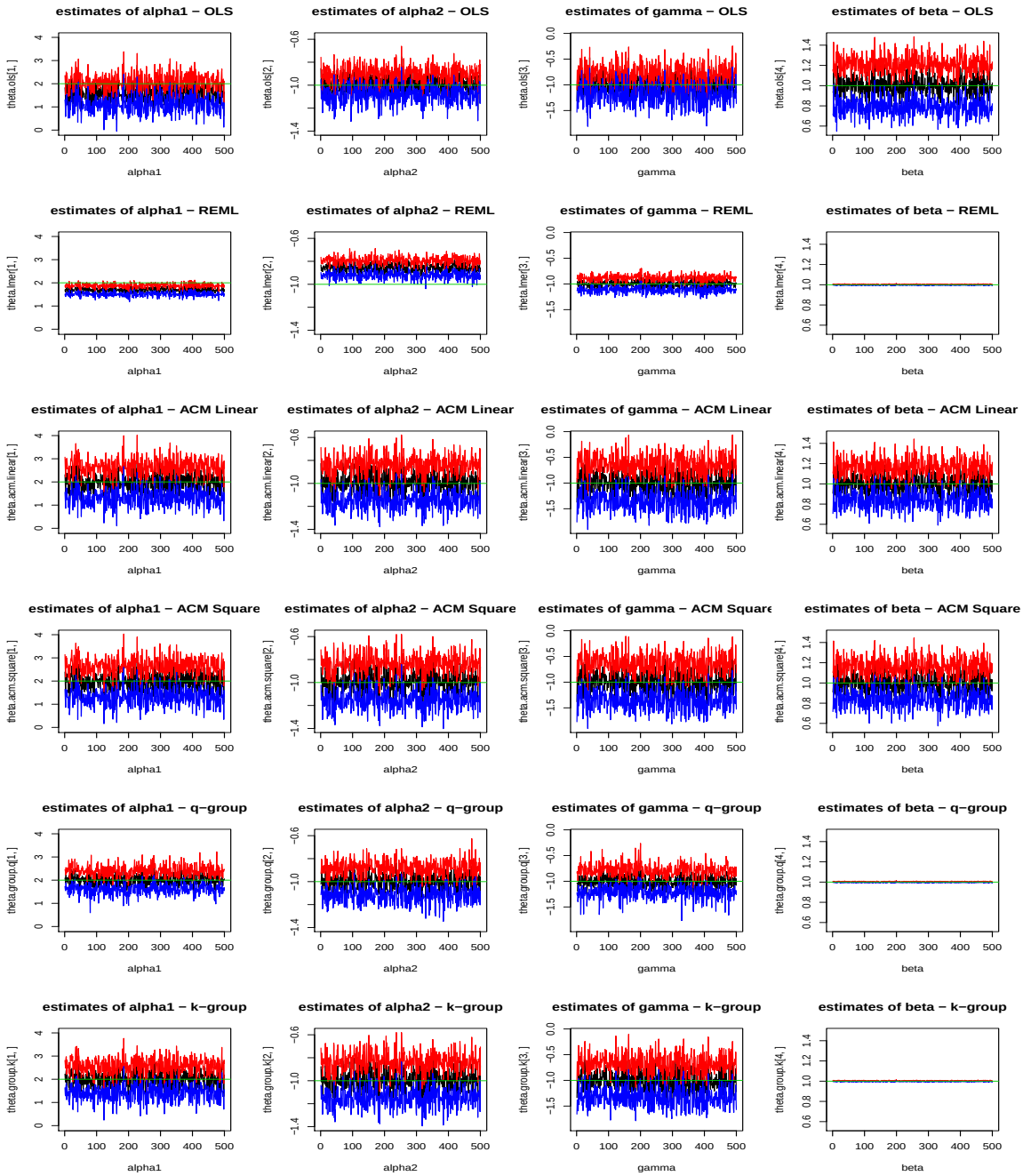

Figure 6.2: Trace plots of simulated estimates and their 95% CIs under missing pattern (M1) in general case, $P(r_{ij} = 0|\boldsymbol{X}_i, \boldsymbol{b}_i) = \frac{1}{1+exp\{-4*b_{i0}\}}$.

The trace plot of simulated estimates and their CIs in all 500 simulations are plotted. In each plot, the green horizontal line indicates the true value of the corresponding parameter. The black line is a trace curve of estimates in all simulations. The red curve is the upper bound of the 95% CI and the blue curve is the lower bound of the same CI.

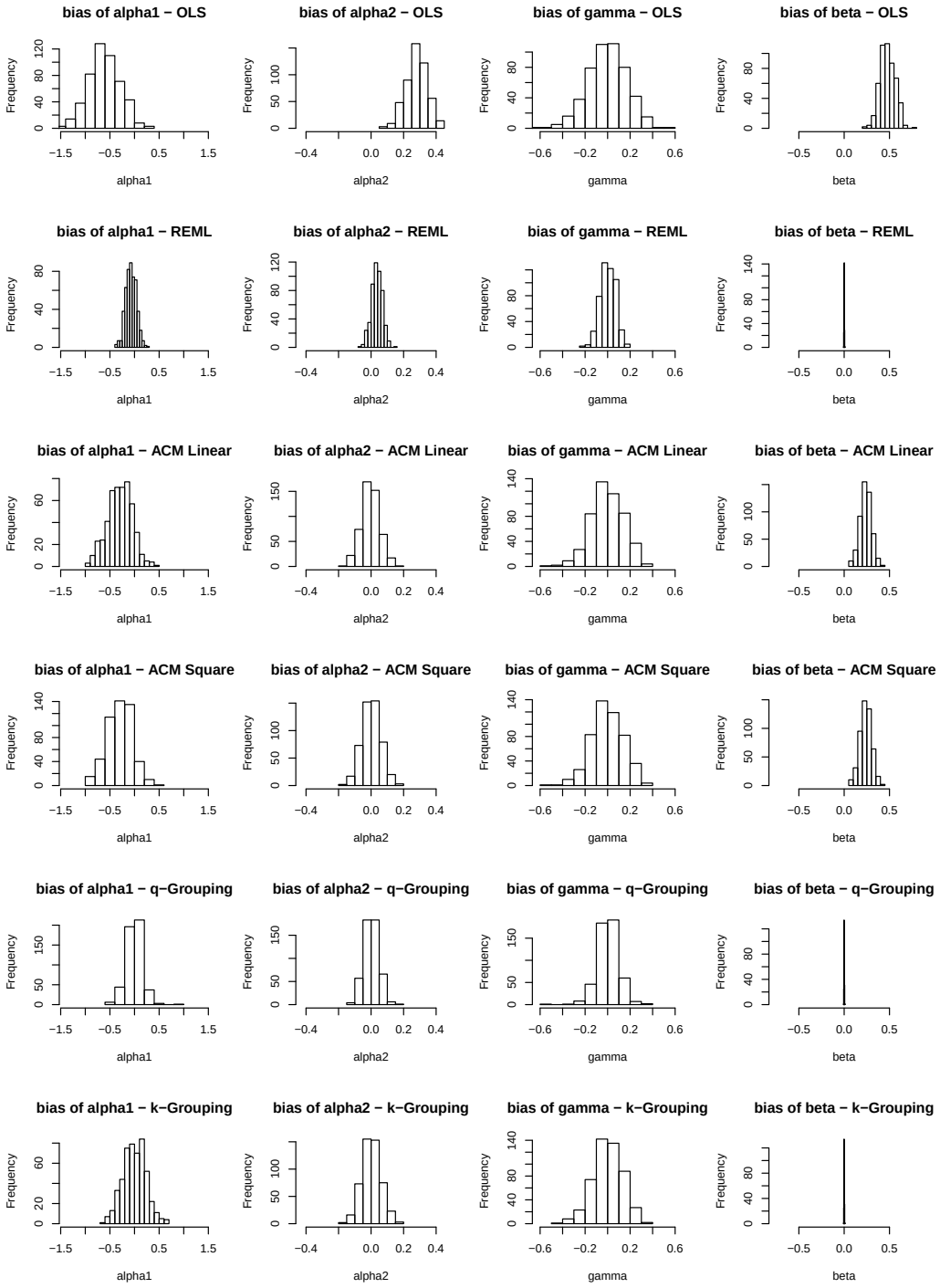

Figure 6.3: Histograms of simulated estimates under missing pattern (M4) in general case, $P(r_{ij} = 0|\boldsymbol{t}_i, b_{i1}) = \frac{1}{1+exp(1-t_{ij}+b_{i1}*t_{ij})}$.

The histogram of simulated estimates in all 500 simulations are plotted. All histograms are plotted in the same range.

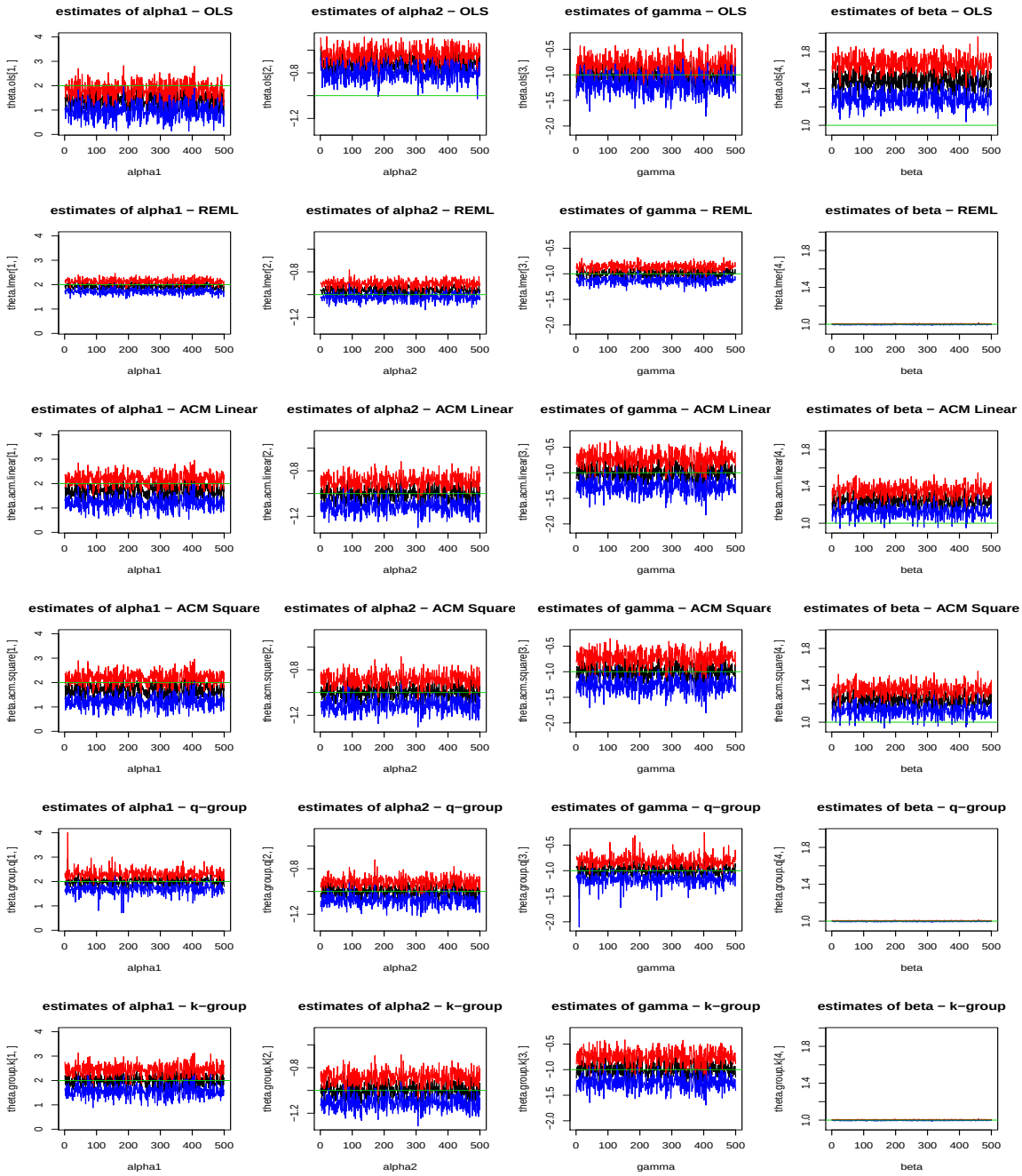

Figure 6.4: Trace plots of simulated estimates and their 95% CIs under missing pattern (M4) in general case, $P(r_{ij} = 0|\boldsymbol{t}_i, b_{i1}) = \frac{1}{1+exp(1-t_{ij}+b_{i1}*t_{ij})}$.

The trace plot of simulated estimates and their CIs in all 500 simulations are plotted. In each plot, the green horizontal line indicates the true value of the corresponding parameter. The black line is a trace curve of estimates in all simulations. The red curve is the upper bound of the 95% CI and the blue curve is the lower bound of the same CI.

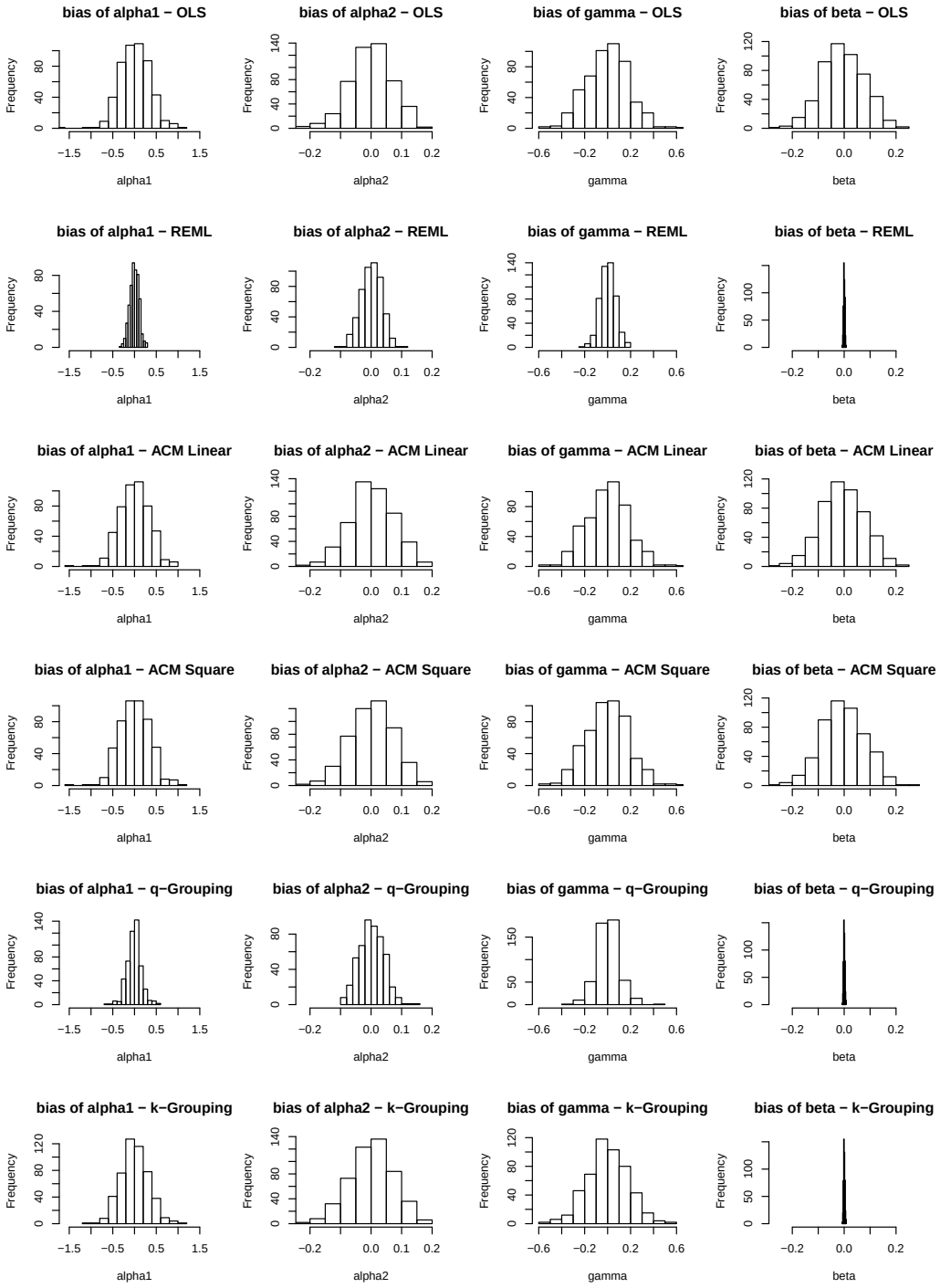

Figure 6.5: Histograms of simulated estimates under missing pattern (M5) in general case, $P(r_{ij} = 0|\boldsymbol{t}_i) = \frac{1}{1+exp\{1-0.8*t_{ij}\}}$.

The histogram of simulated estimates in all 500 simulations are plotted. All histograms are plotted in the same range.

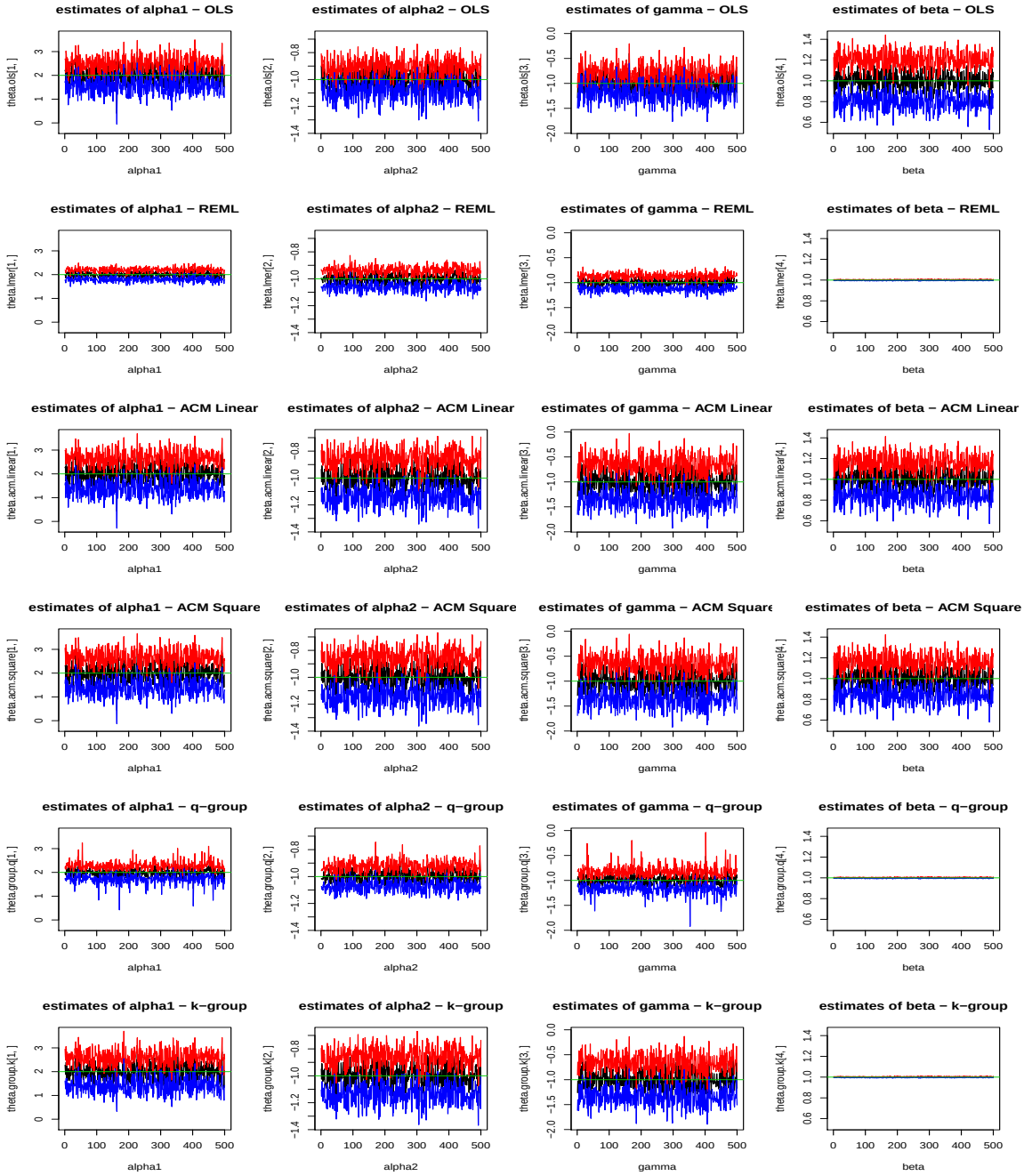

Figure 6.6: Trace plots of simulated estimates and their 95% CIs under missing pattern (M5) in general case, $P(r_{ij} = 0|\boldsymbol{t}_i) = \frac{1}{1+exp\{1-0.8*t_{ij}\}}$.

The trace plot of simulated estimates and their CIs in all 500 simulations are plotted. In each plot, the green horizontal line indicates the true value of the corresponding parameter. The black line is a trace curve of estimates in all simulations. The red curve is the upper bound of the 95% CI and the blue curve is the lower bound of the same CI.

Chapter 7

Application to Real Data Examples: HRS Study

The Health and Retirement Study (HRS) is a national study about the health and economic circumstances of Americans age 51 or older. This study is conducted by Michigan University and founded by NIA (National Institute on Aging) and SSA (Social Security Administration). This study is a longitudinal household survey which was initiated in year 1992. The participants enrolled in this study were followed up biennially, with each follow-up being called as a wave. The most current data contains information form wave 1 to wave 10.

HRS covered many areas including Income and wealth, health and use of health services, work and retire- ment and family connections, with the aim to understand the lives of aging Americans. Some research about statistical methodologies with application to this data set have been published (see, e.g., Cantoni et al. (2007), Kimball et al. (2008), Jiang and Shao (2012)).

RAND corporation combined the data from each wave and provided a cleaned

and easy-to-use version. In the RAND HRS data, there are about 8900 variables and more than 30000 subjects (RAND Center for the Study of Aging (2011)). In our application, we selected a subset of variables with the aim of studying the relation between older people's *individual earning ability (Earn)*, ranging from 0 to 6525 thousand dollars, and the following factors:

(1) Gender: A binary variable with 1 for male, 2 for female.

(2) Edu (*years of education*): A discrete variable indicating the years of education, ranging from 0 to 17(or higher).

(3) Ret (*retirement status*): A categorial variable with three categories: not retired, partially retired and fully retired. In our analysis, we treated partially retired as not retired because the big earning jump usually happens when a person fully retire. 1 is indicating not retired and 2 is for retired.

(4) Wyear (*total years worked*): A continuous variable indicating the total years a person has worked, ranging from 0 to 83.

The individual earning in this study was calculated as the sum of respondent's wage/salary income, bonuses/overtime pay/commissions/tips, 2nd job or military reserve earnings, professional practice or trade income. However, in wave 2, respondents were only asked about the income from all jobs, and in wave 3, respondents were not asked about earning from 2nd job. Due to this inconsistency in the data collection process in early waves, we only included data from wave 4 to wave 7 in our case study.

The purpose of the methodology proposed in this thesis work is to estimate fixed parameters when informative missingness presents in response. A further subset of samples are selected to satisfy our model assumptions.

(1) Participants should have at leave one response in the later 7 waves, as we require at least one observed response for each subject.

(2) Participants did not die during wave 4 to wave 7. The pattern of missingness studied in this thesis is intermittent missingness. Death of participants creates monotonic missingness.

(3) Covariate information on Gender and Edu should be available, as we are dealing with missing values in response, not in covariates. However, it is allowed to have missing values in Ret and/or Wyear when the corresponding Earning measurement is missing.

(4) When the earning response is available in a given wave, the covariates information should also be available. (Instead of throwing away those participants, if the covariate is missing, the corresponding response is treated as missing.)

After this screen, 16313 subjects remains. The distribution of available responses for the remaining population is summarized in Table 7.1. The missing rate in response is about 30%.

k_i	1	2	3	4	5	6	7
No. of participants	1011	1290	1517	3107	1507	2311	5570

Table 7.1: Number of participants grouped by number of observed responses.

7.1 Exploratory Analysis

It is commonly believed that people with higher education is more likely to have higher income. For this HRS data, we would like to examine this trend. Since for each participants, their education status is fixed while they may have multiple income records, average income can be calculated for each participants as a summary measure. However, some participants retired during this study and they are likely to have higher earning before retiring and lower earning after. So, instead of calculating the mean of all earning measurements for each individual, the earning measures are partitioned into two subsets, before retirement and after retirement, and average earning is calculated for each subset. With this partition, for each participant, we have before retirement average and after retirement average. The relation between the two average earning measures are plotted against education status in Figure 7.1. We do see a increasing trend of earning with the increase of education level, no matter before retirement or after. We also notice that the range of earning is much lower in the after retirement plot than in the before retirement plot.

Similarly, we plotted the average earnings against gender. Figure 7.2 shows the boxplots of average earning grouped by gender. Two plots are created for before and after retirement, respectively. In the before retirement plot, we can see that the median earning in males is slightly higher than in females. Also, there are much more very high earnings in males than in female, with most of the high earnings belong to males. In the after retirement plot, both groups have a median close to zero. But the few highest earnings still belong to male.

By comparing the range of earnings before and after retirement in Figure 7.1

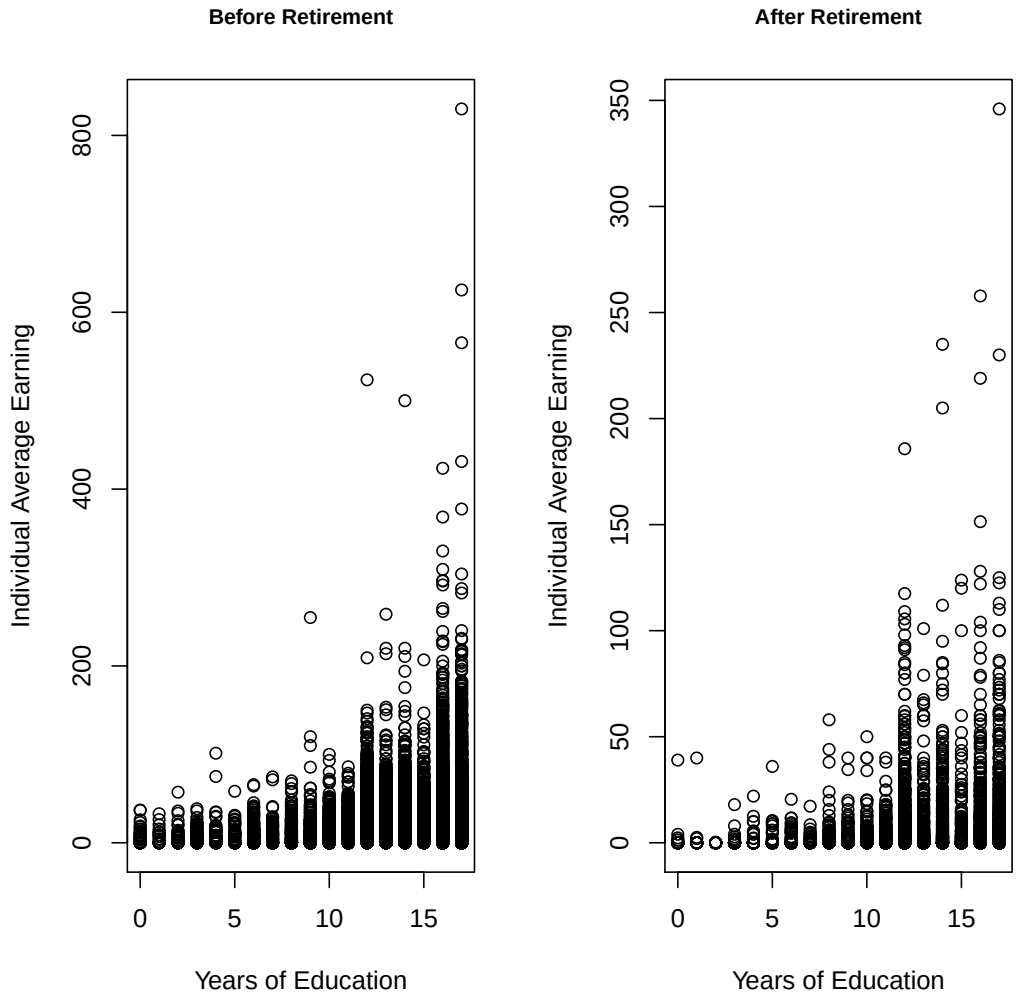

Figure 7.1: The relation between individual earning ability and years of education. In the left plot, the average earning is calculated as the mean earning of each participants before they retire. In the right plot, the average earning is calculated as the mean earning of each participants after they retire.

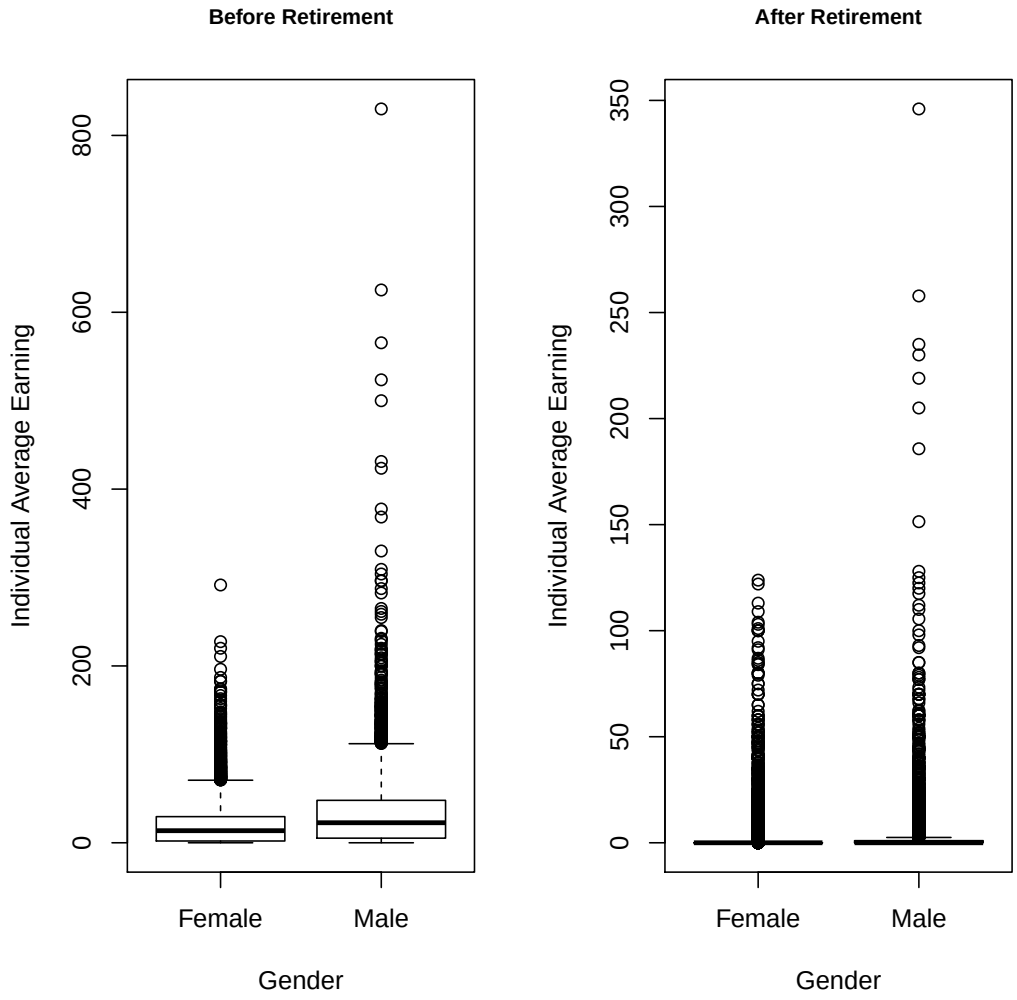

Figure 7.2: The relation between individual earning ability and gender. In the left plot, the average earning is calculated as the mean earning of each participants before they retire. In the right plot, the average earning is calculated as the mean earning of each participants after they retire.

and Figure 7.2, we see an obvious decrease of earning after retirement. Regarding to total years of work, we commonly believe the longer one has worked, the more salary one would gain. In this longitudinal data setting, there is no simple graph which can shown the overall relation between there two variables. We will assume a linear trend in our analysis.

This study is a longitudinal study, where each participants was followed and multiple measurements were collected. For longitudinal data, usually we believe there is subject level random effect besides the random error. As a result, a reasonable model for this study is

$$Earn_{ij} = \alpha + \gamma_1 * Gender_i + \gamma_2 * Edu_i + \beta_1 * Ret_{ij} + \beta_2 * Wyear_{ij} + b_i + e_{ij}. \quad (7.1)$$

A linear mixed-effect model is usually fit by REML, using completely observed data only, without considering missingness existing in response or simply treat the missing mechanism to be MAR. To satisfy our curiosity, an OLS model is fitted with the complete data as well.

(1) Exploratory analysis with OLS:

If the missing mechanism in this data is MCAR and we do not consider the panel structure, a simple ordinary least square (OLS) analysis with complete data will give sensible result. In an OLS analysis, the model fitted is a linear fixed-effect model:

$$Earn_{ij} = \alpha + \gamma_1 * Gender_i + \gamma_2 * Edu_i + \beta_1 * Ret_{ij} + \beta_2 * Wyear_{ij} + e_{ij}.$$

The model fitting result is:

```
             Estimate Std. Error  t value Pr(>|t|)
(Intercept)  14.70982    0.70418   20.889  < 2e-16 ***
GENDER       -7.37692    0.23070  -31.975  < 2e-16 ***
EDU           2.06015    0.03520   58.534  < 2e-16 ***
RET2        -25.20424    0.21225 -118.746  < 2e-16 ***
WYEAR        -0.04295    0.00830   -5.174 2.29e-07 ***
---
Residual standard error: 41750 on 80985 degrees of freedom
Multiple R-squared: 0.1292,Adjusted R-squared: 0.1292
F-statistic:  2404 on 5 and 80985 DF,  p-value: < 2.2e-16.
```

(2) Exploratory analysis with REML:

If the real missing mechanism in this HRS data is MAR, then, as shown in our simulation study, REML analysis will provide unbiased estimates. The mixed-effect model fitting result to the following model is:

$$Earn_{ij} = \alpha + \gamma_1 * Gender_i + \gamma_2 * Edu_i + \beta_1 * Ret_{ij} + \beta_2 * Wyear_{ij} + b_i + e_{ij}.$$

```
Random effects:
 Groups   Name        Variance Std.Dev.
 ID       (Intercept) 464.19   21.545
 Residual             468.48   21.644
Number of obs: 80961, groups: ID, 16313

Fixed effects:
```

	Estimate	Std. Error	t value
(Intercept)	10.40407	1.16636	8.92
GENDER	-8.02748	0.41192	-19.49
EDU	2.23481	0.06130	36.46
RET2	-19.46740	0.22938	-84.87
WYEAR	-0.01802	0.01319	-1.37

The estimates values by REML are different from OLS estimates, which is what we would expect.

7.2 Analysis Under Informative Missingness Assumption

As is usually known, people with very low or very high earning tend to not reveal their total income when asked. This phenomenon may exist in this HRS study. Also, the target of this study is the elder population, who tend to have various health issues and may be experiencing some major changes in life. Unknown personal factors like heath problem and life change can also be reasons for missing responses. These consideration are suggesting that the missingness is not MAR. Because the missing in response may be caused by some unknown personal factors, a more plausible assumption on the missing mechanism is informative missingness.

Let k_i denote the number of observed responses for participants i, based on which we can divide the participants into 7 groups. Figure 7.3 shows the trend

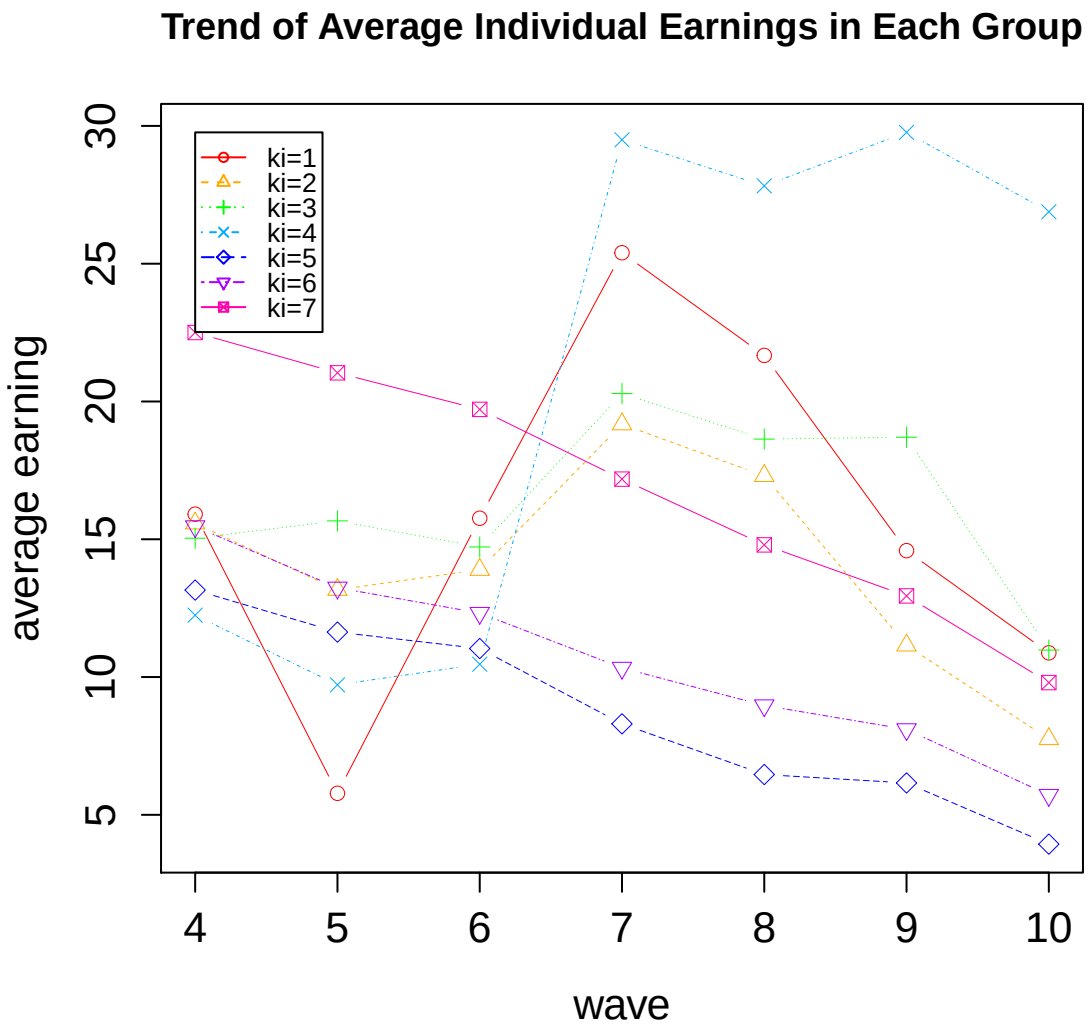

Figure 7.3: The trend of individual earning from wave 4 to wave 7. The points on the curves indicates the mean values of individual earning in each wave. Each curve is calculated in a given group.

of average individual earnings, computed in every wave, of participants in the 7 groups over the duration of this study. This plot shows that participants in different groups manifest different changing patterns of individual earning over the study. This finding is suggesting that the missing pattern is related to the outcome.

Now, we apply the ACM method with linear regression on k_i and our method to this data. Based on our simulation results, if the missing is informative, results from ACM and our method will be different from REML. Furthermore, the ACM method may not agree with our method when k_i is not a sufficient summary statistics, or the linear approximation could not reflect the real trend.

(1) ACM method:

Using k_i as the summary statistics, $E(b_i|k_i)$ is approximated by $\eta * k_i$. So the model fitted here by ACM is

$$Earn_{ij} = \alpha' + \gamma_1 * gender_i + \gamma_2 * edu_i + \beta_1 * Ret_{ij} + \beta_2 * Wyear_{ij} + \eta * k_i + e_{ij}$$

The parameters γ_1, γ_2, β_1 and β_2 are estimated directly, while the estimate for intercept α is by $\hat{\alpha}' + \hat{\eta} * \bar{k}_i$. The standard errors are calculated by bootstrap. The results are summarized in Table 7.2.

(2) Our Q-transformation method without grouping:

In this model, the number of parameters confounded by random effect is just one, i.e. the intercept. So $q = 1$, and thus all participants satisfy $k_i \geq q$ because each participant has at least one observation available. In this simple case, formula is available to calculate the standard errors of estimates. The results are summarized in Table 7.3.

Group	Intercept	Gender	EDU	RET	WYEAR
All	38.1260	-7.1389	2.0949	-24.9719	-0.0068
(sd)	1.0453	0.4302	0.0643	0.3108	0.0129

Table 7.2: Estimates of fixed effect and standard errors of estimate with ACM method.

Group	No. obs	Intercept	Gender	EDU	RET	WYEAR
All	16321	33.9159	-11.2027	2.5024	-15.7706	-0.2542
(sd)		2.3560	0.8023	0.0747	0.3349	0.0506

Table 7.3: Estimates of fixed effect and standard errors of estimate with the proposed Q-transformation method. No grouping is used.

Group	No. obs	Intercept	Gender	EDU	RET	WYEAR
I_1	1011	44.1112	-19.7742	2.7290		
I_2	1290	45.5321	-15.5232	1.8450		
I_3	1517	38.9367	-14.7969	2.5179		
I_4	3107	26.9615	-14.1156	3.8614		
I_5	1507	39.4424	-8.48060	1.3528		
I_6	2311	42.1450	-10.3356	1.5207		
I_7	5570	30.6275	-6.5689	2.2052		
All	16313	35.1622	-11.0081	2.3780	-15.7706	-0.2542
(sd)		2.4035	0.7570	0.0735	0.3412	0.0497

Table 7.4: Estimates of fixed effect and standard errors of estimate with the proposed Q-transformation method with K-grouping. This table provides the number of participants in each group and the local estimates in each group.

(3) Our Q-transformation method with K-grouping:

Even though this data satisfies the simple case requirement, we can still introduce groups based on the observed responses of each participants. $K = 7$ groups can be formed. When grouping is utilized, the standard errors are calculated by bootstrap. The results are summarized in Table 7.4. Local estimates in each group are also provided. We can see that the local estimate are different, which is caused by local biases due to random effect.

7.3 Comparison

For easy comparison, the estimates and standard errors of estimates from all methods are summarized in Table 7.5.

The analysis results from our method are different from OLS and REML, which suggest that the missing mechanism is likely to be informative missing. Our methods give different result from ACM is a sign of the insufficiency of either using k_i as the summary statistic in ACM approximation or approximating the conditional expectation with just a linear function. However, the two estimation procedures of our transformation method, one without grouping and the other with grouping, give highly consistent estimates and standard errors.

From the estimates our method give, elder males on average have earnings about \$11,000 more than females. The higher one's education, the more earnings one make. When changing from working to retired, there is an average of \$15,000 decrease in the earning. The finding for the relation between years worked and earning is a little surprising. As shown in Table 7.5, the earning one can make has a decreasing trend with the years one has worked. This might be caused by the fact

Estimate (s.d.)	OLS	REML	ACM (linear)	Q-tran no grouping	Q-tran K-grouping
Intercept	14.7098	10.40407	38.1260	33.9159	35.1622
	(0.7042)	(1.1664)	(1.0453)	(2.3294)	(2.4035)
Gender	-7.3769	-8.02748	-7.1389	-11.2027	-11.0081
	(0.2307)	(0.4119)	(0.4302)	(0.7340)	(0.7570)
Edu	2.0602	2.23481	2.0949	2.5024	2.3780
	(0.0352)	(0.0613)	(0.0643)	(0.0756)	(0.0735)
Ret	-25.2042	-19.46740	-24.9719	-15.7706	-15.7706
	(0.2123)	(0.2294)	(0.3108)	(0.3427)	(0.3412)
Wyear	-0.0430	-0.01802	-0.0068	-0.2542	-0.2542
	(0.0083)	(0.0132)	(0.0129)	(0.0532)	(0.0497)

Table 7.5: Comparison of estimates and standard errors from all methods.

that some participants actually partially retired before they fully retire, so there is a decrement of earning even though the years they have worked are longer.

Last but not least, all these conclusions are based on assuming an informative missingness mechanism in this HRS data. For real data, we cannot validate what is the true missing mechanism. If this assumption is not true, then the conclusions will no longer hold.

Bibliography

Albert, P. S. and Follmann, D. A. (2000), "Modeling Repeated Count Data Subject to Informative Dropout," *Biometrics*, 56, 667–677.

Ambegaokar, V. and Troyer, M. (2010), "Estimating errors reliably in Monte Carlo simulations of the Ehrenfest model," *American Association of Physics Teachers*, 78, 150–157.

Bock, R. D. and Petersen, A. C. (1975), "A multivariate correction for attenuation," *Biometrika*, 62, 673–678.

Cantoni, E., Field, C., Flemming, J. M., and Ronchetti, E. (2007), "Longitudinal variable selection by cross-validation in the case of many covariates," *STATISTICS IN MEDICINE*, 26, 919–930.

De Gruttola, V. G. and Tu, X. M. (1994), "Modelling Progression of CD4-Lymphocyte Count and Its Relationship to Survival Time," *Biometrics*, 50, 1003–1014.

Follmann, D. A. and Wu, M. C. (1995), "An Approximate Generalized Linear Model with Random Effects for Informative Missing Data," *Biometrics*, 51, 151–168.

Gong, G. and Samaniego, F. J. (1981), "Pseudo Maximum Likelihood Estimation: Theory and Applications," *The Annals of Statistics*, 9, 861–869.

Jiang, D. and Shao, J. (2012), "Semiparametric Pseudo Likelihoods for Longitudinal Data with Outcome-Dependent Nonmonotone Nonresponse," *Statistica Sinica*, 22, 1103–1121.

Kimball, M. S., Sahm, C. R., and Shapiro, M. D. (2008), "Imputing Risk Tolerance From Survey Responses," *Journal of the American Statistical Association*, 103:483, 1028–1038.

Koehler, E., Brown, E., and Haneuse, S. J.-P. A. (2009), "On the Assessment of Monte Carlo Error in Simulation-Based Statistical Analyses," *The American Statistician*, 63, 155–162.

Laird, N. M. and Ware, J. H. (1982), "Random-Effects Models for Longitudinal Data," *Biometrics*, 38, 963–974.

Little, R. J. A. (1993), "Pattern-Mixture Models for Multivariate Incomplete Data," *Journal of the American Statistical Association*, 88:421, 125–134.

— (1995), "Modeling the Drop-Out Mechanism in Repeated-Measures Studies," *Journal of the American Statistical Association*, 90:431, 1112–1121.

Little, R. J. A. and Rubin, D. B. (2002), *Statistical analysis with missing data*, Wiley-Interscience; 2 edition.

Mori, M., Woodworth, G. G., and Woolson, R. F. (1992), "Application of empirical Bayes inference to estimation of rate of change in the presence of informative right censoring," *Statistics in Medicine*, 11, 621–631.

Park, S., Palta, M., Shao, J., and Shen, L. (2002), "Bias adjustment in analysing longitudinal data with informative missingness," *Statistics in Medicine*, 21, 277–291.

Park, T. and Lee, S.-Y. (1997), "A Test of Missing Completely At Random for Longitudinal Data with Missing Observations," *Statistics in Medicine*, 16, 1859–1871.

Patterson, H. D. and Thompson, R. (1971), "Recovery of Inter-Block Information when Block Sizes are Unequal," *Biometrics*, 58, 545–554.

Pulkstenis, E. P., Ten Have, T. R., and Landis, J. R. (1998), "Model for the Analysis of Binary Longitudinal Pain Data Subject to Informative Dropout through Remedication," .

RAND Center for the Study of Aging (2011), "RAND HRS Data Documentation, Version L," .

Schluchter, M. D. (1992), "Methods for the Analysis of Informatively Censored Longitudinal Data," *Statistics in Medicine*, 11, 1861–1870.

Shao, J., ZhiguoXiao, and RuifengXu (2011), "Estimationwithunbalancedpaneldatahavingcovariate measurementerror," *Journal ofStatisticalPlanningandInference*, 141, 800–808.

Ten Have, T. R., Kunselman, A. R., Pulkstenis, E. P., and Landis, J. R. (1998), "Mixed Effects Logistic Regression Models for Longitudinal Binary Response Data with Informative Drop-Out," *Biometrics*, 54, 367–383.

W. A. Thompson, J. (1962), "The Problem of Negative Estimates of Variance Components," *The Annals of Mathematical Statistics*, 33, 273–289.

Wu, M. C. and Bailey, K. (1988), "Analysing Changes in the Presence of Informative Right Censoring Caused by Death and Withdrawal," *Statistics in Medicine*, 7, 337–346.

Wu, M. C. and Bailey, K. R. (1989), "Estimation and comparison of Changes in the Presence of Informative Right Censoring: Conditional Linear Model," *Biometrics*, 45, 939–955.

Wu, M. C. and Carroll, R. J. (1988), "Estimation and Comparison of Changes in the Presence of Informative Right Censoring by Modeling the Censoring Process," *Biometrics*, 44, 175–188.

Wu, M. C. and Follmann, D. A. (1999), "Use of Summary Measures to Adjust for Informative Missingness in Repeated Measures Data with Random Effects," *Biometrics*, 55, 75–84.

Xiao, Z., Shao, J., and Palta, M. (2010), "GMM in linear regression for longitudinal data with multiple covariates measured with error," *Journal of Applied Statistics*, 37, 791–805.

Xiao, Z., Shao, J., Xu, R., and Palta, M. (2007), "Efficiency of GMM Estimation in Panel Data Models with Measurement Error," *Sankhya*, 69, 101–118.

Xu, L. and Shao, J. (2009), "Estimation in Longitudinal or Panel Data Models with Random-Effect-Based Missing Responses," *Biometrics*, 65, 1175–1183.