# Inference Methods for Noisy Correlated Responses with Measurement Error

by

Qihuang Zhang

A thesis
presented to the University of Waterloo
in fulfillment of the
thesis requirement for the degree of
Doctor of Philosophy
in
Statistics

Waterloo, Ontario, Canada, 2020

**Examining Committee Membership**

The following served on the Examining Committee for this thesis. The decision of the Examining Committee is by majority vote.

External Examiner: Dr. Xin Gao
Professor, York University

Supervisor: Dr. Grace Yun Yi
Professor

Internal Member: Dr. Michael Wallace
Assistant Professor

Internal Member: Dr. Yeying Zhu
Associate Professor

Internal-External Member: Dr. Sharon Kirkpatrick
Associate Professor, School of Public Health and Health Systems

## Author's Declaration

I hereby declare that I am the sole author of this thesis. This is a true copy of the thesis, including any required final revisions, as accepted by my examiners.

I understand that my thesis may be made electronically available to the public.

## Abstract

Studying complex relationships between correlated responses and the associated covariates has attracted much research interest. Numerous approaches have been developed to model correlated responses. However, most available methods rely on a crucial condition that response variables need to be precisely measured. In practice, this condition is often violated due to various reasons related to the data collection, study designs, or the nature of the variables. Without taking care of the feature of mismeasurement in variables, inference results are often biased.

Although measurement error and misclassification have been extensively studied in the literature, many problems of mismeasurement in correlated responses remain unexplored. The first problem of interest concerns measurement error and misclassification in the joint modeling of continuous and binary responses. In Chapter 2, we consider the setting with a bivariate outcome vector which contains a continuous component and a binary component both subject to mismeasurement. We propose an induced likelihood approach and describe an EM algorithm to handle measurement error in continuous response and misclassification in binary response simultaneously. The algorithm is fast and can be easily implemented. Simulation studies confirm that the proposed methods successfully remove the bias induced from the response mismeasurement. We implement the proposed methods to mice data arising from a genome-wide association study.

As a complement to the likelihood-based methods discussed in Chapter 2, in Chapter 3, we explore the bivariate generalized estimation equation method with mixed responses subject to measurement error and misclassification. The generalized estimating equation method enjoys robustness to certain model misspecification as well as consistency in the estimation of the mean structure parameters. However, the consistency property relies on the unbiasedness of estimating functions which can break down in the presence of the measurement error and misclassification in responses. We propose an insertion strategy to simultaneously account for measurement error effects in a continuous response and misclassification effects in a binary response. We consider scenarios where either an internal or an external validation subsample is available.

In Chapter 4, we consider a more complex situation where covariates are of a high dimension and may possess a network structure. We start with the case where data are precisely measured and propose a generalized network structure model together with the development of a two-step inferential procedure. In the first step, we employ a Gaussian graphical model to facilitate the network structure, and in the second step, we incorporate the estimated graphical structure of covariates and develop an estimating equation method.

Furthermore, we extend the development to accommodating mismeasured responses. We consider two cases where the information on mismeasurement is known or a validation sample is available. Theoretical results are established and numerical studies are conducted to justify the performance of the proposed methods.

In contrast to error-prone continuous and binary responses considered in the first three chapters, we investigate error-corrupted count data which particularly involve zero-inflated counts, a problem that has received little attention. Zero-inflated count data arise frequently from cancer genomics studies, and it is often of interest to incorporate the feature of excessive zeros in the analysis. However, measurement error in count responses is barely studied, let along the zero-inflated Poisson model with measurement error. In Chapter 5, we propose a novel measurement error model which is unique for addressing error-contaminated count data. We show that ignoring the measurement error effects in analyzing the count response may generally lead to invalid inference results, and meanwhile, we identify situations where ignoring measurement error can still yield consistent estimators. Furthermore, we propose a Bayesian method to address the effects of measurement error under the zero-inflated Poisson model. We develop a data-augmentation algorithm that is easy to implement. Simulation studies are conducted to evaluate the performance of the proposed method. We apply our method to analyze a set of prostate adenocarcinoma genomics data.

Finally, in Chapter 6, we examine another type of correlated responses: time series data. We consider the autoregressive model and establish analytical results for quantifying the biases of the parameter estimation if the measurement error effects are neglected. We propose two measurement error models to describe different processes of data contamination. An estimating equation approach is proposed for the estimation of the model parameters with measurement error effects accounted for. We further discuss forecasting using error-prone times series data. This work is motivated by the need of understanding the ongoing evolving situation of the COVID-19 pandemic. It is important to assess how the mortality rate may change over time, but error-contaminate COVID-19 data present a considerable challenge in uncovering the true development path of the disease.

**Acknowledgements**

First and foremost, I would like to thank my supervisor, Dr. Grace Y. Yi, for guiding, instructing, and empowering me to the completion of the thesis. Under her supervision from my Master's to my Ph.D. studies, I not only learn how to conduct research, but also grow up to be a confident and competent researcher. She always helps me to strive for excellence, deepen the thinking, and push my limits. I feel lucky to have Dr. Yi as my supervisor and cannot imagine to complete my Ph.D. studies without her.

I would also like to thank Dr. Michael Wallace, Dr. Yeying Zhu, Dr. Sharon Kirkpatrick, and Dr. Xin Gao for serving as my committee members and providing thoughtful and invaluable comments.

I also want to express my appreciation for the department faculty and staff for providing such a nice environment and facilities for conducting research and so many meaningful opportunities for me to broaden my horizons.

Also, I want to thank my research friends Di Shu, Li-Pang Chen, Junhan Fang, and Haoxin Zhuang for their useful advice and interesting discussions. I will always remember those fun moments and joyful atmosphere. Thank my friends Yi Chen, Yueshan He, and Lin Lu in Toronto for their support and encouragement to be a better myself.

Last, I would thank my family for their love and constant support. They always give me the power to keep my momentum rolling.

## Dedication

*To my parents, Qi Zhang and Yumei Tang*

# Table of Contents

# List of Tables

# List of Figures

# Chapter 1

# Introduction

Modeling correlated responses are commonly conducted in genomics studies and epidemiology, which has attracted extensive research interests. In this thesis, we focus on handling three classes of correlated responses that have wide application scopes in practice. The first class includes the bivariate mixed responses, with a continuous component and a binary component. The second class contains the zero-inflated count data, which consists of two correlated components, with one component controlling the probability of taking value zero and another component modeling the distribution of a count variable. The third class concerns time series data which are autocorrelated.

Most work of modeling correlated responses relies on the critical condition that the variables are precisely measured, although this condition is usually made implicitly. However, measurement error is almost inevitable in practice. For example, in genetics, a genotype can be misclassified due to sequencing errors. In studies of infectious disease, the number of people infected with a certain disease, e.g., COVID-19, can be underestimated due to the asymptomatic infections.

Despite available discussions of measurement error in responses, there has been relatively little work on exploring the measurement error effects on the analysis of correlated responses (Neuhaus, 2002; Chen et al., 2011). In this thesis, we develop inference methods to address the effects due to measurement error and misclassification in different types of correlated responses, including likelihood-based methods, estimating equations methods, and Bayesian methods.

This thesis research is also motivated to tackle the challenges induced by noisy data arising in applications. In genetic association studies, sometimes the research interest lies in studying the association of a genetic biomarker with mixed responses, which may be error-prone, and this motivates the topics discussed in Chapters 2–3. Meanwhile, understanding the pathway of genetic networks attracts a lot of interest, where numerous candidate genetic variants are associated with multiple traits in a complex manner. This presents a nice scenario of the application of join models discussed in Chapter 4. Chapter 5 examines the zero-inflated Poisson model which is widely applied to handle cancer genomics and microbiome data to account for excessive zeros in count data. The ongoing pandemic of COVID-19 presents a perfect example of measurement error in time series data discussed in Chapter 6. Although our methods developed in this thesis are motivated by the unique features of individual data, the application scope of our methods is very broad.

To better understand our development in the following chapters, in this chapter, we review relevant topics. The remainder is organized as follows. In Section 1.1, we introduce three classes of correlated responses and the approaches that are often used to handle them. In Section 1.2, we explain the measurement error and misclassification mechanisms. In Section 1.3, we discuss the undirected graphical model under the exponential family setting. In Section 1.4, we explain several basic concepts of genome-wide association studies. Finally, we outline the thesis topics in Section 1.5.

## 1.1 Modeling Correlated Responses

In this thesis, we are interested in three classes of correlated responses: mixed continuous and discrete responses, count variables with the zero-inflating feature, and time series data. In the sequel, we separately review each class of correlated responses as well as some associated methods.

### 1.1.1 Mixed Responses and Joint Models

While modeling multiple responses of the same type has been extensively studied in longitudinal studies, modeling mixed type of outcomes, such as continuous and binary responses, has attracted increasing attention. Several models and inference methods were developed, such as generalized estimating equation methods (Liang and Zeger, 1986; Zeger and Liang, 1986), latent variable models (Sammel and Ryan, 1996), and multivariate linear mixed models (MLMM, Sammel et al., 1999). Jointly modeling multiple responses simultaneously has the advantage of boosting the estimation efficiency and the statistical power in testing genetic effects (McCulloch, 2008). There have been various joint models, including those for multiple discrete responses (Chen et al., 2016), latent variable models for mixed continuous and discrete outcomes (Sammel et al., 1997; Teixeira-Pinto and Normand, 2009; Lin et al., 2014), correlated Probit models (Gueorguieva and Agresti, 2001), estimating function methods (Prentice and Zhao, 1991; Fitzmaurice and Laird, 1995), and the Bayesian framework (Wagner and Tüchler, 2010). Further generalizations were explored for handling clustered data (Catalano, 1997; Lin et al., 2010) and high dimensional data (Faes et al., 2008).

Although numerous methods were proposed to incorporate the correlation among responses, these methods can be roughly classified into two categories: likelihood approaches and estimating equation methods. These methods have their advantages and disadvantages. For example, the estimating function method is robust to the model specification by taking the price of the loss of efficiency. On the other hand, the estimators based on the likelihood methods are the most efficient but rely on the correct specification of the full distribution. In this section, we review the generalized linear mixed model with likelihood theory and the generalized estimation equations with estimating function theory.

### Generalized Linear Mixed Model

For $i = 1, \ldots, n$ and $j = 1, \ldots, m$, let $Y_{ij}$ be the $j$th response for the $i$th individual and let $X_i = (X_{i1}^{\mathrm{T}}, \ldots, X_{ip}^{\mathrm{T}})^{\mathrm{T}}$ be the vector of covariates for the $i$th subject. Write $Y_i = (Y_{i1}, \ldots, Y_{im})^{\mathrm{T}}$. The *generalized linear mixed model* (GLMM) can be described in two steps. Assume that conditional on random effects $u_i$ as well as covariates $X_i$, the $Y_{ij}$ are independent and marginally follow a distribution from the exponential family given by

$$f(y_{ij}|x_i, u_i) = \exp\left\{ \frac{y_{ij}\varphi_{ij} - b(\varphi_{ij})}{a(\psi)} + c(y_{ij}; \psi) \right\}, \tag{1.1}$$

where $a(\cdot)$, $b(\cdot)$ and $c(\cdot)$ are known functions, $\varphi_{ij}$ is the canonical parameter, and $\psi$ is a dispersion parameter.

Based on the specification in (1.1), given covariates $X_i$ and random effects $u_i$, the conditional mean $\mu_{uij} = E(Y_{ij}|u_i, X_i)$ equals $b'(\varphi_{ij})$ which is postulated by

$$g(\mu_{uij}) = \beta_0 + \beta_x^{\mathrm{T}} X_{ij} + u_i^{\mathrm{T}} F_{ij},$$

where $b'(\cdot)$ is the first derivative of $b(\cdot)$, $g(\cdot)$ is a link function, $\beta = (\beta_0, \beta_x^{\mathrm{T}})^{\mathrm{T}}$ is the vector of regression parameters, and $F_{ij}$ is a quantity determined by the study design and correlation among the responses. The random effects $u_i$ are assumed to be independent of the covariates $X_i$.

Contrary to the name of *random effects*, the components of $\beta$ are often called *fixed effects*. The $\beta$ parameter in the generalized linear mixed model has a different interpretation from the generalized linear model with fixed effects only. In GLMM, the parameter $\beta$ represents the changes of transformed responses associated with one unit change of covariates for an individual, whereas the $\beta$ in the generalized linear model is interpreted as the changes at the population level.

**Generalized Estimating Equation**

For $i = 1, \ldots, n$ and $j = 1, \ldots, m_i$, let $\mu_{ij} = E(Y_{ij}|X_{ij})$ and $v_{ij} = \text{Var}(Y_{ij}|X_{ij})$ be the conditional mean and variance, respectively, given covariates $X_{ij}$.

The conditional mean $\mu_{ij}$ is modeled by

$$g(\mu_{ij}) = \beta_0 + \beta_x^{\mathrm{T}} X_{ij},$$

where $g(\cdot)$ is a prespecified link function and $\beta = (\beta_0, \beta_x^{\mathrm{T}})^{\mathrm{T}}$ is the vector of regression parameters.

The conditional variance $v_{ij}$ is often modeled by a function of the mean and the dispersion parameters $\psi$. Namely,

$$v_{ij} = h(\mu_{ij}; \psi),$$

where $\psi$ is the dispersion parameter and $h(\cdot)$ is a specified function characterizing the relationship between the conditional variance $v_{ij}$ and the conditional mean $\mu_{ij}$ of $Y_{ij}$ given $X_{ij}$. For instance, the variance function of the binary response is often specified as $h(\mu_{ij}; \psi) = \mu_{ij}(1 - \mu_{ij})$ where $\psi = 1$.

With the only assumptions on the first two moments, the *generalized estimating equation* (GEE) method is a natural way to estimate $\beta$. Let $V_i$ be the conditional variance of $Y_i$ given $X_i$. Define the estimating function

$$U_i(\beta) = D_i V_i^{-1}(Y_i - \mu_i),$$

where $\mu_i = (\mu_{i1}, \ldots, \mu_{im})^{\mathrm{T}}$, and $D_i = \frac{\partial \mu_i^{\mathrm{T}}}{\partial \beta}$.

Then solving

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

for $\beta$ gives a consistent estimator of $\beta$, say $\hat{\beta}$, provided regularity conditions (Liang and Zeger, 1986; Prentice and Zhao, 1991). In addition, $\sqrt{n}(\hat{\beta} - \beta)$ is asymptotically normally distributed with mean 0 and covariance matrix

$$\left\{E\left(\frac{\partial U_i(\beta)}{\partial \beta^{\mathrm{T}}}\right)\right\}^{-1} E\{U_i(\beta)U_i(\beta)^{\mathrm{T}}\} \left\{E\left(\frac{\partial U_i(\beta)}{\partial \beta^{\mathrm{T}}}\right)\right\}^{-1\mathrm{T}}.$$

We comment that the validity of the GEE method hinges on the assumption that

$$E(Y_{ij}|X_i) = E(Y_{ij}|X_{ij})$$

if the working matrix for $V_i$ is not diagonal. A detailed discussion on this assumption can be found in Yi (2017, Section 5.1).

The consistency of the first order generalized estimating equation also requires the mean structure to be correctly specified regardless of whether the covariance structure is correctly specified or not. Sometimes, the association structure may be of scientific interest, and the second-order GEEs are constructed by modeling the second moment (Prentice and Zhao, 1991). Hall and Severini (1998) extended the original GEE model based on quasi-likelihood to improve the efficiency without requiring any covariance specification. Hall (2001) reviewed the relationships between different GEE approaches.

### 1.1.2 Zero-inflated Count Data and Zero-inflated Poisson Model

Count data arise from many studies of genomics (e.g. Fu et al., 2017) and microbiome (e.g. Xu et al., 2020), and they are commonly modeled by a Poisson distribution. On the other hand, count data may contain excessive zeros, which come from two sources, classified as "*structural zeros*" and "*sampling zeros*". The "*structural zeros*" refers to that an individual is not "at risk" for the event and hence has no possibility to have a positive count. The "*sampling zeros*", on the contrary, refers to that the individual is "at-risk" with a positive count, but results in a zero count by chance. For example, the count of the copy number variations (CNVs) is a useful indication of mutations in genes that might be associated with an increased risk of cancer. However, whether or not the CNVs are observed is also determined by whether the relevant pathways are activated. Many subjects have no CNVs simply due to the inactivated pathways, leading to extra "structural" zeros than expected when considering the Poisson distribution.

Viewing data as being generated from a mixture of a point mass at zero and a Poisson distribution, a zero-inflated Poisson model (Lambert, 1992) is commonly used to address the excessive zero issue in the analysis of count data. It basically consists of two correlated components, where each component models a different aspect of zero-inflected count data. Specifically, one component concerns the probability of an individual sampled from an "at-risk" group and another component models the count variable conditional on the "at-risk" group.

To be specific, for $i = 1, \ldots, n$, let $Y_i$ denote the count outcome for subject $i$ taking a non-negative integer value and let $X_i$ denote the associated covariate vector of dimension $p_x$. For $i = 1, \ldots, n$, let $\phi_i = P(A_i = 1|X_i)$ represent the conditional probability of sampling from 'at-risk" group , given $X_i$, and let $\mu_i = E(Y_i|A_i = 1, X_i)$ denote the condition mean of $Y_i$, given being sampled from the 'at-risk" group and the covariate $X_i$, which are assumed to satisfy $0 < \phi_i < 1$, and $\mu_i > 0$. That is, $Y_i$ is sampled from the "non-at-risk" group with probability $1 - \phi_i$, and sampled from the "at-risk" group with probability $\phi_i$, following a Poisson distribution with mean $\mu_i$:

$$
\begin{aligned}
&Y_i = 0, \text{ with probability } 1 - \phi_i, \\
&Y_i \sim \text{Poisson}(\mu_i), \text{ with probability } \phi_i.
\end{aligned}
$$

### 1.1.3 Time Series Data and Autoregressive Model

Time series data arise commonly in epidemiology and infectious disease studies. Such data are taken as the third type of correlated responses in this thesis, where the correlation among the responses is directly reflected by the autocorrelation (or serial correlation). To model time series data, various models have been proposed, such as the classical decomposition model, autoregressive integrated moving average (ARIMA) model, autoregressive conditional heteroskedasticity (ARCH) model, state-space models, etc.

We denote a time series as $\{X_t : t = 1, \ldots, T\}$, where $X_t$ is a random variable and $T$ is a positive integer or infinite. Stationarity is an important assumption for many models. The *strictly stationarity* for time series $X_t$ is defined as

$$
(X_1, \ldots, X_n)^{\mathrm{T}} \stackrel{d}{=} (X_{1+r}, \ldots, X_{n+r})^{\mathrm{T}} \tag{1.2}
$$

for any positive integer $n$ and $r$, where $\stackrel{d}{=}$ means those variables have the same joint distribution.

Sometimes, the assumption described in (1.2) is too strict and unrealistic in reality. We may consider a weaker condition for the stationarity assumption, where time series $\{X_t : t = 1, \ldots, T\}$ is *weakly stationary* if these two conditions are satisfied:

(i) $E(X_t)$ is independent of $t$,

(ii) $\text{Cov}(X_t, X_{t+r})$ is independent of $t$ for each $r$.

In analysis of times series data, $E(X_t)$ and $Cov(X_t, X_{t+r})$ are important to be quantified. We define the autocovariance function

$$\gamma(h) = Cov(X_{t+h}, X_t) \text{ for } h = 0, \pm 1, \cdots,$$

and the autocorrelation function (ACF) is then defined as

$$\rho(h) = \frac{\gamma(h)}{\gamma(0)}.$$

The autocovariance function and the autocorrelation function provide useful measures for the degree of dependence on the serial variables at different time lags, and thus, play important roles in the forecasting of future values.

Due to the importance of the autocovariance function, its properties under the weakly stationarity assumption are well studied:

(1) $\gamma(0) \geq 0$;

(2) $|\gamma(h)| \leq |\gamma(0)|$;

(3) $\gamma(h) = \gamma(-h)$.

Autoregressive models are useful in analyzing time series data, which study the dependence of $X_t$ on $\{X_{t-p}, \ldots, X_{t-1}\}$, and is given by

$$X_t = \phi_0 + \sum_{j=1}^{p} \phi_j X_{t-j} + \epsilon_t,$$

where $p$ is an integer smaller than $T$, $(\epsilon_1, \ldots, \epsilon_t)^{\mathrm{T}}$ is independent of $(X_1, \ldots, X_t)^{\mathrm{T}}$ with each $\epsilon_t$ having zero mean and a variance, say, $\sigma_\epsilon^2$, $\phi_0$ is a constant drift, and $\phi = (\phi_1, \ldots, \phi_p)^{\mathrm{T}}$ is the regression coefficient.

## 1.2 Measurement Error and Misclassification

Measurement error is prevalent in various cases. Sometimes it is because of technique errors. For example, when measuring a length, the last digit of the measurement is usually an estimate. Sometimes it is because of recall bias. In observational epidemiology, people answer questionnaires according to their experience in the past which is error-prone. A detailed discussion on reasons and sources of measurement error is provided by Yi (2017, Section 2.1).

In the literature of measurement error, we often distinguish different types of error-prone variables; the case of error-prone *continuous* variables is called *measurement error* and the case of error-prone *discrete* variables *misclassification*, although sometimes both cases are simply referred to as *measurement error* or *mismeasurement*.

In this section, we review some measurement error models and misclassification models.

### 1.2.1 Measurement Error

For $i = 1, \ldots, n$, let $Y_i$ denote the precisely measured continuous response. Due to measurement error, we do not observe $Y_i$, but instead, we observe a surrogate $Y_i^*$. The relationship between the true response $Y_i$ and the observed surrogates $Y_i^*$ can be described by different measurement error models in the same manner as Yi (2017) by introducing a random variable $e_i$:

1. Classical Additive Error Model:

   $$Y_i^* = Y_i + e_i,$$

   where the error term $e_i$ is often assumed to be independent of the true response $Y_i$.

2. Multiplicative Model:

   $$Y_i^* = Y_i e_i,$$

   where the mean of $e_i$ is assumed to be 1.

3. Linear Regression Model:

   $$Y_i^* = \gamma_0 + \gamma_1 Y_i + \gamma_2^{\mathrm{T}} X_i + e_i,$$

   where $e_i$ is independent of $\{Y_i, X_i\}$ and is often assumed to follow a normal distribution with mean zero and variance $\sigma_e^2$, and $\gamma_0$, $\gamma_1$, and $\gamma_2$ are parameters.

4. General Regression Model:

$$Y_i^* = m(Y_i, X_i; \gamma) + e_i,$$

where $m(\cdot)$ is a prespecified function which can be nonlinear, $\gamma$ is the vector of regression parameters associated with the measurement error model, and $e_i$ is independent of $\{Y_i, X_i\}$.

Although these measurement error models provide a flexible specification of the relationship between the error-prone variable $Y_i$ and its surrogate version $Y_i^*$, model identifiability issues may be a problem. To make the inferences meaningful, a specified model $f(y; \theta)$ must be *identifiable*. That is, if two parameters $\theta_1$ and $\theta_2$ make $f(y; \theta_1) = f(y; \theta_2)$ hold for any all possibly observed $y$ (in a set of probability 1), then

$$\theta_1 = \theta_2.$$

Measurement error in covariates has received extensive research interest. On the other hand, less research work has been directed to measurement error in response, partly because the measurement error in response can be ignored in some scenarios, such as the response model described by a linear regression model together with a certain additive measurement error model. However, measurement error in response is not always ignorable if the measurement error process is nonlinear (Yi, 2017, Page 353).

### 1.2.2 Misclassification

When error-prone variables are discrete, we usually describe it as a misclassification problem. Let $Y_i$ be a binary variable following a Bernoulli distribution. The true response $Y_i$ is not observable but instead we observe the surrogate $Y_i^*$. Let $\pi_{i0} = P(Y_i^* = 1 | Y_i = 0, X_i)$ and $\pi_{i1} = P(Y_i^* = 0 | Y_i = 1, X_i)$ be the misclassification probabilities that may depend on the covariates $X_i$. The relationship between the true response $Y_i$ and the observed surrogate $Y_i^*$ is often modeled by logistic regressions models:

$$\text{logit } \pi_{i1} = \alpha_{01} + \alpha_{x1}^{\mathrm{T}} X_i,$$

and

$$\text{logit } \pi_{i0} = \alpha_{00} + \alpha_{x0}^{\mathrm{T}} X_i,$$

where $\alpha = (\alpha_{01}, \alpha_{x1}^{\mathrm{T}}, \alpha_{00}, \alpha_{x0}^{\mathrm{T}})^{\mathrm{T}}$ is the vector of the regression parameters.

Misclassification in responses and covariates has been studied in the literature (e.g., Neuhaus, 2002; Ramalho, 2002; Prescott and Garthwaite, 2002; Paulino et al., 2003; Chen et al., 2011; Yi et al., 2015; Shu and Yi, 2017). The misclassification in response will generally lead to biased estimation of parameters if no action is properly taken.

## 1.3 Undirected Graphical Model

An *undirected graphical model* (UGM), also called a *Markov random field* (MRF) or a *Markov network*, does not require specification for the edge orientations and is natural to be applied for image analysis and spatial statistics.

For $i = 1, \ldots, n$, suppose $X_i = (X_{i1}, \ldots, X_{ip})^{\mathrm{T}}$ is a random vector for subject $i$. Let $V_i = \{1, 2, \ldots, p\}$ be the index set of the vertices, corresponding to the variables $\{X_{i1}, \ldots, X_{ip}\}$, and let $E_i = V_i \times V_i$ denote the set of edges derived from $V_i$. We use an undirected graph $G_i = (V_i, E_i)$ to describe the relationship among the covariates for subject $i$, where an edge of vertices $s$ and $t$ represents that $X_s$ and $X_t$ are correlated. Since the distribution of random vector $X_i$ is assumed to be the same for each subject, we consider the graph for each individual to be identical. Namely, $G_1 = \cdots = G_n \equiv G$ with $G = (V, E)$.

Markov independence is an important assumption for graphical models. To illustrate this assumption, we first define a cut set $\mathcal{C} \subseteq V$ to be a set of nodes that separate the graph $G$ into two disjoint components $\mathcal{A}$ and $\mathcal{B}$ (Figure 1.1).

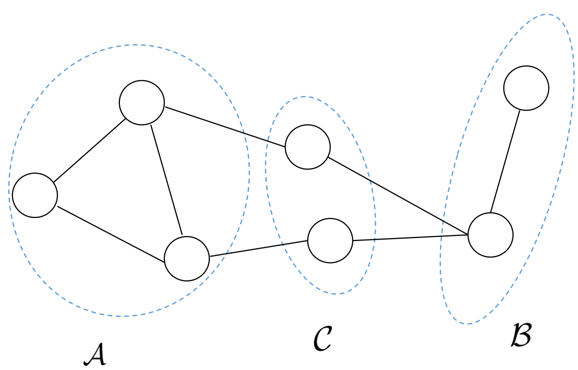


Figure 1.1: *Example of a graph separated by a cut set* $\mathcal{C}$

**Assumption 1 (Markov independence assumption)** *For all cut sets* $\mathcal{C} \subset V$

$$X_{\mathcal{A}} \perp\!\!\!\perp X_{\mathcal{B}} | X_{\mathcal{C}},$$

*where* $X_{\mathcal{A}}$*,* $X_{\mathcal{B}}$ *and* $X_{\mathcal{C}}$ *are the covariates corresponding to the sets* $\mathcal{A}$*,* $\mathcal{B}$ *and* $\mathcal{C}$*, respectively, and* $\perp\!\!\!\perp$ *represents "is conditionally independent of".*

Based on Assumption 1, a graphical model following the exponential family distribution can be constructed by

$$f(x_i; \theta, \Theta) = \exp\left\{\sum_{k\in V} \theta_k B(x_{ik}) + \sum_{(s,t)\in E} \theta_{st} B(x_{is}) B(x_{it}) + \sum_{k\in V} C(x_{ik}) - A(\theta, \Theta)\right\}, \quad (1.3)$$

where $\theta = (\theta_1, \cdots \theta_p)^{\mathrm{T}}$ is the vector of parameters, $\Theta = [\theta_{st}]$ is a $p \times p$ symmetric matrix with diagonal elements to be one, $B(\cdot)$ and $C(\cdot)$ are prespecified functions, and the function $A(\theta, \Theta)$ is the normalizing constant to guarantee (1.3) to be a probability density function.

The density function (1.3) provides a general form, which includes many useful cases. For example, the Gaussian graphical model can be derived with the specification of $B(x_{it}) = \frac{x_{it}}{\sigma_t}$ and $C(x_{it}) = -\frac{x_{it}^2}{2\sigma_t^2}$, where $\sigma_t$ is a dispersion parameter to scale the covariate, and its formulation is given by

$$f(x_i; \theta, \Theta) = \exp\left\{\sum_{k\in V} \frac{1}{\sigma_k} \theta_k x_{ik} + \sum_{(s,t)\in E} \frac{1}{\sigma_s \sigma_t} \theta_{st} x_{is} x_{it} - \sum_{k\in V} \frac{1}{\sigma_k^2} x_{ik}^2 - A(\theta, \Theta)\right\}, \quad (1.4)$$

where $\sigma_k$ is a scale parameter for $X_{ik}$, and $A(\theta, \Theta)$ is the normalizing constant. When the covariates follow a Bernoulli distribution, the Ising model can be derived from (1.3), given by

$$f(x_i; \theta, \Theta) = \exp\left\{\sum_{(s,t)\in E} \theta_{st} x_{is} x_{it} - A(\theta, \Theta)\right\}, \quad (1.5)$$

where $B(x_{ti}) = x_{it}$, $C(x_{it}) = 0$, and $A(\theta, \Theta)$ is a normalizing constant.

There are two methods for the parameter estimation of $\theta$ and $\Theta$. The first method is to estimate the parameter based on the global likelihood. For example, for the Gaussian graphical model, the estimator can be estimated by maximizing the rescaled global log-likelihood $L(\Theta; X)$, which takes the form

$$L(\Theta; X) = \log \mathrm{Det}(\Theta) - \mathrm{Tr}(S\Theta) - \lambda \|\Theta\|_1,$$

where $S = \frac{1}{n}\sum_{i=1}^n x_i x_i^{\mathrm{T}}$ is the empirical covariance matrix, $\|\Theta\|_1 = \sum_{s\neq t} |\theta_{st}|$ is the $\ell_1$-norm of the off-diagonal entries of $\Theta$, $\lambda$ is the shrinkage parameter controlling the strength of the penalty, and $\mathrm{Det}(\cdot)$ and $\mathrm{Tr}(\cdot)$ are, respectively, the determinant and trace of a matrix.

In practice, the parameter can be estimated through the *graphical least absolute shrinkage and selection operator* (LASSO) algorithm (Friedman et al., 2008). Due to the complexity of the computation, the algorithm generally takes a long time. This method seems to be mainly applied for Gaussian graphical models. An alternative method of estimating $\Theta$ is based on the idea of the neighborhood-likelihood.

For a given vertex $s \in V$, we use

$$X_{(-s)} = \{X_t : t \in V \setminus \{s\}\}$$

to denote the collection of all other random variables in the graph except $X_s$. Based on the Markov independence assumption, we define the neighborhood set for $s$

$$\mathcal{N}(s) = \{t \in V : (s,t) \in E\}$$

to be the set of relevant variables for variable $X_s$.

As shown in Figure 1.2, the set $\mathcal{N}(s)$ is a cut set that separates $\{s\}$ from the remaining variables.

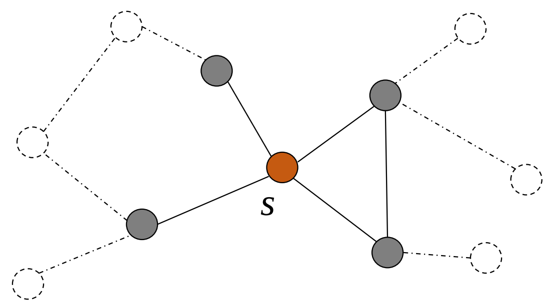


Figure 1.2: *The vertices in gray compose the neighborhood set of vertex s, $\mathcal{N}(s)$*

Write $\theta_{(-s)} = (\theta_{s1}, \ldots, \theta_{s,(s-1)}, \theta_{s,(s+1)}, \ldots, \theta_{sp})^{\mathrm{T}}$ and let $\ell\left(\theta_{(-s)}\right)$ be the log-likelihood for $\theta_{(-s)}$ scaled by $-1/n$,

$$\begin{aligned} \ell\left(\theta_{(-s)}\right) &= -\frac{1}{n} \log \left\{ \prod_{i=1}^{n} P(X_s | X_{(-s)}) \right\} \\ &= -\frac{1}{n} \sum_{i=1}^{n} \left\{ \theta_s B(X_{is}) + \sum_{t \in \mathcal{N}(s)} \theta_{st} B(X_{is}) B(X_{it}) + C(X_{is}) + D(\theta) \right\}, \end{aligned}$$

where $B(\cdot)$ and $C(\cdot)$ are functions specified the same as in (1.3) and $D(\theta)$ is the log-normalization constant.

Then for $s \in V$, we obtain an estimator for $\theta_{(-s)}$, denoted as $\widehat{\theta}_{(-s)}$, by minimizing

$$\ell\left(\theta_{(-s)}\right) + \lambda \|\theta_{(-s)}\|_1,$$

where $\lambda$ is the tuning parameter and $\|\cdot\|_1$ is the $\ell_1$-norm. Although this method is fast in implementation, it has a drawback that the resulting precision matrix $\Theta$ might be asymmetric. Namely, $\hat{\theta}_{st} \neq \hat{\theta}_{ts}$ for some $s, t \in V$, where $\hat{\theta}_{st}$ and $\hat{\theta}_{ts}$ represent the estimates for the $(s,t)$th and the $(t,s)$th entry of $\Theta$. To overcome this problem, extra rules are applied to the estimates. For example, AND rule decides $\widehat{\theta}_{st}$ and $\widehat{\theta}_{ts}$ to be nonzero only if both of $\widehat{\theta}_{st}$ and $\widehat{\theta}_{ts}$ are nonzero; OR rule decides $\widehat{\theta}_{st}$ and $\widehat{\theta}_{ts}$ to be nonzero if either of $\widehat{\theta}_{st}$ and $\widehat{\theta}_{ts}$ is nonzero (Meinshausen and Bühlmann, 2006).

## 1.4 Genome-Wide Association Study

*Genome-wide association studies* (GWAS), also known as *whole-genome association studies* (WGAS), are observational studies searching for causal genetic variants that are associated with the responses of primary interest by scanning over a genome-wide set of genetic variants in different individuals. GWAS often focuses on the associations between *single-nucleotide polymorphisms* (SNPs) and clinical outcomes, such as diseases.

A genome-wide association study is often conducted in two stages (Wason and Dudbridge, 2012). In the first stage, candidate SNPs are selected through a simple model, such as linear regression, to gain computation speed. A set of candidate SNPs are selected according to the strength of the associations. In this stage, due to a large number of tests, several techniques can be applied to address the multiple testing issue, such as Bonferroni correction and the Benjamini-Hochberg Procedure (Benjamini and Hochberg, 1995), etc.

In the second stage, a more advanced approach is applied to study the association between the responses and the candidate SNPs screened from the first stage. There are two purposes for this stage. Firstly, the results getting from the first stage are rudimentary and need further validation. Secondly, advanced approaches can be used to study more complex research problems, such as identifying the possible genetic pathways and the pleiotropy effects to be introduced in Section 1.4.1. In this thesis, our research interest lies in the second stage of the genome-wide association study.

The genome-wide association studies involve several new data features. One important example is the *population stratification*. In the following subsections, we first introduce the basic concepts of statistical genetics and then review several methods concerning the population stratification.

### 1.4.1 Basic Concepts in Statistical Genetics

*Pleiotropy* is a common phenomenon in genetics that one gene is simultaneously associated with multiple traits. It has been widely studied in various types of research including Phenylketonuria (e.g., Penrose, 1951), Schizophrenia (e.g., Navarrete et al., 2013), and many others (e.g., Kraja et al., 2014).

As opposed to the pleiotropy effect, the term *polygenic* refers to the phenomenon that a group of genetic variants is associated with the same phenotype outcome. Examples of human polygenic traits are height, skin color, and eye color.

In genetics, the genes can not only have a complex association with multiple traits but also interact with each other in a complex manner through a collection of molecular regulators. To understand the mechanism of how the genes interact, these interactions are often modeled as a network, which is the so-called *gene regulatory network* (GNR). In the network, each gene will be expressed as a node and an interaction between two genes will be represented by an edge (Yu et al., 2015).

Although there are different types of graphs, the most commonly used graph in the gene regulatory network is the hub graph. In a hub graph, some of the nodes have a number of links that greatly exceed the average number and these nodes are called hub nodes. In GRN, the genes that are highly connected with other genes are called hub genes. Recently, because the hub genes play a key role in biological processes and are informative to uncover the mechanism of diseases, identifying hub genes has attracted a lot of the research interest (Akavia et al., 2010).

### 1.4.2 Population Stratification

The population features, such as the ethnicity for the human being data, can serve as confounders in genetic association studies, due to the non-random mating within populations, which is usually caused by the geological separation. To avoid the spurious association of the genetic variant and the response, several strategies have been developed.

The first method is to adjust the population stratification by the multi-trait mixed model (MTMM). The MTMM is an extension of the linear mixed model by incorporating the relatedness among the subjects into the covariance matrix of random effects. To be specific, the MTMM is defined as

$$Y_i = \beta^{\mathrm{T}} X_i + u_i + \epsilon_i, \qquad \text{for } i = 1, \ldots, n,$$

where $Y_i$ is a continuous trait, $X_i$ includes the environmental covariates and genotype covariates, $\epsilon_i$ is the random error independent of $\{X_i, u_i\}$, $\beta$ is the vector of regression coefficients, $u_i$ is the random effect representing the confounding effect resulting from the subject dependent populations stratification. Write $u = (u_1, \ldots, u_n)^{\mathrm{T}}$. Often, $u$ is assumed to follow $N(0, \sigma_g^2 R)$, where $\sigma_g^2$ is a scale parameter and $R$ is an $n \times n$ positive definite matrix representing the pairwise relatedness among subjects. The relatedness matrix $R$ can be determined by various methods according to the different nature of data. For example, when pedigree data are available, $R$ is determined by the kinship matrix (Lange, 2003, Page 82). An alternative approach is to estimate relatedness matrices from genome-wide SNPs. The detail of specifying the relatedness matrix is to be discussed in Section 2.1.1.

The second method is to control the confounding by including the principal components of the relatedness matrix as fixed effects. To conduct the principal component analysis, the relatedness matrix is decomposed using the eigenvalue decomposition (EVD),

$$R = LDL^{\mathrm{T}},$$

where the columns of $L$ are eigenvectors of $R$, and $D$ is a diagonal matrix of positive eigenvalues of $R$. Let $F = RLD^{-\frac{1}{2}}$ be the matrix of principal components of the genetic information for the subjects, where each row of $F$, denoted as $F_i$, is the principal components for subject $i$. Here, $D^{-\frac{1}{2}}$ represents the diagonal matrix whose diagonal elements are the reciprocal and square root of those diagonal elements of $D$.

The principal components based model can be cast as

$$Y_i = \beta_1^{\mathrm{T}} X_i + \beta_2^{\mathrm{T}} F_i + \epsilon_i,$$

where $\beta = (\beta_1^{\mathrm{T}}, \beta_2^{\mathrm{T}})^{\mathrm{T}}$ is the vector of regression coefficients, $F_i$ is the first few largest principal components of the relatedness matrix $R$.

Compared to the principal component based linear model, the MTMM model can be used to account for higher-rank confounding. On the other hand, controlling confounders by including fixed effects can circumvent the intensive computational burden.

## 1.5 Thesis Topics and the Outline

This thesis tackles several important problems and offers new additions to the literature. The thesis contains seven chapters with the last chapter concluding the thesis and the appendix including additional materials for Chapters 2–6. The topics and the development of Chapters 2–6 are outlined as follows.

### 1.5.1 Latent Variable Model with Bivariate Mixed Responses Subject to Measurement Error and Misclassification

In genetic association studies, due to the concern of population stratification, a multi-trait mixed model (MTMM) is often considered. In the MTMM, random effects are used not only to model the correlation for the multiple outcomes of the same subject but also to adjust for the relatedness among subjects. The MTMM has been widely applied in genome-wide association studies in various setting (Zhang et al., 2010; Korte et al., 2012; Zhou and Stephens, 2014; Furlotte and Eskin, 2015).

While mixed effects models have been widely used, they do not automatically ensure the inference results to be valid without conditions. A critical condition of using such models hinges on the precise measurements of the variables. Measurement error and misclassification, however, are typical features in genetic studies, but they are often ignored in most applications. Even if the practitioners aware the importance of the measurement, they might still ignore the effect of measurement error in genome-wide association studies due to intensive computational burdens, and accounting for the measurement error and misclassification does not only require great efforts of modeling but also introduce a large number of algorithm implementations. This being said, it is important to accommodate the mismeasurement to obtain valid results for genetic studies. Because of the advances in computer technology, implementing time-consuming algorithms does not seem to be an obstacle as before. A few studies, such as Hossain et al. (2009), Smith et al. (2013), and Rekaya et al. (2016), investigated some methods of analyzing genetic data with misclassification in a variable.

In the literature of response mismeasurements, there has been research exploring either *measurement error* in a continuous response (e.g., Buonaccorsi, 1991; Pepe et al., 1994; Buonaccorsi, 1996), or *misclassification* in a discrete response (e.g., Neuhaus, 2002; Ramalho, 2002; Prescott and Garthwaite, 2002; Paulino et al., 2003; Chen et al., 2011). However, no available work has been directed to deal with mixed responses with mismeasurement in continuous and discrete components, although there were a few studies simultaneously addressing mixed types of mismeasurement in covariates (Spiegelman et al., 2000; Yi et al., 2015; Zhang and Yi, 2019).

In Chapter 2, we consider the problem of joint modeling mixed responses with a continuous and a binary variable respectively subject to measurement error and misclassification. We employ the bivariate regression model with a latent variable which features the dependence of the response components as well as the population stratification. We propose two methods, the induced likelihood method and the EM algorithm approach, to account for both measurement error and misclassification of the responses in inferential procedures. A general framework is considered for the specification of the mismeasurement processes. We show that both methods yield valid estimation results.

### 1.5.2 Estimating Equation Approach with Bivariate Mixed Responses Subject to Measurement Error and Misclassification

Correlated mixed types of data, containing both continuous and discrete variables arise commonly from clinical trials and genetic association studies. Many models have been proposed for analyzing such data. In addition to the well-studied likelihood approaches, marginal models have also been widely used to handle correlated mixed types of data due to the advantage of robustness to certain model misspecification since no full distributional assumptions are required. For example, generalized estimating equations, proposed by Liang and Zeger (1986), analyze the longitudinal data marginally and require only modeling of the first and second moments. This method has been further extended by many authors including Prentice and Zhao (1991); Hall and Severini (1998); Pan (2001); Wang and Long (2011); Wang et al. (2012).

Marginal methods are useful in joint modeling of mixed responses, such as a continuous response and a discrete response (e.g., Liu et al., 2009). However, such methods rely on a crucial condition that the variables must be precisely measured. It is well known that the mismeasurement in responses induces both biased parameter estimation and efficiency loss (e.g., Neuhaus, 1999, 2002; Chen et al., 2011).

In Chapter 3, we use the bivariate generalized estimating equation to analyze mixed continuous and discrete responses subject to mismeasurement. We develop an insertion strategy to form unbiased estimating functions to accommodate the effects of measurement error and misclassification in responses. We consider different study designs including the main study/internal validation study and the main study/external study (Spiegelman et al., 2000; Yi et al., 2018). We evaluate the proposed methods both theoretically and numerically.

### 1.5.3 Generalized Network Mixed Model in Discovering Gene Regulatory Network with Mixed Responses Subject to Measurement Error and Misclassification

In genetic analysis, genes can not only have a complex association with multiple traits but also interact with each other in a complex manner through molecular regulators. To understand how these genes may interact, *gene regulatory networks* (GRN) are often employed to describe the associations among genes, where genes are taken as nodes and an interaction between two genes is featured by an edge (Friedman, 2004). While many methods have been proposed for studying gene regulatory networks, a noticeable limitation is that the computational procedures are problem-specific, which hinders their application scope. To overcome this issue, several studies of applying graphical models to construct gene regulatory networks were motivated (e.g., Li et al., 2012; Yu et al., 2015).

Although graphical models have been developed to construct gene regulatory networks, most available work only focused on the modeling of covariates and did not consider how to model the relationship between network structured covariates and a response variable, let alone for the case with mixed bivariate responses with both continuous and discrete components. With the analysis of mixed responses, generalized estimating equation methods are useful because of its robustness of not requiring the specification of the joint distribution of the response variables as well as its flexibility of accommodating different covariance structures of the responses. The validity of such methods, however, is vulnerable to the mismeasurement of the response variables.

While it is well studied that mismeasurement in a discrete response typically breaks down the usual inference methods and ignoring this feature commonly yields erroneous inference results (e.g., Neuhaus, 1999; Chen et al., 2011), to our knowledge, there has not been research on dealing with error-contaminated mixed responses of both discrete and continuous components, let alone for their relationship with covariates of network structures.

In Chapter 4, we tackle this problem and make the following contributions: (1) we propose a new class of generalized network structured models to delineate the relationship between bivariate responses and covariates of a network structure; (2) we develop a two-stage inferential procedure to identify the network structure for covariates and to address the mismeasurement effects in responses of both continuous and discrete components; (3) we rigorously establish the asymptotic results for the proposed estimators and study the efficiency issues for different methods; (4) our methods offer tools for a broad variety of applications to handle error-prone data with complex association structures. For example, they can be applied in genetic studies to simultaneously identify the gene regulatory network and study the association between the gene network and mixed type traits with the effects of mismeasurement accounted for.

To be specific, we develop a generalized network structured model which incorporates the graphical structure in the generalized linear models through a two-step procedure. In the first step, we identify the network structure in the covariates via the Gaussian graphical model. In the second step, we build generalized estimating equations to study the association between the bivariate responses and the network structured covariates selected from Step 1, where the effects due to the contamination in the responses are accommodated for valid inferential procedures. We start with a simple situation where the model parameters for the mismeasurement processes are known; this development highlights the idea of how effects of mismeasurement in the mixed responses can be accounted for in combination with the examination with the network structure for covariates. Furthermore, we extend the development to accommodating the cases where the parameters for the mismeasurement models are unknown and must be estimated from an additional validation subsample.

### 1.5.4 Zero-Inflated Poisson Models with Measurement Error in Response

Research on zero-inflated Poisson models has become active and has attracted various studies from several perspectives. Rodrigues (2003) and Klein et al. (2015) pursued a Bayesian inference analysis with zero-inflated models. Todem et al. (2016) developed a marginal model for the zero-inflated Poisson data. Xiang et al. (2007) and Yang et al. (2010) proposed a score test under the zero-inflated Poisson model.

Measurement error in count data has been scarcely explored, which basically has two challenges. Firstly, the count variable is an integer, and thus the traditional measurement error models, such as the classical additive model is not applicable in this situation. Secondly, the variables are bounded blow but unbounded above, because the observed values are always positive.

In Chapter 5, we propose a measurement error model that is unique for error-corrupted count data by incorporating two possible sources of measurement error. We explore the validity of statistical inference when measurement error in count data is ignored. We develop a Bayesian framework to account for the measurement error effects, which avoids the unidentifiability issue through the inclusion of weakly informative priors.

### 1.5.5 Autoregressive Models with Data Subject to Measurement Error

Time series data are common in the fields of epidemiology, economics, and engineering, and various models and methods have been developed for analyzing such data. The validity of these methods, however, hinges on the condition that time series data are precisely collected. This condition is restrictive in applications. Measurement error is often inevitable. In the study of air pollution, for example, it is difficult or even impossible to precisely obtain the true measurement of the air population.

Some work on time series subject to measurement error is available in the literature. Tanaka (2002) proposed a Lagrange multiplier test to assess the presence of measurement error in time series data. Staudenmayer and Buonaccorsi (2005) explored the classical measurement error model for the autoregressive model. Tripodis and Buonaccorsi (2009) studied measurement error in forecasting using the Kalman filter. Dedecker et al. (2014) considered non-linear AR(1) model with measurement error. Despite available discussions of measurement error in time series, several limitations restrict the application scope of the existing work. Most available methods consider only the autoregressive models without the drift and assume the simplest additive measurement error model. Furthermore, most work involves a complex formulation to adjusted for the measurement error effects, which is not straightforward to implement for practitioners. In addition, to our knowledge, there is no available work addresses measurement error effects on prediction under the autoregressive model.

In Chapter 6, we systematically explore the analysis of error-prone time series data under the autoregressive model. We propose two types of models to delineate measurement error processes: the additive regression models and multiplicative models. These modeling schemes offer us great flexibility in facilitating different applications. We investigate the impact of the naive analysis which ignores the feature of measurement error in the inferential procedures, and we obtain analytical results for characterizing the biases due to the naive analysis. We develop an estimating equation approach to adjust for the measurement error effects on time series analysis. We establish asymptotic results for the proposed estimators and develop the theoretical results for the forecasting of times series in the presence of measurement error. Finally, we describe a block bootstrap algorithm for computing standard errors of the proposed estimators.

Our work is partially motivated by the data of COVID-19, a wide-spread disease that has become a global health challenge and has caused over ten million infections and half million deaths as of August, 2020. Because of the special features of the disease, the data of COVID-19 introduce many new challenges: 1) due to the asymptomatic infected cases and the patients with light symptoms who do not go to hospitals, the reported cases with COVID-19 are typically smaller than the true number of infected cases; 2) due to the limited test resources, many infected cases are not able to be identified instantly; and 3) the varying incubation periods lead to the delay of the identification of the infections. Consequently, the discrepancy between the reported case number and the true case number can be substantial, and ignoring these features and applying the traditional time series analysis method would no longer produce valid results.

In Chapter 6, we apply the developed methods to analyze the COVID-19 data. We are interested in studying how the mortality rate in a region may change over time and describing the trajectory of the death rate. While the mortality rate of a disease is defined as the death number divided by the case number, the determination of the mortality rate of COVID-19 is challenging. In contrast to the standard definition, Baud et al. (2020) estimated mortality rates by dividing the number of deaths on a given day by the number of patients with confirmed COVID-19 infection 14 days before, with the consideration of the maximum incubation time to be 14 days. Due to the unique features of COVID-19, there does not seem to be a precise way to define the mortality rate of COVID-19. In this chapter, we conduct a sensitivity analysis to assess the severity of the pandemic by using different definitions of the mortality rate and considering different ways of modeling measurement error in the data.

Using the data collected from the dashboard developed by Johns Hopkins University (JHU-CSSE, Dong et al., 2020), we analyze the mortality rates of COVID-19 and conduct forecasting of the COVID-19 related mortality rate for the four most populated provinces in Canada, British Columbia, Ontario, Quebec, and Alberta.

# Chapter 2

# Latent Variable Models with Bivariate Mixed Responses Subject to Measurement Error and Misclassification

In this chapter, we focus on the effects of measurement error and misclassification on analysis of the mixed responses postulated with latent variable models. The notation and the setup for the response model, the measurement error model as well as the misclassification models are introduced in Section 2.1. We describe the induce likelihood method in Section 2.2 and the EM algorithm method in Section 2.3. We extend the method to facilitating pedigree data with correlated subjects in Section 2.4. Simulations studies are conducted to evaluate the performance of the two methods in Section 2.5. To illustrate the usage of the methods, in Section 2.6 we conduct numerical analysis using the mice data arising from a genome-wide association study.

## 2.1 Model Setup

### 2.1.1 Response Model

Suppose $n$ subjects are recruited independently in the study. For subject $i = 1, \ldots, n$, two possibly correlated responses $Y_{ij}$ are measured for $j = 1, 2$, where $Y_{i1}$ is a continuous variable, and $Y_{i2}$ is a binary variable. Write $Y_i = (Y_{i1}, Y_{i2})^{\mathrm{T}}$. Let $X_i = (X_{i1}, \ldots, X_{ip_x})^{\mathrm{T}}$ denote the covariate vector for subject $i$, where $p_x$ is the dimension of the covariates. To facilitate the association structure between the mix-type responses $Y_{i1}$ and $Y_{i2}$, we introduce a latent variable $u_i$. Conditional on random effects $u_i$ and covariates $X_i$, we assume that $Y_{i1}$ and $Y_{i2}$ are independent, each having a probability density or mass function from the exponential family

$$f(y_{ij}|u_i, x_i) = \exp\left[\{y_{ij}\eta_j - b_j(\eta_j)\}/d_j(\phi_j) + c_j(y_{ij}, \phi_j)\right],$$

for $i = 1, \ldots, n$ and $j = 1, 2$ where $b_j(\cdot)$, $c_j(\cdot)$ and $d_j(\cdot)$ are known functions, $\eta_j$ is a canonical parameter, and $\phi$ is a dispersion parameter.

Let $\mu_{ij} = E(Y_{ij}|u_i, X_i)$ be the conditional mean of response $Y_{ij}$ for $j = 1, 2$, and then $\mu_{ij} = b_j'(\eta_j)$. To explicitly describe the dependence of $\mu_{ij}$ on random effects and the covariates, we consider a bivariate generalized linear mixed model

$$\begin{bmatrix} g_1(\mu_{i1}) \\ g_2(\mu_{i2}) \end{bmatrix} = \begin{bmatrix} \beta_1^{\mathrm{T}} X_i \\ \beta_2^{\mathrm{T}} X_i \end{bmatrix} + \begin{bmatrix} u_i \\ u_i \end{bmatrix}, \tag{2.1}$$

where $g_1(\cdot)$ and $g_2(\cdot)$ are the link functions, determined by $g_1^{-1}(\cdot) = b_1'(\cdot)$ and $g_2^{-1}(\cdot) = b_2'(\cdot)$; $\beta = (\beta_1^{\mathrm{T}}, \beta_2^{\mathrm{T}})^{\mathrm{T}}$ is the vector of regression coefficients that is of primary interest; and $u_i$ is a random effect. For the continuous response $Y_{i1}$ with a normal distribution, $b_1'(t)$ is taken as $t$; and for the binary response $Y_{i2}$, $b_2'(t) = \exp(t)/(1 + \exp(t))$, yielding that $g_1(t) = t$ and $g_2(t) = \log \frac{t}{1-t}$, respectively, where $t$ represents the argument of functions.

Write $u = (u_1, \ldots, u_n)^{\mathrm{T}}$. Often, $u$ is assumed to follow a normal distribution $N(0, \sigma_g^2 R)$, where $\sigma_g^2$ is an unknown scale parameter, and $R = [R_{jk}]_{n\times n}$ is a specified positive definite matrix with the $(j, k)$ element $R_{jk}$ determined by the study design, where $j, k = 1, \ldots, n$. In applications, a particular specification of $R$ may be imposed to feature a problem-specific association structure. For instance, to reflect the independence among different subjects, Sammel et al. (1997) set $R$ as a diagonal matrix with the diagonal elements being given (such as 1 or other values). In the development in Sections 2-4, we consider the case where $R$ is a given diagonal matrix, and in Section 5, we extend the diagonal $R$ to a blockwise diagonal matrix to reflect the correlation among the subjects.

Model (2.1) is useful for characterizing the dependence of mix-type responses on covariates (Sammel et al., 1997). This model can be conveniently used to analyze genetic data with mix-type responses, where the genotype information may be summarized as the covariates $X_i = (X_{i1}, \ldots, X_{ip_x})^{\mathrm{T}}$, where for $k = 1, \ldots, p_x$, covariate $X_{ik}$ can be continuous (e.g., representing a continuous measurement of an environmental effect), or binary (e.g., representing a clinical treatment); $X_{ik}$ can be ordinal referring to, for example, the genotype.

For instance, we observe the genotypes through genetic markers (called single nucleotide polymorphisms, SNPs) in each locus (the location of a gene on the genome). For $k = 1, \ldots, p_x$, let $X_{ik}^{(1)}$ and $X_{ik}^{(2)}$ stand for the nucleotides of SNP $k$ for subject $i$ inherited from the father and mother respectively. Each SNP consists of two nucleotides, each being one of the two types of alleles, "$A_1$" and "$A_2$". Hence, all possible forms of a SNP are "$A_1A_1$", "$A_1A_2$" and "$A_2A_2$", derived from different combinations of the alleles. Then the covariate $X_{ik}$, representing $k$th SNP for subject $i$, is coded according to the nucleotide level of $A_2$, given by

$$X_{ik} = I(X_{ik}^{(1)} = A_2) + I(X_{ik}^{(2)} = A_2), \tag{2.2}$$

yielding an ordinal variable $X_{ik}$ taking values of 0, 1 and 2, where $I(\cdot)$ is an indicator function.

### 2.1.2 Measurement Error and Misclassification Models

For $i = 1, \ldots, n$, suppose that the response variables $Y_{i1}$ and $Y_{i2}$ are subject to mismeasurement and that their precise measurements may not be observed for every subject. Let $Y_{i1}^*$ and $Y_{i2}^*$ denote the observed measurements of $Y_{i1}$ and $Y_{i2}$, respectively; they are also called surrogate measurements of $Y_{i1}$ and $Y_{i2}$. Let $Z_i = (Z_{i1}, \ldots, Z_{ip_z})^{\mathrm{T}}$ denote the covariate vector involved in the measurement error and misclassification process for subject $i$ where $p_z$ is the dimension of $Z_i$. For ease of exposition, we assume that $Z_i$ is a subset of $X_i$; if this is not the case, we can modify our initial definition of $X_i$ to include $Z_i$ as its part.

To describe the mismeasurement processes, we consider the factorization

$$f(y_{i1}^*, y_{i2}^* | y_{i1}, y_{i2}, u_i, x_i) = f(y_{i1}^* | y_{i2}^*, y_{i1}, y_{i2}, u_i, x_i) f(y_{i2}^* | y_{i1}, y_{i2}, u_i, x_i). \tag{2.3}$$

We assume that

$$f(y_{i1}^* | y_{i2}^*, y_{i1}, y_{i2}, x_i, u_i) = f(y_{i1}^* | y_{i1}, y_{i2}, x_i) = f(y_{i1}^* | y_{i1}, y_{i2}, z_i), \tag{2.4}$$

and

$$f(y_{i2}^* | y_{i1}, y_{i2}, x_i, u_i) = f(y_{i2}^* | y_{i1}, y_{i2}, x_i) = f(y_{i2}^* | y_{i2}, x_i) = f(y_{i2}^* | y_{i2}, z_i). \tag{2.5}$$

Assumptions (2.4) and (2.5) basically say that conditional on the true responses and the covariates, surrogate measurements and random effects are independent. The assumptions also suggest that $Z_i$ completely reflects the dependence on the covariates when featuring the measurement error and misclassification processes. While the last two equalities in

(2.5) are not needed to assume, having them offers us a convenient way to model the misclassification probabilities; see Yi et al. (2015).

Let $\pi_{i0} = P(Y_{i2}^* = 1 | Y_{i2} = 0, Z_i)$ and $\pi_{i1} = P(Y_{i2}^* = 0 | Y_{i2} = 1, Z_i)$ be the misclassification probabilities that may depend on the covariates $Z_i$. We consider logistic models for the misclassification process,

$$\text{logit } \pi_{i1} = \alpha_{01} + \alpha_{z1}^{\mathrm{T}} Z_i,$$

and

$$\text{logit } \pi_{i0} = \alpha_{00} + \alpha_{z0}^{\mathrm{T}} Z_i, \tag{2.6}$$

where $\alpha = (\alpha_{01}, \alpha_{z1}^{\mathrm{T}}, \alpha_{00}, \alpha_{z0}^{\mathrm{T}})^{\mathrm{T}}$ is the vector of the regression parameters.

For the continuous response $Y_{i1}$, we consider a regression model which facilitates possible dependence of $Y_{i1}^*$ on $\{Y_{i1}, Y_{i2}, Z_i\}$,

$$Y_{i1}^* = m(Y_{i1}, Y_{i2}, Z_i; \gamma) + e_i, \tag{2.7}$$

where $e_i$ is the random error independent of $\{Y_{i1}, Y_{i2}, X_i, u_i\}$ and has zero mean and constant variance $\sigma_e^2$, $\gamma$ is the vector of regression coefficients, and $m(\cdot)$ is the mean function that can be linear or nonlinear.

Often, an additive model is considered for (2.7), given by

$$Y_{i1}^* = \gamma_0 + \gamma_1 Y_{i1} + \gamma_2 f(Y_{i2}) + \gamma_3 Z_{i3} + e_i, \tag{2.8}$$

where $f(\cdot)$ is a function of the binary response $Y_{i2}$, $\gamma = (\gamma_0, \gamma_1, \gamma_2, \gamma_3)^{\mathrm{T}}$ is the vector of parameters, and a normal distribution is assumed for $e_i$. We comment that model (2.8) offers the flexibility and convenience of featuring the dependence of surrogate variable $Y_{i1}^*$ on the true responses and covariates, but model identifiability may be a concern if no care is taken. In Appendix A.1, we outline the discussion on this aspect.

## 2.2 Estimation Procedures

### 2.2.1 Induced Likelihood for the Observed Data

To see how the distribution of the observed $Y_i^*$ is different from that of $Y_i$, we derive the conditional distribution of $Y_i^*$ given $\{u_i, X_i\}$. Indeed,

$$
\begin{aligned}
f(y_i^*|u_i,x_i) &= \int_{y_{i1}} \sum_{y_{i2}} f(y_{i1}^*, y_{i2}^*, y_{i1}, y_{i2}|u_i,x_i)\,\mathrm{d}y_{i1} \\
&= \int_{y_{i1}} f(y_{i1}^*|y_{i1}, y_{i2}=1, z_i) f(y_{i2}^*|y_{i2}=1, z_i) f(y_{i1}|u_i,x_i) f(y_{i2}=1|u_i,x_i)\,\mathrm{d}y_{i1} \\
&\quad + \int_{y_{i1}} f(y_{i1}^*|y_{i1}, y_{i2}=0, z_i) f(y_{i2}^*|y_{i2}=0, z_i) f(y_{i1}|u_i,x_i) f(y_{i2}=0|u_i,x_i)\,\mathrm{d}y_{i1},
\end{aligned}
\tag{2.9}
$$

where in the second equality, we use (2.4), (2.5), and the conditional independence of $Y_{i1}$ and $Y_{i2}$ given $\{u_i, X_i, Z_i\}$. Using the model formulations in Section 2.1, we have the following expressions for the terms of (2.9):

$$
\begin{aligned}
f(y_{i1}^*|y_{i1}, y_{i2}, z_i) &= \frac{1}{\sqrt{2\pi\sigma_e^2}} \exp\left[-\frac{\{y_{i1}^* - m(y_{i1}, y_{i2}, z_i;\gamma)\}^2}{2\sigma_e^2}\right]; \\
f(y_{i2}^*|y_{i2}=q, z_i) &= \left\{\frac{\exp(\alpha_{0q} + \alpha_{zq}^{\mathrm{T}} z_i)}{1+\exp(\alpha_{0q} + \alpha_{zq}^{\mathrm{T}} z_i)}\right\}^{q(1-y_{i2}^*)+(1-q)y_{i2}^*} \\
&\quad \times \left\{\frac{1}{1+\exp(\alpha_{0q} + \alpha_{zq}^{\mathrm{T}} z_i)}\right\}^{qy_{i2}^*+(1-q)(1-y_{i2}^*)} \qquad \text{with } q = 0, 1; \\
f(y_{i1}|x_i, u_i) &= \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{-\frac{(y_{i1} - \beta_1^{\mathrm{T}} x_i - u_i)^2}{2\sigma^2}\right\}; \\
f(y_{i2}|x_i, u_i) &= \left\{\frac{\exp(\beta_2^{\mathrm{T}} x_i + u_i)}{1+\exp(\beta_2^{\mathrm{T}} x_i + u_i)}\right\}^{y_{i2}} \left\{\frac{1}{1+\exp(\beta_2^{\mathrm{T}} x_i + u_i)}\right\}^{1-y_{i2}}.
\end{aligned}
$$

Consequently, the conditional distribution of $Y_i^*$, given $X_i$, is given by

$$
\begin{aligned}
f(y_i^*|x_i) &= \int f(y_i^*, u_i|x_i)\,\mathrm{d}u_i \\
&= \int f(y_i^*|u_i, x_i) f(u_i|x_i)\,\mathrm{d}u_i,
\end{aligned}
\tag{2.10}
$$

where $f(u_i|x_i) = \frac{1}{(2\pi\sigma_g^2 R_{ii})^{\frac{1}{2}}} \exp\left(-\frac{u_i^2}{2\sigma_g^2 R_{ii}}\right)$, with $R_{ii}$ being the $i$th diagonal elements of matrix $R$, and $f(y_i^*|u_i, x_i)$ is determined by (2.9).

Let $\theta = (\beta^{\mathrm{T}}, \gamma^{\mathrm{T}}, \alpha^{\mathrm{T}}, \sigma^2, \sigma_e^2, \sigma_g^2)^{\mathrm{T}}$. Inference about $\theta$ can be carried out using the likelihood for the observed data, given by

$$L(\theta) = \prod_{i=1}^{n} f(y_i^* | x_i), \tag{2.11}$$

where $f(y_i^*|x_i)$ is determined by (2.10) with the dependence on parameter $\theta$ suppressed in the notation.

Maximizing $L(\theta)$ with respect to $\theta$ gives the maximum likelihood estimator, say $\hat{\theta}$, of $\theta$. Under regularity conditions, this is equivalent to solving

$$\sum_{i=1}^{n} S_i(\theta) = 0, \tag{2.12}$$

where $S_i(\theta) = \frac{\partial}{\partial \theta} \log f(y_i^*|x_i)$. Typically, (2.12) does not have an analytic solution; solving (2.12) usually requires a numerical method, such as the Newton-Raphson method.

Under regularity conditions, $\hat{\theta}$ is a consistent estimator of $\theta$, and

$$\sqrt{n}(\hat{\theta} - \theta) \xrightarrow{d} N(0, I^{-1}(\theta)) \qquad \text{as} \qquad n \to \infty,$$

where

$$I(\theta) = E\left\{S_i(\theta) S_i^{\mathrm{T}}(\theta)\right\}. \tag{2.13}$$

### 2.2.2 Implementation

The estimation of the parameter $\theta$ is conducted by maximizing (2.11). We realize this by using the Newton-Rhaphson algorithm in combination with the Monte Carlo method. Let $N_u$ and $N_y$ be prespecified large integers.

Step 1. Take an initial parameter value $\theta^{(0)} = ({\beta^{(0)}}^{\mathrm{T}}, {\gamma^{(0)}}^{\mathrm{T}}, {\alpha^{(0)}}^{\mathrm{T}}, {\sigma^2}^{(0)}, {\sigma_e^2}^{(0)}, {\sigma_g^2}^{(0)})^{\mathrm{T}}$ and set the iteration index $t = 0$.

Step 2. At iteration $(t+1)$, for $i = 1, \ldots, n$, independently generate a sequence of values, $\{u_i^{[1]}, \ldots, u_i^{[N_u]}\}$, from $N(0, {\sigma_g^2}^{(t)} R_{ii})$, where ${\sigma_g^2}^{(t)}$ is the parameter $\sigma_g^2$ evaluated at the $t$th iteration and $R_{ii}$ is the $(i,i)$ entry of matrix $R$. For $i = 1, \ldots, n$ and $a = 1, \ldots, N_u$, generate $\{y_{i1}^{[a,1]}, \ldots, y_{i1}^{[a,N_y]}\}$ from $N({\beta^{(t)}}^{\mathrm{T}} X_i + u_i^{[a]}, {\sigma^2}^{(t)})$, where ${\sigma^2}^{(t)}$ and $\beta^{(t)}$ are, respectively, the parameter $\sigma^2$ and $\beta$ evaluated at the $t$th iteration.

Step 3. The likelihood function in (2.11) is approximated by

$$\widetilde{L}(\theta) = \prod_{i=1}^{n} \left\{ \frac{1}{N_u} \sum_{a=1}^{N_u} f(y_i^* | u_i^{[a]}, x_i) \right\}, \tag{2.14}$$

where

$$\begin{aligned} f(y_i^*|u_i^{[a]}, x_i) = {} & \frac{1}{N_y} \sum_{b=1}^{N_y} f(y_{i1}^* | y_{i1}^{[a,b]}, y_{i2} = 1, z_i) f(y_{i2}^* | y_{i2} = 1, z_i) f(y_{i2} = 1 | u_i^{[a]}, x_i) \\ & + \frac{1}{N_y} \sum_{b=1}^{N_y} f(y_{i1}^* | y_{i1}^{[a,b]}, y_{i2} = 0, z_i) f(y_{i2}^* | y_{i2} = 0, z_i) f(y_{i2} = 0 | u_i^{[a]}, x_i). \end{aligned}$$

Step 4. Compute $\widetilde{S}(\theta^{(t)}) = \frac{\partial}{\partial \theta} \log \widetilde{L}(\theta)$ and $\widetilde{I}(\theta^{(t)}) = -\frac{\partial^2}{\partial \theta \partial \theta^{\mathrm{T}}} \log \widetilde{L}(\theta)$, and update $\theta^{(t+1)}$ by

$$\theta^{(t+1)} = \widetilde{I}(\theta^{(t)})^{-1} \widetilde{S}(\theta^{(t)}) + \theta^{(t)}.$$

Step 5. Check if the $\theta^{(t+1)}$ converges by evaluating $|(\theta^{(t+1)} - \theta^{(t)})/(\theta^{(t)} + c_1)| < c_2$, where $c_1$ and $c_2$ are prespecified small tolerance values. Otherwise, $t := t + 1$, and go back to Step 2.

The Monte Carlo approximation from Step 2 to Step 4 has the computation complexity of order $O(nN_uN_y)$. For an accurate approximation, the computation typically requires a large number of replicates such as 10000 for $N_u$ and $N_y$. Alternatively, one may employ the Gaussian quadrature method (James, 1980) to approximate the integral in (2.11), as discussed in Appendix A.2.

## 2.3 EM Algorithm

In this section, we consider an alternative to estimating model parameters using the EM algorithm. The log-likelihood for the complete data $\{(Y_{i1}, Y_{i2}, Y_{i1}^*, Y_{i2}^*, u_i) : i = 1, \ldots, n\}$, given $X_i$, is

$$\sum_{i=1}^{n} \log f(y_{i1}^*, y_{i2}^*, y_{i1}, y_{i2}, u_i | x_i) = \sum_{i=1}^{n} \log f(y_{i1}, y_{i2}, u_i | y_{i1}^*, y_{i2}^*, x_i) + \sum_{i=1}^{n} \log f(y_{i1}^*, y_{i2}^* | x_i).$$

Thus, the E-step of the EM algorithm evaluates

$$Q(\theta, \theta^{(t)}) = \sum_{i=1}^{n} Q_i(\theta, \theta^{(t)}),$$

where

$$Q_i(\theta, \theta^{(t)}) = E_{Y_{i1}, Y_{i2}, u_i | Y_{i1}^*, Y_{i2}^*, X_i; \theta^{(t)}} \left\{ \log f(Y_{i1}^*, Y_{i2}^*, Y_{i1}, Y_{i2}, u_i | X_i; \theta) \right\},$$

and the expectation is taken with respect to the conditional distribution of $(Y_{i1}, Y_{i2}, u_i)$ given $\{Y_{i1}^*, Y_{i2}^*, X_i\}$ with $\theta$ set as $\theta^{(t)}$, the estimate of $\theta$ at iteration $t$.

The M-step is to maximize $Q(\theta, \theta^{(t)})$ with respect to $\theta$, which is equivalent to solving

$$\sum_{i=1}^{n} \frac{\partial Q_i(\theta, \theta^{(t)})}{\partial \theta} = 0 \tag{2.15}$$

for $\theta$, provided regularity conditions. To obtain the solution of (2.15), we may implement the Newton-Raphson algorithm. An updated estimate $\theta^{(t+1)}$ for $\theta$ at iteration $t$ is given by

$$\theta^{(t+1)} = \theta^{(t)} + \left\{ \sum_{i=1}^{n} \frac{\partial^2}{\partial \theta \partial \theta^{\mathrm{T}}} Q_i(\theta, \theta^{(t)}) \right\}^{-1} \left\{ \sum_{i=1}^{n} \frac{\partial}{\partial \theta} Q_i(\theta, \theta^{(t)}) \right\}, \tag{2.16}$$

where $t = 0, 1, 2, \ldots$, and $\theta^{(0)}$ is an initial value of $\theta$.

The integral involved in (2.16) can be approximated by numeric methods, such as the Monte Carlo or Gaussian quadrature algorithm. Repeat through the E and M steps until convergence of $\{\theta^{(t+1)} : t = 0, 1, \ldots\}$, and let $\widehat{\theta}$ denote the resulting limit. The variance estimate of $\widehat{\theta}$ can be obtained following Louis (1982) or using the bootstrap procedure.

## 2.4 Extension to Handling Clustered Data

In genome-wide association studies (GWAS), facilitating the genetic relatedness for subjects in the same clusters (e.g., families) is important to reflect the cluster structure of data. To this end, we extend the preceding development by allowing the matrix $R$ in Section 2.1.1 to feature the inherent relatedness within the same cluster or family.

Let

$$R = \begin{pmatrix} R_1 & & & \\ & R_2 & & \\ & & \ddots & \\ & & & R_{n_f} \end{pmatrix}$$

be a blockwise-diagonal matrix which delineates the population stratification based on the pedigree information, where $n_f$ is the total number of families (or clusters) and $R_i$ is the $n_i \times n_i$ relatedness matrix of family $i$. For $i = 1, \ldots, n_f$, the $(i_1, i_2)$ element $R_{i_1 i_2}$ of $R_i$ may be, for example, defined as the kinship coefficient (Lange, 2003, Page 82) for subjects $i_1$ and $i_2$ in the family $i$, which is the weighted summation of the probabilities of each allele pair for subjects $i_1$ and $i_2$ to be *identical by descent* (IBD) at the same locus $k$. Here an allele pair is taken as *identical by descent* if the pair has the same type of nucleotide and is inherited from the same ancestor.

To be specific, for $k = 1, \ldots, p_x$,

$$R_{i_1 i_2} = \frac{1}{4} \sum_{l=1}^{2} \sum_{s=1}^{2} P\left(X_{i_1 k}^{(l)} = X_{i_2 k}^{(s)}\right),$$

where $P(X_{i_1 k}^{(l)} = X_{i_2 k}^{(s)})$ is assumed to be identical for all the $k$ and represents the common probability that the two alleles, $X_{i_1 k}^{(l)}$ and $X_{i_2 k}^{(s)}$, are inherited from the same ancestor for $l = 1, 2$ and $s = 1, 2$. Here $X_{ik}^{(1)}$ and $X_{ik}^{(2)}$ represent the nucleotides inherited from the father and mother of subject $i$, respectively; in applications, the probabilities $P\left(X_{i_1 k}^{(l)} = X_{i_2 k}^{(s)}\right)$ are often determined by pedigree data. For instance, $R_{i_1 i_2} = 0.5$ if $i_1$ and $i_2$ are monozygotic twins and $R_{i_1 i_2} = 0.25$ if $i_1$ and $i_2$ has a parent-offspring relationship.

For $i = 1, \ldots, n_f$, let $Y_{i11}, \ldots, Y_{in_i 1}$ be the continuous responses and $Y_{i12}, \ldots, Y_{in_i 2}$ be the binary responses of $n_i$ subjects in the $i$th family, and let $Y_{i11}^*, \ldots, Y_{in_i 1}^*$ and $Y_{i12}^*, \ldots, Y_{in_i 2}^*$ be their corresponding surrogate measurements. For $i = 1, \ldots, n_f$ and $r = 1, \ldots, n_i$, we write $Y_{ir} = (Y_{ir1}, Y_{ir2})^{\mathrm{T}}$, $Y_{ir}^* = (Y_{ir1}^*, Y_{ir2}^*)^{\mathrm{T}}$, and $Y_i^* = (Y_{i1}^{*\mathrm{T}}, \ldots, Y_{in_i}^{*\mathrm{T}})^{\mathrm{T}}$. Then for $i = 1, \ldots, n_f$ and $r = 1, \ldots, n_i$, the response model (2.1) and the mismeasurement models in Section 2.1.2 are used to describe $Y_{ir}$ and $Y_{ir}^*$ where the random effect in (2.1) is now denoted as $u_{ir}$.

With this setup, the conditional distribution (2.10) is now modified to be

$$f(y_i^* | x_i) = \int \prod_{r=1}^{n_i} f(y_{ir}^* | \tilde{u}_i, x_i) f(\tilde{u}_i) \, \mathrm{d}\tilde{u}_i,$$

where $\tilde{u}_i = (u_{i1}, \ldots, u_{in_i})^{\mathrm{T}}$ follows a multivariate normal distribution with mean zero and covariance matrix $\sigma_g^2 R_i$. Then, the inference about the model parameter $\theta$ can be carried out using the same procedure in Section 2.2 or 2.3, and the asymptotic distribution for the resulting estimator can be established in a similar manner.

## 2.5 Simulation Studies

We conduct simulation studies to evaluate the performance of the proposed method in terms of parameter estimates and associated variance estimates. In contrast, we also consider three naive methods. In the first naive method (Naive Method 1), we ignore both misclassification and measurement error in response variables and estimate the parameters of the response model by fitting a generalized linear model using R function `glm()` directly to the observed response measurements; in the second naive method (Naive Method 2), we ignore misclassification in the binary response but account for continuous response measurement error; and in the third analysis (Naive Method 3), we ignore continuous response measurement error but just address misclassification in the binary response.

The sample size is set as $n = 1000$ and we consider model (2.1) with $p_x = 2$ and model (2.6) with $p_z = 1$ and covariates $Z_i$ independently generated from the uniform distribution $U(0, 2)$. To generate covariates $X_i$ for model (2.1), we consider two scenarios with different nature in $X_i$. In Scenario 1, covariates are continuous where $X_{i1}$ and $X_{i2}$ are independently generated from $U(-3, 4)$ and $N(0, 1)$, respectively. In Scenario 2, covariates are ordinal representing a genotype shown in model (2.2); specifically, $X_{ij} = X_{ij}^{(1)} + X_{ij}^{(2)}$ for $j = 1, 2$ where $X_{i1}^{(1)}$ and $X_{i1}^{(2)}$ are independently generated from Bernoulli(0.2), and $X_{i2}^{(1)}$ and $X_{i2}^{(2)}$ are independently generated from Bernoulli(0.5); here different success probabilities of the Bernoulli distribution are chosen to reflect different minor allele frequencies (MAF) of genotypes.

### 2.5.1 Performance of the Methods in Sections 2.2 and 2.3: Simulation Design

For $i = 1, \ldots, n$, the random effects $u_i$, featuring the correlation between the continuous and discrete responses $Y_{i1}$ and $Y_{i2}$, are independently generated from $N(0, \sigma_g^2 R_{ii})$, where $R_{ii} = 1$ for $i = 1, \ldots, 500$, and $R_{ii} = 2$ for $i = 501, \ldots, 1000$, and $\sigma_g$ is set as 0.8. The response vector $Y_i = (Y_{i1}, Y_{i2})^{\mathrm{T}}$ is then generated from the joint distribution

$$f(y_{i1}, y_{i2}|u_i) = \frac{1}{\sqrt{2\pi}} \exp\left[-\frac{1}{2}\{y_{i1} - g_1(\mu_{i1})\}^2\right] g_2(\mu_{i2})^{y_{i2}} \{1 - g_2(\mu_{i2})\}^{1-y_{i2}},$$

where $g_1(\mu_{i1})$ and $g_2(\mu_{i2})$ are specified as in model (2.1) with $g_1(t) = t$ and $g_2(t) = \log\left(\frac{t}{1-t}\right)$, and the coefficient $\beta = (\beta_{10}, \beta_{20}, \beta_{11}, \beta_{21}, \beta_{12}, \beta_{22})^{\mathrm{T}}$ is set as $(0.7, 1.5, 0.7, -1.2, 1, 1)^{\mathrm{T}}$.

The surrogate measurement $Y_{i1}^*$ is generated from the measurement error model (2.8) with $\gamma_0 = 0$, $\gamma_1 = 1$, $\gamma_3 = 0$ and $f(x) = 2x - 1$ to transform the values of $Y_{i2}$ from $\{0, 1\}$

into $\{-1, 1\}$. For the misclassification of $Y_{i2}$, we generate the surrogate measurement $Y_{i2}^*$ from the model

$$\text{logit } \pi_{i1} = \alpha_{01} + \alpha_{z1} Z_i,$$

and

$$\text{logit } \pi_{i0} = \alpha_{00} + \alpha_{z0} Z_i,$$

where $\alpha = (\alpha_{01}, \alpha_{z1}, \alpha_{00}, \alpha_{z0})^{\mathrm{T}}$ is the vector of parameters to be specified.

We consider four settings with different degrees of measurement error and misclassification rates. Settings 1 and 2 differ in the value of $\gamma_2$, with $\gamma_2 = 0.001$ for Setting 1, and $\gamma_2 = 1.0$ for Setting 2; in these two settings, $\sigma_e$ is set 0.25 or 0.50 to reflect increasing degrees of measurement error in $Y_{i1}$ and $\alpha$ is set as $(-1.386, 0, -1.386, 0)^{\mathrm{T}}$. In Settings 3 and 4, we take $\sigma_e = 1.0$ and $\gamma_2 = 1.0$ but consider different values for $\alpha$; in Setting 3, we let $\alpha_{z1} = \alpha_{z0} = 0$ and set $\alpha_{01} = \alpha_{00}$ to be $-4.595$ or $-2.197$, respectively, yielding 1% and 10% misclassification proportions; and in Setting 4, we set $\alpha_{01} = \alpha_{00} = -1$ and let $\alpha_{z1} = \alpha_{z0}$ take a value of $-3.5$ or $-1.2$, leading to about 1% and 10% misclassification proportions, respectively.

### 2.5.2 Performance of the Method in Section 2.4: Simulation Design

In this simulation study, we consider the case where subjects are correlated by pairs. For $i = 1, \ldots, 500$, the random effects $(u_{i1}, u_{i2})^{\mathrm{T}}$ are generated from a bivariate normal distribution with mean zero and covariance matrix $\sigma_g^2 R_i$, where $\sigma_g = 0.8$ and $R_i = \left(\begin{smallmatrix} 1 & \rho \\ \rho & 1 \end{smallmatrix}\right)$. Here, $\rho$ is taken as 0.25 or 0.5, respectively, to possibly represent the parent-offspring relationship or monozygotic twin relationship of pairs (Lange, 2003). The response vector for cluster $i$, $Y_i = (Y_{i11}, Y_{i12}, Y_{i21}, Y_{i22})^{\mathrm{T}}$, is then generated from the joint distribution

$$\begin{aligned} f(y_{i11}, y_{i12}, y_{i21}, y_{i22}|u_i) =& \frac{1}{2\pi} \exp\left[ -\frac{1}{2}\{y_{i11} - g_1(\mu_{i11})\}^2 - \frac{1}{2}\{y_{i21} - g_1(\mu_{i21})\}^2 \right] \\ & \times g_2(\mu_{i2})^{y_{i12}} \{1 - g_2(\mu_{i12})\}^{1-y_{i12}} g_2(\mu_{i22})^{y_{i22}} \{1 - g_2(\mu_{i22})\}^{1-y_{i22}}, \end{aligned}$$

where $g_1(\mu_{i11})$, $g_1(\mu_{i21})$, $g_2(\mu_{i12})$ and $g_2(\mu_{i22})$ are specified as in model (2.1) with $g_1(t) = t$ and $g_2(t) = \log\left(\frac{t}{1-t}\right)$, and the coefficient $\beta = (\beta_{10}, \beta_{20}, \beta_{11}, \beta_{21}, \beta_{12}, \beta_{22})^{\mathrm{T}}$ is set as $(0.7, 1.5, 0.7, -1.2, 1, 1)^{\mathrm{T}}$. We comment that the correlation among the components of $Y_i$ is facilitated by the inclusion of random effects $(u_{i1}, u_{i2})^{\mathrm{T}}$ in $(\mu_{i11}, \mu_{i21}, \mu_{i12}, \mu_{i22})^{\mathrm{T}}$.

The surrogate $Y_{i1}^*$ and $Y_{i2}^*$ are generated in the same way as in Section 2.5.1, with $\sigma_e = 0.25$, $\gamma_2 = 1$, $\alpha_{z1} = \alpha_{z0} = 0$ and $\alpha_{z1} = \alpha_{z0} = -1.386$, yielding a misclassification rate about 20%.

### 2.5.3 Simulation Results

We report the results obtained from the proposed methods and Naive Methods 1–3 in Tables 2.1–2.5 here.

In Tables 2.1–2.4, we report the simulation results for Section 2.5.1 and in Table 2.5 we display the results for Section 2.5.2, where "Bias" is calculated as $\frac{1}{N}\sum_{i=1}^{N}\hat{\theta}_k - \theta_k$, "SEE" is the empirical standard error, "SEM" is calculated by (2.13), and "CR" stands for the coverage rate of 95% confidence intervals for a parameter. That is for parameter $\theta_k$, its $CR$ is given by

$$\frac{1}{N}\sum_{i=1}^{N} I(\hat{\theta}_k^{(L)} < \theta_k < \hat{\theta}_k^{(U)}),$$

where $N$ is the number of simulation studies repeated, $\hat{\theta}_k^{(L)}$ and $\hat{\theta}_k^{(U)}$ are calculated as $\hat{\theta}_k - \text{sd}(\hat{\theta}_k) \times Z_{0.975}$ and $\hat{\theta}_k + \text{sd}(\hat{\theta}_k) \times Z_{0.975}$, respectively. Here, $\hat{\theta}_k$ is an estimate of $\theta_k$, $\text{sd}(\hat{\theta}_k)$ is the associated standard error, and $Z_{0.975}$ is the 0.975 quantile of $N(0,1)$.

Simulation results demonstrate that in the presence of mismeasurement in the response components, the naive methods incur various kinds of biases although they may differ in the magnitude for different settings. The naive methods generally produce large finite sample biases and unreliable coverage rates for 95% confidence intervals that way off the nominal level 95%. On the other hand, the two methods that correct for the mismeasurement effects work very well for various settings regardless of whether the subjects are independent or clustered. These methods yield small finite sample biases for the point estimates and fairly good coverage rates for 95% confidence intervals.

## 2.6 Analysis of Mice SNPs Data

In this section, we illustrate our methods by analyzing the outbred Carworth Farms White (CFW) mice data arising from a genome-wide association study.(Parker et al., 2016a,b) This study provided measurements for 1200 mice on complex traits, including behavioral, physiological and gene expression traits. The original data contain measurements of 99787 SNPs for 1200 mice. Mice with missing responses and the SNPs with the minor allele frequency (MAF) lower than 0.05 are removed because such SNPs have low heterozygosity and often lead to false-positive results in association tests (Anderson et al., 2010). We examine the subset with 1157 mice and 77838 SNPs.

For $i = 1, \ldots, 1157$, let $Y_{i1}$ denote the true length of the tibia bone (in $mm$) and let $Y_{i2}$ be a binary outcome where "0" represents a healthy bone and "1" stands for an abnormal bone. The surrogate $Y_{i1}^*$ is obtained in the laboratory and may differ from the true length $Y_{i1}$, and $Y_{i2}^*$ is measured by a subjective classification rule based on the 90 percentile of bone-mineral density (BMD).

To analyze how the true responses are associated with the SNPs using the proposed method with mismeasurement effects accounted for, we carry out three steps of analysis. The first two steps are performed to screen unimportant SNPs to reduce the dimension of SNPs that is substantially larger than the sample size. The third step is to carry out a refined, post-screening analysis by applying the proposed method to the bivariate generalized linear mixed model (2.1) with measurement error effects taken into account.

In Step 1, we conduct the principal component analysis (PCA) (Price et al., 2006). Let $G$ denote the $n \times n$ genomic relationship matrix using the genetic data following the discussion of Section 3.2 in VanRaden (2008). Then we express $G$ using the eigenvalue decomposition (EVD),

$$G = LDL^{\mathrm{T}},$$

where the columns of $L$ are the eigenvectors of $G$, and $D$ is the diagonal matrix of the positive eigenvalues of $G$. Let $F = GLD^{-\frac{1}{2}}$ be the matrix of principal components of the genetic information, with the $i$th row, denoted as $F_i$, representing the principal components for subject $i$. According to the scree plot in Figure 2.1 and using the "elbow" criterion, we include the first five principal components, denoted as $F_{i1}$, $F_{i2}$, $F_{i3}$, $F_{i4}$ and $F_{i5}$, for subject $i$ as the fixed effects when building the response model.

In Step 2, we conduct a genomewide screening procedure by examining each SNP one at a time using a model similar to (2.1). To adjust for the population stratification, we also include five largest principal components $\mathrm{PC}_i = (F_{i1}, F_{i2}, F_{i3}, F_{i4}, F_{i5})^{\mathrm{T}}$ for subject $i$. Let $W_{ij}$ be the $j$th SNP for subject $i$ and $j = 1, \ldots, p_{\text{SNP}}$, where $p_{\text{SNP}}$ is the dimension of SNPs. We repeat the screening for $j = 1, \ldots, p_{\text{SNP}}$ by respectively considering the model with error effects adjusted:

$$g_1(\mu_{i1}) = \beta_{10}^* + \beta_{11}^* W_{ij} + \mathrm{PC}_i^{\mathrm{T}} \cdot \beta_{\mathrm{PC1}}^* + u_i^*; \tag{2.17}$$

$$g_2(\mu_{i2}) = \beta_{20}^* + \beta_{21}^* W_{ij} + \mathrm{PC}_i^{\mathrm{T}} \cdot \beta_{\mathrm{PC2}}^* + u_i^*; \tag{2.18}$$

where $g_1(\cdot)$ is set as the identity function, $g_2(t) = \log\left(\frac{t}{1-t}\right)$, $u_i^*$ is the random effect, and $\beta_{10}^*$, $\beta_{11}^*$, $\beta_{20}^*$, $\beta_{21}^*$, $\beta_{\mathrm{PC1}}^*$ and $\beta_{\mathrm{PC2}}^*$ are parameters.

To feature the misclassification of $Y_{i2}^*$, we use model (2.6) with $Z_i$ taken as the bone-mineral density (BMD). Regarding the measurement error in $Y_{i1}$, following the discussion in Appendix A.1, we consider model (2.8) with $\gamma_0 = 0$, $\gamma_1 = 1$, $\gamma_3 = 0$ and $f(x) = 2x - 1$ for transforming the values of $Y_{i2}$ from $\{0, 1\}$ into $\{-1, 1\}$. Then we perform the Wald test to (2.17) with the null hypothesis $H_0 : \beta_{11}^* = 0$ and to (2.18) with the null hypothesis $H_0 : \beta_{21}^* = 0$, respectively, where we employ the induced likelihood method and the EM algorithm as opposed to the naive method without addressing error-in-variables.

In Figure 2.2, we report the Manhattan plot for each method which displays the resulting distribution of the SNP significant level, where SNPs are placed on the x-axis according to their chromosomal position, and the $-\log_{10}$ of the SNP associated p-values obtained from the Wald test are recorded on the y-axis. Using the significance level $10^{-6}$ as a threshold, we retain three SNPs, *rs31681083* (chromosome 8), *rs33030459* (chromosome 9) and *rs265727287* (chromosome 12) for our post-selection analysis in Step 3.

Finally, in Step 3, we build a final model with form (2.1) where $X_i$ include the three selected SNPs, *rs31681083* ($X_{i1}$), *rs33030459* ($X_{i2}$) and *rs265727287* ($X_{i3}$), as well as the body weight of a mouse ($X_{i4}$) and the five largest principal components $\text{PC} = (F_{i1}, F_{i2}, F_{i3}, F_{i4}, F_{i5})^{\text{T}}$. That is,

$$\begin{bmatrix} g_1(\mu_{i1}) \\ g_2(\mu_{i2}) \end{bmatrix} = \begin{bmatrix} \beta_{10} + \beta_{11}X_{i1} + \beta_{12}X_{i2} + \beta_{13}X_{i3} + \beta_{14}X_{i4} + \text{PC}^{\text{T}} \cdot \beta_{\text{PC1}} \\ \beta_{20} + \beta_{21}X_{i1} + \beta_{22}X_{i2} + \beta_{13}X_{i3} + \beta_{14}X_{i4} + \text{PC}^{\text{T}} \cdot \beta_{\text{PC2}} \end{bmatrix} + \begin{bmatrix} u_i \\ u_i \end{bmatrix} \tag{2.19}$$

with $g_1(t) = t$ and $g_2(t) = \log\left(\frac{t}{1-t}\right)$ as well as random effects $u_i$.

We apply the induced likelihood method and the EM algorithm in contrast the naive method ignoring the error effects to fit model (2.19), and present the results in Table 2.6. Both the induced likelihood method and the EM algorithm produce fairly close results, and they suggest the same evidence for significance or insignificance of each covariates in model (2.19). At the significance level 0.05, the SNPs *rs31681083* ($X_{i1}$), *rs33030459* ($X_{i2}$) and *rs265727287* ($X_{i3}$) are significantly associated with tibia length, and the SNPs *rs31681083* ($X_{i1}$) and *rs33030459* ($X_{i2}$) are significantly associated with the bone condition. It is also observed that the bodyweight ($X_{i4}$) is significantly associated with both tibia length and bone condition. However, in the naive analysis which disregards mismeasurement effects, we obtain different evidence that *rs31681083* ($X_{i1}$), *rs33030459* ($X_{i2}$) and bodyweight ($X_{i4}$) are not significantly associated with the bone condition. It also shows the opposite evidence for the effect of SNP *rs33030459* ($X_{i2}$) on the bone condition from that revealed by the methods of accommodating mismeasurement effects.

The analyses also reveal evidence of misclassification in the binary response $Y_{i2}$, reflected by the estimation results of $\alpha_{z0}$ and $\alpha_{z1}$ in Table 2.6. For healthy bones, a lower BMD is associated with a higher probability of misclassification as the estimate of $\alpha_{z0}$ is negative, and for abnormal bones, a higher BMD is associated with a higher probability of misclassification as the estimate of $\alpha_{z1}$ is positive. In addition, the estimate of $\gamma_2$ is significantly negative, suggesting the measurement error in tibia length ($Y_{i1}$) is negatively dependent on the true bone condition ($Y_{i2}$).

Table 2.1: Simulation results: Settings 1 and 2 of the Scenario 1 in the first simulation study

| $\sigma_e$ | | Naive Method | | | | | | | | | | | | Proposed Methods | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | Naive Method 1 | | | | Naive Method 2 | | | | Naive Method 3 | | | | Induced Likelihood Method | | | | EM Algorithm | | | |
| | | Bias | SEE | SEM | CR% | Bias | SEE | SEM | CR% | Bias | SEE | SEM | CR% | Bias | SEE | SEM | CR% | Bias | SEE | SEM | CR% |
| | | Setting 1: $\gamma = 0.001$ | | | | | | | | | | | | | | | | | | | |
| 0.25 | $\beta_{10}$ | 0.001 | 0.048 | 0.049 | 95.7 | 0.001 | 0.048 | 0.047 | 95.0 | 0.002 | 0.047 | 0.049 | 95.9 | 0.001 | 0.055 | 0.056 | 95.8 | 0.001 | 0.055 | 0.055 | 95.8 |
| | $\beta_{11}$ | 0.001 | 0.023 | 0.024 | 95.6 | 0.020 | 0.028 | 0.034 | 97.1 | 0.000 | 0.022 | 0.024 | 96.7 | 0.001 | 0.058 | 0.055 | 92.5 | 0.002 | 0.059 | 0.055 | 92.4 |
| | $\beta_{12}$ | 0.001 | 0.047 | 0.048 | 95.7 | -0.002 | 0.047 | 0.046 | 94.1 | 0.001 | 0.047 | 0.047 | 95.4 | 0.001 | 0.055 | 0.054 | 95.2 | 0.000 | 0.055 | 0.054 | 95.7 |
| | $\beta_{20}$ | -0.470 | 0.074 | 0.073 | 0.0 | -0.469 | 0.076 | 0.084 | 0.0 | 0.050 | 0.400 | 0.356 | 93.4 | 0.044 | 0.415 | 0.360 | 93.4 | 0.050 | 0.414 | 0.360 | 93.7 |
| | $\beta_{21}$ | 1.016 | 0.039 | 0.038 | 0.0 | 0.931 | 0.042 | 0.048 | 0.0 | -0.084 | 0.370 | 0.299 | 89.9 | -0.075 | 0.393 | 0.327 | 91.9 | -0.085 | 0.388 | 0.329 | 92.5 |
| | $\beta_{22}$ | -0.726 | 0.070 | 0.072 | 0.0 | -0.716 | 0.073 | 0.082 | 0.0 | 0.054 | 0.351 | 0.299 | 92.2 | 0.047 | 0.361 | 0.310 | 91.0 | 0.055 | 0.359 | 0.312 | 91.8 |
| | $\gamma_2$ | - | - | - | - | -1.170 | 0.027 | 0.119 | 0.0 | - | - | - | - | -0.004 | 0.161 | 0.143 | 92.7 | -0.001 | 0.157 | 0.143 | 92.8 |
| | $\alpha_{00}$ | - | - | - | - | - | - | - | - | -0.040 | 0.248 | 0.211 | 94.9 | -0.079 | 0.409 | 0.342 | 95.1 | -0.076 | 0.594 | 1.164 | 95.2 |
| | $\alpha_{01}$ | - | - | - | - | - | - | - | - | -0.063 | 0.827 | 0.205 | 95.2 | -0.078 | 0.396 | 0.327 | 95.4 | -0.099 | 0.984 | 0.283 | 95.1 |
| | $\sigma_g$ | - | - | - | - | -0.191 | 0.027 | 0.138 | 100 | 0.001 | 0.098 | 0.039 | 54.4 | 0.006 | 0.125 | 0.119 | 95.1 | 0.004 | 0.127 | 0.118 | 93.8 |
| 0.5 | $\beta_{10}$ | 0.002 | 0.050 | 0.051 | 95.8 | -0.001 | 0.050 | 0.048 | 95.0 | 0.002 | 0.049 | 0.051 | 96.0 | 0.001 | 0.058 | 0.058 | 95.8 | 0.001 | 0.058 | 0.058 | 95.9 |
| | $\beta_{11}$ | 0.001 | 0.024 | 0.025 | 95.5 | 0.003 | 0.030 | 0.034 | 98.0 | 0.001 | 0.023 | 0.025 | 96.7 | 0.002 | 0.061 | 0.058 | 92.6 | 0.003 | 0.062 | 0.058 | 92.4 |
| | $\beta_{12}$ | 0.001 | 0.049 | 0.050 | 95.4 | -0.001 | 0.049 | 0.047 | 94.0 | 0.001 | 0.048 | 0.049 | 95.6 | 0.000 | 0.057 | 0.057 | 95.5 | -0.000 | 0.058 | 0.057 | 95.5 |
| | $\beta_{20}$ | -0.470 | 0.074 | 0.073 | 0.0 | -0.468 | 0.077 | 0.084 | 0.0 | 0.051 | 0.407 | 0.360 | 94.0 | 0.048 | 0.419 | 0.365 | 94.0 | 0.051 | 0.418 | 0.364 | 94.2 |
| | $\beta_{21}$ | 1.016 | 0.039 | 0.038 | 0.0 | 0.933 | 0.042 | 0.049 | 0.0 | -0.086 | 0.383 | 0.306 | 90.9 | -0.081 | 0.402 | 0.336 | 91.5 | -0.087 | 0.399 | 0.337 | 92.3 |
| | $\beta_{22}$ | -0.726 | 0.070 | 0.072 | 0.0 | -0.715 | 0.073 | 0.082 | 0.0 | 0.056 | 0.360 | 0.303 | 91.7 | 0.052 | 0.372 | 0.316 | 90.8 | 0.057 | 0.369 | 0.317 | 91.1 |
| | $\gamma_2$ | - | - | - | - | -1.167 | 0.029 | 0.123 | 0.0 | - | - | - | - | -0.002 | 0.165 | 0.151 | 92.2 | 0.001 | 0.166 | 0.151 | 92.1 |
| | $\alpha_{00}$ | - | - | - | - | - | - | - | - | -0.049 | 0.344 | 0.213 | 95.0 | -0.074 | 0.406 | 0.355 | 95.1 | -0.089 | 0.747 | 0.715 | 95.1 |
| | $\alpha_{01}$ | - | - | - | - | - | - | - | - | -0.056 | 0.532 | 0.210 | 94.7 | -0.074 | 0.395 | 0.331 | 95.4 | -0.102 | 0.980 | 1.126 | 95.1 |
| | $\sigma_g$ | - | - | - | - | -0.099 | 0.023 | 0.140 | 100 | 0.002 | 0.104 | 0.042 | 54.4 | 0.005 | 0.129 | 0.129 | 95.7 | 0.002 | 0.134 | 0.128 | 95.2 |
| | | Setting 2: $\gamma$=1 | | | | | | | | | | | | | | | | | | | |
| 0.25 | $\beta_{10}$ | 0.167 | 0.058 | 0.058 | 19.4 | 0.164 | 0.057 | 0.058 | 20 | 0.167 | 0.057 | 0.057 | 17.5 | 0.001 | 0.057 | 0.058 | 95.9 | 0.000 | 0.057 | 0.058 | 95.6 |
| | $\beta_{11}$ | -0.348 | 0.025 | 0.028 | 0.0 | -0.336 | 0.027 | 0.045 | 0.0 | -0.350 | 0.025 | 0.032 | 0.0 | -0.001 | 0.047 | 0.049 | 95.4 | 0.002 | 0.049 | 0.049 | 94.5 |
| | $\beta_{12}$ | 0.186 | 0.056 | 0.056 | 9.2 | 0.182 | 0.055 | 0.057 | 10 | 0.186 | 0.057 | 0.055 | 8.7 | 0.001 | 0.055 | 0.056 | 95.2 | 0.001 | 0.055 | 0.056 | 95.4 |
| | $\beta_{20}$ | -0.470 | 0.074 | 0.073 | 0.0 | -0.466 | 0.078 | 0.090 | 0.0 | -0.074 | 0.255 | 0.303 | 96.5 | 0.016 | 0.194 | 0.193 | 95.7 | 0.017 | 0.192 | 0.192 | 96.0 |
| | $\beta_{21}$ | 1.016 | 0.039 | 0.038 | 0.0 | 0.925 | 0.044 | 0.056 | 0.0 | 0.144 | 0.198 | 0.178 | 79.2 | -0.028 | 0.163 | 0.165 | 95.0 | -0.025 | 0.163 | 0.164 | 94.9 |
| | $\beta_{22}$ | -0.726 | 0.070 | 0.072 | 0.0 | -0.714 | 0.074 | 0.088 | 0.0 | -0.141 | 0.227 | 0.231 | 87.9 | 0.015 | 0.196 | 0.197 | 95.7 | 0.014 | 0.198 | 0.196 | 95.2 |
| | $\gamma_2$ | - | - | - | - | -1.021 | 0.035 | 0.163 | 0.0 | - | - | - | - | -0.004 | 0.115 | 0.121 | 95.8 | 0.005 | 0.121 | 0.120 | 94.6 |
| | $\alpha_{00}$ | - | - | - | - | - | - | - | - | -0.266 | 0.219 | 0.209 | 84.9 | -0.002 | 0.132 | 0.132 | 94.6 | -0.008 | 0.134 | 0.133 | 94.7 |
| | $\alpha_{01}$ | - | - | - | - | - | - | - | - | -0.253 | 0.215 | 0.204 | 84.8 | -0.004 | 0.131 | 0.131 | 94.6 | -0.008 | 0.133 | 0.131 | 94.7 |
| | $\sigma_g$ | - | - | - | - | 0.017 | 0.031 | 0.173 | 100 | 0.414 | 0.084 | 0.039 | 0.0 | 0.007 | 0.118 | 0.122 | 95.6 | -0.001 | 0.127 | 0.122 | 94.2 |
| 0.5 | $\beta_{10}$ | 0.167 | 0.060 | 0.060 | 21.9 | 0.162 | 0.058 | 0.062 | 26 | 0.168 | 0.059 | 0.059 | 19.3 | 0.003 | 0.059 | 0.060 | 95.7 | 0.000 | 0.060 | 0.061 | 95.5 |
| | $\beta_{11}$ | -0.348 | 0.026 | 0.029 | 0.0 | -0.343 | 0.030 | 0.049 | 0.0 | -0.350 | 0.026 | 0.033 | 0.0 | -0.003 | 0.049 | 0.052 | 95.9 | 0.003 | 0.052 | 0.052 | 94.8 |
| | $\beta_{12}$ | 0.186 | 0.057 | 0.058 | 9.8 | 0.182 | 0.057 | 0.061 | 14 | 0.186 | 0.057 | 0.057 | 10.2 | 0.003 | 0.057 | 0.058 | 95.3 | 0.001 | 0.057 | 0.058 | 95.5 |
| | $\beta_{20}$ | -0.470 | 0.074 | 0.073 | 0.0 | -0.462 | 0.080 | 0.091 | 0.0 | -0.061 | 0.264 | 0.307 | 96.8 | 0.014 | 0.198 | 0.201 | 95.9 | 0.016 | 0.198 | 0.199 | 96.1 |
| | $\beta_{21}$ | 1.016 | 0.039 | 0.038 | 0.0 | 0.910 | 0.049 | 0.059 | 0.0 | 0.124 | 0.209 | 0.182 | 80.9 | -0.034 | 0.169 | 0.173 | 95.4 | -0.026 | 0.171 | 0.172 | 94.9 |
| | $\beta_{22}$ | -0.726 | 0.070 | 0.072 | 0.0 | -0.709 | 0.076 | 0.089 | 0.0 | -0.127 | 0.236 | 0.234 | 87.6 | 0.017 | 0.203 | 0.203 | 95.4 | 0.014 | 0.205 | 0.202 | 95.0 |
| | $\gamma_2$ | - | - | - | - | -1.024 | 0.042 | 0.179 | 0.0 | - | - | - | - | -0.010 | 0.120 | 0.128 | 96.2 | 0.006 | 0.129 | 0.128 | 94.8 |
| | $\alpha_{00}$ | - | - | - | - | - | - | - | - | -0.259 | 0.232 | 0.212 | 85.9 | -0.001 | 0.136 | 0.135 | 94.7 | -0.009 | 0.137 | 0.135 | 94.7 |
| | $\alpha_{01}$ | - | - | - | - | - | - | - | - | -0.248 | 0.220 | 0.206 | 85.8 | -0.004 | 0.134 | 0.133 | 94.8 | -0.008 | 0.136 | 0.133 | 94.6 |
| | $\sigma_g$ | - | - | - | - | 0.034 | 0.037 | 0.189 | 100 | 0.427 | 0.093 | 0.041 | 0.0 | 0.016 | 0.120 | 0.130 | 96.2 | -0.002 | 0.134 | 0.132 | 95.2 |

Table 2.2: Simulation results: Settings 3 and 4 of the Scenario 1 in the first simulation study

| $\alpha_0$ | $\alpha_z$ | | Naive Method | | | | | | | | | | | | Proposed Methods | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | Naive Method 1 | | | | Naive Method 2 | | | | Naive Method 3 | | | | Induced Likelihood Method | | | | EM Algorithm | | | |
| | | | Bias | SEE | SEM | CR% | Bias | SEE | SEM | CR% | Bias | SEE | SEM | CR% | Bias | SEE | SEM | CR% | Bias | SEE | SEM | CR% |
| | | | Setting 3 | | | | | | | | | | | | | | | | | | | |
| -4.597 | 0.0 | $\beta_{10}$ | 0.168 | 0.066 | 0.066 | 27.9 | 0.138 | 0.076 | 0.076 | 54.2 | 0.169 | 0.067 | 0.064 | 25.1 | 0.001 | 0.055 | 0.056 | 95.8 | 0.004 | 0.062 | 0.064 | 95.8 |
| | | $\beta_{11}$ | -0.348 | 0.029 | 0.032 | 0.0 | -0.200 | 0.078 | 0.075 | 19.2 | -0.349 | 0.030 | 0.036 | 0.0 | 0.001 | 0.058 | 0.055 | 92.5 | 0.002 | 0.061 | 0.061 | 95.6 |
| | | $\beta_{12}$ | 0.186 | 0.063 | 0.064 | 17.5 | 0.145 | 0.074 | 0.076 | 50.8 | 0.186 | 0.064 | 0.062 | 16.3 | 0.001 | 0.055 | 0.054 | 95.2 | 0.000 | 0.063 | 0.064 | 95.1 |
| | | $\beta_{20}$ | -0.141 | 0.100 | 0.100 | 69.7 | -0.122 | 0.107 | 0.134 | 89.9 | 0.117 | 0.145 | 0.172 | 93.3 | 0.044 | 0.415 | 0.360 | 93.4 | 0.009 | 0.136 | 0.140 | 95.1 |
| | | $\beta_{21}$ | 0.306 | 0.072 | 0.071 | 2.9 | 0.113 | 0.107 | 0.117 | 86.6 | -0.257 | 0.104 | 0.119 | 40.3 | -0.075 | 0.393 | 0.327 | 91.9 | -0.020 | 0.122 | 0.126 | 95.9 |
| | | $\beta_{22}$ | -0.210 | 0.104 | 0.103 | 46.1 | -0.179 | 0.116 | 0.140 | 77.0 | 0.127 | 0.152 | 0.161 | 90.8 | 0.047 | 0.361 | 0.310 | 91.0 | 0.013 | 0.143 | 0.143 | 95.5 |
| | | $\gamma_2$ | - | - | - | - | -0.604 | 0.185 | 0.196 | 9.3 | - | - | - | - | -0.004 | 0.161 | 0.143 | 92.7 | 0.001 | 0.153 | 0.153 | 95.5 |
| | | $\alpha_{00}$ | - | - | - | - | - | - | - | - | -2.512 | 5.326 | 1.705 | 99.7 | -0.079 | 0.409 | 0.342 | 95.1 | -0.669 | 4.298 | 3.098 | 94.9 |
| | | $\alpha_{01}$ | - | - | - | - | - | - | - | - | -2.292 | 4.636 | 1.515 | 98.4 | -0.078 | 0.396 | 0.327 | 95.4 | -0.732 | 4.683 | 2.084 | 94.3 |
| | | $\sigma_g$ | - | - | - | - | 0.255 | 0.104 | 0.184 | 97.0 | 0.677 | 0.052 | 0.042 | 0.0 | 0.006 | 0.125 | 0.119 | 95.1 | -0.001 | 0.155 | 0.156 | 96.4 |
| -2.197 | 0.0 | $\beta_{10}$ | 0.168 | 0.066 | 0.066 | 27.9 | 0.166 | 0.065 | 0.073 | 35.1 | 0.169 | 0.065 | 0.065 | 25.7 | 0.009 | 0.065 | 0.067 | 95.6 | 0.002 | 0.067 | 0.067 | 95.2 |
| | | $\beta_{11}$ | -0.348 | 0.029 | 0.032 | 0.0 | -0.308 | 0.030 | 0.068 | 0.5 | -0.349 | 0.029 | 0.036 | 0.0 | -0.014 | 0.059 | 0.061 | 95.6 | 0.003 | 0.063 | 0.062 | 94.7 |
| | | $\beta_{12}$ | 0.186 | 0.063 | 0.064 | 17.5 | 0.182 | 0.063 | 0.073 | 28.9 | 0.186 | 0.063 | 0.063 | 17.0 | 0.011 | 0.063 | 0.065 | 95.4 | 0.001 | 0.064 | 0.065 | 95.6 |
| | | $\beta_{20}$ | -0.350 | 0.085 | 0.082 | 1.9 | -0.349 | 0.087 | 0.108 | 7.2 | 0.043 | 0.208 | 0.225 | 96.4 | 0.018 | 0.184 | 0.181 | 95.2 | 0.012 | 0.180 | 0.179 | 95.4 |
| | | $\beta_{21}$ | 0.752 | 0.050 | 0.048 | 0.0 | 0.644 | 0.052 | 0.082 | 0.1 | -0.098 | 0.162 | 0.149 | 91.2 | -0.036 | 0.155 | 0.160 | 96.5 | -0.022 | 0.159 | 0.160 | 95.2 |
| | | $\beta_{22}$ | -0.545 | 0.080 | 0.082 | 0.1 | -0.537 | 0.082 | 0.108 | 0.0 | 0.023 | 0.190 | 0.194 | 95.5 | 0.022 | 0.171 | 0.179 | 96.2 | 0.012 | 0.177 | 0.178 | 95.1 |
| | | $\gamma_2$ | - | - | - | - | -0.915 | 0.033 | 0.217 | 0.0 | - | - | - | - | -0.047 | 0.145 | 0.155 | 94.8 | 0.004 | 0.158 | 0.155 | 94.1 |
| | | $\alpha_{00}$ | - | - | - | - | - | - | - | - | -0.287 | 0.311 | 0.285 | 92.9 | -0.052 | 0.221 | 0.215 | 96.2 | -0.022 | 0.217 | 0.211 | 95.4 |
| | | $\alpha_{01}$ | - | - | - | - | - | - | - | - | -0.276 | 0.285 | 0.274 | 92.9 | -0.070 | 0.212 | 0.213 | 95.4 | -0.028 | 0.214 | 0.207 | 94.7 |
| | | $\sigma_g$ | - | - | - | - | 0.212 | 0.039 | 0.213 | 100.0 | 0.585 | 0.085 | 0.045 | 0.0 | 0.046 | 0.138 | 0.156 | 95.2 | -0.002 | 0.161 | 0.160 | 96.1 |
| | | | Setting 4 | | | | | | | | | | | | | | | | | | | |
| -1.000 | -3.5 | $\beta_{10}$ | 0.166 | 0.063 | 0.066 | 28.1 | 0.158 | 0.065 | 0.075 | 44.8 | 0.166 | 0.065 | 0.064 | 27.1 | 0.006 | 0.063 | 0.065 | 95.1 | -0.000 | 0.064 | 0.065 | 95.1 |
| | | $\beta_{11}$ | -0.348 | 0.029 | 0.032 | 0.0 | -0.286 | 0.044 | 0.075 | 1.1 | -0.349 | 0.029 | 0.036 | 0.0 | -0.009 | 0.059 | 0.061 | 94.8 | 0.002 | 0.061 | 0.061 | 94.3 |
| | | $\beta_{12}$ | 0.184 | 0.063 | 0.064 | 15.7 | 0.172 | 0.065 | 0.076 | 36.4 | 0.185 | 0.064 | 0.062 | 14.5 | 0.006 | 0.062 | 0.064 | 93.9 | -0.001 | 0.065 | 0.064 | 93.4 |
| | | $\beta_{20}$ | -0.243 | 0.095 | 0.091 | 23.9 | -0.237 | 0.099 | 0.122 | 51.7 | 0.125 | 0.165 | 0.176 | 92.0 | 0.023 | 0.153 | 0.145 | 93.9 | 0.011 | 0.148 | 0.143 | 94.3 |
| | | $\beta_{21}$ | 0.526 | 0.062 | 0.058 | 0.0 | 0.393 | 0.084 | 0.097 | 3.7 | -0.263 | 0.114 | 0.123 | 41.9 | -0.029 | 0.126 | 0.128 | 95.5 | -0.016 | 0.128 | 0.127 | 94.8 |
| | | $\beta_{22}$ | -0.382 | 0.091 | 0.092 | 1.7 | -0.369 | 0.095 | 0.122 | 12.3 | 0.131 | 0.156 | 0.172 | 91.9 | 0.022 | 0.142 | 0.149 | 96.0 | 0.007 | 0.145 | 0.148 | 95.4 |
| | | $\gamma_2$ | - | - | - | - | -0.797 | 0.089 | 0.212 | 0.0 | - | - | - | - | -0.033 | 0.149 | 0.154 | 95.0 | 0.003 | 0.154 | 0.155 | 94.7 |
| | | $\alpha_{00}$ | - | - | - | - | - | - | - | - | -0.069 | 0.424 | 0.495 | 98.4 | -0.090 | 0.393 | 0.458 | 98.3 | 0.028 | 0.437 | 0.454 | 96.1 |
| | | $\alpha_{z0}$ | - | - | - | - | - | - | - | - | -0.364 | 0.915 | 1.680 | 98.5 | -0.042 | 0.993 | 1.406 | 94.6 | -0.383 | 1.661 | 1.485 | 93.8 |
| | | $\alpha_{01}$ | - | - | - | - | - | - | - | - | -0.050 | 0.432 | 0.487 | 97.7 | -0.073 | 0.398 | 0.455 | 96.9 | 0.053 | 0.435 | 0.449 | 95.3 |
| | | $\alpha_{z1}$ | - | - | - | - | - | - | - | - | -0.406 | 0.914 | 1.676 | 99.2 | -0.106 | 0.988 | 1.415 | 95.5 | -0.493 | 1.803 | 1.516 | 95.4 |
| | | $\sigma_g$ | - | - | - | - | 0.246 | 0.059 | 0.201 | 100.0 | 0.673 | 0.054 | 0.043 | 0.0 | 0.036 | 0.135 | 0.152 | 95.0 | -0.002 | 0.150 | 0.156 | 96.4 |
| -1.000 | -1.2 | $\beta_{10}$ | 0.166 | 0.063 | 0.066 | 28.1 | 0.165 | 0.063 | 0.073 | 36.1 | 0.166 | 0.064 | 0.065 | 27.1 | 0.009 | 0.064 | 0.067 | 95.7 | 0.000 | 0.066 | 0.067 | 95.8 |
| | | $\beta_{11}$ | -0.348 | 0.029 | 0.032 | 0.0 | -0.311 | 0.030 | 0.067 | 0.0 | -0.349 | 0.029 | 0.036 | 0.0 | -0.015 | 0.057 | 0.062 | 96.1 | 0.001 | 0.063 | 0.062 | 94.2 |
| | | $\beta_{12}$ | 0.184 | 0.063 | 0.064 | 15.7 | 0.180 | 0.063 | 0.073 | 27.9 | 0.185 | 0.063 | 0.063 | 15.5 | 0.006 | 0.065 | 0.065 | 93.5 | 0.000 | 0.067 | 0.066 | 93.7 |
| | | $\beta_{20}$ | -0.372 | 0.085 | 0.080 | 1.2 | -0.371 | 0.088 | 0.106 | 4.2 | 0.058 | 0.212 | 0.226 | 96.6 | 0.022 | 0.180 | 0.183 | 95.8 | 0.013 | 0.187 | 0.179 | 94.7 |
| | | $\beta_{21}$ | 0.806 | 0.051 | 0.045 | 0.0 | 0.699 | 0.052 | 0.079 | 0.0 | -0.113 | 0.162 | 0.151 | 90.8 | -0.041 | 0.159 | 0.163 | 95.0 | -0.017 | 0.171 | 0.160 | 94.0 |
| | | $\beta_{22}$ | -0.585 | 0.081 | 0.079 | 0.0 | -0.576 | 0.083 | 0.104 | 0.0 | 0.034 | 0.198 | 0.199 | 95.9 | 0.025 | 0.181 | 0.183 | 96.8 | 0.009 | 0.185 | 0.180 | 94.7 |
| | | $\gamma_2$ | - | - | - | - | -0.939 | 0.030 | 0.221 | 0.0 | - | - | - | - | -0.045 | 0.146 | 0.156 | 94.5 | 0.000 | 0.158 | 0.157 | 94.5 |
| | | $\alpha_{00}$ | - | - | - | - | - | - | - | - | -0.035 | 0.324 | 0.328 | 97.0 | -0.015 | 0.288 | 0.289 | 96.7 | 0.009 | 0.347 | 0.296 | 96.4 |
| | | $\alpha_{z0}$ | - | - | - | - | - | - | - | - | -0.299 | 0.501 | 0.491 | 97.2 | -0.023 | 0.358 | 0.366 | 95.2 | -0.060 | 0.516 | 0.397 | 95.5 |
| | | $\alpha_{01}$ | - | - | - | - | - | - | - | - | -0.026 | 0.324 | 0.323 | 95.8 | -0.011 | 0.288 | 0.284 | 95.0 | 0.008 | 0.289 | 0.287 | 94.8 |
| | | $\alpha_{z1}$ | - | - | - | - | - | - | - | - | -0.310 | 0.515 | 0.484 | 95.8 | -0.014 | 0.350 | 0.357 | 95.2 | -0.092 | 0.767 | 0.408 | 95.6 |
| | | $\sigma_g$ | - | - | - | - | 0.200 | 0.039 | 0.218 | 100.0 | 0.585 | 0.085 | 0.045 | 0.0 | 0.048 | 0.139 | 0.156 | 94.8 | -0.006 | 0.158 | 0.163 | 96.7 |

Table 2.3: Simulation results: Settings 1 and 2 of the Scenario 2 in the first simulation study

| $\sigma_e$ | | Naive Method | | | | | | | | | | | | Proposed Methods | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | Naive Method 1 | | | | Naive Method 2 | | | | Naive Method 3 | | | | Induced Likelihood Method | | | | EM Algorithm | | | |
| | | Bias | SEE | SEM | CR% | Bias | SEE | SEM | CR% | Bias | SEE | SEM | CR% | Bias | SEE | SEM | CR% | Bias | SEE | SEM | CR% |
| | | Setting 1: $\gamma = 0.001$ | | | | | | | | | | | | | | | | | | | |
| 0.25 | $\beta_{10}$ | -0.004 | 0.093 | 0.089 | 93.9 | 0.064 | 0.104 | 0.100 | 90.0 | -0.005 | 0.085 | 0.088 | 93.6 | -0.005 | 0.148 | 0.156 | 95.6 | -0.023 | 0.093 | 0.145 | 95.0 |
| | $\beta_{11}$ | -0.001 | 0.087 | 0.084 | 94.3 | -0.042 | 0.091 | 0.090 | 93.7 | -0.001 | 0.079 | 0.084 | 94.9 | 0.004 | 0.105 | 0.106 | 94.9 | 0.020 | 0.076 | 0.113 | 93.4 |
| | $\beta_{12}$ | 0.004 | 0.071 | 0.067 | 94.0 | 0.033 | 0.074 | 0.072 | 92.0 | 0.004 | 0.063 | 0.068 | 94.1 | 0.002 | 0.083 | 0.081 | 93.6 | -0.009 | 0.056 | 0.085 | 91.6 |
| | $\beta_{20}$ | -0.839 | 0.129 | 0.127 | 0.0 | -0.706 | 0.158 | 0.156 | 1.1 | -0.045 | 0.675 | 0.818 | 94.7 | 0.116 | 0.716 | 0.878 | 94.5 | 0.001 | 0.527 | 0.831 | 92.1 |
| | $\beta_{21}$ | 0.754 | 0.119 | 0.119 | 0.0 | 0.672 | 0.143 | 0.144 | 0.2 | -0.080 | 0.430 | 0.459 | 92.4 | -0.016 | 0.470 | 0.497 | 93.6 | -0.048 | 0.342 | 0.551 | 94.0 |
| | $\beta_{22}$ | -0.666 | 0.101 | 0.099 | 0.0 | -0.608 | 0.122 | 0.118 | 0.4 | 0.066 | 0.387 | 0.426 | 92.7 | 0.004 | 0.421 | 0.446 | 93.0 | 0.049 | 0.328 | 0.492 | 93.7 |
| | $\gamma_2$ | - | - | - | - | -1.216 | 0.129 | 0.137 | 0.0 | - | - | - | - | -0.009 | 0.204 | 0.201 | 90.8 | 0.044 | 0.152 | 0.194 | 87.0 |
| | $\alpha_{00}$ | - | - | - | - | - | - | - | - | -0.017 | 0.204 | 0.260 | 93.3 | -0.080 | 0.340 | 0.427 | 92.9 | 0.027 | 0.165 | 0.369 | 90.9 |
| | $\alpha_{01}$ | - | - | - | - | - | - | - | - | -0.887 | 2.285 | 1.329 | 80.7 | -0.038 | 1.017 | 1.805 | 92.4 | -0.096 | 1.456 | 1.349 | 81.9 |
| | $\sigma_g$ | - | - | - | - | -0.189 | 0.021 | 0.148 | 99.9 | 0.007 | 0.094 | 0.040 | 53.6 | -0.006 | 0.138 | 0.135 | 94.7 | -0.001 | 0.089 | 0.132 | 94.3 |
| 0.5 | $\beta_{10}$ | -0.005 | 0.096 | 0.093 | 94.2 | -0.084 | 0.149 | 0.154 | 93.9 | -0.005 | 0.088 | 0.092 | 93.4 | -0.004 | 0.154 | 0.165 | 96.0 | -0.020 | 0.096 | 0.153 | 95.4 |
| | $\beta_{11}$ | 0.000 | 0.089 | 0.088 | 94.6 | 0.062 | 0.107 | 0.105 | 90.9 | 0.000 | 0.082 | 0.088 | 95.2 | 0.002 | 0.112 | 0.112 | 95.0 | 0.022 | 0.082 | 0.119 | 93.0 |
| | $\beta_{12}$ | 0.004 | 0.073 | 0.070 | 93.6 | -0.041 | 0.094 | 0.094 | 93.5 | 0.005 | 0.066 | 0.071 | 94.0 | 0.004 | 0.087 | 0.085 | 93.8 | -0.008 | 0.060 | 0.090 | 91.4 |
| | $\beta_{20}$ | -0.839 | 0.129 | 0.127 | 0.0 | 0.033 | 0.076 | 0.074 | 92.8 | -0.048 | 0.695 | 0.848 | 95.5 | 0.111 | 0.733 | 0.914 | 94.4 | -0.013 | 0.546 | 0.862 | 92.8 |
| | $\beta_{21}$ | 0.754 | 0.119 | 0.119 | 0.0 | -0.708 | 0.158 | 0.157 | 1.1 | -0.094 | 0.450 | 0.478 | 93.0 | -0.015 | 0.474 | 0.513 | 94.0 | -0.042 | 0.362 | 0.574 | 94.2 |
| | $\beta_{22}$ | -0.666 | 0.101 | 0.099 | 0.0 | 0.673 | 0.144 | 0.144 | 0.2 | 0.072 | 0.395 | 0.442 | 92.4 | 0.005 | 0.422 | 0.458 | 93.5 | 0.049 | 0.336 | 0.509 | 94.5 |
| | $\gamma_2$ | - | - | - | - | -0.609 | 0.122 | 0.118 | 0.5 | - | - | - | - | -0.014 | 0.216 | 0.213 | 91.1 | 0.044 | 0.160 | 0.206 | 87.0 |
| | $\alpha_{00}$ | - | - | - | - | -1.210 | 0.136 | 0.145 | 0.0 | -0.016 | 0.208 | 0.275 | 93.3 | -0.081 | 0.345 | 0.464 | 92.9 | 0.029 | 0.167 | 0.381 | 90.3 |
| | $\alpha_{01}$ | - | - | - | - | - | - | - | - | -1.101 | 2.507 | 1.430 | 78.6 | -0.020 | 0.993 | 1.867 | 93.2 | -0.109 | 1.510 | 1.323 | 81.6 |
| | $\sigma_g$ | - | - | - | - | -0.092 | 0.159 | 0.165 | 95.2 | 0.008 | 0.101 | 0.043 | 54.7 | -0.002 | 0.148 | 0.146 | 95.0 | 0.001 | 0.094 | 0.143 | 94.9 |
| | | Setting 2: $\gamma = 1$ | | | | | | | | | | | | | | | | | | | |
| 0.25 | $\beta_{10}$ | 0.521 | 0.116 | 0.108 | 0.3 | 0.458 | 0.120 | 0.116 | 2.3 | 0.525 | 0.110 | 0.103 | 0.2 | 0.001 | 0.136 | 0.131 | 93.7 | -0.002 | 0.095 | 0.131 | 93.8 |
| | $\beta_{11}$ | -0.319 | 0.110 | 0.102 | 14.7 | -0.279 | 0.108 | 0.103 | 23.6 | -0.318 | 0.102 | 0.095 | 11.0 | 0.005 | 0.107 | 0.105 | 93.3 | 0.008 | 0.069 | 0.106 | 94.5 |
| | $\beta_{12}$ | 0.234 | 0.086 | 0.082 | 19.2 | 0.204 | 0.084 | 0.085 | 33.1 | 0.234 | 0.078 | 0.081 | 18.5 | 0.003 | 0.085 | 0.081 | 93.8 | -0.003 | 0.057 | 0.081 | 93.8 |
| | $\beta_{20}$ | -0.839 | 0.129 | 0.127 | 0.0 | -0.717 | 0.155 | 0.158 | 0.9 | 0.236 | 0.316 | 0.427 | 95.5 | -0.003 | 0.292 | 0.284 | 95.1 | -0.027 | 0.186 | 0.281 | 95.0 |
| | $\beta_{21}$ | 0.754 | 0.119 | 0.119 | 0.0 | 0.685 | 0.140 | 0.145 | 0.2 | 0.153 | 0.263 | 0.276 | 88.4 | -0.020 | 0.262 | 0.254 | 94.6 | -0.014 | 0.174 | 0.251 | 94.8 |
| | $\beta_{22}$ | -0.666 | 0.101 | 0.099 | 0.0 | -0.620 | 0.118 | 0.119 | 0.3 | -0.158 | 0.241 | 0.251 | 86.7 | 0.025 | 0.233 | 0.232 | 95.4 | 0.022 | 0.160 | 0.231 | 95.6 |
| | $\gamma_2$ | - | - | - | - | -0.795 | 0.121 | 0.172 | 0.0 | - | - | - | - | -0.008 | 0.133 | 0.137 | 94.9 | 0.008 | 0.085 | 0.138 | 94.9 |
| | $\alpha_{00}$ | - | - | - | - | - | - | - | - | -0.271 | 0.153 | 0.181 | 70.0 | 0.003 | 0.105 | 0.105 | 94.8 | 0.000 | 0.073 | 0.105 | 95.1 |
| | $\alpha_{01}$ | - | - | - | - | - | - | - | - | -2.485 | 2.824 | 1.493 | 99.8 | -0.006 | 0.327 | 0.352 | 94.9 | 0.010 | 0.211 | 0.340 | 95.3 |
| | $\sigma_g$ | - | - | - | - | -0.122 | 0.135 | 0.197 | 99.2 | 0.445 | 0.083 | 0.041 | 0.0 | 0.008 | 0.130 | 0.132 | 94.9 | -0.001 | 0.087 | 0.134 | 95.8 |
| 0.5 | $\beta_{10}$ | 0.521 | 0.118 | 0.111 | 0.4 | 0.457 | 0.123 | 0.120 | 2.9 | 0.524 | 0.112 | 0.106 | 0.1 | 0.004 | 0.140 | 0.138 | 94.3 | -0.002 | 0.103 | 0.138 | 94.1 |
| | $\beta_{11}$ | -0.318 | 0.113 | 0.105 | 17.1 | -0.277 | 0.111 | 0.107 | 26.7 | -0.319 | 0.104 | 0.099 | 13.0 | 0.004 | 0.110 | 0.111 | 94.4 | 0.008 | 0.073 | 0.111 | 94.6 |
| | $\beta_{12}$ | 0.234 | 0.087 | 0.084 | 21.3 | 0.204 | 0.085 | 0.087 | 35.6 | 0.233 | 0.080 | 0.084 | 20.3 | 0.001 | 0.088 | 0.085 | 94.3 | -0.003 | 0.059 | 0.085 | 94.2 |
| | $\beta_{20}$ | -0.839 | 0.129 | 0.127 | 0.0 | -0.717 | 0.155 | 0.159 | 0.9 | 0.248 | 0.323 | 0.441 | 95.0 | -0.004 | 0.300 | 0.297 | 96.4 | -0.024 | 0.198 | 0.293 | 95.4 |
| | $\beta_{21}$ | 0.754 | 0.119 | 0.119 | 0.0 | 0.685 | 0.140 | 0.145 | 0.2 | 0.139 | 0.270 | 0.279 | 89.0 | -0.016 | 0.265 | 0.261 | 95.5 | -0.013 | 0.172 | 0.259 | 95.1 |
| | $\beta_{22}$ | -0.666 | 0.101 | 0.099 | 0.0 | -0.619 | 0.118 | 0.119 | 0.3 | -0.147 | 0.245 | 0.254 | 87.8 | 0.026 | 0.238 | 0.239 | 95.7 | 0.032 | 0.163 | 0.238 | 95.9 |
| | $\gamma_2$ | - | - | - | - | -0.793 | 0.125 | 0.179 | 0.0 | - | - | - | - | -0.007 | 0.142 | 0.147 | 94.3 | 0.007 | 0.092 | 0.148 | 94.0 |
| | $\alpha_{00}$ | - | - | - | - | -1.210 | 0.136 | 0.145 | 0.0 | -0.262 | 0.156 | 0.184 | 73.0 | 0.002 | 0.107 | 0.108 | 95.4 | 0.001 | 0.074 | 0.108 | 95.6 |
| | $\alpha_{01}$ | - | - | - | - | - | - | - | - | -2.677 | 2.994 | 1.516 | 100.0 | -0.011 | 0.355 | 0.368 | 94.2 | 0.012 | 0.215 | 0.362 | 95.2 |
| | $\sigma_g$ | - | - | - | - | -0.126 | 0.142 | 0.206 | 99.7 | 0.452 | 0.094 | 0.043 | 0.1 | 0.009 | 0.136 | 0.141 | 95.6 | 0.002 | 0.094 | 0.143 | 95.9 |

Table 2.4: Simulation Results: settings 3 and 4 of the scenario 2 in the first simulation study

| $\alpha_0$ | $\alpha_z$ | | Naive Method | | | | | | | | | | | | Proposed Methods | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | Naive Method 1 | | | | Naive Method 2 | | | | Naive Method 3 | | | | Induced Likelihood Method | | | | EM Algorithm | | | |
| | | | Bias | SEE | SEM | CR% | Bias | SEE | SEM | CR% | Bias | SEE | SEM | CR% | Bias | SEE | SEM | CR% | Bias | SEE | SEM | CR% |
| | | | Setting 3 | | | | | | | | | | | | | | | | | | | |
| -4.595 | 0.0 | $\beta_{10}$ | 0.521 | 0.128 | 0.123 | 1.6 | 0.414 | 0.151 | 0.171 | 27.5 | 0.505 | 0.131 | 0.115 | 0.8 | -0.009 | 0.143 | 0.143 | 95.1 | -0.003 | 0.143 | 0.143 | 95.2 |
| | | $\beta_{11}$ | -0.316 | 0.122 | 0.116 | 23.9 | -0.318 | 0.131 | 0.141 | 34.3 | -0.272 | 0.126 | 0.107 | 20.5 | 0.011 | 0.116 | 0.118 | 95.3 | 0.016 | 0.115 | 0.118 | 95.6 |
| | | $\beta_{12}$ | 0.234 | 0.094 | 0.093 | 28.5 | 0.134 | 0.105 | 0.117 | 82.8 | 0.228 | 0.098 | 0.090 | 28.6 | 0.001 | 0.092 | 0.092 | 94.6 | -0.004 | 0.091 | 0.092 | 94.7 |
| | | $\beta_{20}$ | -0.302 | 0.145 | 0.146 | 46.2 | -0.157 | 0.159 | 0.227 | 95.5 | 0.378 | 0.221 | 0.301 | 80.0 | -0.015 | 0.208 | 0.228 | 96.7 | -0.036 | 0.208 | 0.228 | 96.7 |
| | | $\beta_{21}$ | 0.233 | 0.146 | 0.139 | 60.5 | 0.256 | 0.160 | 0.202 | 79.0 | -0.235 | 0.228 | 0.222 | 83.3 | -0.024 | 0.199 | 0.193 | 94.6 | -0.042 | 0.198 | 0.193 | 94.7 |
| | | $\beta_{22}$ | -0.190 | 0.126 | 0.126 | 66.2 | -0.045 | 0.138 | 0.183 | 98.1 | 0.183 | 0.187 | 0.196 | 90.4 | 0.024 | 0.169 | 0.172 | 95.8 | 0.027 | 0.169 | 0.172 | 95.6 |
| | | $\gamma_2$ | - | - | - | - | -0.638 | 0.142 | 0.237 | 6.1 | - | - | - | - | 0.014 | 0.174 | 0.178 | 94.9 | 0.015 | 0.170 | 0.178 | 95.1 |
| | | $\alpha_{00}$ | - | - | - | - | - | - | - | - | 0.124 | 0.281 | 2.012 | 99.4 | 0.079 | 1.051 | 1.181 | 92.0 | 0.038 | 0.836 | 0.917 | 92.1 |
| | | $\alpha_{01}$ | - | - | - | - | - | - | - | - | 0.429 | 0.607 | 4.130 | 99.0 | -0.733 | 2.480 | 2.833 | 86.7 | 1.320 | 1.313 | 1.651 | 86.4 |
| | | $\sigma_g$ | - | - | - | - | 0.309 | 0.087 | 0.224 | 98.4 | 0.737 | 0.047 | 0.045 | 0.0 | 0.001 | 0.179 | 0.178 | 96.1 | 0.015 | 0.169 | 0.174 | 96.4 |
| -2.197 | 0.0 | $\beta_{10}$ | 0.521 | 0.128 | 0.123 | 1.6 | 0.494 | 0.127 | 0.148 | 5.4 | 0.521 | 0.129 | 0.117 | 1.0 | 0.000 | 0.154 | 0.153 | 94.9 | -0.005 | 0.156 | 0.153 | 94.8 |
| | | $\beta_{11}$ | -0.316 | 0.122 | 0.116 | 23.9 | -0.323 | 0.121 | 0.128 | 27.3 | -0.319 | 0.123 | 0.109 | 20.7 | 0.004 | 0.123 | 0.123 | 94.1 | 0.009 | 0.122 | 0.123 | 94.3 |
| | | $\beta_{12}$ | 0.234 | 0.094 | 0.093 | 28.5 | 0.206 | 0.094 | 0.107 | 52.5 | 0.238 | 0.095 | 0.092 | 28.9 | 0.003 | 0.096 | 0.095 | 94.3 | -0.001 | 0.096 | 0.095 | 94.4 |
| | | $\beta_{20}$ | -0.579 | 0.135 | 0.134 | 1.6 | -0.514 | 0.138 | 0.197 | 19.1 | 0.307 | 0.297 | 0.354 | 89.0 | -0.006 | 0.280 | 0.271 | 94.5 | -0.019 | 0.280 | 0.271 | 94.4 |
| | | $\beta_{21}$ | 0.535 | 0.131 | 0.126 | 1.5 | 0.553 | 0.134 | 0.177 | 8.5 | -0.111 | 0.271 | 0.248 | 91.9 | -0.020 | 0.246 | 0.230 | 93.8 | -0.019 | 0.245 | 0.230 | 93.8 |
| | | $\beta_{22}$ | -0.479 | 0.111 | 0.109 | 1.0 | -0.408 | 0.111 | 0.151 | 16.7 | 0.061 | 0.230 | 0.224 | 94.7 | 0.025 | 0.206 | 0.210 | 96.5 | 0.025 | 0.207 | 0.210 | 96.2 |
| | | $\gamma_2$ | - | - | - | - | -0.887 | 0.030 | 0.232 | 0.0 | - | - | - | - | 0.000 | 0.172 | 0.182 | 96.0 | 0.012 | 0.175 | 0.182 | 95.2 |
| | | $\alpha_{00}$ | - | - | - | - | - | - | - | - | -0.297 | 0.279 | 0.261 | 87.4 | -0.011 | 0.176 | 0.178 | 96.5 | -0.002 | 0.179 | 0.177 | 95.8 |
| | | $\alpha_{01}$ | - | - | - | - | - | - | - | - | -0.914 | 0.808 | 2.012 | 100.0 | -0.039 | 1.604 | 1.042 | 93.7 | 0.014 | 0.867 | 0.865 | 93.9 |
| | | $\sigma_g$ | - | - | - | - | 0.261 | 0.038 | 0.230 | 100.0 | 0.651 | 0.089 | 0.047 | 0.0 | 0.007 | 0.163 | 0.176 | 96.9 | 0.002 | 0.170 | 0.177 | 97.0 |
| | | | Setting 4 | | | | | | | | | | | | | | | | | | | |
| -1.000 | -3.5 | $\beta_{10}$ | 0.494 | 0.081 | 0.084 | 0.0 | 0.464 | 0.086 | 0.143 | 2.0 | 0.500 | 0.086 | 0.084 | 0.0 | 0.002 | 0.110 | 0.110 | 95.3 | 0.009 | 0.139 | 0.143 | 94.9 |
| | | $\beta_{11}$ | -0.390 | 0.123 | 0.121 | 10.9 | -0.390 | 0.122 | 0.158 | 25.6 | -0.397 | 0.132 | 0.113 | 9.5 | 0.002 | 0.122 | 0.123 | 95.2 | 0.006 | 0.118 | 0.118 | 94.1 |
| | | $\beta_{12}$ | 0.305 | 0.068 | 0.068 | 0.6 | 0.280 | 0.074 | 0.105 | 14.3 | 0.307 | 0.070 | 0.073 | 0.7 | -0.003 | 0.078 | 0.078 | 94.1 | -0.006 | 0.091 | 0.091 | 95.3 |
| | | $\beta_{20}$ | -0.426 | 0.095 | 0.093 | 0.8 | -0.373 | 0.107 | 0.173 | 36.6 | 0.394 | 0.166 | 0.180 | 40.3 | 0.005 | 0.165 | 0.163 | 94.1 | -0.017 | 0.218 | 0.216 | 94.3 |
| | | $\beta_{21}$ | 0.343 | 0.130 | 0.126 | 22.3 | 0.343 | 0.133 | 0.198 | 61.7 | -0.304 | 0.224 | 0.220 | 72.0 | -0.007 | 0.189 | 0.186 | 95.1 | -0.023 | 0.196 | 0.193 | 95.4 |
| | | $\beta_{22}$ | -0.287 | 0.078 | 0.079 | 6.2 | -0.246 | 0.088 | 0.133 | 54.1 | 0.238 | 0.134 | 0.143 | 63.4 | 0.012 | 0.127 | 0.125 | 95.0 | 0.021 | 0.179 | 0.173 | 95.2 |
| | | $\gamma_2$ | - | - | - | - | -0.709 | 0.067 | 0.277 | 2.1 | - | - | - | - | 0.001 | 0.170 | 0.169 | 94.7 | 0.002 | 0.170 | 0.175 | 94.6 |
| | | $\alpha_{00}$ | - | - | - | - | - | - | - | - | -0.041 | 0.531 | 0.504 | 95.2 | 0.009 | 0.432 | 0.416 | 95.1 | 0.061 | 0.430 | 0.408 | 94.1 |
| | | $\alpha_{z0}$ | - | - | - | - | - | - | - | - | -0.866 | 2.236 | 2.096 | 98.1 | -0.089 | 1.777 | 1.444 | 92.8 | -1.329 | 2.009 | 1.832 | 94.9 |
| | | $\alpha_{01}$ | - | - | - | - | - | - | - | - | -0.475 | 1.250 | 1.321 | 98.6 | 0.072 | 0.877 | 0.804 | 96.2 | 0.128 | 3.005 | 2.232 | 98.7 |
| | | $\alpha_{z1}$ | - | - | - | - | - | - | - | - | -0.146 | 5.021 | 4.139 | 99.7 | -0.416 | 5.080 | 2.927 | 90.1 | -0.561 | 5.178 | 6.885 | 90.4 |
| | | $\sigma_g$ | - | - | - | - | 0.400 | 0.058 | 0.280 | 99.7 | 0.800 | 0.045 | 0.050 | 0.0 | 0.013 | 0.179 | 0.179 | 95.9 | 0.011 | 0.165 | 0.173 | 96.7 |
| -1.000 | -1.2 | $\beta_{10}$ | 0.494 | 0.081 | 0.084 | 0.0 | 0.481 | 0.080 | 0.127 | 0.3 | 0.499 | 0.083 | 0.087 | 0.0 | 0.002 | 0.116 | 0.116 | 94.9 | 0.010 | 0.142 | 0.147 | 95.1 |
| | | $\beta_{11}$ | -0.390 | 0.123 | 0.121 | 10.9 | -0.391 | 0.122 | 0.146 | 19.3 | -0.394 | 0.125 | 0.116 | 10.2 | 0.001 | 0.130 | 0.128 | 94.6 | 0.001 | 0.119 | 0.120 | 94.5 |
| | | $\beta_{12}$ | 0.305 | 0.068 | 0.068 | 0.6 | 0.294 | 0.067 | 0.094 | 5.5 | 0.307 | 0.069 | 0.075 | 1.0 | 0.001 | 0.081 | 0.081 | 95.0 | -0.004 | 0.093 | 0.093 | 95.2 |
| | | $\beta_{20}$ | -0.646 | 0.087 | 0.087 | 0.0 | -0.612 | 0.090 | 0.151 | 0.1 | 0.280 | 0.226 | 0.227 | 77.9 | 0.006 | 0.211 | 0.200 | 93.5 | -0.010 | 0.243 | 0.235 | 93.7 |
| | | $\beta_{21}$ | 0.520 | 0.123 | 0.121 | 1.3 | 0.521 | 0.126 | 0.183 | 10.6 | -0.170 | 0.244 | 0.241 | 90.3 | -0.004 | 0.216 | 0.214 | 95.2 | -0.018 | 0.215 | 0.208 | 94.2 |
| | | $\beta_{22}$ | -0.449 | 0.072 | 0.073 | 0.0 | -0.422 | 0.074 | 0.116 | 1.5 | 0.117 | 0.159 | 0.158 | 89.7 | 0.007 | 0.151 | 0.146 | 94.3 | 0.018 | 0.193 | 0.188 | 93.5 |
| | | $\gamma_2$ | - | - | - | - | -0.821 | 0.031 | 0.274 | 0.0 | - | - | - | - | -0.003 | 0.173 | 0.172 | 95.5 | -0.002 | 0.172 | 0.179 | 95.3 |
| | | $\alpha_{00}$ | - | - | - | - | - | - | - | - | -0.056 | 0.320 | 0.331 | 97.1 | -0.006 | 0.264 | 0.266 | 96.8 | 0.031 | 0.293 | 0.280 | 94.2 |
| | | $\alpha_{z0}$ | - | - | - | - | - | - | - | - | -0.402 | 0.780 | 0.597 | 97.4 | -0.014 | 0.379 | 0.365 | 96.1 | -1.102 | 0.654 | 0.590 | 54.4 |
| | | $\alpha_{01}$ | - | - | - | - | - | - | - | - | -0.025 | 0.840 | 0.767 | 97.2 | 0.036 | 0.521 | 0.481 | 95.4 | 0.222 | 2.290 | 1.503 | 97.2 |
| | | $\alpha_{z1}$ | - | - | - | - | - | - | - | - | -0.943 | 4.205 | 1.906 | 99.3 | -0.046 | 1.461 | 0.862 | 94.6 | -1.306 | 4.113 | 4.295 | 98.7 |
| | | $\sigma_g$ | - | - | - | - | -0.117 | 0.143 | 0.212 | 100.0 | 0.732 | 0.075 | 0.052 | 0.0 | 0.012 | 0.183 | 0.181 | 96.2 | 0.007 | 0.165 | 0.175 | 96.7 |

Table 2.5: Simulation results: the second simulation study

| | Naive Method | | | | | | | | | | | | Proposed Methods | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Naive Method 1 | | | | Naive Method 2 | | | | Naive Method 3 | | | | Induced Likelihood Method | | | | EM Algorithm | | | |
| | Bias | SEE | SEM | CR% | Bias | SEE | SEM | CR% | Bias | SEE | SEM | CR% | Bias | SEE | SEM | CR% | Bias | SEE | SEM | CR% |
| | Parent-offspring relationship | | | | | | | | | | | | | | | | | | | |
| $\beta_{10}$ | 0.552 | 0.104 | 0.098 | 0.0 | 0.295 | 0.111 | 0.118 | 27.6 | 0.556 | 0.109 | 0.098 | 0.0 | 0.008 | 0.124 | 0.125 | 95.8 | 0.025 | 0.123 | 0.132 | 96.6 |
| $\beta_{11}$ | -0.323 | 0.091 | 0.093 | 5.8 | -0.172 | 0.089 | 0.098 | 58.2 | -0.325 | 0.094 | 0.086 | 4.8 | 0.000 | 0.096 | 0.097 | 96.2 | -0.013 | 0.096 | 0.102 | 95.9 |
| $\beta_{12}$ | 0.224 | 0.074 | 0.074 | 15.0 | 0.116 | 0.071 | 0.078 | 69.4 | 0.223 | 0.079 | 0.072 | 14.4 | -0.007 | 0.075 | 0.074 | 93.9 | 0.007 | 0.074 | 0.077 | 95.5 |
| $\beta_{20}$ | -0.532 | 0.145 | 0.136 | 4.6 | -0.429 | 0.167 | 0.169 | 27.2 | 0.251 | 0.269 | 0.328 | 94.1 | -0.005 | 0.265 | 0.239 | 92.1 | 0.025 | 0.300 | 0.291 | 93.3 |
| $\beta_{21}$ | 0.516 | 0.128 | 0.127 | 1.8 | 0.458 | 0.143 | 0.151 | 13.8 | 0.027 | 0.226 | 0.241 | 97.5 | -0.007 | 0.217 | 0.212 | 95.4 | -0.036 | 0.252 | 0.257 | 95.9 |
| $\beta_{22}$ | -0.467 | 0.119 | 0.110 | 1.6 | -0.428 | 0.128 | 0.127 | 9.6 | -0.070 | 0.202 | 0.215 | 94.3 | 0.011 | 0.195 | 0.195 | 95.8 | 0.046 | 0.233 | 0.238 | 96.5 |
| $\gamma_2$ | - | - | - | - | -0.415 | 0.113 | 0.177 | 25.2 | - | - | - | - | 0.001 | 0.142 | 0.149 | 94.5 | -0.041 | 0.113 | 0.151 | 98.2 |
| $\alpha_{00}$ | - | - | - | - | - | - | - | - | -0.484 | 0.271 | 0.289 | 71.1 | -0.023 | 0.159 | 0.156 | 94.9 | -0.006 | 0.107 | 0.106 | 95.3 |
| $\alpha_{01}$ | - | - | - | - | - | - | - | - | -2.163 | 0.875 | 3.955 | 100.0 | -0.148 | 0.773 | 1.036 | 94.1 | -0.103 | 1.108 | 0.604 | 95.3 |
| $\sigma_g$ | - | - | - | - | -0.209 | 0.187 | 0.619 | 99.8 | 0.701 | 0.045 | 0.064 | 0.0 | -0.020 | 0.209 | 0.338 | 97.0 | 0.157 | 0.120 | 0.200 | 95.5 |
| | Monozygotic twin relationship | | | | | | | | | | | | | | | | | | | |
| $\beta_{10}$ | 0.552 | 0.109 | 0.098 | 0.0 | 0.335 | 0.109 | 0.099 | 9.4 | 0.552 | 0.107 | 0.104 | 0.0 | 0.007 | 0.110 | 0.109 | 94.8 | -0.001 | 0.118 | 0.119 | 96.1 |
| $\beta_{11}$ | -0.323 | 0.093 | 0.093 | 5.8 | -0.196 | 0.087 | 0.088 | 40.8 | -0.324 | 0.092 | 0.084 | 5.0 | 0.001 | 0.088 | 0.089 | 95.0 | 0.001 | 0.095 | 0.096 | 94.8 |
| $\beta_{12}$ | 0.223 | 0.076 | 0.074 | 17.0 | 0.133 | 0.070 | 0.072 | 56.4 | 0.223 | 0.075 | 0.070 | 12.5 | -0.006 | 0.071 | 0.069 | 92.6 | 0.000 | 0.073 | 0.073 | 94.1 |
| $\beta_{20}$ | -0.531 | 0.144 | 0.136 | 3.6 | -0.382 | 0.167 | 0.163 | 35.2 | 0.083 | 0.250 | 0.368 | 98.8 | -0.001 | 0.255 | 0.229 | 91.8 | 0.018 | 0.291 | 0.277 | 93.2 |
| $\beta_{21}$ | 0.516 | 0.130 | 0.127 | 1.4 | 0.432 | 0.149 | 0.149 | 19.0 | 0.142 | 0.215 | 0.231 | 89.9 | -0.005 | 0.210 | 0.203 | 95.2 | -0.025 | 0.258 | 0.243 | 94.2 |
| $\beta_{22}$ | -0.470 | 0.118 | 0.110 | 1.8 | -0.414 | 0.131 | 0.126 | 10.4 | -0.169 | 0.187 | 0.201 | 85.1 | -0.001 | 0.189 | 0.188 | 95.6 | 0.018 | 0.225 | 0.226 | 95.7 |
| $\gamma_2$ | - | - | - | - | -0.506 | 0.089 | 0.097 | 0.0 | - | - | - | - | -0.000 | 0.099 | 0.100 | 94.8 | -0.002 | 0.099 | 0.104 | 96.0 |
| $\alpha_{00}$ | - | - | - | - | - | - | - | - | -0.871 | 0.523 | 0.561 | 74.2 | -0.024 | 0.158 | 0.156 | 94.8 | -0.005 | 0.104 | 0.106 | 95.3 |
| $\alpha_{01}$ | - | - | - | - | - | - | - | - | -2.732 | 0.315 | 5.631 | 100.0 | -0.164 | 0.782 | 1.087 | 93.8 | -0.091 | 0.607 | 0.419 | 95.9 |
| $\sigma_g$ | - | - | - | - | -0.034 | 0.110 | 0.133 | 98.8 | 0.597 | 0.089 | 0.095 | 0.0 | -0.005 | 0.105 | 0.106 | 95.8 | 0.096 | 0.099 | 0.108 | 85.6 |

Table 2.6: Analysis results for the mice SNPs data

| | Naive Analysis | | | Induced Likelihood Method | | | EM Algorithm | | |
|---|---|---|---|---|---|---|---|---|---|
| Parameter | Estimate | S.E | p-value | Estimate | S.E | p-value | Estimate | S.E | p-value |
| | Estimates for Response Models - Tibia Length | | | | | | | | |
| $\beta_{10}$ | 17.141 | 0.121 | <0.001 | 17.451 | 0.161 | <0.001 | 17.469 | 0.166 | <0.001 |
| $\beta_{11}$ | -0.058 | 0.022 | 0.009 | -0.152 | 0.035 | <0.001 | -0.154 | 0.036 | <0.001 |
| $\beta_{12}$ | -0.109 | 0.032 | 0.001 | -0.204 | 0.043 | <0.001 | -0.210 | 0.045 | <0.001 |
| $\beta_{13}$ | 0.113 | 0.020 | <0.001 | 0.131 | 0.026 | <0.001 | 0.131 | 0.027 | <0.001 |
| $\beta_{14}$ | 0.047 | 0.004 | <0.001 | 0.037 | 0.006 | <0.001 | 0.036 | 0.006 | <0.001 |
| $\beta_{15}$ | -0.002 | 0.001 | 0.005 | -0.002 | 0.001 | 0.027 | -0.003 | 0.001 | 0.013 |
| $\beta_{16}$ | 0.002 | 0.001 | 0.156 | -0.001 | 0.002 | 0.505 | -0.002 | 0.002 | 0.355 |
| $\beta_{17}$ | 0.002 | 0.001 | 0.207 | -0.004 | 0.002 | 0.027 | -0.005 | 0.002 | 0.024 |
| $\beta_{18}$ | 0.000 | 0.001 | 0.695 | -0.008 | 0.002 | <0.001 | -0.008 | 0.002 | <0.001 |
| $\beta_{19}$ | -0.003 | 0.001 | 0.047 | -0.008 | 0.002 | <0.001 | -0.008 | 0.002 | <0.001 |
| | Estimates for Response Models - Tiebia Length | | | | | | | | |
| $\beta_{20}$ | -1.364 | 1.052 | 0.195 | 6.360 | 2.696 | 0.018 | 6.137 | 2.575 | 0.017 |
| $\beta_{21}$ | -0.305 | 0.188 | 0.105 | -1.590 | 0.532 | 0.003 | -1.513 | 0.501 | 0.003 |
| $\beta_{22}$ | 0.097 | 0.272 | 0.720 | -2.245 | 0.851 | 0.008 | -2.187 | 0.799 | 0.006 |
| $\beta_{23}$ | -0.170 | 0.171 | 0.320 | 0.363 | 0.393 | 0.356 | 0.336 | 0.376 | 0.371 |
| $\beta_{24}$ | -0.020 | 0.038 | 0.605 | -0.219 | 0.097 | 0.024 | -0.207 | 0.093 | 0.025 |
| $\beta_{25}$ | -0.004 | 0.008 | 0.617 | 0.001 | 0.015 | 0.946 | -0.003 | 0.014 | 0.832 |
| $\beta_{26}$ | -0.002 | 0.010 | 0.830 | -0.045 | 0.021 | 0.035 | -0.048 | 0.021 | 0.019 |
| $\beta_{27}$ | -0.040 | 0.011 | <0.001 | -0.102 | 0.031 | 0.001 | -0.096 | 0.030 | 0.001 |
| $\beta_{28}$ | -0.022 | 0.010 | 0.029 | -0.138 | 0.038 | <0.001 | -0.132 | 0.036 | <0.001 |
| $\beta_{29}$ | 0.028 | 0.011 | 0.011 | -0.102 | 0.031 | 0.001 | -0.095 | 0.029 | 0.001 |
| | Estimates for Mismeasurement Models | | | | | | | | |
| $\gamma_2$ | - | - | - | -0.232 | 0.045 | <0.001 | -0.248 | 0.052 | <0.001 |
| $\alpha_{00}$ | - | - | - | 17.794 | 3.954 | <0.001 | 17.490 | 3.529 | <0.001 |
| $\alpha_{01}$ | - | - | - | -12.210 | 1.467 | <0.001 | -12.329 | 1.534 | <0.001 |
| $\alpha_{z0}$ | - | - | - | -0.171 | 0.039 | <0.001 | -0.168 | 0.035 | <0.001 |
| $\alpha_{z1}$ | - | - | - | 0.099 | 0.013 | <0.001 | 0.100 | 0.014 | <0.001 |
| $\sigma_g$ | - | - | - | 0.021 | 2.123 | 0.992 | 0.181 | 0.282 | 0.522 |

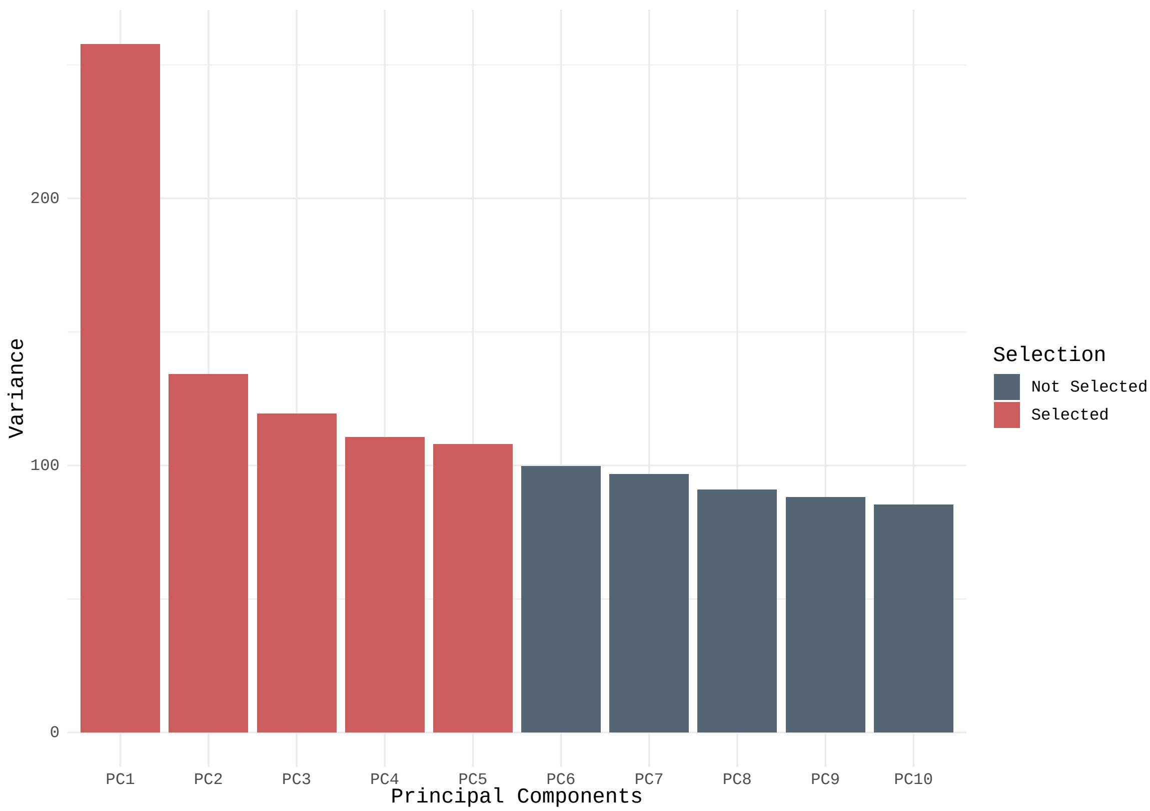


Figure 2.1: *The scree plot of the principal component analysis*

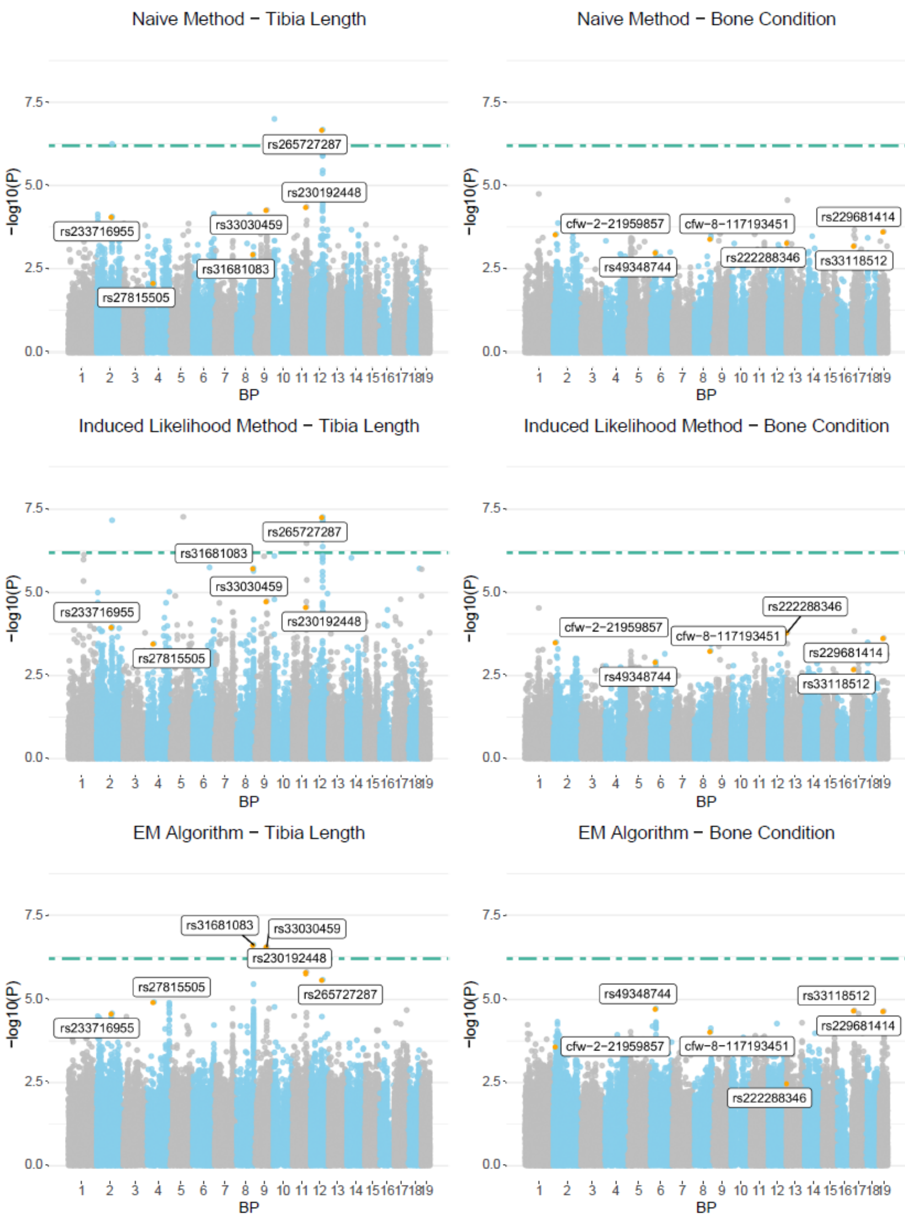


Figure 2.2: *The Manhattan plots of the genome-wide association studies for the three methods and two responses. The x-axis shows the base-pair position (BP, the location of a SNP) on genome which is divided as 19 chromosomes labeled from 1 to 19. The y-axis is the* $-\log_{10}$ *scale of the p-value. Horizontal green dash lines mark the significant level* $10^{-6}$. *The orange dots show the SNPs discussed in the text and their labels are marked in the small boxes.*

# Chapter 3

# Estimating Equation Approach with Bivariate Mixed Responses Subject to Measurement Error and Misclassification

In contrast to likelihood-based approaches developed in Chapter 2, in this chapter, we confine our attention to the marginal modeling, where we explore estimation equation approaches to handle measurement error and misclassification in responses. In Section 3.1, we present the basic notation and the model setup. In Section 3.2, we first introduce the measurement error model and the misclassification model, and then we develop an insertion strategy for estimation of the model parameters to account for the effects of measurement error and misclassification in responses. In Section 3.3, we extend the method to the scenario where either external or internal validation data are available. Simulations studies are conducted in Section 3.4 to evaluate the performance of the proposed methods. In Section 3.5 we apply the proposed method to analyze the mice data arising from a genome-wide association study.

## 3.1 Model Setup and Framework

### 3.1.1 Response Model

We consider the case with bivariate responses for which one component is continuous and one component is binary. For $i = 1, \ldots, n$, let $Y_i = (Y_{i1}, Y_{i2})^{\mathrm{T}}$, where $Y_{i1}$ denotes the continuous response, and $Y_{i2}$ represents the binary response, and $n$ is the number of subjects. Let $X_i = (X_{i1}, \ldots, X_{ip})^{\mathrm{T}}$ denote the covariate vector for subject $i$, where $p$ is a positive integer. For $i = 1, \ldots, n$ and $j = 1, 2$, let $\mu_{ij} = E(Y_{ij}|X_i)$ be the conditional mean of the $Y_{ij}$, given $X_i$, and let $v_{ij} = \mathrm{Var}(Y_{ij}|X_i)$ be the conditional variance of $Y_{ij}$ given covariates $X_i$.

We assume $Y_i$ and $Y_{i'}$ are independent for any $i \neq i'$, but $Y_{i1}$ and $Y_{i2}$ could be correlated. A bivariate generalized linear model is employed to characterize the dependence of $\mu_{ij}$ on $X_i$ for $j = 1, 2$:

$$\begin{aligned} g_1(\mu_{i1}) &= \beta_1^{\mathrm{T}} X_i; \\ g_2(\mu_{i2}) &= \beta_2^{\mathrm{T}} X_i, \end{aligned} \tag{3.1}$$

where $\beta = (\beta_1^{\mathrm{T}}, \beta_2^{\mathrm{T}})^{\mathrm{T}}$ is the vector of regression parameters, and $g_1(\cdot)$ and $g_2(\cdot)$ are link functions. For example, one may specify $g_1(t) = t$ and $g_2(t) = \log\{t/(1-t)\}$.

We assume that for $j = 1, 2$,

$$v_{ij} = h(\mu_{ij}; \psi_j), \tag{3.2}$$

where $\psi_j$ is the dispersion parameter and $h(\cdot)$ is a specified function characterizing the relationship between the conditional variance $v_{ij}$ and the conditional mean $\mu_{ij}$ of $Y_{ij}$ given $X_i$. For instance, the variance functions of the continuous and binary response are often specified, respectively, as

$$v_{i1}(\mu_{i1}) = \psi_1,$$

and

$$v_{i2}(\mu_{i2}) = \mu_{i2}(1 - \mu_{i2}),$$

where $\psi_1$ is often further reparameterized as $\sigma^2$ because of its non-negative property.

### 3.1.2 Estimating Equation Method

Let $V_{i1} = \mathrm{Var}(Y_i|X_i)$ be the conditional covariance matrix of the response vector $Y_i$, given $X_i$. The covariance matrix $V_{i1}$ is decomposed as

$$V_{i1} = B_i^{\frac{1}{2}} C_i B_i^{\frac{1}{2}}, \tag{3.3}$$

where $B_i = \text{diag}\{v_{ij} : j = 1, 2\}$ and $C_i$ is the correlation matrix $\left(\begin{smallmatrix} 1 & \lambda \\ \lambda & 1 \end{smallmatrix}\right)$ of the response vector $Y_i$, given $X_i$, with the parameter $\lambda$ bounded in $[-1, 1]$. Let $\phi = (\psi_1, \lambda)^{\mathrm{T}}$, and $\theta = (\beta^{\mathrm{T}}, \phi^{\mathrm{T}})^{\mathrm{T}}$.

For $i = 1, \ldots, n$, let

$$U_{i1}(\theta) = D_{1i}^{\mathrm{T}} V_{i1}^{-1}(Y_i - \mu_i), \tag{3.4}$$

where $D_{i1} = \frac{\partial \mu_i}{\partial \beta^{\mathrm{T}}}$ is a $2 \times 2p$ matrix. Then $U_{i1}(\theta)$ is an unbiased estimating function which can be used to estimate $\beta$ if the parameter $\phi$ were known.

To estimate $\phi$, we construct a second set of estimating functions. For $i = 1, \ldots, n$ and $j, k = 1, 2$, let $v_{ijk}$ denote the $(j, k)$th element of $V_{i1}$. Define $\xi_i = (v_{ijk} : 1 \le j \le k \le 2)^{\mathrm{T}}$, and $S_i = \{(Y_{ij} - \mu_{ij})(Y_{ik} - \mu_{ik}) : 1 \le j \le k \le 2\}^{\mathrm{T}}$. Let

$$U_{i2}(\theta) = D_{2i}^{\mathrm{T}} V_{i2}^{-1}(S_i - \xi_i), \tag{3.5}$$

where $D_{i2} = \frac{\partial \xi_i}{\partial \phi^{\mathrm{T}}}$, and $V_{i2}$ is a $3 \times 3$ weight matrix. Then $U_{i2}(\theta)$ is an unbiased estimating function of $\phi$ for any given $\beta$. This estimating function is the most efficient in the class of all estimating functions of form (3.5) if the weight matrix $V_{i2}$ is set as the covariance matrix of $S_i$. However, such a specification requires the modeling of the third and fourth moments of $Y_{ij}$, which is often difficult or of no interest. In practice, $V_{i2}$ is often specified as a diagonal matrix (e.g., Hall, 2001; Yi and Cook, 2002). Although such a specification may incur some efficiency loss, it allows us to keep the model assumptions minimal, thus protecting us against model misspecification.

Let $U_i(\theta) = \left(U_{i1}^{\mathrm{T}}(\theta), U_{i2}^{\mathrm{T}}(\theta)\right)^{\mathrm{T}}$. By the estimating function theory (e.g., Liang and Zeger, 1986; Godambe, 1991; Newey and McFadden, 1994; Yi, 2017, Section 1.3.2), under regularity conditions, solving

$$\sum_{i=1}^{n} U_i(\theta) = 0$$

for $\theta$ gives a consistent estimator, say, $\tilde{\theta}$, of $\theta$, and $\sqrt{n}(\tilde{\theta} - \theta)$ has an asymptotic normal distribution with mean zero and covariance matrix

$$\left\{E\left(\frac{\partial U_i(\theta)}{\partial \theta^{\mathrm{T}}}\right)\right\}^{-1} E\{U_i(\theta)U_i^{\mathrm{T}}(\theta)\} \left\{E\left(\frac{\partial U_i(\theta)}{\partial \theta^{\mathrm{T}}}\right)\right\}^{-1\mathrm{T}}.$$

## 3.2 Methodology

### 3.2.1 Measurement Error and Misclassification Models

Suppose that for $i = 1, \ldots, n$, the response variables $Y_{i1}$ and $Y_{i2}$ are subject to mismeasurement and their precise measurements are not observed for every subject, but instead, surrogate measurements $Y_{i1}^*$ and $Y_{i2}^*$ are observed, respectively, for $Y_{i1}$ and $Y_{i2}$.

To describe the mismeasurement processes, we consider the factorization

$$f(y_{i1}^*, y_{i2}^*|y_{i1}, y_{i2}, x_i) = f(y_{i1}^*|y_{i2}^*, y_{i1}, y_{i2}, x_i) f(y_{i2}^*|y_{i1}, y_{i2}, x_i), \tag{3.6}$$

for which we assume that

$$f(y_{i1}^*|y_{i2}^*, y_{i1}, y_{i2}, x_i) = f(y_{i1}^*|y_{i1}, y_{i2}, x_i)$$

and

$$f(y_{i2}^*|y_{i1}, y_{i2}, x_i) = f(y_{i2}^*|y_{i2}, x_i). \tag{3.7}$$

Let $Z_i = (Z_{i1}, \ldots, Z_{ip_z})^{\mathrm{T}}$ be the covariates involved in the misclassification. For ease of exposition, we assume that $Z_i$ is a subset of $X_i$; if this is not the case, we can modify our initial definition of $X_i$ to include $Z_i$ as its part. Let $\pi_{i0} = P(Y_{i2}^* = 1|Y_{i2} = 0, Z_i)$ and $\pi_{i1} = P(Y_{i2}^* = 0|Y_{i2} = 1, Z_i)$ be the misclassification probabilities that may depend on the covariates. We consider logistic models,

$$\text{logit } \pi_{i1} = \alpha_{01} + \alpha_{z1}^{\mathrm{T}} Z_i$$

and

$$\text{logit } \pi_{i0} = \alpha_{00} + \alpha_{z0}^{\mathrm{T}} Z_i, \tag{3.8}$$

where $\alpha = (\alpha_{01}, \alpha_{z1}^{\mathrm{T}}, \alpha_{00}, \alpha_{z0}^{\mathrm{T}})^{\mathrm{T}}$ is the vector of the regression parameters.

For the continuous response $Y_{i1}$, we consider a regression model which facilitates possible dependence of $Y_{i1}^*$ on $\{Y_{i1}, Y_{i2}, Z_i\}$

$$Y_{i1}^* = \gamma_0 + \gamma_1 Y_{i1} + \gamma_2 Y_{i2} + \gamma_3^{\mathrm{T}} Z_i + e_i, \tag{3.9}$$

where $e_i$ is the random error which is independent of $\{Y_{i1}, Y_{i2}, Z_i\}$ and has zero mean and constant variance $\sigma_e^2$, $\gamma = (\gamma_0, \gamma_1, \gamma_2, \gamma_3^{\mathrm{T}})^{\mathrm{T}}$ is the vector of regression coefficients. Often, a normal distribution is assumed for the $e_i$.

Let $\eta = (\gamma^{\mathrm{T}}, \alpha^{\mathrm{T}})^{\mathrm{T}}$ denote the vector of parameters associated with the models (3.8) and (3.9).

### 3.2.2 Estimating Equation Method in the Presence of Mismeasurement

Without addressing measurement error and misclassification in the response, simply replacing $Y_{ij}$ with $Y_{ij}^*$ in the estimating functions (3.4) and (3.5) results in estimating functions that are no longer unbiased, and the resultant estimators may be inconsistent. To account for the mismeasurement effects, we develop a two-step procedure to construct new estimating functions, say $U_{i1}^{**}(\theta)$ and $U_{i2}^{**}(\theta)$, which are expressed in terms of the observed measurements $Y_{i1}^*$ and $Y_{i2}^*$ together with the covariates and the model parameters and satisfy

$$E\{U_{i1}^{**}(\theta)\} = 0, \text{ and } E\{U_{i2}^{**}(\theta)\} = 0.$$

To this end, we develop a two-step procedure to correct for the effects of misclassification in $Y_{i2}$ and those of measurement error in $Y_{i1}$ sequentially. In Step 1, we define

$$Y_{i2}^{**} = \frac{Y_{i2}^* - \pi_{i0}}{1 - \pi_{i0} - \pi_{i1}},$$

where $\pi_{i0}$ and $\pi_{i1}$ are the misclassification rates postulated by (3.8). It is readily seen that $E(Y_{i2}^{**}|Y_{i2}, X_{i1}) = Y_{i2}$.

Then we modify (3.4) and (3.5) by replacing $Y_{i2}$ and $Y_{i2}^{**}$ and define

$$U_{i1}^*(\theta) = D_{1i}^{\mathrm{T}} V_{i1}^{-1} \begin{pmatrix} Y_{i1} - \mu_{i1} \\ Y_{i2}^{**} - \mu_{i2} \end{pmatrix}, \text{ and } U_{i2}^*(\theta) = D_{2i}^{\mathrm{T}} V_{i2}^{-1} \begin{pmatrix} Y_{i1}^2 - 2\mu_{i1} Y_{i1} - \mu_{i1}^2 - \xi_{i1} \\ Y_{i1} Y_{i2}^{**} - \mu_{i1} Y_{i2}^{**} - \mu_{i2} Y_{i1} - \xi_{i2} \\ Y_{i2}^{**} - 2\mu_{i2} Y_{i2}^{**} + \mu_{i2}^2 - \xi_{i3} \end{pmatrix}, \tag{3.10}$$

for which we use $Y_{i2}^2 = Y_{i2}$ for a binary variable $Y_{i2}$ taking the value of either 0 or 1.

In Step 2, we further modify (3.10) by replacing $Y_{i1}$ with the observed variables in order to obtain $U_{i1}^*(\theta)$ and $U_{i2}^*(\theta)$. To this end, define

$$Y_{i1}^{**} = \frac{Y_{i1}^* - \gamma_0 - \gamma_2 Y_{i2}^{**} - \gamma_3^{\mathrm{T}} Z_i}{\gamma_1},$$

$$Y_{i11}^{**} = Y_{i1}^{**2} - \frac{\sigma_e^2}{\gamma_1^2} - \frac{\gamma_2^2}{\gamma_1^2}\Delta_i,$$

and

$$Y_{i12}^{**} = Y_{i1}^{**} Y_{i2}^{**} + \frac{\gamma_2}{\gamma_1}\Delta_i,$$

where

$$\Delta_i = \frac{\Delta_{i0}^{1-Y_{i2}^*}\Delta_{i1}^{Y_{i2}^*} - \Delta_{i0}\pi_{i1} - \Delta_{i1}\pi_{i0}}{1 - \pi_{i1} - \pi_{i0}},$$

$$\Delta_{i0} = \frac{\pi_{i0} - \pi_{i0}^2}{(1 - \pi_{i1} - \pi_{i0})^2},$$

and

$$\Delta_{i1} = \frac{\pi_{i1} - \pi_{i1}^2}{(1 - \pi_{i1} - \pi_{i0})^2}.$$

Let $U_i^{**}(\theta)$ be $U_i^*(\theta) = (U_{i1}^{*\mathrm{T}}(\theta), U_{i2}^{*\mathrm{T}}(\theta))^{\mathrm{T}}$ with $Y_{i1}$, $Y_{i1}^2$, $Y_{i1}Y_{i2}^{**}$ replaced by $Y_{i1}^{**}$, $Y_{i11}^{**}$, $Y_{i12}^{**}$, respectively. In Appendix B.1, we show that $E(Y_{i1}^{**}|Y_{i1}, Y_{i2}) = Y_{i1}$, $E(Y_{i11}^{**}|Y_{i1}, Y_{i2}, X_i) = Y_{i1}^2$, and $E(Y_{i12}^{**}|Y_{i1}, Y_{i2}, X_i) = Y_{i1}Y_{i2}$, thus yielding

$$E[U_i^{**}(\theta)|Y_{i1}, Y_{i2}, X_i] = U_i(\theta).$$

The unbiasedness of $U_i^{**}(\theta)$ is immediate from that of $U_i(\theta)$, thus $U_i^{**}(\theta)$ may be used to obtain a consistent estimator of $\theta$ because it is expressed in terms of the observed data. To do so, we note that however, parameter $\eta$ for the misclassification and measurement error models are involved in $U_i^{**}(\theta)$. To explicitly spell out the dependence on $\eta$, we write $U_i^{**}(\theta)$ as $U_i^{**}(\theta, \eta)$. If $\eta$ is known, say taking a value $\eta_0$, then by the estimating function theory, under regularity conditions (e.g., Godambe, 1991; Newey and McFadden, 1994; Yi, 2017, Section 1.3.2), solving

$$\sum_{i=1}^{n} U_i^{**}(\theta, \eta_0) = 0 \tag{3.11}$$

gives a consistent estimator, say $\hat{\theta}$, of $\theta = (\beta^{\mathrm{T}}, \phi^{\mathrm{T}})^{\mathrm{T}}$, and that $\sqrt{n}(\hat{\theta} - \theta)$ has an asymptotic normal distribution with mean zero and covariance matrix

$$\left\{ E\left( \frac{\partial U_i^{**}(\theta, \eta_0)}{\partial \theta^{\mathrm{T}}} \right) \right\}^{-1} E\{U_i^{**}(\theta, \eta_0) U_i^{**\mathrm{T}}(\theta, \eta_0)\} \left\{ E\left( \frac{\partial U_i^{**}(\theta, \eta_0)}{\partial \theta^{\mathrm{T}}} \right) \right\}^{-1\mathrm{T}}.$$

## 3.3 Estimation Methods with Validation Data

In many applications, the parameter $\eta$ for the measurement error and misclassification models is usually unknown, and is estimated from additional validation data. We now consider two types of validation studies, *internal validation* and *external validation*. Let $\mathcal{M}$ denote the index set of the subjects in the main study, where $\{(y_{i1}^*, y_{i2}^*, x_i) : i \in \mathcal{M}\}$ is available. Let $\mathcal{V}$ represent the index set of the subjects in the validation data. For internal validation, the validation data contain $\{(y_{i1}^*, y_{i2}^*, y_{i1}, y_{i2}, x_i) : i \in \mathcal{V}\}$ with $\mathcal{V} \subset \mathcal{M}$; for external validation, the validation data contain $\{(y_{i1}^*, y_{i2}^*, y_{i1}, y_{i2}, z_i) : i \in \mathcal{V}\}$ with $\mathcal{M} \cap \mathcal{V} = \emptyset$. Let $m$ denote the size of the validation subsample $\mathcal{V}$.

### 3.3.1 External Validation

Estimation of $\eta$ can be carried out by maximizing the conditional likelihood function

$$L(\eta) = \Pi_{i=1}^{n} L_i(y_{i1}^*, y_{i2}^*|y_{i1}, y_{i2}, x_i; \eta),$$

with respect to $\eta$, where $L_i(y_{i1}^*, y_{i2}^*|y_{i1}, y_{i2}, x_i; \eta)$ is the likelihood function contributed from the $i$th individual and is determined by (3.8) and (3.9).

Let

$$S_i(\eta) = \partial \log L_i(y_{i1}^*, y_{i2}^*|y_{i1}, y_{i2}, x_i; \eta)/\partial\eta \tag{3.12}$$

denote the score function of parameter $\eta$.

With external validation data, we consider estimation function

$$U^{(\mathrm{E})}(\theta, \eta) = \sum_{i\in\mathcal{M}} \begin{pmatrix} U_{i1}^{**}(\theta, \eta) \\ U_{i2}^{**}(\theta, \eta) \\ 0 \end{pmatrix} + \sum_{i\in\mathcal{V}} \begin{pmatrix} 0 \\ 0 \\ S_i(\eta) \end{pmatrix}, \tag{3.13}$$

where $S_i(\eta)$ is the score function determined by (3.12). Then solving

$$U^{(\mathrm{E})}(\theta, \eta) = 0$$

gives an estimator of $(\theta^{\mathrm{T}}, \eta^{\mathrm{T}})^{\mathrm{T}}$, denoted as $(\hat{\theta}_{\mathrm{E}}^{\mathrm{T}}, \hat{\eta}_{\mathrm{E}}^{\mathrm{T}})^{\mathrm{T}}$.

Assume that regularity conditions hold and that the ratio $m/n$ approaches a positive constant $\rho$ as $n \to \infty$. In Appendix B.2, we show that $(\hat{\theta}_{\mathrm{E}}^{\mathrm{T}}, \hat{\eta}_{\mathrm{E}}^{\mathrm{T}})^{\mathrm{T}}$ is a consistent estimator of $(\theta^{\mathrm{T}}, \eta^{\mathrm{T}})^{\mathrm{T}}$, and $\sqrt{n}\left\{(\hat{\theta}_{\mathrm{E}}^{\mathrm{T}}, \hat{\eta}_{\mathrm{E}}^{\mathrm{T}})^{\mathrm{T}} - (\theta^{\mathrm{T}}, \eta^{\mathrm{T}})^{\mathrm{T}}\right\}$ has an asymptotic normal distribution with mean zero and covariance matrix $\frac{1}{1+\rho}\Gamma_{\mathrm{E}}^{-1}\Sigma_{\mathrm{E}}(\Gamma_{\mathrm{E}}^{-1})^{\mathrm{T}}$, where

$$\begin{aligned} \Gamma_{\mathrm{E}} &= -\tfrac{1}{1+\rho}\begin{bmatrix} E\left(\frac{\partial U_{i1}^{**}}{\partial\theta^{\mathrm{T}}}\right) & E\left(\frac{\partial U_{i1}^{**}}{\partial\eta^{\mathrm{T}}}\right) \\ E\left(\frac{\partial U_{i2}^{**}}{\partial\theta^{\mathrm{T}}}\right) & E\left(\frac{\partial U_{i2}^{**}}{\partial\theta^{\mathrm{T}}}\right) \\ 0 & 0 \end{bmatrix} - \tfrac{\rho}{1+\rho}\begin{bmatrix} 0 & 0 \\ 0 & 0 \\ 0 & E\left(\frac{\partial S_i}{\partial\eta^{\mathrm{T}}}\right) \end{bmatrix}; \\ \Sigma_{\mathrm{E}} &= \tfrac{1}{1+\rho}\begin{bmatrix} E\left(U_{i1}^{**}U_{i1}^{**\mathrm{T}}\right) & E\left(U_{i1}^{**}U_{i2}^{**\mathrm{T}}\right) & 0 \\ E\left(U_{i2}^{**}U_{i1}^{**\mathrm{T}}\right) & E\left(U_{i2}^{**}U_{i2}^{**\mathrm{T}}\right) & 0 \\ 0 & 0 & 0 \end{bmatrix} + \tfrac{\rho}{1+\rho}\begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & E\left(S_iS_i^{\mathrm{T}}\right) \end{bmatrix}. \end{aligned} \tag{3.14}$$

### 3.3.2 Internal Validation

To account for the effects of measurement error and misclassification in responses, we construct the estimating functions

$$U^{(\mathrm{I})}(\theta, \eta) = \sum_{i\in\mathcal{M}\backslash\mathcal{V}} \begin{pmatrix} U_{i1}^{**}(\theta, \eta) \\ U_{i2}^{**}(\theta, \eta) \\ 0 \end{pmatrix} + \sum_{i\in\mathcal{V}} \begin{pmatrix} U_{i1}(\theta, \eta) \\ U_{i2}(\theta, \eta) \\ S_i(\eta) \end{pmatrix}, \tag{3.15}$$

where $U_{i1}(\theta, \eta)$ and $U_{i2}(\theta, \eta)$ are the estimating functions under the true model as (3.4) and (3.5), and $S_i(\eta)$ is the score function determined by (3.12). Here and elsewhere, 0 may represent the real number zero, a zero vector, or a zero matrix whose meaning is clear in each context. One can obtain an estimator, $(\hat{\theta}_{\mathrm{I}}^{\mathrm{T}}, \hat{\eta}_{\mathrm{I}}^{\mathrm{T}})^{\mathrm{T}}$ for $(\theta^{\mathrm{T}}, \eta^{\mathrm{T}})^{\mathrm{T}}$, by solving equation

$$U^{(\mathrm{I})}(\theta, \eta) = 0 \tag{3.16}$$

with respect to $\theta$ and $\eta$.

Since $S_i(\eta)$ does not depend on $\theta$, solving (3.16) is equivalent to a two-step procedure. First obtain $\hat{\eta}_{\mathrm{I}}$ by solving $\sum_{i\in\mathcal{V}} S_i(\eta) = 0$. Then solve the equation

$$U^{(\mathrm{I})}(\theta, \eta) = \sum_{i\in\mathcal{M}\setminus\mathcal{V}} \begin{pmatrix} U_{i1}^{**}(\theta, \hat{\eta}_{\mathrm{I}}) \\ U_{i2}^{**}(\theta, \hat{\eta}_{\mathrm{I}}) \end{pmatrix} + \sum_{i\in\mathcal{V}} \begin{pmatrix} U_{i1}(\theta, \hat{\eta}_{\mathrm{I}}) \\ U_{i2}(\theta, \hat{\eta}_{\mathrm{I}}) \end{pmatrix} = 0,$$

to obtain an estimator of $\theta$, denoted as $\hat{\theta}_{\mathrm{I}} = (\hat{\beta}_{\mathrm{I}}^{\mathrm{T}}, \hat{\phi}_{\mathrm{I}}^{\mathrm{T}})^{\mathrm{T}}$.

Assume that regularity conditions hold and that the ratio $m/n$ approaches a positive constant $\rho$ as $n \to \infty$. In Appendix B.3, we show that $(\hat{\theta}_{\mathrm{I}}^{\mathrm{T}}, \hat{\eta}_{\mathrm{I}}^{\mathrm{T}})^{\mathrm{T}}$ is a consistent estimator of $(\theta^{\mathrm{T}}, \eta^{\mathrm{T}})^{\mathrm{T}}$, and $\sqrt{n}\left\{(\hat{\theta}_{\mathrm{I}}^{\mathrm{T}}, \hat{\eta}_{\mathrm{I}}^{\mathrm{T}})^{\mathrm{T}} - (\theta^{\mathrm{T}}, \eta^{\mathrm{T}})^{\mathrm{T}}\right\}$ has an asymptotic normal distribution with mean zero and covariance matrix $\Gamma_{\mathrm{I}}^{-1}\Sigma_{\mathrm{I}}(\Gamma_{\mathrm{I}}^{-1})^{\mathrm{T}}$, where

$$\begin{aligned}
\Gamma_{\mathrm{I}} =& -(1-\rho)\begin{bmatrix} E\left(\frac{\partial U_{i1}^{**}}{\partial\theta^{\mathrm{T}}}\right) & E\left(\frac{\partial U_{i1}^{**}}{\partial\eta^{\mathrm{T}}}\right) \\ E\left(\frac{\partial U_{i2}^{**}}{\partial\theta^{\mathrm{T}}}\right) & E\left(\frac{\partial U_{i2}^{**}}{\partial\eta^{\mathrm{T}}}\right) \\ 0 & 0 \end{bmatrix} - \rho\begin{bmatrix} E\left(\frac{\partial U_{i1}}{\partial\theta^{\mathrm{T}}}\right) & 0 \\ E\left(\frac{\partial U_{i2}}{\partial\theta^{\mathrm{T}}}\right) & 0 \\ 0 & E\left(\frac{\partial S_i}{\partial\eta^{\mathrm{T}}}\right) \end{bmatrix}; \\
\Sigma_{\mathrm{I}} =& (1-\rho)\begin{bmatrix} E\left(U_{i1}^{**}U_{i1}^{**\mathrm{T}}\right) & E\left(U_{i1}^{**}U_{i2}^{**\mathrm{T}}\right) & 0 \\ E\left(U_{i2}^{**}U_{i1}^{**\mathrm{T}}\right) & E\left(U_{i2}^{**}U_{i2}^{**\mathrm{T}}\right) & 0 \\ 0 & 0 & 0 \end{bmatrix} \\
&+ \rho\begin{bmatrix} E\left(U_{i1}U_{i1}^{\mathrm{T}}\right) & E\left(U_{i1}U_{i2}^{\mathrm{T}}\right) & E\left(U_{i1}S_i^{\mathrm{T}}\right) \\ E\left(U_{i2}U_{i1}^{\mathrm{T}}\right) & E\left(U_{i2}U_{i2}^{\mathrm{T}}\right) & E\left(U_{i2}S_i^{\mathrm{T}}\right) \\ E\left(S_iU_{i1}^{\mathrm{T}}\right) & E\left(S_iU_{i2}^{\mathrm{T}}\right) & E\left(S_iS_i^{\mathrm{T}}\right) \end{bmatrix}.
\end{aligned} \tag{3.17}$$

### 3.3.3 Weighted Estimator with Internal Validation Data

Estimation of $(\theta^{\mathrm{T}}, \eta^{\mathrm{T}})^{\mathrm{T}}$ based on (3.15) basically treats the validation data and non-validation data equally. To improve the efficiency of parameter estimation, we may attach suitable weights to adjust contributions from the validation sample and the non-validation sample.

Let $W = \text{diag}(w_1, \ldots, w_q)$ be a diagonal matrix, where $0 \le w_j \le 1$ for $j = 1, \ldots, p_\theta$ and $w_j = 0$ for $j = (p_\theta + 1), \ldots, q$. Here $q = p_\theta + p_\eta$, and $p_\theta$ and $p_\eta$ represent, respectively, the dimension of $\theta$ and $\eta$. We modify the estimating function (3.15) as

$$U^{(\mathrm{W})}(\theta,\eta) = \sum_{i \in \mathcal{M}\setminus\mathcal{V}} W \begin{pmatrix} U_{i1}^{**}(\theta,\eta) \\ U_{i2}^{**}(\theta,\eta) \\ 0 \end{pmatrix} + \sum_{i \in \mathcal{V}} (I_q - W) \begin{pmatrix} U_{i1}(\theta,\eta) \\ U_{i2}(\theta,\eta) \\ S_i(\eta) \end{pmatrix},$$

where $U_{i1}(\theta,\eta)$, $U_{i2}(\theta,\eta)$, and $S_i(\eta)$ are defined in the same way as in (3.15), and $I_q$ is the $q \times q$ identity matrix. An estimator of $(\theta^{\mathrm{T}}, \eta^{\mathrm{T}})^{\mathrm{T}}$, denoted $(\hat{\theta}_w^{\mathrm{T}}, \hat{\eta}_w^{\mathrm{T}})^{\mathrm{T}}$, is obtained by solving the equation

$$U^{(\mathrm{W})}(\theta,\eta) = 0$$

for $\theta$ and $\eta$.

Assume regularity conditions hold and that the ratio $m/n$ approaches a positive constant $\rho$ as $n \to \infty$. Similar to the estimator obtained from (3.15), we can show that $(\hat{\theta}_w^{\mathrm{T}}, \hat{\eta}_w^{\mathrm{T}})^{\mathrm{T}}$ is a consistent estimator of $(\theta^{\mathrm{T}}, \eta^{\mathrm{T}})^{\mathrm{T}}$, and $\sqrt{n}\left\{(\hat{\theta}_w^{\mathrm{T}}, \hat{\eta}_w^{\mathrm{T}})^{\mathrm{T}} - (\theta^{\mathrm{T}}, \eta^{\mathrm{T}})^{\mathrm{T}}\right\}$ has an asymptotic normal distribution with mean zero and covariance matrix $\Gamma_{\mathrm{W}}^{-1}\Sigma_{\mathrm{W}}\Gamma_{\mathrm{W}}^{-1\mathrm{T}}$, where

$$\begin{aligned}
\Gamma_{\mathrm{W}} =& -(1-\rho)W \begin{bmatrix} E\left(\frac{\partial U_{i1}^{**}}{\partial \theta^{\mathrm{T}}}\right) & E\left(\frac{\partial U_{i1}^{**}}{\partial \eta^{\mathrm{T}}}\right) \\ E\left(\frac{\partial U_{i2}^{**}}{\partial \theta^{\mathrm{T}}}\right) & E\left(\frac{\partial U_{i2}^{**}}{\partial \theta^{\mathrm{T}}}\right) \\ 0 & 0 \end{bmatrix} - \rho(I-W) \begin{bmatrix} E\left(\frac{\partial U_{i1}}{\partial \theta^{\mathrm{T}}}\right) & 0 \\ E\left(\frac{\partial U_{i2}}{\partial \theta^{\mathrm{T}}}\right) & 0 \\ 0 & E\left(\frac{\partial S_i}{\partial \eta^{\mathrm{T}}}\right) \end{bmatrix}; \\
\Sigma_{\mathrm{W}} =& (1-\rho)W \begin{bmatrix} E\left(U_{i1}^{**}U_{i1}^{**\mathrm{T}}\right) & E\left(U_{i1}^{**}U_{i2}^{**\mathrm{T}}\right) & 0 \\ E\left(U_{i2}^{**}U_{i1}^{**\mathrm{T}}\right) & E\left(U_{i2}^{**}U_{i2}^{**\mathrm{T}}\right) & 0 \\ 0 & 0 & 0 \end{bmatrix} W \\
& + \rho(I-W) \begin{bmatrix} E\left(U_{i1}U_{i1}^{\mathrm{T}}\right) & E\left(U_{i1}U_{i2}^{\mathrm{T}}\right) & E\left(U_{i1}S_i^{\mathrm{T}}\right) \\ E\left(U_{i2}U_{i1}^{\mathrm{T}}\right) & E\left(U_{i2}U_{i2}^{\mathrm{T}}\right) & E\left(U_{i2}S_i^{\mathrm{T}}\right) \\ E\left(S_iU_{i1}^{\mathrm{T}}\right) & E\left(S_iU_{i2}^{\mathrm{T}}\right) & E\left(S_iS_i^{\mathrm{T}}\right) \end{bmatrix} (I-W).
\end{aligned}$$

The optimal weights can be obtained by minimizing $\mathrm{Tr}\left(\Gamma_{\mathrm{W}}^{-1}\Sigma_{\mathrm{W}}\Gamma_{\mathrm{W}}^{-1\mathrm{T}}\right)$ with respect to $\{w_1, \ldots, w_\theta\}$ with the constraints $w_j = 0$ for $j = (p_\theta + 1), \ldots, q$, where $\mathrm{Tr}(A)$ is the trace of matrix $A$. Although the idea is straightforward, this optimization is computationally difficult. Alternatively, we develop an optimal weighted estimator based on linear combinations of two simple estimators discussed as follows.

The first estimator of $(\theta^{\mathrm{T}}, \eta^{\mathrm{T}})^{\mathrm{T}}$ is obtained using the validation data only. Let $(\widehat{\theta}_{\mathrm{I}}^{(0)\mathrm{T}}, \widehat{\eta}_{\mathrm{I}}^{(0)\mathrm{T}})^{\mathrm{T}}$

be the resulting estimator by solving the equation

$$\sum_{i \in \mathcal{V}} \begin{pmatrix} U_{i1}(\theta, \eta) \\ U_{i2}(\theta, \eta) \\ S_i(\eta) \end{pmatrix} = 0 \tag{3.18}$$

for $\theta$ and $\eta$.

The second estimator of $\theta$, denoted as $\widehat{\theta}_{\mathrm{I}}^{(1)}$, solves the estimating equation constructed from the non-validation data,

$$\sum_{i \in \mathcal{M} \setminus \mathcal{V}} \begin{pmatrix} U_{i1}^{**}(\theta, \widehat{\eta}_{\mathrm{I}}^{(0)}) \\ U_{i2}^{**}(\theta, \widehat{\eta}_{\mathrm{I}}^{(0)}) \end{pmatrix} = 0 \tag{3.19}$$

for $\theta$, where $U_{ij}^{**}(\theta, \widehat{\eta}_{\mathrm{I}}^{(0)})$ is determined by $U_{ij}^{**}(\theta, \eta)$ in (3.11) with $\eta$ replaced by $\widehat{\eta}_{\mathrm{I}}^{(0)}$.

Under regularity conditions, both $\widehat{\theta}_{\mathrm{I}}^{(0)}$ and $\widehat{\theta}_{\mathrm{I}}^{(1)}$ are consistent estimators for $\theta$. We consider a weighted estimator to be a linear combination of $\widehat{\theta}_{\mathrm{I}}^{(0)}$ and $\widehat{\theta}_{\mathrm{I}}^{(1)}$:

$$\widetilde{\theta}_{\mathrm{I}}(\Omega) = \Omega \widehat{\theta}_{\mathrm{I}}^{(1)} + (I_{p_\theta} - \Omega) \widehat{\theta}_{\mathrm{I}}^{(0)}, \tag{3.20}$$

where $\Omega = \mathrm{diag}(\omega_1, \ldots, \omega_{p_\theta})$ is a diagonal matrix with constants $0 \leq \omega_j \leq 1$ for $j = 1, \ldots, p_\theta$.

To find the optimal weights, we target to minimize the asymptotic variance for each element of $\widetilde{\theta}_{\mathrm{I}}(\Omega)$. For $r = 1, \ldots, p_\theta$, let $\widetilde{\theta}_{Ir}(\Omega)$, $\hat{\theta}_{Ir}^{(0)}$ and $\hat{\theta}_{Ir}^{(1)}$ be the $r$th component of $\widetilde{\theta}_{\mathrm{I}}(\Omega)$, $\hat{\theta}_{\mathrm{I}}^{(0)}$ and $\hat{\theta}_{\mathrm{I}}^{(1)}$, respectively. The variance of $\widetilde{\theta}_{Ir}(\Omega)$ is given by

$$\begin{aligned} \mathrm{Var}\left(\widetilde{\theta}_{Ir}(\Omega)\right) =& \omega_r^2 \left(\mathrm{Var}(\hat{\theta}_{Ir}^{(0)}) + \mathrm{Var}(\hat{\theta}_{Ir}^{(1)}) - 2\mathrm{Cov}(\hat{\theta}_{Ir}^{(0)}, \hat{\theta}_{Ir}^{(1)})\right) \\ & - 2\omega_r \left(\mathrm{Var}(\hat{\theta}_{Ir}^{(0)}) - \mathrm{Cov}(\hat{\theta}_{Ir}^{(0)}, \hat{\theta}_{Ir}^{(1)})\right) + \mathrm{Var}(\hat{\theta}_{Ir}^{(1)}), \end{aligned}$$

which is minimized at

$$\omega_r^* = \frac{\mathrm{Var}(\hat{\theta}_{Ir}^{(0)}) - \mathrm{Cov}(\hat{\theta}_{Ir}^{(0)}, \hat{\theta}_{Ir}^{(1)})}{\mathrm{Var}(\hat{\theta}_{Ir}^{(0)}) + \mathrm{Var}(\hat{\theta}_{Ir}^{(1)}) - 2\mathrm{Cov}(\hat{\theta}_{Ir}^{(0)}, \hat{\theta}_{Ir}^{(1)})}.$$

Let $\Omega^* = \mathrm{diag}(\omega_1^*, \ldots, \omega_{p_\theta}^*)$. Then the estimator $\widetilde{\theta}_{\mathrm{I}}^* = \Omega^* \widehat{\theta}_{\mathrm{I}}^{(1)} + (I - \Omega^*) \widehat{\theta}_{\mathrm{I}}^{(0)}$ is the optimal estimator among the linear combinations of form (3.20).

In practice, $\omega_r^*$ is estimated by

$$\widehat{\omega}_r^* = \frac{\widehat{\text{Var}}(\hat{\theta}_{Ir}^{(1)}) - \widehat{\text{Cov}}(\hat{\theta}_{Ir}^{(0)}, \hat{\theta}_{Ir}^{(1)})}{\widehat{\text{Var}}(\hat{\theta}_{Ir}^{(0)}) + \widehat{\text{Var}}(\hat{\theta}_{Ir}^{(1)}) - 2\widehat{\text{Cov}}(\hat{\theta}_{Ir}^{(0)}, \hat{\theta}_{Ir}^{(1)})},$$

where $\widehat{\text{Var}}(\hat{\theta}_{Ir}^{(1)})$, $\widehat{\text{Var}}(\hat{\theta}_{Ir}^{(0)})$ and $\widehat{\text{Cov}}(\hat{\theta}_{Ir}^{(0)}, \hat{\theta}_{Ir}^{(1)})$ are estimates for $\text{Var}(\hat{\theta}_{Ir}^{(1)})$, $\text{Var}(\hat{\theta}_{Ir}^{(0)})$ and $\text{Cov}(\hat{\theta}_{Ir}^{(0)}, \hat{\theta}_{Ir}^{(1)})$ by stacking the estimating functions in (3.18) and (3.19). The details are presented in Appendix B.4.

We comment that in practice, the resulting weights $\widehat{\omega}_r^*$ may not satisfy the constraint that $0 \leq \widehat{\omega}_r^* \leq 1$. If $\widehat{\omega}_r^* < 0$, we set $\widehat{\omega}_r^*$ to be 0 and if $\widehat{\omega}_r^* > 1$, we specify $\widehat{\omega}_r^*$ to be 1.

## 3.4 Simulation Studies

We conduct simulation studies to evaluate the performance of the proposed methods in terms of parameter estimation and associated variance estimation. Similar to the simulation studies in Chapter 2, for the sake of comparison, we consider three naive methods, where either measurement error or misclassification, or both are ignored.

We consider the sample size $n = 1000$. The $X_{i1}$ is independently generated from $U(-3, 4)$, and the $X_{i2}$ is independently generated from $N(0, 1)$. The response vector $Y_i = (Y_{i1}, Y_{i2})^{\text{T}}$ is generated from the model

$$\begin{aligned} g_1(\mu_{i1}) &= \beta_{10} + \beta_{11}X_{i1} + \beta_{12}X_{i2}, \\ g_2(\mu_{i2}) &= \beta_{20} + \beta_{21}X_{i1} + \beta_{22}X_{i2}, \end{aligned}$$

where the coefficient vector $\theta = (\beta_{10}, \beta_{20}, \beta_{11}, \beta_{21}, \beta_{12}, \beta_{22})^{\text{T}}$ is set as $(0.7, 0.7, 1.5, -1.5, -1, 1)^{\text{T}}$, $g_1(t) = t$, and $g_2(t) = \log\left(\frac{t}{1-t}\right)$. That is, $Y_{i1}$ is generated by $N(\mu_{i1}, \sigma^2)$ where $\sigma^2$ is set as 1, and $Y_{i2}$ is independently generated from Bernoulli$(\mu_{i2})$.

The surrogate measurement $Y_{i1}^*$ is generated from the measurement error model, $Y_{i1}^* = Y_{i1} + \gamma Y_{i2} + e_i$, where $e_i$ is a centered normal random error with variance $\sigma_e^2$ and is independent of $\{Y_{i1}, Y_{i2}\}$. For the misclassification of $Y_{i2}$, we generate the surrogate measurement $Y_{i2}^*$ by misclassification models (3.8). The values of $\alpha$, $\gamma$, $\sigma_e^2$ are specified in Section 3.4.1.

For each estimator, we report the finite sample biases (denoted as "bias"), the standard error (denoted as "SEE"), the model-based standard error (denoted as "SEM"), or the coverage rate (denoted as "CR").

### 3.4.1 Simulation 1: Evaluation for the Case with Known Mismeasurement Parameters

In this subsection, we consider the case where the parameters of measurement error and misclassification models are known to the method in Section 3.2.2, taking the values as in the specifications of generating the random variables.

To study the performance of the methods, we consider three settings. In Setting 1, fix $\alpha = (-1.386, 0, -1.386, 0)^{\mathrm{T}}$ and $\gamma = 0.8$, and compare the performance of the naive models and the proposed model under different degrees of measurement error, where $\sigma_e$ are set as 0.1, 0.5, 0.7. In Setting 2, fix $\sigma_e = 0.1$ and $\gamma = 0.8$, and evaluate the performance using the data simulated with different misclassification rates, where $\alpha = (\alpha_{01}, \alpha_{x1}, \alpha_{00}, \alpha_{x0})^{\mathrm{T}}$ is set as $(-4.595, 0, -4.595, 0)^{\mathrm{T}}$, $(-2.197, 0, -2.197, 0)^{\mathrm{T}}$, or $(-1.386, 0, -1.386, 0)^{\mathrm{T}}$, yielding the misclassification rates $\pi_0$ and $\pi_1$ as 1%, 10% or 20%, respectively. In Setting 3, fix $\sigma_e = 0.1$ and $\alpha = (-1.386, 0, -1.386, 0)^{\mathrm{T}}$, and evaluate the methods for different measurement error mechanisms which are independent of the binary outcome $Y_{i2}$ ($\gamma = 0$), negatively associated with the binary outcome ($\gamma = -0.8$), or positively associated with the binary outcome ($\gamma = 0.8$).

The results are presented in Tables 3.1–3.3. Different naive methods may perform differently in both the point estimation and the variance estimation, but they all produce large biases in the point estimation and poor coverage rates. Conversely, the proposed method successfully corrects the biases due to the response mismeasurement, yielding reasonably small finite sample biases and coverage rates in good agreement with the nominal value 95%.

### 3.4.2 Simulation 2: Evaluation of the Case with Validation Data

In this simulation study, we compare the performance of the methods for three scenarios. In the first scenario, we consider the same case as in Section 3.4.1 where the mismeasurement parameter $\eta$ is known. In the second and the third scenarios, we evaluate the performance for the methods described in Sections 3.3.1 and 3.3.2 where $\eta$ is unavailable but estimated from either external validation data or internal validation data. We also display the results of the method using the true measurements $Y_{i1}$ and $Y_{i2}$ for comparisons.

We consider the same three settings as in Simulation 1. The results are reported in Tables 3.4–3.6. As expected, the method using true response measurements produces the best results with the smallest finite sample biases and model-based standard errors as well as the best coverage rates of 95% confidence intervals. On the other hand, the proposed methods perform quite well for different scenarios. Finite sample biases are close to those produced from the method with the true response measurements; model-based standard errors agree fairly well with empirical standard errors and coverage rates of 95% confidence intervals are in good agreement with the nominal level 95%.

### 3.4.3 Simulation 3: Evaluation of the Proposed Method with Internal Validation Data with Different Sample Sizes and Different Weights

In this subsection we compare the estimator described in Section 3.3.2 and the weighted estimators described in Section 3.3.3. We also consider four different weights to compare the estimates of using validation data only ($\widehat{\theta}_{\mathrm{I}}^{(0)}$), using non-validation data only ($\widehat{\theta}_{\mathrm{I}}^{(1)}$), using equal weights for validation data and non-validation data ($\widehat{\theta}_{\mathrm{I}}$), and optimal weighted estimator ($\widetilde{\theta}_{\mathrm{I}}^{*}$). Our assessment focuses on examining the impact on the performance of the proposed estimator of the sample size, the sample size ratio between the validation data and non-validation data, and the weight choice. We consider two scenarios. In Scenario 1, we fix the total sample size to be 1500 and let the sample size ratio vary as 2:1, 1:1, 1:2. In Scenario 2, we fix the ratio to be 1:2 and let the total sample size be 1500 and 3000.

The results for Scenario 1 are presented in Figures 3.1–3.2 and the results for Scenario 2 are presented in Table 3.7. It is clear that the estimator with optimal weights described in Section 3.3.3 and the estimator described in Section 3.3.2 perform the best among all the estimators in terms of both finite sample biases and standard errors. Moreover, the former estimator greatly outperforms the latter one. The efficiency gain of using the estimator with optimal weights over the estimator in Section 3.3.2 can be as large as 58%, shown by the estimate of $\beta_{21}$ with $n = 1500$.

## 3.5 Application to Mice SNPs Data

To illustrate the usage of the proposed method, we analyze data arising from a genome-wide association study of outbred Carworth Farms White mice data (Parker et al., 2016b). This study provided measurements with complex traits, including behavioral, physiological, and gene expression traits.

For $i = 1, \ldots, 1128$, let $Y_{i1}$ be the weight of the tibialis anterior muscle (in $mg$), and let $Y_{i2}$ be the binary outcome where "0" represents a healthy tibia bone and "1" stands for abnormal tibia bone, which is defined as the 90% quantile of the bone-mineral density. Due to the concern of data quality, the true measurements of the responses $Y_{i1}$ and $Y_{i2}$ for 464 subjects are not available but their surrogate measurements $Y_{i1}^*$ and $Y_{i2}^*$ are available, where $Y_{i1}^*$ is the predicted tibialis anterior muscle weights based on muscle from other body parts of the mice and $Y_{i2}^*$ is the bone condition judged by subjective observations from technicians. Precise measurements of the responses $Y_{i1}$ and $Y_{i2}$ together with their surrogates $Y_{i1}^*$ and $Y_{i2}^*$ are available for the remaining 664 subjects, which are taken as the validation data. Covariates include a continuous variable measuring the SNPs *rs27338905* ($X_{i1}$), and the first two principal components of genetics data ($X_{i2}$) and ($X_{i3}$) for subject $i$ which are described below in detail. Our main interest lies in studying the association of SNPs *rs27338905* with two physiological traits. We employ the model (3.1) with $g_1(t) = t$ and $g_2(t) = \log\{t/(1-t)\}$ to facilitate the dependence of the responses on the covariates.

To account for the effect of population stratification (Price et al., 2006), similar to Section 2.6 in Chapter 2, we conduct principal component analysis. According to the scree plot in figure 3.3 and based on the "elbow" criterion, we include the first two principal components ($X_{i2}$) and ($X_{i3}$) as fixed effects in the response model.

We consider two settings for the misclassification and measurement error models. In Setting 1, we consider the body weight ($X_{i4}$) to be the covariates in model (3.8) to feature the misclassification of $Y_{i2}^*$. For the measurement error model, we consider (3.9) with the covariates chosen to be the body weight ($X_{i4}$). In Setting 2, we consider model (3.8) to be postulated by constants $\alpha_{00}$ and $\alpha_{01}$, and an additional constraint that $\gamma_2 = 0$ is imposed for the measurement error model (3.9).

We analyze the data using the proposed optimal weighted estimator with internal validation data, described in Section 3.3.3. We compare the results with the naive model where the mismeasurement is ignored. The results are presented in Table 3.8. The proposed method with two settings produces similar estimation results. Under the significance level of 0.05, the estimates of $\beta_{11}$ and $\beta_{21}$ suggest that SNP *rs27338905* is significantly associated with the weight of tibialis anterior muscle but is not associated with the bone condition. The estimates of $\beta_{12}$, $\beta_{13}$, $\beta_{22}$ and $\beta_{23}$ show that the effects of population stratification are not significant. However, the naive method produces somewhat different findings; there is no evidence showing the effects of SNP *rs27338905* on the weight of the tibialis anterior muscle.

Regarding the results for the parameters of the mismeasurement models, the measurement error process is not influenced by the bone condition ($Y_{i2}$), as indicated by the estimates of $\gamma_2$. The bodyweight of mice is only involved in the measurement error process but not the misclassification process because the estimate of $\gamma_3$ is significant but the estimates of $\alpha_{x1}$ and $\alpha_{x0}$ are not. This suggests that the simpler specification of mismeasurement models in Setting 2 is perhaps adequate and there is no need to consider the more complicated models in Setting 1.

Table 3.1: Results for Simulation 1 where parameters for mismeasurement models are assumed known: Setting 1 with different degrees of measurement error ($\sigma_e = 0.1, 0.5$ and $0.7$)

| | | Naive Method | | | | | | | | | | | | Proposed | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | Naive Scenario 1 | | | | Naive Scenario 2 | | | | Naive Scenario 3 | | | | | | | |
| $\sigma_e$ | | Bias | SEE | SEM | CR% | Bias | SEE | SEM | CR% | Bias | SEE | SEM | CR% | Bias | SEE | SEM | CR% |
| 0.1 | $\beta_{10}$ | 0.469 | 0.034 | 0.034 | 0.0 | 0.012 | 0.031 | 0.034 | 94.3 | 0.466 | 0.040 | 0.040 | 0.4 | -0.004 | 0.034 | 0.034 | 94.7 |
| | $\beta_{11}$ | -0.146 | 0.016 | 0.016 | 0.0 | -0.027 | 0.016 | 0.016 | 61.0 | -0.137 | 0.023 | 0.020 | 3.8 | 0.001 | 0.016 | 0.017 | 93.8 |
| | $\beta_{12}$ | 0.075 | 0.033 | 0.033 | 37.1 | 0.012 | 0.033 | 0.033 | 93.9 | 0.063 | 0.049 | 0.042 | 62.8 | -0.000 | 0.036 | 0.034 | 94.2 |
| | $\beta_{20}$ | -0.314 | 0.085 | 0.084 | 4.9 | -0.313 | 0.082 | 0.084 | 4.6 | 0.015 | 0.190 | 0.152 | 90.3 | 0.047 | 0.201 | 0.204 | 96.9 |
| | $\beta_{21}$ | 0.686 | 0.054 | 0.050 | 0.0 | 0.684 | 0.052 | 0.054 | 0.0 | -0.018 | 0.195 | 0.180 | 92.4 | -0.086 | 0.220 | 0.249 | 96.1 |
| | $\beta_{22}$ | -0.511 | 0.089 | 0.084 | 0.1 | -0.509 | 0.087 | 0.086 | 0.1 | 0.002 | 0.212 | 0.180 | 90.5 | 0.068 | 0.243 | 0.253 | 97.1 |
| | $\sigma$ | - | - | - | - | 0.029 | 0.022 | 0.023 | 77.1 | 0.044 | 0.023 | 0.023 | 59.4 | 0.001 | 0.024 | 0.025 | 94.9 |
| | $\xi$ | - | - | - | - | -0.137 | 0.033 | 0.034 | 2.4 | 0.303 | 0.057 | 0.049 | 0.1 | -0.002 | 0.060 | 0.063 | 95.3 |
| 0.5 | $\beta_{10}$ | 0.469 | 0.037 | 0.037 | 0.0 | 0.012 | 0.035 | 0.037 | 94.1 | 0.467 | 0.043 | 0.043 | 0.3 | -0.004 | 0.036 | 0.038 | 95.3 |
| | $\beta_{11}$ | -0.146 | 0.018 | 0.018 | 0.0 | -0.027 | 0.018 | 0.018 | 66.6 | -0.137 | 0.025 | 0.022 | 4.8 | 0.001 | 0.019 | 0.018 | 93.6 |
| | $\beta_{12}$ | 0.075 | 0.037 | 0.036 | 45.4 | 0.012 | 0.036 | 0.036 | 93.8 | 0.063 | 0.051 | 0.046 | 67.7 | -0.001 | 0.038 | 0.037 | 93.3 |
| | $\beta_{20}$ | -0.455 | 0.076 | 0.074 | 0.0 | -0.313 | 0.082 | 0.084 | 4.8 | 0.020 | 0.191 | 0.154 | 90.3 | 0.047 | 0.200 | 0.204 | 96.8 |
| | $\beta_{21}$ | 0.986 | 0.041 | 0.039 | 0.0 | 0.684 | 0.052 | 0.054 | 0.0 | -0.029 | 0.199 | 0.185 | 92.8 | -0.085 | 0.220 | 0.249 | 96.1 |
| | $\beta_{22}$ | -0.712 | 0.077 | 0.073 | 0.0 | -0.509 | 0.087 | 0.086 | 0.1 | 0.013 | 0.217 | 0.184 | 91.0 | 0.068 | 0.244 | 0.252 | 97.1 |
| | $\sigma$ | - | - | - | - | 0.029 | 0.028 | 0.028 | 83.5 | 0.155 | 0.026 | 0.026 | 0.0 | 0.002 | 0.029 | 0.030 | 94.6 |
| | $\xi$ | - | - | - | - | -0.137 | 0.035 | 0.037 | 4.5 | 0.275 | 0.056 | 0.048 | 0.1 | -0.002 | 0.065 | 0.069 | 95.4 |
| 0.7 | $\beta_{10}$ | 0.470 | 0.040 | 0.041 | 0.0 | 0.013 | 0.038 | 0.040 | 94.0 | 0.466 | 0.046 | 0.046 | 0.4 | -0.004 | 0.040 | 0.041 | 95.1 |
| | $\beta_{11}$ | -0.146 | 0.020 | 0.020 | 0.0 | -0.027 | 0.019 | 0.019 | 70.9 | -0.136 | 0.026 | 0.023 | 5.2 | 0.001 | 0.020 | 0.020 | 94.0 |
| | $\beta_{12}$ | 0.075 | 0.040 | 0.040 | 51.6 | 0.011 | 0.038 | 0.039 | 93.8 | 0.059 | 0.053 | 0.049 | 70.8 | -0.001 | 0.041 | 0.040 | 93.3 |
| | $\beta_{20}$ | -0.456 | 0.076 | 0.074 | 0.0 | -0.313 | 0.082 | 0.084 | 4.9 | 0.024 | 0.191 | 0.157 | 91.0 | 0.047 | 0.200 | 0.203 | 96.8 |
| | $\beta_{21}$ | 0.986 | 0.041 | 0.039 | 0.0 | 0.684 | 0.052 | 0.054 | 0.0 | -0.036 | 0.201 | 0.188 | 92.9 | -0.085 | 0.221 | 0.249 | 96.1 |
| | $\beta_{22}$ | -0.712 | 0.078 | 0.073 | 0.0 | -0.509 | 0.087 | 0.086 | 0.1 | 0.019 | 0.221 | 0.187 | 91.6 | 0.067 | 0.243 | 0.252 | 97.1 |
| | $\sigma$ | - | - | - | - | 0.028 | 0.033 | 0.034 | 87.0 | 0.255 | 0.028 | 0.029 | 0.0 | 0.002 | 0.034 | 0.035 | 94.6 |
| | $\xi$ | - | - | - | - | -0.137 | 0.038 | 0.041 | 7.4 | 0.253 | 0.056 | 0.048 | 0.4 | -0.002 | 0.070 | 0.075 | 95.7 |

Table 3.2: Results for Simulation 1 where parameters for mismeasurement models are assumed known: Setting 2 with different misclassification levels ($\alpha = -4.595, -2.197$ and $-1.386$)

| | | Naive Method | | | | | | | | | | | | Proposed | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | Naive Scenario 1 | | | | Naive Scenario 2 | | | | Naive Scenario 3 | | | | | | | |
| $\alpha_0$ | | Bias | SEE | SEM | CR% | Bias | SEE | SEM | CR% | Bias | SEE | SEM | CR% | Bias | SEE | SEM | CR% |
| -4.595 | $\beta_{10}$ | 0.469 | 0.037 | 0.037 | 0.0 | 0.000 | 0.031 | 0.033 | 94.5 | 0.467 | 0.034 | 0.034 | 0.0 | -0.001 | 0.031 | 0.033 | 94.3 |
| | $\beta_{11}$ | -0.146 | 0.018 | 0.018 | 0.0 | -0.003 | 0.016 | 0.016 | 93.6 | -0.144 | 0.016 | 0.016 | 0.0 | -0.000 | 0.016 | 0.016 | 94.6 |
| | $\beta_{12}$ | 0.075 | 0.037 | 0.036 | 45.5 | 0.001 | 0.032 | 0.032 | 95.6 | 0.073 | 0.034 | 0.034 | 42.3 | 0.000 | 0.032 | 0.032 | 95.5 |
| | $\beta_{20}$ | -0.045 | 0.104 | 0.108 | 93.1 | -0.045 | 0.100 | 0.107 | 93.1 | -0.006 | 0.109 | 0.112 | 94.7 | 0.015 | 0.114 | 0.121 | 96.2 |
| | $\beta_{21}$ | 0.109 | 0.085 | 0.084 | 71.8 | 0.109 | 0.082 | 0.088 | 74.0 | 0.023 | 0.095 | 0.099 | 94.2 | -0.020 | 0.104 | 0.111 | 95.5 |
| | $\beta_{22}$ | -0.082 | 0.116 | 0.113 | 88.1 | -0.081 | 0.113 | 0.114 | 88.4 | -0.027 | 0.121 | 0.120 | 91.9 | 0.017 | 0.128 | 0.133 | 94.3 |
| | $\sigma$ | - | - | - | - | 0.002 | 0.022 | 0.023 | 95.3 | 0.037 | 0.022 | 0.023 | 65.2 | -0.001 | 0.022 | 0.023 | 95.3 |
| | $\xi$ | - | - | - | - | -0.017 | 0.040 | 0.041 | 92.4 | 0.302 | 0.040 | 0.040 | 0.0 | -0.001 | 0.043 | 0.045 | 95.5 |
| -2.197 | $\beta_{10}$ | 0.469 | 0.037 | 0.037 | 0.0 | 0.012 | 0.031 | 0.034 | 94.3 | 0.466 | 0.040 | 0.040 | 0.4 | -0.004 | 0.034 | 0.034 | 94.7 |
| | $\beta_{11}$ | -0.146 | 0.018 | 0.018 | 0.0 | -0.027 | 0.016 | 0.016 | 61.0 | -0.137 | 0.023 | 0.020 | 3.8 | 0.001 | 0.016 | 0.017 | 93.8 |
| | $\beta_{12}$ | 0.075 | 0.037 | 0.036 | 45.5 | 0.012 | 0.033 | 0.033 | 93.9 | 0.063 | 0.049 | 0.042 | 62.8 | -0.000 | 0.036 | 0.034 | 94.2 |
| | $\beta_{20}$ | -0.180 | 0.165 | 0.096 | 49.0 | -0.313 | 0.082 | 0.084 | 4.6 | 0.015 | 0.190 | 0.152 | 90.3 | 0.047 | 0.201 | 0.204 | 96.9 |
| | $\beta_{21}$ | 0.398 | 0.297 | 0.067 | 35.9 | 0.684 | 0.052 | 0.054 | 0.0 | -0.018 | 0.195 | 0.180 | 92.4 | -0.086 | 0.220 | 0.249 | 96.1 |
| | $\beta_{22}$ | -0.296 | 0.238 | 0.099 | 44.1 | -0.509 | 0.087 | 0.086 | 0.1 | 0.002 | 0.212 | 0.180 | 90.5 | 0.068 | 0.243 | 0.253 | 97.1 |
| | $\sigma$ | - | - | - | - | 0.029 | 0.022 | 0.023 | 77.1 | 0.044 | 0.023 | 0.023 | 59.4 | 0.001 | 0.024 | 0.025 | 94.9 |
| | $\xi$ | - | - | - | - | -0.137 | 0.033 | 0.034 | 2.4 | 0.303 | 0.057 | 0.049 | 0.1 | -0.002 | 0.060 | 0.063 | 95.3 |
| -1.386 | $\beta_{10}$ | 0.469 | 0.037 | 0.037 | 0.0 | 0.026 | 0.033 | 0.034 | 88.7 | 0.460 | 0.059 | 0.048 | 7.9 | -0.007 | 0.040 | 0.039 | 94.4 |
| | $\beta_{11}$ | -0.146 | 0.018 | 0.018 | 0.0 | -0.057 | 0.016 | 0.017 | 6.3 | -0.119 | 0.033 | 0.025 | 17.3 | 0.005 | 0.020 | 0.019 | 93.8 |
| | $\beta_{12}$ | 0.075 | 0.037 | 0.036 | 45.5 | 0.028 | 0.034 | 0.033 | 86.4 | 0.048 | 0.071 | 0.055 | 65.1 | -0.021 | 0.044 | 0.040 | 93.8 |
| | $\beta_{20}$ | -0.272 | 0.192 | 0.089 | 32.7 | -0.455 | 0.075 | 0.074 | 0.0 | 0.061 | 0.292 | 0.192 | 84.2 | 0.122 | 0.330 | 0.338 | 96.5 |
| | $\beta_{21}$ | 0.594 | 0.369 | 0.058 | 23.9 | 0.984 | 0.039 | 0.041 | 0.0 | -0.111 | 0.320 | 0.249 | 84.6 | -0.244 | 0.361 | 0.448 | 94.0 |
| | $\beta_{22}$ | -0.435 | 0.280 | 0.090 | 29.4 | -0.709 | 0.075 | 0.074 | 0.0 | 0.074 | 0.339 | 0.243 | 83.2 | 0.195 | 0.393 | 0.433 | 95.4 |
| | $\sigma$ | - | - | - | - | 0.054 | 0.023 | 0.024 | 35.4 | 0.119 | 0.027 | 0.023 | 47.7 | 0.036 | 0.029 | 0.029 | 89.2 |
| | $\xi$ | - | - | - | - | -0.228 | 0.029 | 0.031 | 0.0 | 0.289 | 0.081 | 0.057 | 2.9 | -0.008 | 0.080 | 0.093 | 95.8 |

Table 3.3: Results for Simulation 1 where parameters for mismeasurement models are assumed known: The setting 3 with different values of $\gamma$ ($\gamma = 0, -0.8$ and $0.8$)

| | | Naive Method | | | | | | | | | | | | Proposed | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | Naive Scenario 1 | | | | Naive Scenario 2 | | | | Naive Scenario 3 | | | | | | | |
| $\gamma$ | | Bias | SEE | SEM | CR% | Bias | SEE | SEM | CR% | Bias | SEE | SEM | CR% | Bias | SEE | SEM | CR% |
| 0.0 | $\beta_{10}$ | -0.001 | 0.033 | 0.033 | 94.2 | -0.001 | 0.031 | 0.033 | 94.2 | -0.003 | 0.032 | 0.033 | 94.1 | -0.003 | 0.032 | 0.033 | 94.4 |
| | $\beta_{11}$ | -0.000 | 0.016 | 0.016 | 94.4 | -0.000 | 0.016 | 0.016 | 94.3 | 0.000 | 0.017 | 0.016 | 93.0 | 0.000 | 0.017 | 0.016 | 93.4 |
| | $\beta_{12}$ | 0.000 | 0.032 | 0.032 | 94.9 | 0.000 | 0.032 | 0.032 | 95.1 | -0.000 | 0.034 | 0.032 | 93.8 | -0.000 | 0.034 | 0.032 | 94.4 |
| | $\beta_{20}$ | -0.314 | 0.085 | 0.084 | 4.9 | -0.314 | 0.082 | 0.084 | 4.6 | 0.047 | 0.200 | 0.168 | 92.6 | 0.047 | 0.201 | 0.204 | 96.9 |
| | $\beta_{21}$ | 0.686 | 0.054 | 0.050 | 0.0 | 0.686 | 0.052 | 0.054 | 0.0 | -0.086 | 0.221 | 0.207 | 94.2 | -0.086 | 0.221 | 0.249 | 96.1 |
| | $\beta_{22}$ | -0.511 | 0.089 | 0.084 | 0.1 | -0.511 | 0.087 | 0.086 | 0.1 | 0.068 | 0.241 | 0.209 | 93.8 | 0.068 | 0.241 | 0.253 | 97.1 |
| | $\sigma$ | - | - | - | - | -0.001 | 0.022 | 0.023 | 94.7 | 0.006 | 0.022 | 0.022 | 94.6 | 0.001 | 0.022 | 0.023 | 93.9 |
| | $\xi$ | - | - | - | - | -0.000 | 0.032 | 0.034 | 96.2 | -0.000 | 0.056 | 0.047 | 90.8 | -0.000 | 0.056 | 0.059 | 96.5 |
| -0.8 | $\beta_{10}$ | -0.470 | 0.033 | 0.034 | 0.0 | -0.014 | 0.032 | 0.034 | 91.5 | -0.467 | 0.043 | 0.040 | 0.3 | -0.002 | 0.034 | 0.034 | 94.7 |
| | $\beta_{11}$ | 0.146 | 0.016 | 0.016 | 0.0 | 0.027 | 0.016 | 0.016 | 59.8 | 0.137 | 0.024 | 0.020 | 3.8 | -0.000 | 0.017 | 0.017 | 93.2 |
| | $\beta_{12}$ | -0.075 | 0.034 | 0.033 | 39.8 | -0.012 | 0.033 | 0.033 | 92.2 | -0.064 | 0.050 | 0.043 | 63.5 | 0.000 | 0.034 | 0.034 | 93.8 |
| | $\beta_{20}$ | -0.314 | 0.085 | 0.084 | 4.9 | -0.314 | 0.082 | 0.084 | 5.1 | 0.018 | 0.185 | 0.152 | 90.3 | 0.046 | 0.200 | 0.204 | 96.7 |
| | $\beta_{21}$ | 0.686 | 0.054 | 0.050 | 0.0 | 0.684 | 0.052 | 0.054 | 0.0 | -0.018 | 0.197 | 0.180 | 92.2 | -0.085 | 0.221 | 0.249 | 96.1 |
| | $\beta_{22}$ | -0.511 | 0.089 | 0.084 | 0.1 | -0.509 | 0.087 | 0.086 | 0.1 | 0.003 | 0.212 | 0.181 | 90.1 | 0.068 | 0.240 | 0.253 | 97.2 |
| | $\sigma$ | - | - | - | - | 0.028 | 0.023 | 0.023 | 76.9 | 0.042 | 0.023 | 0.023 | 59.1 | -0.000 | 0.025 | 0.025 | 94.7 |
| | $\xi$ | - | - | - | - | 0.136 | 0.032 | 0.034 | 1.9 | -0.304 | 0.055 | 0.048 | 0.0 | 0.001 | 0.060 | 0.063 | 95.8 |
| 0.8 | $\beta_{10}$ | 0.469 | 0.034 | 0.034 | 0.0 | 0.012 | 0.031 | 0.034 | 94.3 | 0.466 | 0.040 | 0.040 | 0.4 | -0.004 | 0.034 | 0.034 | 94.7 |
| | $\beta_{11}$ | -0.146 | 0.016 | 0.016 | 0.0 | -0.027 | 0.016 | 0.016 | 61.0 | -0.137 | 0.023 | 0.020 | 3.8 | 0.001 | 0.016 | 0.017 | 93.8 |
| | $\beta_{12}$ | 0.075 | 0.033 | 0.033 | 37.1 | 0.012 | 0.033 | 0.033 | 93.9 | 0.063 | 0.049 | 0.042 | 62.8 | -0.000 | 0.036 | 0.034 | 94.2 |
| | $\beta_{20}$ | -0.314 | 0.085 | 0.084 | 4.9 | -0.313 | 0.082 | 0.084 | 4.6 | 0.015 | 0.190 | 0.152 | 90.3 | 0.047 | 0.201 | 0.204 | 96.9 |
| | $\beta_{21}$ | 0.686 | 0.054 | 0.050 | 0.0 | 0.684 | 0.052 | 0.054 | 0.0 | -0.018 | 0.195 | 0.180 | 92.4 | -0.086 | 0.220 | 0.249 | 96.1 |
| | $\beta_{22}$ | -0.511 | 0.089 | 0.084 | 0.1 | -0.509 | 0.087 | 0.086 | 0.1 | 0.002 | 0.212 | 0.180 | 90.5 | 0.068 | 0.243 | 0.253 | 97.1 |
| | $\sigma$ | - | - | - | - | 0.029 | 0.022 | 0.023 | 77.1 | 0.044 | 0.023 | 0.023 | 59.4 | 0.001 | 0.024 | 0.025 | 94.9 |
| | $\xi$ | - | - | - | - | -0.137 | 0.033 | 0.034 | 2.4 | 0.303 | 0.057 | 0.049 | 0.1 | -0.002 | 0.060 | 0.063 | 95.3 |

Table 3.4: Results for Simulation 2: Setting 2 with different degree of measurement error when interval validation data is available ($\sigma_e = 0.1, 0.5$ and $0.7$)

| | | Method with True Measurements | | | | Proposed Method | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | | | | Known Parameter | | | | Internal Validation | | | | External Validation | | | |
| $\sigma_e$ | | Bias | SEE | SEM | CR% | Bias | SEE | SEM | CR% | Bias | SEE | SEM | CR% | Bias | SEE | SEM | CR% |
| 0.1 | $\beta_{10}$ | -0.000 | 0.027 | 0.027 | 95.1 | -0.004 | 0.034 | 0.034 | 94.7 | -0.002 | 0.033 | 0.034 | 94.3 | 0.001 | 0.028 | 0.029 | 95.5 |
| | $\beta_{11}$ | -0.000 | 0.013 | 0.013 | 95.2 | 0.001 | 0.016 | 0.017 | 93.8 | 0.002 | 0.017 | 0.016 | 93.7 | 0.000 | 0.014 | 0.014 | 94.5 |
| | $\beta_{12}$ | 0.001 | 0.026 | 0.026 | 95.5 | -0.000 | 0.036 | 0.034 | 94.2 | -0.004 | 0.034 | 0.033 | 94.9 | -0.000 | 0.028 | 0.027 | 94.3 |
| | $\beta_{20}$ | 0.004 | 0.089 | 0.092 | 95.8 | 0.047 | 0.201 | 0.204 | 96.9 | 0.037 | 0.182 | 0.198 | 97.0 | 0.031 | 0.198 | 0.208 | 95.5 |
| | $\beta_{21}$ | -0.007 | 0.075 | 0.076 | 95.9 | -0.086 | 0.220 | 0.249 | 96.1 | -0.071 | 0.194 | 0.224 | 94.9 | -0.070 | 0.231 | 0.250 | 95.1 |
| | $\beta_{22}$ | 0.009 | 0.099 | 0.098 | 94.9 | 0.068 | 0.243 | 0.253 | 97.1 | 0.056 | 0.211 | 0.219 | 96.4 | 0.055 | 0.230 | 0.236 | 94.6 |
| | $\sigma$ | - | - | - | - | 0.001 | 0.024 | 0.025 | 94.9 | 0.003 | 0.022 | 0.024 | 95.7 | -0.000 | 0.020 | 0.020 | 93.9 |
| | $\xi$ | - | - | - | - | -0.002 | 0.060 | 0.063 | 95.3 | -0.001 | 0.054 | 0.055 | 96.2 | -0.003 | 0.055 | 0.054 | 94.3 |
| | $\gamma_2$ | - | - | - | - | - | - | - | - | 0.000 | 0.006 | 0.006 | 94.9 | -0.000 | 0.004 | 0.004 | 95.6 |
| | $\sigma_e$ | - | - | - | - | - | - | - | - | -0.000 | 0.003 | 0.003 | 94.5 | -0.000 | 0.002 | 0.002 | 94.8 |
| | $\alpha_1$ | - | - | - | - | - | - | - | - | -0.025 | 0.208 | 0.215 | 95.7 | -0.006 | 0.123 | 0.123 | 94.4 |
| | $\alpha_0$ | - | - | - | - | - | - | - | - | -0.021 | 0.204 | 0.212 | 95.0 | -0.005 | 0.122 | 0.121 | 94.4 |
| 0.5 | $\beta_{10}$ | -0.000 | 0.027 | 0.027 | 95.1 | -0.004 | 0.036 | 0.038 | 95.3 | -0.003 | 0.035 | 0.037 | 94.6 | 0.001 | 0.032 | 0.034 | 94.8 |
| | $\beta_{11}$ | -0.000 | 0.013 | 0.013 | 95.2 | 0.001 | 0.019 | 0.018 | 93.6 | 0.003 | 0.018 | 0.017 | 94.6 | 0.000 | 0.016 | 0.016 | 94.2 |
| | $\beta_{12}$ | 0.001 | 0.026 | 0.026 | 95.5 | -0.001 | 0.038 | 0.037 | 93.3 | -0.005 | 0.035 | 0.035 | 93.8 | -0.001 | 0.031 | 0.030 | 94.3 |
| | $\beta_{20}$ | 0.004 | 0.089 | 0.092 | 95.8 | 0.047 | 0.200 | 0.204 | 96.8 | 0.037 | 0.182 | 0.198 | 97.0 | 0.030 | 0.198 | 0.208 | 95.4 |
| | $\beta_{21}$ | -0.007 | 0.075 | 0.076 | 95.9 | -0.085 | 0.220 | 0.249 | 96.1 | -0.071 | 0.194 | 0.224 | 94.8 | -0.069 | 0.232 | 0.250 | 95.0 |
| | $\beta_{22}$ | 0.009 | 0.099 | 0.098 | 94.9 | 0.068 | 0.244 | 0.252 | 97.1 | 0.056 | 0.211 | 0.220 | 96.4 | 0.055 | 0.230 | 0.236 | 94.6 |
| | $\sigma$ | - | - | - | - | 0.002 | 0.029 | 0.030 | 94.6 | 0.004 | 0.026 | 0.027 | 95.5 | -0.000 | 0.025 | 0.025 | 95.4 |
| | $\xi$ | - | - | - | - | -0.002 | 0.065 | 0.069 | 95.4 | -0.002 | 0.058 | 0.059 | 95.9 | -0.002 | 0.060 | 0.059 | 94.8 |
| | $\gamma_2$ | - | - | - | - | - | - | - | - | 0.002 | 0.032 | 0.032 | 94.9 | -0.000 | 0.018 | 0.018 | 95.6 |
| | $\sigma_e$ | - | - | - | - | - | - | - | - | -0.001 | 0.016 | 0.016 | 94.5 | -0.000 | 0.009 | 0.009 | 94.8 |
| | $\alpha_1$ | - | - | - | - | - | - | - | - | -0.025 | 0.208 | 0.215 | 95.7 | -0.006 | 0.123 | 0.123 | 94.4 |
| | $\alpha_0$ | - | - | - | - | - | - | - | - | -0.021 | 0.204 | 0.212 | 95.0 | -0.005 | 0.122 | 0.121 | 94.4 |
| 0.7 | $\beta_{10}$ | -0.000 | 0.027 | 0.027 | 95.1 | -0.004 | 0.040 | 0.041 | 95.1 | -0.003 | 0.038 | 0.040 | 95.4 | 0.001 | 0.037 | 0.038 | 94.7 |
| | $\beta_{11}$ | -0.000 | 0.013 | 0.013 | 95.2 | 0.001 | 0.020 | 0.020 | 94.0 | 0.003 | 0.019 | 0.019 | 95.0 | 0.000 | 0.017 | 0.017 | 95.1 |
| | $\beta_{12}$ | 0.001 | 0.026 | 0.026 | 95.5 | -0.001 | 0.041 | 0.040 | 93.3 | -0.006 | 0.037 | 0.036 | 92.8 | -0.001 | 0.034 | 0.033 | 94.8 |
| | $\beta_{20}$ | 0.004 | 0.089 | 0.092 | 95.8 | 0.047 | 0.200 | 0.203 | 96.8 | 0.037 | 0.181 | 0.198 | 97.2 | 0.030 | 0.198 | 0.208 | 95.4 |
| | $\beta_{21}$ | -0.007 | 0.075 | 0.076 | 95.9 | -0.085 | 0.221 | 0.249 | 96.1 | -0.070 | 0.194 | 0.224 | 94.7 | -0.069 | 0.233 | 0.250 | 94.9 |
| | $\beta_{22}$ | 0.009 | 0.099 | 0.098 | 94.9 | 0.067 | 0.243 | 0.252 | 97.1 | 0.055 | 0.210 | 0.220 | 96.4 | 0.054 | 0.230 | 0.236 | 94.6 |
| | $\sigma$ | - | - | - | - | 0.002 | 0.034 | 0.035 | 94.6 | 0.004 | 0.030 | 0.031 | 95.2 | -0.000 | 0.030 | 0.030 | 94.4 |
| | $\xi$ | - | - | - | - | -0.002 | 0.070 | 0.075 | 95.7 | -0.002 | 0.061 | 0.063 | 96.0 | -0.002 | 0.064 | 0.064 | 95.0 |
| | $\gamma_2$ | - | - | - | - | - | - | - | - | 0.003 | 0.045 | 0.045 | 94.9 | -0.001 | 0.025 | 0.026 | 95.6 |
| | $\sigma_e$ | - | - | - | - | - | - | - | - | -0.001 | 0.022 | 0.022 | 94.5 | -0.000 | 0.012 | 0.013 | 94.8 |
| | $\alpha_1$ | - | - | - | - | - | - | - | - | -0.025 | 0.208 | 0.215 | 95.7 | -0.006 | 0.123 | 0.123 | 94.4 |
| | $\alpha_0$ | - | - | - | - | - | - | - | - | -0.021 | 0.204 | 0.212 | 95.0 | -0.005 | 0.122 | 0.121 | 94.4 |

Table 3.5: Results for Simulation 2: Setting 2 with different misclassification levels with interval validation data ($\alpha = -4.595, -2.197$ and $-1.386$)

| | | Method with True Measurements | | | | Proposed Method | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | | | | Known Parameter | | | | Internal Validation | | | | External Validation | | | |
| $\alpha$ | | Bias | SEE | SEM | CR% | Bias | SEE | SEM | CR% | Bias | SEE | SEM | CR% | Bias | SEE | SEM | CR% |
| -4.595 | $\beta_{10}$ | -0.000 | 0.027 | 0.027 | 95.1 | -0.001 | 0.031 | 0.033 | 94.3 | -0.001 | 0.031 | 0.033 | 94.2 | 0.000 | 0.025 | 0.027 | 95.9 |
| | $\beta_{11}$ | -0.000 | 0.013 | 0.013 | 95.2 | -0.000 | 0.016 | 0.016 | 94.6 | 0.000 | 0.016 | 0.016 | 94.2 | 0.000 | 0.013 | 0.013 | 95.0 |
| | $\beta_{12}$ | 0.001 | 0.026 | 0.026 | 95.5 | 0.000 | 0.032 | 0.032 | 95.5 | 0.000 | 0.032 | 0.032 | 95.7 | -0.000 | 0.025 | 0.026 | 95.7 |
| | $\beta_{20}$ | 0.004 | 0.089 | 0.092 | 95.8 | 0.015 | 0.114 | 0.121 | 96.2 | 0.013 | 0.114 | 0.122 | 96.5 | 0.009 | 0.105 | 0.104 | 94.7 |
| | $\beta_{21}$ | -0.007 | 0.075 | 0.076 | 95.9 | -0.020 | 0.104 | 0.111 | 95.5 | -0.019 | 0.104 | 0.109 | 97.0 | -0.018 | 0.095 | 0.097 | 94.3 |
| | $\beta_{22}$ | 0.009 | 0.099 | 0.098 | 94.9 | 0.017 | 0.128 | 0.133 | 94.3 | 0.016 | 0.122 | 0.130 | 94.3 | 0.015 | 0.112 | 0.112 | 95.2 |
| | $\sigma$ | - | - | - | - | -0.001 | 0.022 | 0.023 | 95.3 | -0.001 | 0.021 | 0.023 | 95.3 | -0.001 | 0.018 | 0.019 | 94.5 |
| | $\xi$ | - | - | - | - | -0.001 | 0.043 | 0.045 | 95.5 | -0.001 | 0.042 | 0.044 | 95.1 | -0.000 | 0.037 | 0.037 | 93.8 |
| | $\gamma_2$ | - | - | - | - | - | - | - | - | 0.000 | 0.006 | 0.006 | 94.4 | -0.000 | 0.004 | 0.004 | 95.6 |
| | $\sigma_e$ | - | - | - | - | - | - | - | - | -0.000 | 0.003 | 0.003 | 94.3 | -0.000 | 0.002 | 0.002 | 94.8 |
| | $\alpha_1$ | - | - | - | - | - | - | - | - | -1.767 | 0.566 | 0.693 | 96.4 | -0.085 | 0.368 | 0.385 | 95.8 |
| | $\alpha_0$ | - | - | - | - | - | - | - | - | -1.612 | 0.572 | 0.678 | 94.6 | -0.086 | 0.364 | 0.373 | 95.6 |
| -2.197 | $\beta_{10}$ | -0.000 | 0.027 | 0.027 | 95.1 | -0.004 | 0.034 | 0.034 | 94.7 | -0.002 | 0.033 | 0.034 | 94.3 | 0.001 | 0.028 | 0.029 | 95.5 |
| | $\beta_{11}$ | -0.000 | 0.013 | 0.013 | 95.2 | 0.001 | 0.016 | 0.017 | 93.8 | 0.002 | 0.017 | 0.016 | 93.7 | 0.000 | 0.014 | 0.014 | 94.5 |
| | $\beta_{12}$ | 0.001 | 0.026 | 0.026 | 95.5 | -0.000 | 0.036 | 0.034 | 94.2 | -0.004 | 0.034 | 0.033 | 94.9 | -0.000 | 0.028 | 0.027 | 94.3 |
| | $\beta_{20}$ | 0.004 | 0.089 | 0.092 | 95.8 | 0.047 | 0.201 | 0.204 | 96.9 | 0.037 | 0.182 | 0.198 | 97.0 | 0.031 | 0.198 | 0.208 | 95.5 |
| | $\beta_{21}$ | -0.007 | 0.075 | 0.076 | 95.9 | -0.086 | 0.220 | 0.249 | 96.1 | -0.071 | 0.194 | 0.224 | 94.9 | -0.070 | 0.231 | 0.250 | 95.1 |
| | $\beta_{22}$ | 0.009 | 0.099 | 0.098 | 94.9 | 0.068 | 0.243 | 0.253 | 97.1 | 0.056 | 0.211 | 0.219 | 96.4 | 0.055 | 0.230 | 0.236 | 94.6 |
| | $\sigma$ | - | - | - | - | 0.001 | 0.024 | 0.025 | 94.9 | 0.003 | 0.022 | 0.024 | 95.7 | -0.000 | 0.020 | 0.020 | 93.9 |
| | $\xi$ | - | - | - | - | -0.002 | 0.060 | 0.063 | 95.3 | -0.001 | 0.054 | 0.055 | 96.2 | -0.003 | 0.055 | 0.054 | 94.3 |
| | $\gamma_2$ | - | - | - | - | - | - | - | - | 0.000 | 0.006 | 0.006 | 94.9 | -0.000 | 0.004 | 0.004 | 95.6 |
| | $\sigma_e$ | - | - | - | - | - | - | - | - | -0.000 | 0.003 | 0.003 | 94.5 | -0.000 | 0.002 | 0.002 | 94.8 |
| | $\alpha_1$ | - | - | - | - | - | - | - | - | -0.025 | 0.208 | 0.215 | 95.7 | -0.006 | 0.123 | 0.123 | 94.4 |
| | $\alpha_0$ | - | - | - | - | - | - | - | - | -0.021 | 0.204 | 0.212 | 95.0 | -0.005 | 0.122 | 0.121 | 94.4 |
| -1.386 | $\beta_{10}$ | -0.000 | 0.027 | 0.027 | 95.1 | -0.007 | 0.040 | 0.039 | 94.5 | -0.003 | 0.039 | 0.037 | 94.7 | 0.003 | 0.036 | 0.034 | 95.1 |
| | $\beta_{11}$ | -0.000 | 0.013 | 0.013 | 95.2 | 0.005 | 0.020 | 0.019 | 94.0 | 0.008 | 0.019 | 0.018 | 94.6 | 0.010 | 0.017 | 0.016 | 94.2 |
| | $\beta_{12}$ | 0.001 | 0.026 | 0.026 | 95.5 | -0.021 | 0.044 | 0.039 | 94.2 | -0.021 | 0.039 | 0.035 | 94.2 | -0.009 | 0.033 | 0.031 | 94.8 |
| | $\beta_{20}$ | 0.004 | 0.089 | 0.092 | 95.8 | 0.122 | 0.330 | 0.337 | 96.4 | 0.103 | 0.276 | 0.313 | 96.5 | 0.102 | 0.322 | 0.342 | 94.6 |
| | $\beta_{21}$ | -0.007 | 0.075 | 0.076 | 95.9 | -0.244 | 0.361 | 0.446 | 93.9 | -0.224 | 0.325 | 0.390 | 94.2 | -0.245 | 0.376 | 0.420 | 90.9 |
| | $\beta_{22}$ | 0.009 | 0.099 | 0.098 | 94.9 | 0.195 | 0.393 | 0.430 | 95.3 | 0.167 | 0.335 | 0.352 | 94.9 | 0.188 | 0.362 | 0.393 | 92.4 |
| | $\sigma$ | - | - | - | - | 0.036 | 0.029 | 0.029 | 89.7 | 0.031 | 0.025 | 0.026 | 93.2 | 0.037 | 0.025 | 0.024 | 90.5 |
| | $\xi$ | - | - | - | - | -0.008 | 0.080 | 0.093 | 95.7 | -0.003 | 0.069 | 0.077 | 96.9 | -0.009 | 0.072 | 0.080 | 94.6 |
| | $\gamma_2$ | - | - | - | - | - | - | - | - | 0.000 | 0.006 | 0.006 | 94.9 | -0.000 | 0.004 | 0.004 | 95.7 |
| | $\sigma_e$ | - | - | - | - | - | - | - | - | -0.000 | 0.003 | 0.003 | 94.6 | -0.000 | 0.002 | 0.002 | 94.6 |
| | $\alpha_1$ | - | - | - | - | - | - | - | - | -0.010 | 0.151 | 0.160 | 95.9 | -0.007 | 0.089 | 0.092 | 94.9 |
| | $\alpha_0$ | - | - | - | - | - | - | - | - | -0.005 | 0.153 | 0.158 | 95.1 | -0.002 | 0.085 | 0.091 | 95.2 |

Table 3.6: Results for Simulation 2: Setting 3 with different values of $\gamma$ with validation data ($\gamma = 0.0, -0.8$ and 0.8)

| | | Method with True Measurements | | | | Proposed Method | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | | | | Known Parameter | | | | Internal Validation | | | | External Validation | | | |
| $\gamma$ | | Bias | SEE | SEM | CR% | Bias | SEE | SEM | CR% | Bias | SEE | SEM | CR% | Bias | SEE | SEM | CR% |
| 0.0 | $\beta_{10}$ | -0.000 | 0.027 | 0.027 | 95.1 | -0.003 | 0.032 | 0.033 | 94.4 | -0.002 | 0.031 | 0.033 | 94.0 | 0.001 | 0.025 | 0.027 | 96.2 |
| | $\beta_{11}$ | -0.000 | 0.013 | 0.013 | 95.2 | 0.000 | 0.017 | 0.016 | 93.4 | 0.001 | 0.016 | 0.016 | 94.0 | -0.000 | 0.013 | 0.013 | 95.3 |
| | $\beta_{12}$ | 0.001 | 0.026 | 0.026 | 95.5 | -0.000 | 0.034 | 0.032 | 94.4 | -0.003 | 0.033 | 0.032 | 94.3 | 0.001 | 0.027 | 0.026 | 95.2 |
| | $\beta_{20}$ | 0.004 | 0.089 | 0.092 | 95.8 | 0.047 | 0.201 | 0.204 | 96.9 | 0.037 | 0.182 | 0.198 | 97.0 | 0.031 | 0.199 | 0.208 | 95.5 |
| | $\beta_{21}$ | -0.007 | 0.075 | 0.076 | 95.9 | -0.086 | 0.221 | 0.249 | 96.1 | -0.070 | 0.195 | 0.225 | 94.8 | -0.070 | 0.231 | 0.250 | 95.1 |
| | $\beta_{22}$ | 0.009 | 0.099 | 0.098 | 94.9 | 0.068 | 0.241 | 0.253 | 97.1 | 0.056 | 0.212 | 0.220 | 96.4 | 0.055 | 0.230 | 0.237 | 94.5 |
| | $\sigma$ | - | - | - | - | 0.001 | 0.022 | 0.023 | 93.9 | 0.002 | 0.022 | 0.022 | 94.7 | -0.001 | 0.018 | 0.018 | 94.2 |
| | $\xi$ | - | - | - | - | -0.000 | 0.056 | 0.059 | 96.5 | -0.001 | 0.050 | 0.051 | 95.7 | -0.002 | 0.049 | 0.048 | 95.0 |
| | $\gamma_2$ | - | - | - | - | - | - | - | - | 0.000 | 0.006 | 0.006 | 94.9 | -0.000 | 0.004 | 0.004 | 95.6 |
| | $\sigma_e$ | - | - | - | - | - | - | - | - | -0.000 | 0.003 | 0.003 | 94.5 | -0.000 | 0.002 | 0.002 | 94.8 |
| | $\alpha_1$ | - | - | - | - | - | - | - | - | -0.025 | 0.208 | 0.215 | 95.7 | -0.006 | 0.123 | 0.123 | 94.4 |
| | $\alpha_0$ | - | - | - | - | - | - | - | - | -0.021 | 0.204 | 0.212 | 95.0 | -0.005 | 0.122 | 0.121 | 94.4 |
| -0.8 | $\beta_{10}$ | -0.000 | 0.027 | 0.027 | 95.1 | -0.002 | 0.034 | 0.034 | 94.7 | -0.001 | 0.033 | 0.034 | 94.5 | 0.001 | 0.028 | 0.029 | 95.9 |
| | $\beta_{11}$ | -0.000 | 0.013 | 0.013 | 95.2 | -0.000 | 0.017 | 0.017 | 93.2 | 0.000 | 0.017 | 0.016 | 94.1 | -0.001 | 0.013 | 0.014 | 95.9 |
| | $\beta_{12}$ | 0.001 | 0.026 | 0.026 | 95.5 | 0.000 | 0.034 | 0.034 | 93.8 | -0.001 | 0.034 | 0.033 | 94.5 | 0.002 | 0.028 | 0.028 | 95.7 |
| | $\beta_{20}$ | 0.004 | 0.089 | 0.092 | 95.8 | 0.046 | 0.200 | 0.204 | 96.7 | 0.037 | 0.182 | 0.198 | 97.1 | 0.030 | 0.199 | 0.208 | 95.4 |
| | $\beta_{21}$ | -0.007 | 0.075 | 0.076 | 95.9 | -0.085 | 0.221 | 0.249 | 96.1 | -0.070 | 0.195 | 0.225 | 94.9 | -0.069 | 0.231 | 0.250 | 95.1 |
| | $\beta_{22}$ | 0.009 | 0.099 | 0.098 | 94.9 | 0.068 | 0.240 | 0.253 | 97.2 | 0.056 | 0.210 | 0.219 | 96.4 | 0.054 | 0.230 | 0.237 | 94.5 |
| | $\sigma$ | - | - | - | - | -0.000 | 0.025 | 0.025 | 94.7 | 0.001 | 0.023 | 0.024 | 94.7 | -0.001 | 0.020 | 0.020 | 93.6 |
| | $\xi$ | - | - | - | - | 0.001 | 0.060 | 0.063 | 95.8 | -0.000 | 0.055 | 0.056 | 95.8 | -0.000 | 0.054 | 0.054 | 94.4 |
| | $\gamma_2$ | - | - | - | - | - | - | - | - | 0.000 | 0.006 | 0.006 | 94.9 | -0.000 | 0.004 | 0.004 | 95.6 |
| | $\sigma_e$ | - | - | - | - | - | - | - | - | -0.000 | 0.003 | 0.003 | 94.5 | -0.000 | 0.002 | 0.002 | 94.8 |
| | $\alpha_1$ | - | - | - | - | - | - | - | - | -0.025 | 0.208 | 0.215 | 95.7 | -0.006 | 0.123 | 0.123 | 94.4 |
| | $\alpha_0$ | - | - | - | - | - | - | - | - | -0.021 | 0.204 | 0.212 | 95.0 | -0.005 | 0.122 | 0.121 | 94.4 |
| 0.8 | $\beta_{10}$ | -0.000 | 0.027 | 0.027 | 95.1 | -0.004 | 0.034 | 0.034 | 94.7 | -0.002 | 0.033 | 0.034 | 94.3 | 0.001 | 0.028 | 0.029 | 95.5 |
| | $\beta_{11}$ | -0.000 | 0.013 | 0.013 | 95.2 | 0.001 | 0.016 | 0.017 | 93.8 | 0.002 | 0.017 | 0.016 | 93.7 | 0.000 | 0.014 | 0.014 | 94.5 |
| | $\beta_{12}$ | 0.001 | 0.026 | 0.026 | 95.5 | -0.000 | 0.036 | 0.034 | 94.2 | -0.004 | 0.034 | 0.033 | 94.9 | -0.000 | 0.028 | 0.027 | 94.3 |
| | $\beta_{20}$ | 0.004 | 0.089 | 0.092 | 95.8 | 0.047 | 0.201 | 0.204 | 96.9 | 0.037 | 0.182 | 0.198 | 97.0 | 0.031 | 0.198 | 0.208 | 95.5 |
| | $\beta_{21}$ | -0.007 | 0.075 | 0.076 | 95.9 | -0.086 | 0.220 | 0.249 | 96.1 | -0.071 | 0.194 | 0.224 | 94.9 | -0.070 | 0.231 | 0.250 | 95.1 |
| | $\beta_{22}$ | 0.009 | 0.099 | 0.098 | 94.9 | 0.068 | 0.243 | 0.253 | 97.1 | 0.056 | 0.211 | 0.219 | 96.4 | 0.055 | 0.230 | 0.236 | 94.6 |
| | $\sigma$ | - | - | - | - | 0.001 | 0.024 | 0.025 | 94.9 | 0.003 | 0.022 | 0.024 | 95.7 | -0.000 | 0.020 | 0.020 | 93.9 |
| | $\xi$ | - | - | - | - | -0.002 | 0.060 | 0.063 | 95.3 | -0.001 | 0.054 | 0.055 | 96.2 | -0.003 | 0.055 | 0.054 | 94.3 |
| | $\gamma_2$ | - | - | - | - | - | - | - | - | 0.000 | 0.006 | 0.006 | 94.9 | -0.000 | 0.004 | 0.004 | 95.6 |
| | $\sigma_e$ | - | - | - | - | - | - | - | - | -0.000 | 0.003 | 0.003 | 94.5 | -0.000 | 0.002 | 0.002 | 94.8 |
| | $\alpha_1$ | - | - | - | - | - | - | - | - | -0.025 | 0.208 | 0.215 | 95.7 | -0.006 | 0.123 | 0.123 | 94.4 |
| | $\alpha_0$ | - | - | - | - | - | - | - | - | -0.021 | 0.204 | 0.212 | 95.0 | -0.005 | 0.122 | 0.121 | 94.4 |

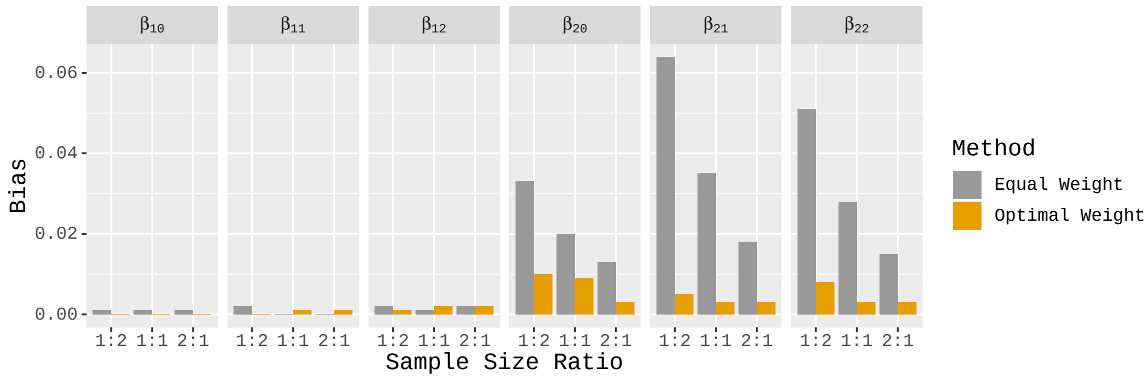


Figure 3.1: *Biases of the estimates in Simulation 3: Scenario 1 with different sample size ratios between validation data and non-validation data. Equal Weight: the proposed method with internal validation data described in Section 3.3.2. Optimal Weight: the proposed weighted estimator with optimal weights as described in Section 3.3.3.*

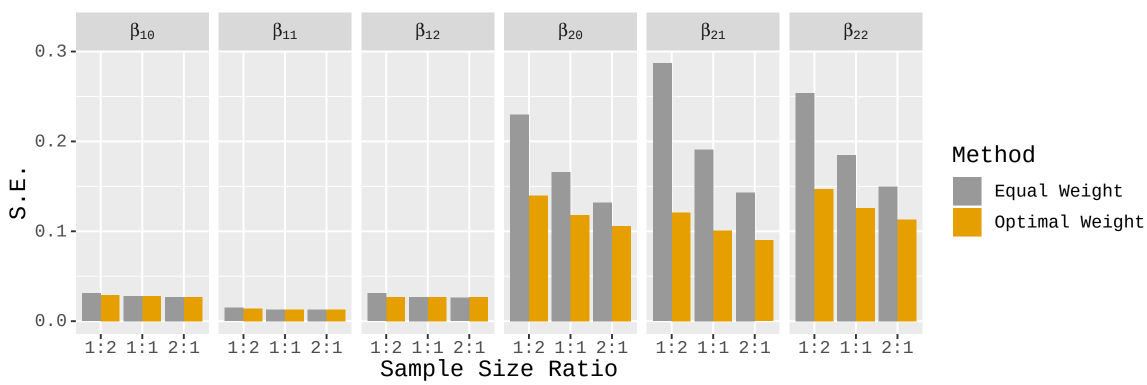


Figure 3.2: *Standard error of the estimates in Simulation 3: Scenario 1 with different sample size ratios between validation data and non-validation data. Equal Weight: the proposed method with internal validation data described in Section 3.3.2. Optimal Weight: the proposed weighted estimator with optimal weights as described in Section 3.3.3.*

Table 3.7: Results for Simulation 3: Scenario 2 with different weights and sample sizes

| | | n=1500 | | | | | n=3000 | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Parameter | Estimator | Bias | SEE | SEM | CR% | ARE [a] | Bias | SEE | SEM | CR% | ARE [a] |
| $\beta_{10}$ | Naive | 0.469 | 0.028 | 0.028 | 0.0 | - | 0.469 | 0.019 | 0.020 | 0.0 | - |
| | $\widehat{\theta}_I^{(0)}$ | 0.001 | 0.046 | 0.046 | 94.5 | 63% | -0.002 | 0.025 | 0.027 | 95.2 | 74% |
| | $\widehat{\theta}_I^{(1)}$ | -0.005 | 0.038 | 0.037 | 95.5 | 78% | 0.001 | 0.028 | 0.029 | 95.5 | 69% |
| | $\widehat{\theta}_I$ | -0.001 | 0.029 | 0.031 | 95.4 | 94% | -0.001 | 0.022 | 0.021 | 94.9 | 95% |
| | $\tilde{\theta}_I^*$ | 0.000 | 0.029 | 0.029 | 94.5 | - | -0.001 | 0.019 | 0.020 | 96.0 | - |
| $\beta_{11}$ | Naive | -0.146 | 0.013 | 0.013 | 0.0 | - | -0.146 | 0.009 | 0.009 | 0.0 | - |
| | $\widehat{\theta}_I^{(0)}$ | -0.000 | 0.022 | 0.022 | 94.8 | 64% | -0.000 | 0.012 | 0.013 | 94.8 | 69% |
| | $\widehat{\theta}_I^{(1)}$ | 0.007 | 0.017 | 0.017 | 95.1 | 82% | 0.000 | 0.014 | 0.014 | 94.5 | 64% |
| | $\widehat{\theta}_I$ | 0.002 | 0.013 | 0.015 | 94.2 | 93% | 0.000 | 0.011 | 0.010 | 94.5 | 90% |
| | $\tilde{\theta}_I^*$ | -0.000 | 0.014 | 0.014 | 94.8 | - | -0.000 | 0.009 | 0.009 | 95.3 | - |
| $\beta_{12}$ | Naive | 0.077 | 0.026 | 0.027 | 19.2 | - | 0.074 | 0.019 | 0.019 | 2.8 | - |
| | $\widehat{\theta}_I^{(0)}$ | 0.002 | 0.045 | 0.045 | 94.7 | 6% | -0.001 | 0.025 | 0.026 | 93.7 | 73% |
| | $\widehat{\theta}_I^{(1)}$ | -0.011 | 0.035 | 0.034 | 93.8 | 79% | -0.000 | 0.028 | 0.027 | 94.3 | 70% |
| | $\widehat{\theta}_I$ | -0.002 | 0.027 | 0.031 | 94.7 | 87% | -0.002 | 0.022 | 0.019 | 94.9 | 100% |
| | $\tilde{\theta}_I^*$ | 0.001 | 0.028 | 0.027 | 95.0 | - | -0.001 | 0.020 | 0.019 | 95.0 | - |
| $\beta_{20}$ | Naive | -0.319 | 0.067 | 0.069 | 0.2 | - | -0.321 | 0.050 | 0.048 | 0.0 | - |
| | $\widehat{\theta}_I^{(0)}$ | 0.014 | 0.155 | 0.160 | 95.3 | 88% | 0.003 | 0.088 | 0.092 | 94.8 | 90% |
| | $\widehat{\theta}_I^{(1)}$ | 0.081 | 0.284 | 0.296 | 96.1 | 47% | 0.031 | 0.198 | 0.208 | 95.5 | 40% |
| | $\widehat{\theta}_I$ | 0.033 | 0.186 | 0.230 | 96.50 | 61% | 0.013 | 0.148 | 0.148 | 96.1 | 56% |
| | $\tilde{\theta}_I^*$ | -0.010 | 0.130 | 0.140 | 95.9 | - | -0.005 | 0.081 | 0.083 | 94.2 | - |
| $\beta_{21}$ | Naive | 0.692 | 0.046 | 0.041 | 0.0 | - | 0.693 | 0.032 | 0.029 | 0.0 | - |
| | $\widehat{\theta}_I^{(0)}$ | -0.027 | 0.126 | 0.133 | 95.2 | 91% | -0.007 | 0.074 | 0.076 | 94.0 | 93% |
| | $\widehat{\theta}_I^{(1)}$ | -0.161 | 0.302 | 0.358 | 93.3 | 34% | -0.070 | 0.231 | 0.250 | 95.1 | 28% |
| | $\widehat{\theta}_I$ | -0.064 | 0.204 | 0.287 | 94.60 | 42% | -0.033 | 0.184 | 0.175 | 94.5 | 41% |
| | $\tilde{\theta}_I^*$ | 0.005 | 0.118 | 0.121 | 94.0 | - | 0.005 | 0.072 | 0.071 | 93.1 | - |
| $\beta_{22}$ | Naive | -0.516 | 0.071 | 0.068 | 0.0 | - | -0.517 | 0.050 | 0.048 | 0.0 | - |
| | $\widehat{\theta}_I^{(0)}$ | 0.028 | 0.168 | 0.171 | 95.2 | 86% | 0.008 | 0.095 | 0.097 | 95.0 | 92% |
| | $\widehat{\theta}_I^{(1)}$ | 0.119 | 0.287 | 0.319 | 94.9 | 46% | 0.055 | 0.230 | 0.236 | 94.6 | 38% |
| | $\widehat{\theta}_I$ | 0.051 | 0.201 | 0.254 | 95.9 | 58% | 0.027 | 0.164 | 0.160 | 94.9 | 56% |
| | $\tilde{\theta}_I^*$ | -0.008 | 0.144 | 0.147 | 93.5 | - | -0.005 | 0.087 | 0.089 | 94.2 | - |

[a] ARE: Average relative efficiency comparing to the optimal weights.

Table 3.8: The implementation of the GEE in the genetic study under different specification of measurement error variance and misclassfication rate

| | Naive Method | | | Proposed Method[a] | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | | | | Setting 1 | | | Setting 2 | | |
| | Estimate | S.E. | p-value | Estimate | S.E. | p-value | Estimate | S.E. | p-value |
| | Estimates for Response Model | | | | | | | | |
| $\beta_{10}$ | **57.958** | **0.540** | **<0.001** | **59.001** | **0.555** | **<0.001** | **57.508** | **0.482** | **<0.001** |
| $\beta_{11}$ | 0.471 | 0.375 | 0.209 | **1.105** | **0.388** | **0.004** | 0.924 | 0.339 | 0.006 |
| $\beta_{12}$ | 0.001 | 0.013 | 0.964 | -0.000 | 0.019 | 0.986 | -0.000 | 0.015 | 0.997 |
| $\beta_{13}$ | -0.012 | 0.022 | 0.599 | -0.008 | 0.023 | 0.720 | -0.008 | 0.021 | 0.692 |
| $\beta_{20}$ | **-2.156** | **0.540** | **<0.001** | **-2.615** | **0.313** | **<0.001** | **-2.535** | **0.286** | **<0.001** |
| $\beta_{21}$ | 0.203 | 0.375 | 0.588 | 0.273 | 0.202 | 0.176 | 0.259 | 0.183 | 0.157 |
| $\beta_{22}$ | -0.001 | 0.013 | 0.958 | 0.003 | 0.009 | 0.718 | 0.002 | 0.008 | 0.816 |
| $\beta_{23}$ | 0.003 | 0.022 | 0.896 | -0.009 | 0.011 | 0.400 | -0.006 | 0.010 | 0.520 |
| | Estimates for Mismeasurement Models | | | | | | | | |
| $\gamma_0$ | - | - | - | **17.278** | **1.096** | **<0.001** | **17.285** | **1.097** | **<0.001** |
| $\gamma_1$ | - | - | - | **0.560** | **0.016** | **<0.001** | **0.560** | **0.016** | **<0.001** |
| $\gamma_2$ | - | - | - | - | - | - | -0.371 | 0.390 | 0.341 |
| $\gamma_3$ | - | - | - | **0.334** | **0.042** | **<0.001** | **0.334** | **0.042** | **<0.001** |
| $\sigma_e$ | - | - | - | **2.966** | **0.069** | **<0.001** | **2.967** | **0.069** | **<0.001** |
| $\alpha_{01}$ | - | - | - | -0.201 | 0.260 | 0.441 | -2.060 | 2.442 | 0.399 |
| $\alpha_{x1}$ | - | - | - | - | - | - | 0.071 | 0.093 | 0.444 |
| $\alpha_{00}$ | - | - | - | **-4.446** | **0.380** | **<0.001** | 0.685 | 7.498 | 0.927 |
| $\alpha_{x0}$ | - | - | - | - | - | - | -0.206 | 0.311 | 0.509 |

[a] Optimal weights are chosen for the internal validation data.

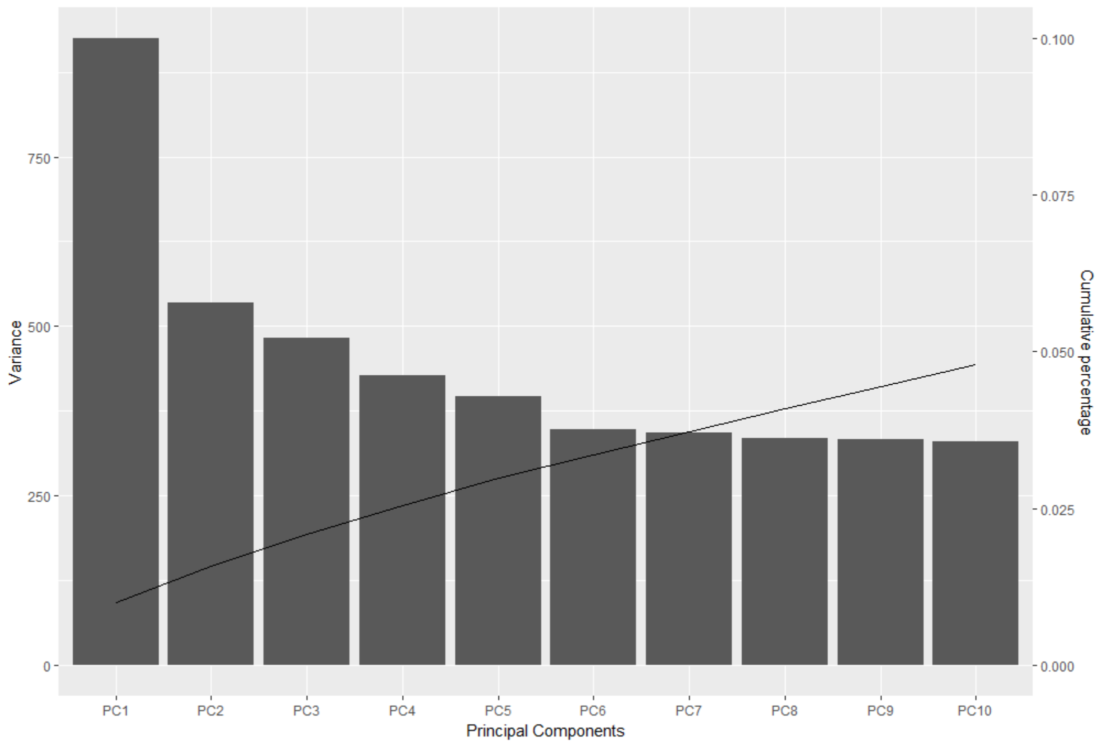


Figure 3.3: *The screeplot of the principal component analysis of the genotype data. The top 10 principal components are presented. The bar refers to the variance of each principal components. The solid line refers to the cumulative percentage of the variance.*

# Chapter 4

# Generalized Network Structured Models with Mixed Responses subject to Measurement Error and Misclassification

As a continuation of the previous two chapters, in this chapter, we further consider settings where covariates are of a high dimension and are associated with a network structure. In Section 4.1, we start with the saturated response model and discuss the estimation of the model parameters under the framework of generalized estimating equations. In Section 4.2, we develop the generalized network structured model (GNSM), describe a two-step implementation procedure of GNSM, and present the theoretical results for the proposed estimators. In Sections 4.3 and 4.4, we further extend the GNSM to the augmented GNSM by accounting for the effects due to measurement error and misclassification in the response variables, and we also discuss efficiency issues for the proposed estimators. Simulations studies are conducted in Section 4.5 to evaluate the performance of GNSM in regards to both variable selection and parameter estimation. In Section 4.6, we apply the augmented GNSM to a mice data set arising from a genome-wide association study.

## 4.1 Notation and Framework

Suppose $n$ independent subjects are recruited for the study. For subject $i = 1, \ldots, n$, correlated responses $Y_{i1}$ and $Y_{i2}$ are measured, where $Y_{i1}$ denotes the continuous response, and $Y_{i2}$ denotes the binary response. Define $Y_i = (Y_{i1}, Y_{i2})^{\mathrm{T}}$. Let $X_i = (X_{i1}, \ldots, X_{ip})^{\mathrm{T}}$ denote the covariate vector for subject $i$, where $p$ is the number of covariates. For $i = 1, \ldots, n$ and $j = 1, 2$, let $\mu_{ij} = E(Y_{ij}|X_i)$ be the conditional mean of the $Y_{ij}$, given $X_i$, and let $v_{ij} = \mathrm{Var}(Y_{ij}|X_i)$ be the conditional variance of $Y_{ij}$ given covariates $X_i$.

### 4.1.1 Saturated Response Model

To characterize the relationship among the covariates $\{X_{i1}, \ldots, X_{ip}\}$, we use a *graph*, denoted as $G_i = (V_i, \widetilde{E}_i)$, where $V_i = \{1, \cdots, p\}$ includes all the indices of covariates and $\widetilde{E}_i = V_i \times V_i$ is an index set of all pairs of covariates. A covariate $X_{ij}$ is represented by a *vertex* of the graph $G_i$ if $j \in V_i$. A pair of predictors $\{X_{is}, X_{it}\}$ is linked by an *edge* of the graph $G_i$ if $(s, t) \in E_i$, and $X_{is}$ and $X_{it}$ are conditional dependent, given the remaining variables; let $E_i$ denote the set of all pairs $(s, t)$ if $X_{is}$ and $X_{it}$ are linked by an edge. We assume that all subjects have the same covariate dependence structures. Namely, $G_1 = G_2 = \cdots = G_n \equiv G$. We now let $V$, $\widetilde{E}$ and $E$ denote the vertices, $V \times V$, and edges of the graph, respectively.

We first consider a saturated model which includes all main effects and interactions,

$$\begin{aligned} g_1(\mu_{i1}) &= \beta_{1,0} + \sum_{k \in V} \beta_{1,k} X_{ik} + \sum_{(s,t) \in \widetilde{E}} \beta_{1,st} X_{is} X_{it}; \\ g_2(\mu_{i2}) &= \beta_{2,0} + \sum_{k \in V} \beta_{2,k} X_{ik} + \sum_{(s,t) \in \widetilde{E}} \beta_{2,st} X_{is} X_{it}, \end{aligned} \tag{4.1}$$

where $\beta = (\beta_{\mathrm{M}}^{\mathrm{T}}, \beta_{\mathrm{P}}^{\mathrm{T}})^{\mathrm{T}}$ with $\beta_{\mathrm{M}} = (\beta_{1,0}, \beta_{2,0}, \beta_{1,k}, \beta_{2,k} : k \in V)^{\mathrm{T}}$ and $\beta_{\mathrm{P}} = (\beta_{1,st}, \beta_{2,st} : (s,t) \in \widetilde{E})^{\mathrm{T}}$, and $g_1(\cdot)$ and $g_2(\cdot)$ are link functions. For example, one may specify $g_1(t) = t$ and $g_2(t) = \log\{t/(1-t)\}$. Let $p_{\mathrm{s}}$ be the dimension of $\beta$.

### 4.1.2 Estimating Equation

In this chapter, we start with the same estimation procedure as in described in Section 3.1.2. We use the notation $\widetilde{U}_{i1}(\beta, \phi)$ and $\widetilde{U}_{i2}(\beta, \phi)$ to denote the estimation equations constructed based on the saturated model (4.1).

Let $\mu_i = (\mu_{i1}, \mu_{i2})^{\mathrm{T}}$. For $i = 1, \ldots, n$, define the estimating functions

$$\widetilde{U}_{i1}(\beta, \phi) = D_{i1}^{\mathrm{T}} V_{i1}^{-1} (Y_i - \mu_i); \tag{4.2}$$

$$\widetilde{U}_{i2}(\beta, \phi) = D_{i2}^{\mathrm{T}} V_{i2}^{-1} (S_i - \xi_i), \tag{4.3}$$

where $D_{i1} = \frac{\partial \mu_i}{\partial \beta}$, $V_{i1}$ is given by (3.3), $\xi_i = (v_{ijk} : 1 \le j \le k \le 2)^{\mathrm{T}}$, $D_{i2} = \frac{\partial \xi_i}{\partial \phi^{\mathrm{T}}}$, $S_i = \{(Y_{ij} - \mu_{ij})(Y_{ik} - \mu_{ik}) : 1 \le j \le k \le 2\}^{\mathrm{T}}$, and $V_{i2}$ is a $3 \times 3$ weight matrix as defined in Section 3.1.2.

Let $\widetilde{U}_i(\beta, \phi) = \left(\widetilde{U}_{i1}^{\mathrm{T}}(\beta, \phi), \widetilde{U}_{i2}^{\mathrm{T}}(\beta, \phi)\right)^{\mathrm{T}}$. By the estimating function theory (e.g., Liang and Zeger, 1986; Godambe, 1991; Newey and McFadden, 1994; Yi, 2017, Section 1.3.2), under regularity conditions, solving

$$\sum_{i=1}^{n} \widetilde{U}_i(\beta, \phi) = 0$$

for $(\beta^{\mathrm{T}}, \phi^{\mathrm{T}})^{\mathrm{T}}$ gives an estimator, say, $(\widetilde{\beta}^{\mathrm{T}}, \widetilde{\phi}^{\mathrm{T}})^{\mathrm{T}}$, of $(\beta^{\mathrm{T}}, \phi^{\mathrm{T}})^{\mathrm{T}}$.

## 4.2 Generalized Network Structured Model

### 4.2.1 Model Form

To focus on modeling the pairwise associations among the components of $X_{ij}$, we consider the graphical model (Hastie et al., 2015, Section 11)

$$f(x_i; \Theta) = \exp\left\{ -\frac{1}{2} \sum_{(s,t) \in \widetilde{E}} \theta_{st} x_{is} x_{it} - \frac{1}{2} \sum_{k \in V} x_{ik}^2 - A(\Theta) \right\}, \tag{4.4}$$

where $\Theta = [\theta_{st}]$ is a $p \times p$ symmetric matrix with diagonal elements to be one and $(s, t)$ element to be $\theta_{st}$, and $A(\Theta) = -\frac{1}{2} \log \det |\frac{\Theta}{2\pi}|$ is the normalizing constant. This model basically implies that $X_i$ follows a multivariate Gaussian distribution with zero mean and covariance matrix $\Sigma$, where $\Theta = \Sigma^{-1}$; and $\Theta$ is also known as the *precision matrix*.

In model (4.4), a nonzero parameter $\theta_{st}$ implies that $X_{is}$ and $X_{it}$ are conditionally dependent, given other covariates. In applications, not every paired covariates components in $X_i$ are necessarily correlated. That is, the edge set $E$ is not necessarily identical to $\widetilde{E}$ but $E = \left\{(s, t) \in \widetilde{E} : \theta_{st} \neq 0\right\}$. To feature the dependence of the responses on the covariates with a network structure, we propose a generalized network structured model

$$\begin{aligned} g_1(\mu_{i1}) &= \beta_{1,0} + \sum_{k \in V} \beta_{1,k} X_{ik} + \sum_{(s,t) \in E} \beta_{1,st} X_{is} X_{it}; \\ g_2(\mu_{i2}) &= \beta_{2,0} + \sum_{k \in V} \beta_{2,k} X_{ik} + \sum_{(s,t) \in E} \beta_{2,st} X_{is} X_{it}, \end{aligned} \tag{4.5}$$

where $\beta_{\text{M}} = (\beta_{1,0}, \beta_{2,0}, \beta_{1,k}, \beta_{2,k} : k \in V)^{\text{T}}$ and $\beta_{\text{I}} = (\beta_{1,st}, \beta_{2,st} : (s,t) \in E)^{\text{T}}$ are the regression coefficients. To differentiate the parameters in the saturated model (4.1), we let $\beta_{\text{II}} = (\beta_{1,st}, \beta_{2,st} : (s,t) \in \widetilde{E} \setminus E)^{\text{T}}$.

### 4.2.2 Estimation Procedure

To determine the model form (4.5) as well as to estimate the associated parameters, we need first to determine the set $E$. This essentially is equivalent to selecting active interaction terms in the saturated model (4.1). In this section, we describe a two-stage procedure. In Stage 1, we determine the dependence structure of the covariates via the graphical model (4.4). In Stage 2, we use the estimating equation method to estimate the associated model parameters. These two stages are respectively described in the following two subsections in detail.

#### Stage 1: Identification of the Covariates Network Structure $E$

To identify $E$, we maximize the penalized log-likelihood function

$$\ell(\Theta) = \sum_{i=1}^{n} \log f(x_i; \Theta) - \lambda \|\Theta\|, \tag{4.6}$$

where $\lambda$ is a positive tuning parameter controlling the sparsity of the resulting parameter matrix and $\|\cdot\|$ is a penalizing norm function. A widely used norm is the $\ell_1$-norm, yielding the penalty of the Least Absolute Shrinkage and Selection Operator (Tibshirani, 1996).

Directly maximizing (4.6), such as the Graphical LASSO Algorithm, requires a computationally intensive algorithm (Friedman et al., 2008). In practice, a simpler estimation method is often carried out by a neighborhood-based likelihood derived from (4.4). Specifically, for every $s \in V$, let $X_{i,V\setminus\{s\}}$ denote the $(p-1)$-dimensional subvector of $X_i$ with its $s$th component removed, i.e., $X_{i,V\setminus\{s\}} = (X_{i1}, \cdots, X_{i,s-1}, X_{i,s+1}, \cdots, X_{ip})^{\text{T}}$. Then the conditional probability density function for $X_{is}$, given $X_{i,V\setminus\{s\}}$, is given by

$$f\left(x_{is} | x_{i,V\setminus\{s\}}; \theta_{(-s)}\right) = \exp\left\{-\frac{1}{2} x_{is} \left(\sum_{t \in V\setminus\{s\}} \theta_{st} x_{it}\right) - \frac{1}{2} x_{is}^2 - D\left(\sum_{t \in V\setminus\{s\}} \theta_{st} x_{it}\right)\right\}, \tag{4.7}$$

where $D(\cdot) = \frac{1}{2}\log 2\pi + \frac{1}{8}(\sum_{t\in V\setminus\{s\}} \theta_{st} x_{it})^2$ is the normalizing constant ensuring the integration of (4.7) equal one, and $\theta_{(-s)} = (\theta_{s1}, \cdots, \theta_{s,s-1}, \theta_{s,s+1}, \cdots, \theta_{sp})^{\text{T}}$ is a $(p-1)$-dimensional vector of parameters indicating the relationship of $X_{is}$ with all other predictors $X_{it}$ for $t \in \{1, \cdots, s-1, s+1, \cdots, p\}$ associated with (4.7).

Let $\ell\left(\theta_{(-s)}\right)$ be the log-likelihood for $\theta_{(-s)}$ multiplied by $-\frac{1}{n}$,

$$\begin{aligned}\ell\left(\theta_{(-s)}\right) &= -\frac{1}{n}\log\left\{\prod_{i=1}^{n} f\left(x_{is}|x_{i,V\backslash\{s\}};\theta_{(-s)}\right)\right\} \\ &= \frac{1}{n}\sum_{i=1}^{n}\left\{\frac{1}{2}x_{is}\left(\sum_{t\in V\backslash\{s\}}\theta_{st}x_{it}\right)+D\left(\sum_{t\in V\backslash\{s\}}\theta_{st}x_{it}\right)\right\}.\end{aligned}$$

Then an estimator of $\theta_{(-s)}$ can be obtained as

$$\widehat{\theta}_{(-s)} = \underset{\theta_{(-s)}}{\arg\min}\left\{\ell\left(\theta_{(-s)}\right)+\lambda\left\|\theta_{(-s)}\right\|_1\right\},$$

where $\lambda$ is a tuning parameter and $\|\cdot\|_1$ is the $\ell_1$-norm.

The preceding procedure is repeated for all $s \in V$ to yield an estimator $\widehat{\theta}_s$ for all $s \in V$. Let $\mathcal{N}(s) = \{t : (s,t) \in E\}$ denote the neighbor set for $s \in V$. To determine an estimated set of edges, we define

$$\widehat{\mathcal{N}}(s) = \left\{t \in V : \widehat{\theta}_{st} \neq 0\right\}$$

as the estimated neighbor set for $s \in V$. It is worth noting that, for $(s,t) \in E$, the estimates $\widehat{\theta}_{st}$ and $\widehat{\theta}_{ts}$ are not necessarily identical or both equal to zero at the same time although $\theta_{st}$ and $\theta_{ts}$ are constrained to be equal. Therefore, $s \in \widehat{\mathcal{N}}(t)$ does not imply $t \in \widehat{\mathcal{N}}(s)$, and vice versa. To overcome this discrepancy, we apply the OR rule (Meinshausen and Bühlmann, 2006; Hastie et al., 2015, Page 255) when determining the inclusion of edge $(s,t)$ in the estimated edge set $\widehat{E}$ by either $s \in \widehat{\mathcal{N}}(t)$ or $t \in \widehat{\mathcal{N}}(s)$. Namely, we take

$$\widehat{E} = \left\{(s,t) : s \in \widehat{\mathcal{N}}(t) \text{ or } t \in \widehat{\mathcal{N}}(s)\right\}$$

as the estimated set of the edges.

The preceding procedure requires a suitable choice of the tuning parameter $\lambda$. Several methods, including the cross-validation (e.g., the BIC), the stability approach to regularization selection (StARS) (Liu et al., 2010), and the rotation information criterion (Zhao et al., 2012), may be employed to determine the optimal value of $\lambda$. In this paper, we use the rotation information criterion. To be specific, we take a set of candidate values for $\lambda$, such as an equally spaced sequence from 0 to a certain positive value. We first arrange the sample data in an array and then shuffle the data by randomly rotating the order of subjects (rows) for each variable (columns). This procedure creates a reshuffled dataset so that the association between paired variables is minimal. Then we implement our method to this reshuffled dataset and find the smallest value of $\lambda$ such that all edges are regularized to 0. We repeat this procedure several times (such as 10 times) using the R package `huge` (Zhao et al., 2012) and select the resulting smallest value of $\lambda$.

Under regular conditions in Meinshausen and Bühlmann (2006), we have that as $n \to \infty$,

$$P\left(\widehat{E} = E\right) \longrightarrow 1.$$

That is, the estimated set of edges $\widehat{E}$ approximates the true network structure $E$ in probability. This results is available in Ravikumar et al. (2010, Section 2.2) and Theorem 5(b) of Yang et al. (2015).

### Stage 2: Estimation of Model Parameters

Once the network structure for $X$ is identified, estimation of the model parameters in model (4.5) can proceed in the same manner as in Section 4.1.2, with modifications of estimating functions (4.2) and (4.3) to reflect the difference in the parameters for models (4.1) and (4.5). Let $(\widehat{\beta}_{\mathrm{M}}^{\mathrm{T}}, \widehat{\beta}_{\mathrm{I}}^{\mathrm{T}})^{\mathrm{T}}$ and $\widehat{\phi}$, respectively, denote the resultant estimators of $(\beta_{\mathrm{M}}^{\mathrm{T}}, \beta_{\mathrm{I}}^{\mathrm{T}})^{\mathrm{T}}$ and $\phi$. Let $U_{i1}(\beta_{\mathrm{M}}, \beta_{\mathrm{I}}, \phi)$ and $U_{i2}(\beta_{\mathrm{M}}, \beta_{\mathrm{I}}, \phi)$ denote, respectively, the estimating functions by modifying $\widetilde{U}_{i1}(\beta, \phi)$ and $\widetilde{U}_{i2}(\beta, \phi)$ in (4.2) and (4.3). Let $U_i(\beta_{\mathrm{M}}, \beta_{\mathrm{I}}, \phi) = \left(U_{i1}^{\mathrm{T}}(\beta_{\mathrm{M}}, \beta_{\mathrm{I}}, \phi), U_{i2}^{\mathrm{T}}(\beta_{\mathrm{M}}, \beta_{\mathrm{I}}, \phi)\right)^{\mathrm{T}}$.

Then solving

$$\sum_{i=1}^{n} U_i(\beta_{\mathrm{M}}, \beta_{\mathrm{I}}, \phi) = 0 \tag{4.8}$$

for $(\beta_{\mathrm{M}}^{\mathrm{T}}, \beta_{\mathrm{I}}^{\mathrm{T}}, \phi^{\mathrm{T}})^{\mathrm{T}}$ gives an estimator, say, $(\widehat{\beta}_{\mathrm{M}}^{\mathrm{T}}, \widehat{\beta}_{\mathrm{I}}^{\mathrm{T}}, \widehat{\phi}^{\mathrm{T}})^{\mathrm{T}}$.

We comment that the true edge set $E$ in model (4.5) is unknown and is estimated via the procedure in Section 4.2.2, thus inducing extra uncertainty in implementing (4.8) for the estimation of parameters. Consistent with the comments after (4.5) on the expression of the parameters in models (4.1) and (4.5), we set $\widehat{\beta}_{\mathrm{II}} = 0$ as the estimator for the subvector of $\beta_{\mathrm{II}}$ which includes the coefficients corresponding to the covariates in the unselected edge set $\widetilde{E} \setminus \widehat{E}$. It is noted that $\widehat{\beta}_{\mathrm{II}}$ may not be identical to $\beta_{\mathrm{II}}$: their dimension can even be different due to the variability induced in estimating $E$ in Section 4.2.2. We now establish theoretical results for the estimator obtained from preceding Stages 1 and 2.

**Theorem 4.1** *Let $\beta_{\mathrm{II}(0)}$ be the true value of $\beta_{\mathrm{II}}$. Under regularity conditions, there exists some constant $c > 0$ such that*

$$P(\widehat{\beta}_{\mathrm{II}} = \beta_{\mathrm{II}(0)}) \geq 1 - O\left(\exp(-cn)\right).$$

This theorem suggests that as $n \to \infty$, with the probability approaching 1, the estimator $\widehat{\beta}_{\mathrm{II}}$ has the oracle property.

Next, we establish asymptotic properties for the estimator $(\widehat{\beta}_{\mathrm{M}}^{\mathrm{T}}, \widehat{\beta}_{\mathrm{I}}^{\mathrm{T}}, \widehat{\phi}^{\mathrm{T}})^{\mathrm{T}}$ in the following theorem; the proof of the results is presented in Appendix C.3.

**Theorem 4.2** *Let $(\beta_{\mathrm{M}(0)}^{\mathrm{T}}, \beta_{\mathrm{I}(0)}^{\mathrm{T}}, \phi_0^{\mathrm{T}})^{\mathrm{T}}$ denote the true value of the parameters $(\beta_{\mathrm{M}}^{\mathrm{T}}, \beta_{\mathrm{I}}^{\mathrm{T}}, \phi^{\mathrm{T}})^{\mathrm{T}}$. Under regularity conditions, we have the following results:*

(i) $(\widehat{\beta}_{\mathrm{M}}^{\mathrm{T}}, \widehat{\beta}_{\mathrm{I}}^{\mathrm{T}}, \widehat{\phi}^{\mathrm{T}})^{\mathrm{T}}$ *is a consistent estimator of* $(\beta_{\mathrm{M}(0)}^{\mathrm{T}}, \beta_{\mathrm{I}(0)}^{\mathrm{T}}, \phi_0^{\mathrm{T}})^{\mathrm{T}}$.

(ii) $\sqrt{n}\{(\widehat{\beta}_{\mathrm{M}}^{\mathrm{T}}, \widehat{\beta}_{\mathrm{I}}^{\mathrm{T}}, \widehat{\phi}^{\mathrm{T}})^{\mathrm{T}} - (\beta_{\mathrm{M}(0)}^{\mathrm{T}}, \beta_{\mathrm{I}(0)}^{\mathrm{T}}, \phi_0^{\mathrm{T}})^{\mathrm{T}}\}$ *has the asymptotic normal distribution with mean zero and covariance matrix*

$$\Gamma_0^{-1}\Sigma_0\Gamma_0^{-1\mathrm{T}}, \tag{4.9}$$

*where* $\Gamma_0 = \left\{E\left(\frac{\partial U_i(\beta_{\mathrm{M}},\beta_{\mathrm{I}},\phi)}{\partial \beta_{\mathrm{M}}^{\mathrm{T}}}\right) \quad E\left(\frac{\partial U_i(\beta_{\mathrm{M}},\beta_{\mathrm{I}},\phi)}{\partial \beta_{\mathrm{I}}^{\mathrm{T}}}\right) \quad E\left(\frac{\partial U_i(\beta_{\mathrm{M}},\beta_{\mathrm{I}},\phi)}{\partial \phi^{\mathrm{T}}}\right)\right\}\Bigg|_{\substack{\beta_{\mathrm{M}}=\beta_{\mathrm{M}(0)}\\ \beta_{\mathrm{I}}=\beta_{\mathrm{I}(0)}\\ \phi=\phi_0}}$ *and*

$\Sigma_0 = E\{U_i(\beta_{\mathrm{M}(0)}, \beta_{\mathrm{I}(0)}, \phi_0)U_i^{\mathrm{T}}(\beta_{\mathrm{M}(0)}, \beta_{\mathrm{I}(0)}, \phi_0)\}$.

## 4.3 Generalized Network Structured Model with Measurement Error and Misclassification

Suppose that the response variables $Y_{i1}$ and $Y_{i2}$ are subject to mismeasurement and their precise measurements are not observed for every subject $i = 1, \ldots, n$, but instead, surrogate measurements $Y_{i1}^*$ and $Y_{i2}^*$ are observed, respectively, for $Y_{i1}$ and $Y_{i2}$. To describe the mismeasurement processes, we consider the same factorization described in (3.6) and the assumptions described in (3.7).

### 4.3.1 Measurement Error and Misclassification Models

Let $\pi_{i0} = P(Y_{i2}^* = 1|Y_{i2} = 0, Z_i)$ and $\pi_{i1} = P(Y_{i2}^* = 0|Y_{i2} = 1, Z_i)$ be the misclassification probabilities that may depend on the covariates. We consider the same misclassification models (3.8) as described in Section 3.2.1. For the continuous response $Y_{i1}$, we consider a regression model which facilitates possible dependence of $Y_{i1}^*$ on $\{Y_{i1}, Y_{i2}, Z_i\}$, as given by (3.9). Let $\eta = (\gamma^T, \alpha^T)^T$ denote the vector of parameters associated with (3.8) and (3.9).

### 4.3.2 Estimation Procedures with a Given Nuisance Parameter $\eta$

The presence of mismeasurement in $Y_i$ does not affect the first step of identifying the network structure in $X_i$ described in Section 4.2.2. However, if no action is taken to address measurement error and misclassification in the responses, simply replacing $Y_{ij}$ with $Y_{ij}^*$ in the estimating functions (4.2) and (4.3) would yield biased estimating functions, and hence, possibly resulting in inconsistent estimators.

To account for the mismeasurement effects, we construct valid estimating functions, say $U_i^*(\beta_M, \beta_I, \phi)$ expressed in terms of the observed measurements $Y_{i1}^*$ and $Y_{i2}^*$ together with the covariates and the model parameters, such that

$$E\{U_i^*(\beta_M, \beta_I, \phi)\} = 0.$$

To this end, we first define that $\Delta_{i0} = \frac{\pi_{i0}-\pi_{i0}^2}{(1-\pi_{i1}-\pi_{i0})^2}$, $\Delta_{i1} = \frac{\pi_{i1}-\pi_{i1}^2}{(1-\pi_{i1}-\pi_{i0})^2}$, and $\Delta_i = \frac{\Delta_{i0}^{1-Y_{i2}^*}\Delta_{i1}^{Y_{i2}^*}-\Delta_{i0}\pi_{i1}-\Delta_{i1}\pi_{i0}}{1-\pi_{i1}-\pi_{i0}}$, where $\pi_{i0}$ and $\pi_{i1}$ are the misclassification rates postulated by (3.8). Let $Y_{i2}^{**} = \frac{Y_{i2}^*-\pi_{i0}}{1-\pi_{i0}-\pi_{i1}}$, $Y_{i1}^{**} = \frac{Y_{i1}^*-\gamma_0-\gamma_2 Y_{i2}^{**}-\gamma_3^T Z_i}{\gamma_1}$, $Y_{i11}^{**} = Y_{i1}^{**2} - \frac{\sigma_e^2}{\gamma_1^2} - \frac{\gamma_2^2}{\gamma_1^2}\Delta_i$, and $Y_{i12}^{**} = Y_{i1}^{**}Y_{i2}^{**} + \frac{\gamma_2}{\gamma_1}\Delta_i$. In Section 3.2.2, it has been shown that

$$E(Y_{ik}^{**}|Y_{i1}, Y_{i2}) = Y_{ik} \quad \text{and} \quad E(Y_{i1k}^{**}|Y_{i1}, Y_{i2}, X_i) = Y_{i1}Y_{ik} \qquad \text{for } k = 1, 2. \tag{4.10}$$

Let $U_i^*(\beta_M, \beta_I, \phi)$ be $U_i(\beta_M, \beta_I, \phi)$ in (4.8) with $Y_{i1}$, $Y_{i2}$, $Y_{i1}^2$, $Y_{i1}Y_{i2}$ replaced by $Y_{i1}^{**}$, $Y_{i2}^{**}$, $Y_{i11}^{**}$ and $Y_{i12}^{**}$, respectively. Then by (4.10),

$$E[U_i^*(\beta_M, \beta_I, \phi)|Y_{i1}, Y_{i2}, X_i] = U_i(\beta_M, \beta_I, \phi),$$

and thus, by the unbiasedness of $U_i(\beta_M, \beta_I, \phi)$, $U_i^*(\beta_M, \beta_I, \phi)$ is an unbiased estimating function. If the parameter $\eta$ for the misclassification and measurement error models is known or estimated from an additional study, then solving

$$\sum_{i=1}^{n} U_i^*(\beta_{\mathrm{M}}, \beta_{\mathrm{I}}, \phi) = 0 \tag{4.11}$$

for $\beta_{\mathrm{M}}$, $\beta_{\mathrm{I}}$ and $\phi$ gives an estimator, say $(\widehat{\beta}_{\mathrm{M}}^{\mathrm{T}}, \widehat{\beta}_{\mathrm{I}}^{\mathrm{T}}, \widehat{\phi}^{\mathrm{T}})^{\mathrm{T}}$, of $(\beta_{\mathrm{M}}^{\mathrm{T}}, \beta_{\mathrm{I}}^{\mathrm{T}}, \phi^{\mathrm{T}})^{\mathrm{T}}$.

**Theorem 4.3** *Under regularity conditions including those of Newey and McFadden (1994), Yi (2017, Section 1.3.2) and Meinshausen and Bühlmann (2006), the estimator* $(\widehat{\beta}_{\mathrm{M}}^{\mathrm{T}}, \widehat{\beta}_{\mathrm{I}}^{\mathrm{T}}, \widehat{\phi}^{\mathrm{T}})^{\mathrm{T}}$ *is consistent, and* $\sqrt{n}\{(\widehat{\beta}_{\mathrm{M}}^{\mathrm{T}}, \widehat{\beta}_{\mathrm{I}}^{\mathrm{T}}, \widehat{\phi}^{\mathrm{T}})^{\mathrm{T}} - (\beta_{\mathrm{M(0)}}^{\mathrm{T}}, \beta_{\mathrm{I(0)}}^{\mathrm{T}}, \phi_0^{\mathrm{T}})^{\mathrm{T}}\}$ *has an asymptotic normal distribution with mean zero and covariance matrix*

$$\Gamma^{-1}\Sigma\Gamma^{-1\mathrm{T}}, \tag{4.12}$$

*where* $\Gamma = \left\{ E\left(\frac{\partial U_i^*(\beta_{\mathrm{M}},\beta_{\mathrm{I}},\phi,\eta_0)}{\partial \beta_{\mathrm{M}}^{\mathrm{T}}}\right) \quad E\left(\frac{\partial U_i^*(\beta_{\mathrm{M}},\beta_{\mathrm{I}},\phi,\eta_0)}{\partial \beta_{\mathrm{I}}^{\mathrm{T}}}\right) \quad E\left(\frac{\partial U_i^*(\beta_{\mathrm{M}},\beta_{\mathrm{I}},\phi,\eta_0)}{\partial \phi^{\mathrm{T}}}\right) \right\}\Big|_{\substack{\beta_{\mathrm{M}}=\beta_{\mathrm{M(0)}} \\ \beta_{\mathrm{I}}=\beta_{\mathrm{I(0)}} \\ \phi=\phi_0}}$ *and* $\Sigma = E\{U_i^*(\beta_{\mathrm{M(0)}}, \beta_{\mathrm{I(0)}}, \phi_0, \eta_0) U_i^{*\mathrm{T}}(\beta_{\mathrm{M(0)}}, \beta_{\mathrm{I(0)}}, \phi_0, \eta_0)\}$.

## 4.4 Estimation Procedures with Validation Data

### 4.4.1 External Validation

To incorporate estimation of $\eta$ in the estimation of $(\beta_{\mathrm{M}}^{\mathrm{T}}, \beta_{\mathrm{I}}^{\mathrm{T}}, \phi^{\mathrm{T}})^{\mathrm{T}}$, consider the likelihood function contributed from subject $i$ in the validation sample:

$$L_i(y_{i1}^*, y_{i2}^* | y_{i1}, y_{i2}, x_i; \eta) = f(y_{i1}^* | y_{i1}, y_{i2}, z_i) f(y_{i2}^* | y_{i1}, y_{i2}, z_i),$$

where the index $i \in \mathcal{V}$, $f(y_{i1}^*|y_{i1}, y_{i2}, z_i)$ is determined by (3.9) with the form $\frac{1}{\sqrt{2\pi}\sigma_e} \exp\left\{-\frac{(y_{i1}^* - \gamma_0 - \gamma_1 y_{i1} - \gamma_2 y_{i2} - \gamma_3 z_i)^2}{2\sigma_e^2}\right\}$; and determined by (3.8), $f(y_{i2}^*|y_{i1}, y_{i2}, z_i)$ equals

$$\left\{\frac{\exp(\alpha_{00} + \alpha_{z0}^{\mathrm{T}} z_i)}{1 + \exp(\alpha_{00} + \alpha_{z0}^{\mathrm{T}} z_i)}\right\}^{(1-y_{i2})y_{i2}^*} \left\{\frac{1}{1 + \exp(\alpha_{00} + \alpha_{z0}^{\mathrm{T}} z_i)}\right\}^{(1-y_{i2})(1-y_{i2}^*)}$$
$$\times \left\{\frac{\exp(\alpha_{01} + \alpha_{z1}^{\mathrm{T}} z_i)}{1 + \exp(\alpha_{01} + \alpha_{z1}^{\mathrm{T}} z_i)}\right\}^{y_{i2}(1-y_{i2}^*)} \left\{\frac{1}{1 + \exp(\alpha_{01} + \alpha_{z1}^{\mathrm{T}} z_i)}\right\}^{y_{i2} y_{i2}^*}.$$

Let

$$S_i(\eta) = \partial \log L_i(y_{i1}^*, y_{i2}^* | y_{i1}, y_{i2}, x_i; \eta)/\partial\eta \qquad \text{for } i \in \mathcal{V}, \tag{4.13}$$

and construct the estimating function

$$U^{\text{(EV)}}(\beta_{\text{M}}, \beta_{\text{I}}, \phi, \eta) = \sum_{i\in\mathcal{M}} \begin{pmatrix} U_i^*(\beta_{\text{M}}, \beta_{\text{I}}, \phi, \eta) \\ 0 \end{pmatrix} + \sum_{i\in\mathcal{V}} \begin{pmatrix} 0 \\ S_i(\eta) \end{pmatrix}, \tag{4.14}$$

where $S_i(\eta)$ is the score function determined by (4.13), and $U_i^*(\beta_{\text{M}}, \beta_{\text{I}}, \phi, \eta)$ is the estimating equation in (4.11) with the dependence on $\eta$ explicitly spelled out. Then solving

$$U^{\text{(EV)}}(\beta_{\text{M}}, \beta_{\text{I}}, \phi, \eta) = 0$$

for $\beta_{\text{M}}$, $\beta_{\text{I}}$, $\phi$ and $\eta$ gives an estimator of $\beta_{\text{M}}, \beta_{\text{I}}, \phi$ and $\eta$, denoted as $\widehat{\beta}_{\text{M}}^{\text{(EV)}}$, $\widehat{\beta}_{\text{I}}^{\text{(EV)}}$, $\widehat{\phi}^{\text{(EV)}}$, and $\widehat{\eta}^{\text{(EV)}}$, respectively.

Since $S_i(\eta)$ does not depend on $(\beta_{\text{M}}^{\text{T}}, \beta_{\text{I}}^{\text{T}}, \phi^{\text{T}})^{\text{T}}$, solving (4.14) is equivalent to a two-step procedure. First obtain $\hat{\eta}^{\text{(EV)}}$ by solving $\sum_{i\in\mathcal{V}} S_i(\eta) = 0$. Then solve the equation

$$\sum_{i\in\mathcal{M}} U_i^*(\beta_{\text{M}}, \beta_{\text{I}}, \phi, \hat{\eta}^{\text{(EV)}}) = 0$$

for $\beta_{\text{M}}$, $\beta_{\text{I}}$ and $\phi$ to obtain estimators of $\beta_{\text{M}}, \beta_{\text{I}}$ and $\phi$, denoted as $\widehat{\beta}_{\text{M}}^{\text{(EV)}}$, $\widehat{\beta}_{\text{I}}^{\text{(EV)}}$ and $\widehat{\phi}^{\text{(EV)}}$, respectively.

**Theorem 4.4** *Assume that regularity conditions in Theorem 4.3 hold and that the ratio $m/n$ approaches a positive constant $\rho$ as $n \to \infty$, we have the following results:*

(i) $\sqrt{n}\left\{(\widehat{\beta}_{\text{M}}^{\text{(EV)T}}, \widehat{\beta}_{\text{I}}^{\text{(EV)T}}, \widehat{\phi}^{\text{(EV)T}}, \widehat{\eta}^{\text{(EV)T}})^{\text{T}} - (\beta_{\text{M(0)}}^{\text{T}}, \beta_{\text{I(0)}}^{\text{T}}, \phi_0^{\text{T}}, \eta_0^{\text{T}})^{\text{T}}\right\}$ *has an asymptotic normal distribution with mean zero and covariance matrix* $\frac{1}{1+\rho}\Gamma_{\text{(EV)}}^{-1}\Sigma_{\text{(EV)}}(\Gamma_{\text{(EV)}}^{-1})^{\text{T}}$, *where*

$$\begin{aligned}
\Gamma_{\text{(EV)}} &= \tfrac{1}{1+\rho}\begin{bmatrix} E\left(\frac{\partial U_i^*}{\partial\beta_{\text{M}}^{\text{T}}}\right) & E\left(\frac{\partial U_i^*}{\partial\beta_{\text{I}}^{\text{T}}}\right) & E\left(\frac{\partial U_i^*}{\partial\phi^{\text{T}}}\right) & E\left(\frac{\partial U_i^*}{\partial\eta^{\text{T}}}\right) \\ 0 & 0 & 0 & 0 \end{bmatrix} + \tfrac{\rho}{1+\rho}\begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & E\left(\frac{\partial S_i}{\partial\eta^{\text{T}}}\right) \end{bmatrix}; \\
\Sigma_{\text{(EV)}} &= \tfrac{1}{1+\rho}\begin{bmatrix} E\left(U_i^* U_i^{*\text{T}}\right) & 0 \\ 0 & 0 \end{bmatrix} + \tfrac{\rho}{1+\rho}\begin{bmatrix} 0 & 0 \\ 0 & E\left(S_i S_i^{\text{T}}\right) \end{bmatrix}.
\end{aligned} \tag{4.15}$$

(ii) $\sqrt{n}\left\{(\widehat{\beta}_{\text{M}}^{\text{(EV)T}}, \widehat{\beta}_{\text{I}}^{\text{(EV)T}}, \widehat{\phi}^{\text{(EV)T}})^{\text{T}} - (\beta_{\text{M(0)}}^{\text{T}}, \beta_{\text{I(0)}}^{\text{T}}, \phi_0^{\text{T}})^{\text{T}}\right\}$ *has an asymptotic normal distribution with mean zero and covariance matrix* $(1+\rho)\Gamma_{\text{(EV)}\beta}^{-1}\Sigma_{\text{(EV)}\beta}\Gamma_{\text{(EV)}\beta}^{-1\text{T}}$, *where*

$$\begin{aligned}
\Gamma_{\text{(EV)}\beta} &= \left\{E\left(\frac{\partial U_i^*(\beta_{\text{M}}, \beta_{\text{I}}, \phi, \widehat{\eta}^{\text{(EV)}})}{\partial\beta_{\text{M}}^{\text{T}}}\right) \quad E\left(\frac{\partial U_i^*(\beta_{\text{M}}, \beta_{\text{I}}, \phi, \widehat{\eta}^{\text{(EV)}})}{\partial\beta_{\text{I}}^{\text{T}}}\right) \quad E\left(\frac{\partial U_i^*(\beta_{\text{M}}, \beta_{\text{I}}, \phi, \widehat{\eta}^{\text{(EV)}})}{\partial\phi^{\text{T}}}\right)\right\}\Bigg|_{\substack{\beta_{\text{M}}=\beta_{\text{M(0)}} \\ \beta_{\text{I}}=\beta_{\text{I(0)}} \\ \phi=\phi_0}}; \\
\Sigma_{\text{(EV)}\beta} &= E\{U_i^*(\beta_{\text{M(0)}}, \beta_{\text{I(0)}}, \phi_0, \widehat{\eta}^{\text{(EV)}})U_i^{*\text{T}}(\beta_{\text{M(0)}}, \beta_{\text{I(0)}}, \phi_0, \widehat{\eta}^{\text{(EV)}})\}.
\end{aligned} \tag{4.16}$$

The proof of Theorem 4.4(i) is presented in Appendix C.4, and Theorem 4.4(ii) can be readily derived from Theorem 4.4(i) by matrix calculation.

### 4.4.2 Internal Validation

With internal validation data, we consider the estimating function

$$U^{\text{(IV)}}(\beta_{\text{M}}, \beta_{\text{I}}, \phi, \eta) = \sum_{i \in \mathcal{M} \setminus \mathcal{V}} \begin{pmatrix} U_i^*(\beta_{\text{M}}, \beta_{\text{I}}, \phi, \eta) \\ 0 \end{pmatrix} + \sum_{i \in \mathcal{V}} \begin{pmatrix} U_i(\beta_{\text{M}}, \beta_{\text{I}}, \phi) \\ S_i(\eta) \end{pmatrix}, \tag{4.17}$$

where for $i \in \mathcal{M} \setminus \mathcal{V}$, $U_i^*(\beta_{\text{M}}, \beta_{\text{I}}, \phi, \eta)$ is the estimating equation in (4.11) with the dependence on $\eta$ explicitly spelled out and for $i \in \mathcal{V}$, $U_i(\beta_{\text{M}}, \beta_{\text{I}}, \phi)$ is the estimating function in (4.8), and $S_i(\eta)$ is determined by (4.13). Then solving equation

$$U^{\text{(IV)}}(\beta_{\text{M}}, \beta_{\text{I}}, \phi, \eta) = 0 \tag{4.18}$$

for $\beta_{\text{M}}$, $\beta_{\text{I}}$, $\phi$ and $\eta$ yields estimators for them, respectively denoted as $\widehat{\beta}_{\text{M}}^{\text{(IV)}}$, $\widehat{\beta}_{\text{I}}^{\text{(IV)}}$, $\widehat{\phi}^{\text{(IV)}}$, and $\widehat{\eta}^{\text{(IV)}}$.

Similar to that in Section 4.4.1, solving (4.18) is equivalent to a two-step procedure. First obtain $\hat{\eta}^{\text{(IV)}}$ by solving $\sum_{i \in \mathcal{V}} S_i(\eta) = 0$. Then solve the equation

$$\sum_{i \in \mathcal{M} \setminus \mathcal{V}} U_i^*(\beta_{\text{M}}, \beta_{\text{I}}, \phi, \hat{\eta}^{\text{(IV)}}) + \sum_{i \in \mathcal{V}} U_i(\beta_{\text{M}}, \beta_{\text{I}}, \phi) = 0$$

for $\beta_{\text{M}}$, $\beta_{\text{I}}$ and $\phi$ to obtain estimators of $\beta_{\text{M}}$, $\beta_{\text{I}}$ and $\phi$, denoted as $\widehat{\beta}_{\text{M}}^{\text{(IV)}}$, $\widehat{\beta}_{\text{I}}^{\text{(IV)}}$, $\widehat{\phi}^{\text{(IV)}}$, respectively.

**Theorem 4.5** *Assume that regularity conditions in Theorem 4.3 hold and that the ratio $m/n$ approaches a positive constant $\rho$ as $n \to \infty$, we have the following results:*

(i) $\sqrt{n}\left\{(\widehat{\beta}_{\text{M}}^{\text{(IV)}\text{T}}, \widehat{\beta}_{\text{I}}^{\text{(IV)}\text{T}}, \widehat{\phi}^{\text{(IV)}\text{T}}, \widehat{\eta}^{\text{(IV)}\text{T}})^{\text{T}} - (\beta_{\text{M}(0)}^{\text{T}}, \beta_{\text{I}(0)}^{\text{T}}, \phi_0^{\text{T}}, \eta_0^{\text{T}})^{\text{T}}\right\}$ *has an asymptotic normal distribution with mean zero and covariance matrix* $\Gamma_{\text{(IV)}}^{-1}\Sigma_{\text{(IV)}}(\Gamma_{\text{(IV)}}^{-1})^{\text{T}}$, *where*

$$\begin{aligned} \Gamma_{\text{(IV)}} &= (1-\rho)\begin{bmatrix} E\left(\frac{\partial U_i^*}{\partial \beta_{\text{M}}^{\text{T}}}\right) & E\left(\frac{\partial U_i^*}{\partial \beta_{\text{I}}^{\text{T}}}\right) & E\left(\frac{\partial U_i^*}{\partial \phi^{\text{T}}}\right) & E\left(\frac{\partial U_i^*}{\partial \eta^{\text{T}}}\right) \\ 0 & 0 & 0 & 0 \end{bmatrix} + \rho \begin{bmatrix} E\left(\frac{\partial U_i}{\partial \beta_{\text{M}}^{\text{T}}}\right) & E\left(\frac{\partial U_i}{\partial \theta_{\text{I}}^{\text{T}}}\right) & E\left(\frac{\partial U_i}{\partial \phi^{\text{T}}}\right) & 0 \\ 0 & 0 & 0 & E\left(\frac{\partial S_i}{\partial \eta^{\text{T}}}\right) \end{bmatrix}; \\ \Sigma_{\text{(IV)}} &= (1-\rho)\begin{bmatrix} E\left(U_i^* U_i^{*\text{T}}\right) & 0 \\ 0 & 0 \end{bmatrix} + \rho \begin{bmatrix} E\left(U_i U_i^{\text{T}}\right) & E\left(U_i S_i^{\text{T}}\right) \\ E\left(S_i U_i^{\text{T}}\right) & E\left(S_i S_i^{\text{T}}\right) \end{bmatrix}. \end{aligned} \tag{4.19}$$

(ii) $\sqrt{n}\left\{(\widehat{\beta}_{\mathrm{M}}^{(\mathrm{IV})\mathrm{T}}, \widehat{\beta}_{\mathrm{I}}^{(\mathrm{IV})\mathrm{T}}, \widehat{\phi}^{(\mathrm{IV})\mathrm{T}})^{\mathrm{T}} - (\beta_{\mathrm{M}(0)}^{\mathrm{T}}, \beta_{\mathrm{I}(0)}^{\mathrm{T}}, \phi_0^{\mathrm{T}})^{\mathrm{T}}\right\}$ *has an asymptotic normal distribution with mean zero and covariance matrix* $\Gamma_{(\mathrm{IV})\beta}^{-1}\Sigma_{(\mathrm{IV})\beta}(\Gamma_{(\mathrm{IV})\beta}^{-1})^{\mathrm{T}}$*, where*

$$\Gamma_{(\mathrm{IV})\beta} = (1-\rho)\left[E\left(\frac{\partial U_i^*}{\partial \beta_{\mathrm{M}}^{\mathrm{T}}}\right) \quad E\left(\frac{\partial U_i^*}{\partial \beta_{\mathrm{I}}^{\mathrm{T}}}\right) \quad E\left(\frac{\partial U_i^*}{\partial \phi^{\mathrm{T}}}\right)\right] + \rho\left[E\left(\frac{\partial U_i}{\partial \beta_{\mathrm{M}}^{\mathrm{T}}}\right) \quad E\left(\frac{\partial U_i}{\partial \theta_{\mathrm{I}}^{\mathrm{T}}}\right) \quad E\left(\frac{\partial U_i}{\partial \phi^{\mathrm{T}}}\right)\right], \tag{4.20}$$

$$\Sigma_{(\mathrm{IV})\beta} = (1-\rho)E\left(U_i^* U_i^{*\mathrm{T}}\right) + \rho E\left(U_i U_i^{\mathrm{T}}\right) - \rho E\left(U_i S_i^{\mathrm{T}}\right)\left\{E\left(S_i S_i^{\mathrm{T}}\right)\right\}^{-1} E\left(S_i U_i^{\mathrm{T}}\right).$$

The proof of Theorem 4.5(i) is presented in Appendix C.5, and Theorem 4.5(ii) can be readily derived from Theorem 4.5(i) by matrix calculation. We comment that Theorem 4.4(ii) and Theorem 4.5(ii) have appealing implications in that they extend the estimator in Theorem 4.3 to more realistic scenarios with unknown parameter $\eta$ associated with the measurement error and misclassification models to be estimated from additional data sources. The estimators in Theorem 4.4(ii), Theorem 4.5(ii) and Theorem 4.3 are all consistent estimators of $(\beta_{\mathrm{M}}^{\mathrm{T}}, \beta_{\mathrm{I}}^{\mathrm{T}}, \phi^{\mathrm{T}})^{\mathrm{T}}$, but they differ in the efficiency because of the nuisance parameter $\eta$. The estimator $(\widehat{\beta}_{\mathrm{M}}^{(\mathrm{EV})\mathrm{T}}, \widehat{\beta}_{\mathrm{I}}^{(\mathrm{EV})\mathrm{T}}, \widehat{\phi}^{(\mathrm{EV})\mathrm{T}})^{\mathrm{T}}$ in Theorem 4.4(ii) is less efficient than the estimator $(\widehat{\beta}_{\mathrm{M}}^{\mathrm{T}}, \widehat{\beta}_{\mathrm{I}}^{\mathrm{T}}, \widehat{\phi}^{\mathrm{T}})^{\mathrm{T}}$ in Theorem 4.3 if $\eta_0$ is set as $\widehat{\eta}^{(\mathrm{E})}$ associated with the asymptotic covariance matrix in Theorem 4.4(ii), because the asymptotic covariance matrix for the former estimator is $(1+\rho)$ times of that of the latter estimator. On the contrary, the estimator in Theorem 4.5(ii) is more efficient than that in Theorem 4.3, provided certain conditions, as shown in

**Theorem 4.6** *Let* $\Delta = E(U_i S_i^{\mathrm{T}})\{E(S_i S_i^{\mathrm{T}})\}^{-1}E(S_i U_i^{\mathrm{T}})$. *Consider* $\Sigma_0$, $\Gamma_0$, $\Sigma$ *and* $\Gamma$ *that are defined in Theorems 4.2 and 4.3, respectively. Assume the regularity conditions of Theorem 4.5. If*

$$\Gamma_0^{-1}\Sigma_0\Gamma_0^{-1\mathrm{T}} \le \Gamma^{-1}\Sigma\Gamma^{-1\mathrm{T}}, \tag{4.21}$$

*and*

$$\Gamma_0^{-1}\Sigma_0\Gamma_0^{-1\mathrm{T}} + \Gamma_0^{-1}\Sigma\Gamma_0^{-1\mathrm{T}} \le \Gamma^{-1}\Sigma\Gamma_0^{-1\mathrm{T}} + \Gamma_0^{-1}\Sigma\Gamma^{-1\mathrm{T}} + \Gamma_0^{-1}\Delta\Gamma_0^{-1\mathrm{T}}, \tag{4.22}$$

*then we have*

$$Avar\{(\widehat{\beta}_{\mathrm{M}}^{(\mathrm{IV})\mathrm{T}}, \widehat{\beta}_{\mathrm{I}}^{(\mathrm{IV})\mathrm{T}}, \widehat{\phi}^{(\mathrm{IV})\mathrm{T}})^{\mathrm{T}}\} \le Avar\{(\widehat{\beta}_{\mathrm{M}}^{\mathrm{T}}, \widehat{\beta}_{\mathrm{I}}^{\mathrm{T}}, \widehat{\phi}^{\mathrm{T}})^{\mathrm{T}}\}, \tag{4.23}$$

*for every* $\rho \in (0, 1]$, *where* $Avar(\cdot)$ *represents the asymptotic covariance matrix of an estimator, and the inequality* $\le$ *is the Loewner order.*

The proof of Theorem 4.6 is outlined in Appendix C.6. This theorem says that under some conditions, the estimators in Section 4.4.2, with nuisance parameter $\eta$ estimated from an internal validation subsample, are more efficient than the estimators in Section 4.3.2, with $\eta$ being given. Such a result appears somewhat counterintuitive as one may expect estimation of $\eta$ would induce additional variability for estimators of $\beta_{\text{M}}$, $\beta_{\text{I}}$ and $\phi$. However, this phenomenon arises commonly in the context of using estimating functions (instead of the likelihood method) for estimation, as discussed by Newey and McFadden (1994, Chapter 6). The condition (4.21) compares the asymptotic covariance matrix for two estimators derived from different scenarios. This condition requires the estimator in Theorem 4.2, derived from the true response measurements, to be more efficient than the estimator in Theorem 4.3, obtained from surrogates measurements, which is often true when $Y_i$ is less variable than $Y_i^*$. To understand condition (4.22), we look at the two terms at the left-hand side first where the first term represents the asymptotic covariance matrix in Theorem 4.2(ii). For the first term of the left-hand side of (4.22), we replace the left $\Gamma_0$ with $\Gamma$ and $\Sigma_0$ with $\Sigma$; for the second term of the left-hand side of (4.22), we replace the right $\Gamma_0$ with $\Gamma$, then (4.22) requires that the difference of such changes cannot exceed $\Gamma_0 \Delta \Gamma_0^{\text{T}}$, a non-negative definite matrix which involves the variability of $S_i$ (i,e, $E(S_i S_i^{\text{T}})$), the covariance between $S_i$ and $U_i$ (i.e., $E(S_i U_i^{\text{T}})$) and the sensitivity of $U_i$ (i.e., $\Gamma_0$). The efficiency gain stated in Theorem 4.6 holds for any value $\rho \in (0, 1]$, meaning that using any reasonably large internal validation subsample always increase efficiency relative to the case with $\eta$ being given, provided certain conditions discussed earlier.

## 4.5 Simulation Studies

### 4.5.1 Simulation 1: Comparison of the GNSM with Ordinary LASSO

In this subsection, we conduct simulation studies to evaluate the performance of the proposed method for the variable selection (Section 4.2.2) and the parameter estimation (Section 4.2.2), where no mismeasurement is present.

To evaluate the performance of the methods, we consider different graphs, displayed in Figure 4.1, with different dependent structures of the covariates, which are characterized by varying degrees of nodes. Here the *degree* of a node is defined as the number of edges connected to this node. In the *hub* graph, two nodes have a higher degree than the other four nodes. The *scale-free* graph is generated by the Barabási-Albert algorithm (Barabási and Albert, 1999), where we start with an initial graph with only two connected nodes and then randomly connect a new node to only one existing node successively. In the *block*

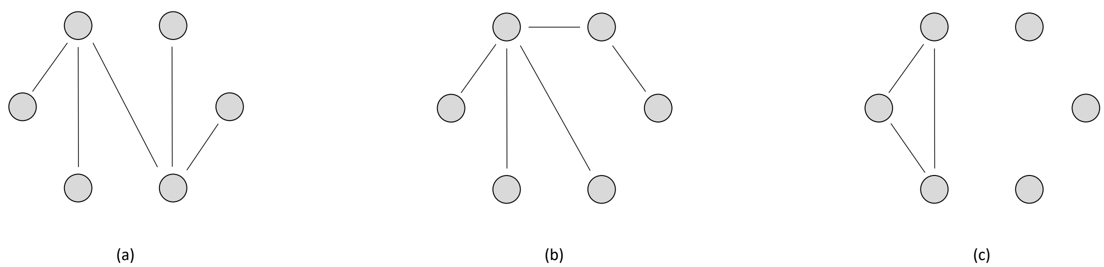


Figure 4.1: *Illustration of different graphs. (a): hub graph. (b): scale-free graph. (c): block graph*

graph, nodes are classified into several blocks in which the degrees of nodes within each block are the same.

The covariates $X_i$ are generated from a multivariate normal distribution with mean zero and covariance matrix $\Sigma = \Theta^{-1}$, where the precision matrix $\Theta$ is, respectively, given by

$$\Theta_1 = \begin{bmatrix} 1 & \theta_{12} & \theta_{13} & \theta_{14} & 0 & 0 \\ \theta_{21} & 1 & 0 & 0 & 0 & 0 \\ \theta_{31} & 0 & 1 & 0 & 0 & 0 \\ \theta_{41} & 0 & 0 & 1 & \theta_{45} & \theta_{46} \\ 0 & 0 & 0 & \theta_{54} & 1 & 0 \\ 0 & 0 & 0 & \theta_{64} & 0 & 1 \end{bmatrix}, \Theta_2 = \begin{bmatrix} 1 & \theta_{12} & \theta_{13} & \theta_{14} & 0 & \theta_{16} \\ \theta_{21} & 1 & 0 & 0 & 0 & 0 \\ \theta_{31} & 0 & 1 & 0 & 0 & 0 \\ \theta_{41} & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & \theta_{56} \\ \theta_{61} & 0 & 0 & 0 & \theta_{65} & 1 \end{bmatrix}, \text{ and } \Theta_3 = \begin{bmatrix} 1 & \theta_{12} & \theta_{13} & 0 & 0 & 0 \\ \theta_{21} & 1 & \theta_{23} & 0 & 0 & 0 \\ \theta_{31} & \theta_{32} & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix}$$

for the hub, scale-free, and block graphs. Here the $\theta_{ij}$ take a value either 0.2 or $-0.2$ for the connected edges, and 0 otherwise.

The responses $Y_i$ are generated from the joint distribution

$$f(y_{i1}, y_{i2}) = \left[ \Phi \left\{ \frac{g_2(\mu_{i2}) + \rho_c \left( \frac{y_{i1} - g_1(\mu_{i1})}{\sigma} \right)}{\sqrt{1-\rho_c^2}} \right\} \right]^{y_{i2}} \left[ 1 - \Phi \left\{ \frac{g_2(\mu_{i2}) + \rho_c \left( \frac{y_{i1} - g_1(\mu_{i1})}{\sigma} \right)}{\sqrt{1-\rho_c^2}} \right\} \right]^{1-y_{i2}}$$
$$\times \frac{1}{\sqrt{2\pi}\sigma} \exp\left\{ -\frac{(y_{i1} - g_1(\mu_{i1}))^2}{2\sigma^2} \right\},$$

where $\Phi(t)$ is the cumulative distribution function for the standard normal distribution, $g_1(\mu_{i1})$ and $g_2(\mu_{i2})$ are specified as in model (4.5) with $E$ indicated by each graph in Figure 4.1, $\rho_c$ determines the correlation among $Y_{i1}$ and $Y_{i2}$, and we set $g_1(t) = t$ and

$g_2(t) = \Phi^{-1}(t)$. The joint distribution was discussed by de Leon and Wu (2011). We set $\beta_{\text{II}}$ to be a zero vector of dimension 22 or 18. Let $\beta_{\text{M}}$ be of dimension 12, and let $\beta_{\text{I}}$ be of dimension 6 or 10; the values of $\beta_{\text{M}}$ and $\beta_{\text{I}}$ are recorded in Table 4.1 which fall in the intervals $[-0.7, -0.1] \cup [0.1, 0.7]$. The sample size $n$ is set as 50, 200 or 1000. The simulations are run for 1000 times for each parameter configuration.

To show the performance of the methods, we evaluate the results for the network identification described in Stage 1 of Section 4.2.2 and the results of parameter estimation described in Stage 2 of Section 4.2.2 using different measures. For the procedure in Stage 1, we define two measures of variable selection, the true positive rate

$$\text{TPR}(\lambda) = \frac{\left|\{(s,t) : (s,t) \in E \text{ and } (s,t) \in \widehat{E}\}\right|}{\left|\{(s,t) : (s,t) \in E\}\right|}$$

and false negative rate

$$\text{FNR}(\lambda) = \frac{\left|\{(s,t) : (s,t) \notin E \text{ and } (s,t) \in \widehat{E}\}\right|}{\left|\{(s,t) : (s,t) \notin E\}\right|},$$

where $|\mathcal{A}|$ is the size of the set $\mathcal{A}$, and $\widehat{E}$ is the estimated edge set for a given tuning parameter $\lambda$. For the estimation procedure in Stage 2, we consider two measures,

$$\left\|\widehat{\beta} - \beta\right\|_1 = \sum_{j \in \mathcal{I}} \sum_{t=1}^{T} |\hat{\beta}_j^{(t)} - \beta_j^{(t)}| \quad \text{and} \quad \left\|\widehat{\beta} - \beta\right\|_2 = \sum_{j \in \mathcal{I}} \sum_{t=1}^{T} \sqrt{|\hat{\beta}_j^{(t)} - \beta_j^{(t)}|^2},$$

where $\hat{\beta}_j^{(t)}$ and $\beta_j^{(t)}$ represent, respectively, the $j$th component of $\widehat{\beta}$ and $\beta$ in the $t$th simulation, $T$ is the total number of simulations, and $\mathcal{I}$ is the index set of $\beta_{\text{M}}$ and $\beta_{\text{I}}$.

To see how the choice of the tuning parameter $\lambda$ may affect the results obtained from (4.6), for a given $\lambda$ in the interval $[0, 0.8]$, we plot $TPR(\lambda)$ against $FNR(\lambda)$ in Figure 4.2, where we report the results for the sample size $n$=50, 200 and 1000. Figure 4.2 shows that for a given $\lambda$, the performance of the method in Section 4.2.2 improves as the sample size increases.

In Table 4.2, we report the results obtained from the estimation procedure in Section 4.2.2, which clearly demonstrate the improved performance of the proposed GNSM method as the sample size increases regardless of the graph types.

### 4.5.2 Simulation 2: Augmented GNSM with Measurement Error and Misclassification in Responses

In this subsection, we evaluate the performance of the proposed estimators when the mixed responses are subject to both measurement error and misclassification.

The covariates and the true responses are generated in the same way as in Section 4.5.1. The surrogate measurement $Y_{i1}^*$ is generated from the measurement error model, $Y_{i1}^* = Y_{i1} + \gamma Y_{i2} + e_i$, where $\gamma$ is set as 0.5, $e_i$ follows a normal distribution with mean zero and variance $\sigma_e^2$ and is independent of $\{Y_{i1}, Y_{i2}\}$. We set $\sigma_e^2$ to be 0.2 or 0.7, reflecting different degrees of measurement error. The surrogate measurement $Y_{i2}^*$ is generated from the misclassification models (3.8), where $Z_i$ is generated from Uniform$(-2, 3)$, and the parameter $\alpha$ is set as $(-4, -1)^{\mathrm{T}}$, $(-3, 0)^{\mathrm{T}}$ and $(-3, 1)^{\mathrm{T}}$, respectively yielding the misclassification rate of 1%, 5% and 10%. The sample size $n$ is taken as 1000, and we take the generated data $\{(y_{i1}^*, y_{i2}^*, x_i) : i = 1, \ldots, n\}$ as the main study data.

To simulate validation data, we generate a validation sample of size 500 using the same method as for generating the main study data. For the internal validation data, we keep all the measurements $\{(y_{j1}^*, y_{j2}^*, y_{j1}, y_{j2}, x_j, z_j) : j = 1, \ldots, 500\}$ as validation data; and for external validation sample, we take $\{(y_{j1}^*, y_{j2}^*, y_{j1}, y_{j2}, z_j) : j = 1, \ldots, 500\}$ as validation data.

Simulation studies are run 1000 times for each parameter configuration. We compare the performance of the augmented GNSM (Section 4.3.2) with the naive GNSM (Section 4.2.2) where the effects of measurement error and misclassification are ignored. To display the results, we separately report the results for $g_1(\mu_{i1})$ and $g_2(\mu_{i2})$ which respectively describe the continuous and binary responses. Let $\mathcal{I}$ be the index set of $\beta_{\mathrm{M}}$ and $\beta_{\mathrm{I}}$ as in Section 4.5.1. Let $\widehat{\beta}^{(t)j}$ denote the estimate for the $j$th component of $(\beta_{\mathrm{M}}^{\mathrm{T}}, \beta_{\mathrm{I}}^{\mathrm{T}})^{\mathrm{T}}$ at the $t$th simulation. We report the average bias (denoted "avgBias") by calculating $\frac{1}{|\mathcal{I}|}\sum_{j\in\mathcal{I}} |\bar{\widehat{\beta}}^j - \beta^j|$, the average empirical standard error (denoted "avgSEE") by calculating $\frac{1}{|\mathcal{I}|}\sum_{j\in\mathcal{I}} \mathrm{esd}^j$, average model standard error (denoted "avgSEM") by calculating $\frac{1}{|\mathcal{I}|}\sum_{j\in\mathcal{I}} \mathrm{msd}^j$, and the average coverage rate (denoted "avgCR") by calculating $\frac{1}{|\mathcal{I}|}\sum_{j\in\mathcal{I}} \mathrm{CR}^j$, where $\bar{\widehat{\beta}}^j = \frac{1}{T}\sum_{t=1}^{T} \widehat{\beta}^{(t)j}$, $\mathrm{esd}^j$ is the empirical standard error of the $j$th estimator, $\mathrm{msd}^j$ stands for the standard error of the $j$th estimator estimated by proposed model, $CR^j$ is computed as

$$CR_l^j = \frac{1}{T}\sum_{t=1}^{T} I(\hat{\beta}^{(t)j(L)} < \beta_l^j < \hat{\beta}^{(t)j(U)}),$$

with $\hat{\beta}^{(t)j(L)}$ and $\hat{\beta}^{(t)j(U)}$ respectively representing the lower and upper bounds of the 95% confidence interval at simulation $t$, and $T$ is the number of simulations taken as 1000 for each setting.

The results are presented in Tables 4.3 and 4.4. Simulation results clearly show that in the presence of mismeasurement in responses, the naive GNSM generally produces large finite sample biases and unreliable coverage rates for 95% confidence intervals. On the contrary, the augmented GNSM method adjusts for the mismeasurement effects and produces good results with small finite sample biases for the point estimates and fairly good coverage rates for 95% confidence intervals.

The estimators produced with the availability of an external validation sample have higher standard errors than those obtained under the scenario where the true parameters are known. On the other hand, the estimators resulted from the interval validation method are the most efficient among the three methods, confirming the results in Theorem 4.6.

## 4.6 Sensitivity Analysis of Mice SNPs Data

In this section, we apply the proposed method to analyze the outbred Carworth Farms White (CFW) mice data arising from a genome-wide association study (Parker et al., 2016b). The data set includes measurements of 1200 mice on behavioral, physiological, and gene expression traits. It is interesting to study the association between a set of candidate SNPs as well as their possible interactions with two bone morphology traits, defined as the length of the tibia and the bone condition. To be specific, the covariates include 20 candidate SNPs which were shown to be potentially associated with physiological traits, reported in Supplementary Table 2 of Parker et al. (2016b) and were scaled to have zero mean and unit standard error. Let $Y_{i1}$ denote the length of the tibia bone (in $mm$) and let $Y_{i2}$ be a binary outcome where "0" represents a healthy bone and "1" stands for an abnormal bone. The measurements of $Y_{i1}$ and $Y_{i2}$ are error-prone, where measurement error may be involved with the continuous responses $Y_{i1}$ due to laboratory error and variation, and misclassification may occur in classifying the value of $Y_{i2}$ which is based on the 90 percentile of bone-mineral density (BMD) of the sample. Consequently, the available measurements are taken as surrogate measurements, denoted as $Y_{i1}^*$ and $Y_{i2}^*$, of the true responses $Y_{i1}$ and $Y_{i2}$.

To analyze the data by accommodating possibly existing association structures in the covariates as well as addressing the mismeasurement effects in responses, we employ the two-step procedure for the proposed augmented generalized structured network model to conduct inferences. In the first step, we fit a Gaussian graphical model to the covariates using the method of Section 4.2.2 with the optimal $\lambda$ determined by the rotation information criterion. The identified association structure among the covariates is displayed on the left-hand side of Figure 4.3, which shows only four identified edges. On the right-hand side of Figure 4.3, we plot the sparsity level against the tuning parameter $\lambda$, where the sparsity level is defined as the number of selected edges divided by the total number of edges in the saturated graph. It is seen that the sparsity is fairly insensitive to the choice of tuning parameter around the neighbor of our optimal $\lambda$.

In the second step, we implement the estimation method described in Section 4.2.2 by incorporating the covariate association structure identified in the first step, where the response model is given by (4.5) with $g_1(t) = t$ and $g_2(t) = \log \frac{t}{1-t}$, and the measurement error model and the misclassification model are specified as (3.9) and (3.8), respectively. To show how inference results may be affected by mismeasurement effects, we conduct sensitivity analysis by considering different degrees of mismeasurement in $Y_{i1}$ and $Y_{i2}$. For model (3.9), we take $\sigma_e = 0.77$ according to Lynch et al. (2019); in addition, we set $\sigma_e$ to be 0.72 or 0.82. For model (3.8), we consider $\alpha_0 = \alpha_1 = -2.5, -1.5$, or $-0.5$, respectively yielding tiny (5%), moderate (10%), and substantial (20%) misclassification rates.

The analysis results are presented in Tables 4.5–4.7. The estimation results and the inference conclusions are not sensitive to the different degrees of measurement error and misclassification rates we consider. The SNPs rs25203010 and rs265727287 are significantly associated with tibia length, which is consistent with the finding in Parker et al. (2016b). For the bone condition responses, rs33583459, rs29477109, and rs265727287 are identified to be the significant factors as their p-values are smaller than 0.01. The four interaction terms are strongly associated with the responses, indicating that the network structure plays an important role in studying the relationship between the candidate SNPs and the responses.

Table 4.1: The true parameter values for the data generation model in the simulation studies

| Graph | The continuous component | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | $\beta_{11}$ | $\beta_{12}$ | $\beta_{13}$ | $\beta_{14}$ | $\beta_{15}$ | $\beta_{16}$ | $\beta_{1,(1,2)}$ | $\beta_{1,(1,3)}$ | $\beta_{1,(1,4)}$ | $\beta_{1,(1,5)}$ | $\beta_{1,(1,6)}$ |
| Hub | 0.644 | 0.675 | 0.331 | 0.430 | 0.139 | -0.442 | -0.458 | 0.409 | -0.681 | 0 | 0 |
| Scale-free | 0.644 | 0.675 | 0.331 | 0.430 | 0.139 | -0.442 | -0.458 | 0.409 | -0.681 | 0 | 0.518 |
| Block | 0.644 | 0.675 | 0.331 | 0.430 | 0.139 | -0.442 | -0.458 | 0.409 | 0 | 0 | 0 |
| | $\beta_{1,(2,3)}$ | $\beta_{1,(2,4)}$ | $\beta_{1,(2,5)}$ | $\beta_{1,(2,6)}$ | $\beta_{1,(3,4)}$ | $\beta_{1,(3,5)}$ | $\beta_{1,(3,6)}$ | $\beta_{1,(4,5)}$ | $\beta_{1,(4,6)}$ | $\beta_{1,(5,6)}$ | |
| Hub | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0.662 | 0.633 | 0 | |
| Scale-free | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0.484 | |
| Block | -0.390 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | |
| | The binary component | | | | | | | | | | |
| | $\beta_{21}$ | $\beta_{22}$ | $\beta_{23}$ | $\beta_{24}$ | $\beta_{25}$ | $\beta_{26}$ | $\beta_{2,(1,2)}$ | $\beta_{2,(1,3)}$ | $\beta_{2,(1,4)}$ | $\beta_{2,(1,5)}$ | $\beta_{2,(1,6)}$ |
| Hub | -0.168 | -0.270 | -0.579 | -0.266 | 0.143 | -0.256 | -0.396 | -0.599 | -0.160 | 0 | 0 |
| Scale-free | -0.168 | -0.270 | -0.579 | -0.266 | 0.143 | -0.256 | -0.396 | -0.599 | -0.160 | 0 | -0.516 |
| Block | -0.168 | -0.270 | -0.579 | -0.266 | 0.143 | -0.256 | -0.396 | -0.599 | 0 | 0 | 0 |
| | $\beta_{2,(2,3)}$ | $\beta_{2,(2,4)}$ | $\beta_{2,(2,5)}$ | $\beta_{2,(2,6)}$ | $\beta_{2,(3,4)}$ | $\beta_{2,(3,5)}$ | $\beta_{2,(3,6)}$ | $\beta_{2,(4,5)}$ | $\beta_{2,(4,6)}$ | $\beta_{2,(5,6)}$ | |
| Hub | 0 | 0 | 0 | 0 | 0 | 0 | 0 | -0.223 | -0.445 | 0 | |
| Scale-free | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | -0.665 | |
| Block | -0.658 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | |

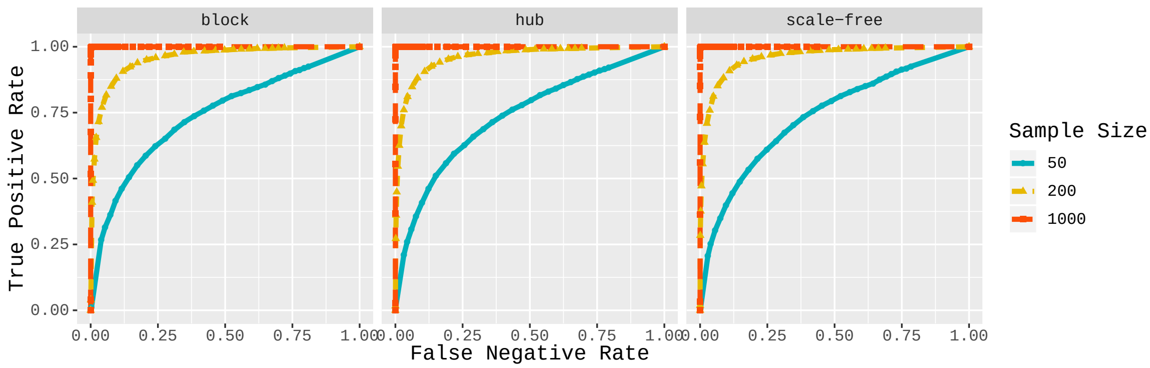


Figure 4.2: *Results for Simulation 1: The plot of true positive rate against the false negative rate obtained from the proposed GNSM for different values of tuning parameter* $\lambda$

Table 4.2: Results for Simulation 1: The bias of the estimators of $\beta$ with different responses types, sample sizes, graph types

| | | Continuous Component | | Discrete Component | |
|---|---|---|---|---|---|
| $n$ | Graph | $\|\cdot\|_1$ | $\|\cdot\|_2$ | $\|\cdot\|_1$ | $\|\cdot\|_2$ |
| 50 | block | 6.369 | 3.562 | 7.636 | 3.195 |
| | hub | 7.291 | 3.727 | 9.274 | 3.651 |
| | scale-free | 6.914 | 3.420 | 9.128 | 3.579 |
| 200 | block | 1.366 | 0.752 | 1.345 | 0.550 |
| | hub | 0.772 | 0.312 | 1.836 | 0.684 |
| | scale-free | 0.691 | 0.279 | 1.806 | 0.676 |
| 1000 | block | 0.228 | 0.093 | 0.561 | 0.228 |
| | hub | 0.273 | 0.101 | 0.742 | 0.274 |
| | scale-free | 0.269 | 0.099 | 0.755 | 0.278 |

Table 4.3: Results for Simulation 2 with $\sigma_e = 0.2$

| | | | | | | Proposed Method | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | Naive Method | | | | Known Parameter | | | | Internal Validation | | | | External Validation | | | |
| Graph | $\pi$ [a] | avgBias | avgSEE | avgSEM | avgCI% | avgBias | avgSEE | avgSEM | avgCI% | avgBias | avgSEE | avgSEM | avgCI% | avgBias | avgSEE | avgSEM | avgCI% |
| | | Continuous Response | | | | | | | | | | | | | | | |
| Block | 1% | 0.040 | 0.042 | 0.034 | 67.7 | 0.001 | 0.032 | 0.035 | 95.7 | 0.003 | 0.026 | 0.026 | 94.5 | 0.002 | 0.032 | 0.033 | 94.4 |
| | 5% | 0.040 | 0.042 | 0.034 | 67.4 | 0.001 | 0.032 | 0.035 | 95.7 | 0.001 | 0.026 | 0.026 | 94.4 | 0.004 | 0.032 | 0.033 | 94.4 |
| | 10% | 0.040 | 0.042 | 0.034 | 67.3 | 0.002 | 0.033 | 0.035 | 95.3 | 0.001 | 0.027 | 0.027 | 94.6 | 0.003 | 0.033 | 0.034 | 94.6 |
| Hub | 1% | 0.036 | 0.042 | 0.034 | 70.6 | 0.001 | 0.032 | 0.034 | 95.9 | 0.001 | 0.026 | 0.026 | 94.6 | 0.001 | 0.032 | 0.033 | 94.3 |
| | 5% | 0.037 | 0.042 | 0.034 | 70.5 | 0.001 | 0.032 | 0.034 | 95.7 | 0.001 | 0.026 | 0.026 | 94.6 | 0.001 | 0.032 | 0.033 | 94.3 |
| | 10% | 0.037 | 0.042 | 0.034 | 70.4 | 0.001 | 0.032 | 0.034 | 95.3 | 0.001 | 0.026 | 0.027 | 94.6 | 0.004 | 0.033 | 0.034 | 94.3 |
| Scale-free | 1% | 0.036 | 0.042 | 0.034 | 72.2 | 0.001 | 0.032 | 0.034 | 95.7 | 0.000 | 0.026 | 0.026 | 94.4 | 0.001 | 0.032 | 0.033 | 94.3 |
| | 5% | 0.037 | 0.042 | 0.034 | 72.0 | 0.001 | 0.032 | 0.034 | 95.6 | 0.001 | 0.026 | 0.026 | 94.5 | 0.001 | 0.032 | 0.033 | 94.3 |
| | 10% | 0.037 | 0.042 | 0.034 | 71.9 | 0.001 | 0.033 | 0.035 | 95.1 | 0.001 | 0.026 | 0.027 | 94.6 | 0.003 | 0.034 | 0.036 | 94.4 |
| | | Discrete Response | | | | | | | | | | | | | | | |
| Block | 1% | 0.021 | 0.079 | 0.081 | 93.7 | 0.007 | 0.085 | 0.089 | 95.3 | 0.005 | 0.069 | 0.071 | 95.3 | 0.009 | 0.086 | 0.091 | 95.6 |
| | 5% | 0.069 | 0.076 | 0.079 | 83.3 | 0.011 | 0.095 | 0.099 | 95.4 | 0.004 | 0.075 | 0.078 | 95.3 | 0.009 | 0.098 | 0.102 | 95.6 |
| | 10% | 0.139 | 0.073 | 0.076 | 53.1 | 0.018 | 0.119 | 0.114 | 93.6 | 0.012 | 0.097 | 0.102 | 96.3 | 0.024 | 0.135 | 0.148 | 96.7 |
| Hub | 1% | 0.019 | 0.080 | 0.082 | 94.0 | 0.009 | 0.086 | 0.089 | 95.3 | 0.006 | 0.069 | 0.072 | 94.9 | 0.009 | 0.087 | 0.095 | 94.9 |
| | 5% | 0.064 | 0.078 | 0.079 | 84.7 | 0.011 | 0.097 | 0.100 | 95.1 | 0.007 | 0.076 | 0.078 | 94.9 | 0.013 | 0.100 | 0.103 | 95.4 |
| | 10% | 0.126 | 0.075 | 0.076 | 59.1 | 0.020 | 0.123 | 0.116 | 92.9 | 0.015 | 0.098 | 0.104 | 96.0 | 0.030 | 0.138 | 0.152 | 96.8 |
| Scale-free | 1% | 0.022 | 0.080 | 0.083 | 93.6 | 0.006 | 0.088 | 0.091 | 95.2 | 0.007 | 0.071 | 0.073 | 94.7 | 0.011 | 0.091 | 0.098 | 95.1 |
| | 5% | 0.071 | 0.078 | 0.081 | 83.4 | 0.009 | 0.099 | 0.102 | 94.8 | 0.007 | 0.078 | 0.081 | 95.3 | 0.013 | 0.102 | 0.106 | 95.4 |
| | 10% | 0.140 | 0.076 | 0.077 | 53.6 | 0.019 | 0.126 | 0.120 | 93.0 | 0.016 | 0.101 | 0.109 | 96.2 | 0.032 | 0.143 | 0.161 | 96.9 |

[a] The misclassification rate $\pi$ is set as $\pi_1 = \pi_0$.

Table 4.4: Results for Simulation 2 with $\sigma_e = 0.7$

| | | | | | | Proposed Method | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | Naive Method | | | | Known Parameter | | | | Internal Validation | | | | External Validation | | | |
| Graph | $\pi$ [a] | avgBias | avgSEE | avgSEM | avgCI% | avgBias | avgSEE | avgSEM | avgCI% | avgBias | avgSEE | avgSEM | avgCI% | avgBias | avgSEE | avgSEM | avgCI% |
| | | Continuous Response | | | | | | | | | | | | | | | |
| Block | 1 | 0.040 | 0.047 | 0.041 | 75.1 | 0.006 | 0.038 | 0.041 | 95.3 | 0.001 | 0.030 | 0.030 | 94.7 | 0.005 | 0.040 | 0.040 | 94.3 |
| | 5 | 0.040 | 0.047 | 0.041 | 74.9 | 0.003 | 0.038 | 0.041 | 95.2 | 0.001 | 0.030 | 0.030 | 94.4 | 0.004 | 0.040 | 0.040 | 94.4 |
| | 10 | 0.040 | 0.047 | 0.040 | 74.9 | 0.007 | 0.039 | 0.041 | 94.9 | 0.003 | 0.030 | 0.031 | 94.6 | 0.011 | 0.040 | 0.041 | 94.4 |
| Hub | 1 | 0.037 | 0.047 | 0.040 | 77.5 | 0.002 | 0.038 | 0.040 | 95.4 | 0.003 | 0.030 | 0.030 | 94.5 | 0.016 | 0.039 | 0.039 | 94.6 |
| | 5 | 0.037 | 0.047 | 0.040 | 77.3 | 0.003 | 0.038 | 0.040 | 95.3 | 0.006 | 0.030 | 0.030 | 94.4 | 0.019 | 0.039 | 0.039 | 94.6 |
| | 10 | 0.037 | 0.047 | 0.040 | 77.3 | 0.006 | 0.039 | 0.040 | 94.9 | 0.001 | 0.030 | 0.030 | 94.6 | 0.016 | 0.040 | 0.041 | 94.5 |
| Scale-free | 1 | 0.036 | 0.047 | 0.040 | 78.2 | 0.001 | 0.039 | 0.040 | 95.2 | 0.001 | 0.030 | 0.030 | 94.3 | 0.001 | 0.039 | 0.039 | 94.1 |
| | 5 | 0.037 | 0.047 | 0.040 | 78.0 | 0.002 | 0.039 | 0.040 | 95.1 | 0.001 | 0.030 | 0.030 | 94.4 | 0.004 | 0.039 | 0.039 | 94.1 |
| | 10 | 0.036 | 0.047 | 0.040 | 78.0 | 0.004 | 0.040 | 0.041 | 94.7 | 0.001 | 0.030 | 0.031 | 94.6 | 0.010 | 0.040 | 0.043 | 94.3 |
| | | Discrete Response | | | | | | | | | | | | | | | |
| Block | 1 | 0.021 | 0.079 | 0.081 | 93.8 | 0.007 | 0.086 | 0.089 | 95.4 | 0.005 | 0.069 | 0.071 | 95.3 | 0.009 | 0.086 | 0.091 | 95.5 |
| | 5 | 0.069 | 0.076 | 0.079 | 83.3 | 0.011 | 0.095 | 0.099 | 95.4 | 0.004 | 0.075 | 0.078 | 95.3 | 0.009 | 0.098 | 0.102 | 95.6 |
| | 10 | 0.139 | 0.073 | 0.076 | 53.2 | 0.018 | 0.119 | 0.114 | 93.6 | 0.012 | 0.097 | 0.102 | 96.3 | 0.024 | 0.134 | 0.148 | 96.7 |
| Hub | 1 | 0.018 | 0.080 | 0.082 | 94.1 | 0.009 | 0.086 | 0.089 | 95.3 | 0.006 | 0.069 | 0.072 | 94.9 | 0.008 | 0.088 | 0.095 | 94.8 |
| | 5 | 0.063 | 0.078 | 0.079 | 84.9 | 0.011 | 0.097 | 0.100 | 95.1 | 0.007 | 0.076 | 0.078 | 94.9 | 0.012 | 0.100 | 0.103 | 95.4 |
| | 10 | 0.126 | 0.075 | 0.076 | 59.3 | 0.020 | 0.123 | 0.116 | 92.9 | 0.015 | 0.098 | 0.104 | 96.0 | 0.029 | 0.138 | 0.151 | 96.7 |
| Scale-free | 1 | 0.022 | 0.081 | 0.083 | 93.6 | 0.006 | 0.088 | 0.091 | 95.1 | 0.007 | 0.071 | 0.073 | 94.7 | 0.011 | 0.090 | 0.098 | 95.1 |
| | 5 | 0.071 | 0.078 | 0.081 | 83.5 | 0.009 | 0.099 | 0.102 | 94.8 | 0.007 | 0.078 | 0.081 | 95.3 | 0.013 | 0.102 | 0.106 | 95.4 |
| | 10 | 0.139 | 0.076 | 0.077 | 53.8 | 0.019 | 0.126 | 0.120 | 93.0 | 0.016 | 0.101 | 0.109 | 96.2 | 0.031 | 0.143 | 0.160 | 96.9 |

[a] The misclassification rate $\pi$ is set as $\pi_1 = \pi_0$.

Table 4.5: Results of the sensitivity analysis for the mice SNPs data when $\sigma_e = 0.72$

| Response | Tiebia Length | | | | | | Bone Condition | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | 5% | | 10% | | 20% | | 5% | | 10% | | 20% | |
| SNPnames | $\beta$ (S.E.) | p-value | $\beta$ (S.E.) | p-value | $\beta$ (S.E.) | p-value | $\beta$ (S.E.) | p-value | $\beta$ (S.E.) | p-value | $\beta$ (S.E.) | p-value |
| rs45690064 | 0.106 (0.349) | 0.761 | 0.106 (0.349) | 0.761 | 0.106 (0.349) | 0.761 | 0.028 (0.096) | 0.772 | 0.073 (0.161) | 0.650 | 0.059 (0.203) | 0.770 |
| rs27338905 | -0.386 (0.347) | 0.266 | -0.386 (0.347) | 0.266 | -0.386 (0.347) | 0.266 | 0.095 (0.101) | 0.345 | 0.212 (0.169) | 0.209 | 0.181 (0.213) | 0.395 |
| rs32962338 | -0.547 (0.362) | 0.131 | -0.547 (0.362) | 0.131 | -0.547 (0.362) | 0.131 | 0.182 (0.095) | 0.054 | 0.245 (0.160) | 0.125 | 0.287 (0.203) | 0.158 |
| rs33583459 | -0.123 (0.345) | 0.722 | -0.123 (0.345) | 0.722 | -0.123 (0.345) | 0.722 | 0.655 (0.109) | <**0.001** | 1.281 (0.256) | <**0.001** | 1.146 (0.389) | **0.003** |
| rs224051056 | 0.086 (0.344) | 0.802 | 0.086 (0.344) | 0.802 | 0.086 (0.344) | 0.802 | 0.033 (0.097) | 0.734 | 0.036 (0.156) | 0.819 | 0.060 (0.199) | 0.762 |
| rs33217671 | -0.230 (0.358) | 0.521 | -0.230 (0.358) | 0.521 | -0.230 (0.358) | 0.521 | 0.241 (0.101) | 0.017 | 0.465 (0.190) | 0.014 | 0.439 (0.236) | 0.063 |
| rs38916331 | -0.045 (0.340) | 0.896 | -0.045 (0.340) | 0.896 | -0.045 (0.340) | 0.896 | 0.013 (0.107) | 0.903 | -0.001 (0.180) | 0.997 | 0.035 (0.222) | 0.873 |
| rs47869247 | -0.091 (0.354) | 0.796 | -0.091 (0.354) | 0.796 | -0.091 (0.354) | 0.796 | 0.035 (0.095) | 0.716 | 0.059 (0.155) | 0.704 | 0.048 (0.196) | 0.808 |
| rs217439518 | -0.202 (0.345) | 0.559 | -0.202 (0.345) | 0.559 | -0.202 (0.345) | 0.559 | 0.062 (0.101) | 0.540 | 0.111 (0.166) | 0.505 | 0.129 (0.208) | 0.534 |
| rs29477109 | -0.017 (0.355) | 0.962 | -0.017 (0.355) | 0.962 | -0.017 (0.355) | 0.962 | -0.458 (0.107) | <**0.001** | -0.820 (0.233) | <**0.001** | -0.783 (0.285) | **0.006** |
| rs252503010 | -5.065 (1.598) | **0.002** | -5.065 (1.598) | **0.002** | -5.065 (1.598) | **0.002** | 1.479 (0.642) | 0.021 | 0.574 (0.466) | 0.219 | 1.128 (0.882) | 0.201 |
| rs265727287 | 8.857 (1.627) | <**0.001** | 8.857 (1.627) | <**0.001** | 8.857 (1.627) | <**0.001** | -1.879 (0.445) | <**0.001** | -2.607 (0.586) | <**0.001** | -2.645 (0.977) | **0.007** |
| rs246035173 | 0.698 (0.481) | 0.147 | 0.698 (0.481) | 0.147 | 0.698 (0.481) | 0.147 | -0.094 (0.123) | 0.445 | -0.171 (0.215) | 0.427 | -0.184 (0.251) | 0.463 |
| rs231489766 | -0.449 (0.497) | 0.366 | -0.449 (0.497) | 0.366 | -0.449 (0.497) | 0.366 | 0.108 (0.137) | 0.428 | 0.095 (0.225) | 0.673 | 0.164 (0.287) | 0.569 |
| rs46826545 | 0.395 (0.411) | 0.336 | 0.395 (0.411) | 0.336 | 0.395 (0.411) | 0.336 | -0.089 (0.110) | 0.419 | -0.174 (0.185) | 0.348 | -0.173 (0.231) | 0.452 |
| rs51809856 | 0.751 (0.455) | 0.099 | 0.751 (0.455) | 0.099 | 0.751 (0.455) | 0.099 | -0.099 (0.116) | 0.390 | -0.257 (0.218) | 0.238 | -0.197 (0.254) | 0.440 |
| rs6279141 | -0.141 (0.429) | 0.743 | -0.141 (0.429) | 0.743 | -0.141 (0.429) | 0.743 | 0.053 (0.106) | 0.619 | 0.058 (0.186) | 0.754 | 0.091 (0.227) | 0.688 |
| rs30535702 | 0.087 (0.357) | 0.808 | 0.087 (0.357) | 0.808 | 0.087 (0.357) | 0.808 | -0.223 (0.101) | 0.028 | -0.498 (0.193) | **0.010** | -0.383 (0.243) | 0.115 |
| rs30201629 | 0.180 (0.381) | 0.636 | 0.180 (0.381) | 0.636 | 0.180 (0.381) | 0.636 | -0.046 (0.098) | 0.641 | -0.048 (0.141) | 0.731 | -0.081 (0.183) | 0.655 |
| rs30549753 | -0.495 (0.334) | 0.138 | -0.495 (0.334) | 0.138 | -0.495 (0.334) | 0.138 | 0.129 (0.090) | 0.156 | 0.094 (0.125) | 0.452 | 0.206 (0.173) | 0.236 |
| Interaction 1 [a] | 7.178 (0.203) | <**0.001** | 7.178 (0.203) | <**0.001** | 7.178 (0.203) | <**0.001** | -1.699 (0.598) | **0.004** | -2.082 (0.417) | <**0.001** | -1.526 (0.338) | <**0.001** |
| Interaction 2 [b] | 3.249 (0.334) | <**0.001** | 3.249 (0.334) | <**0.001** | 3.249 (0.334) | <**0.001** | -0.650 (0.170) | <**0.001** | -0.969 (0.279) | **0.001** | -1.090 (0.384) | **0.004** |
| Interaction 3 [c] | 2.612 (0.316) | <**0.001** | 2.612 (0.316) | <**0.001** | 2.612 (0.316) | <**0.001** | -0.507 (0.150) | **0.001** | -0.529 (0.190) | **0.005** | -0.777 (0.318) | 0.014 |
| Interaction 4 [d] | 2.946 (0.345) | <**0.001** | 2.946 (0.345) | <**0.001** | 2.946 (0.345) | <**0.001** | -0.459 (0.128) | <**0.001** | -0.668 (0.229) | **0.004** | -0.808 (0.315) | **0.010** |

[a] rs252503010 × rs265727287 [b] rs246035173 × rs231489766 [c] rs231489766 × rs46826545 [d] rs51809856 × rs6279141

Table 4.6: Results of the sensitivity analysis for the mice SNPs data when $\sigma_e = 0.77$

| Response | Tiebia Length | | | | | | Bone Condition | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | 5% | | 10% | | 20% | | 5% | | 10% | | 20% | |
| SNPnames | $\beta$ (S.E.) | p-value | $\beta$ (S.E.) | p-value | $\beta$ (S.E.) | p-value | $\beta$ (S.E.) | p-value | $\beta$ (S.E.) | p-value | $\beta$ (S.E.) | p-value |
| rs45690064 | 0.106 (0.349) | 0.761 | 0.106 (0.349) | 0.761 | 0.106 (0.349) | 0.761 | 0.028 (0.096) | 0.772 | 0.057 (0.166) | 0.732 | 0.070 (0.205) | 0.732 |
| rs27338905 | -0.386 (0.347) | 0.266 | -0.386 (0.347) | 0.266 | -0.386 (0.347) | 0.266 | 0.095 (0.101) | 0.345 | 0.213 (0.171) | 0.214 | 0.198 (0.217) | 0.362 |
| rs32962338 | -0.547 (0.362) | 0.131 | -0.547 (0.362) | 0.131 | -0.547 (0.362) | 0.131 | 0.182 (0.095) | 0.054 | 0.237 (0.164) | 0.149 | 0.287 (0.206) | 0.164 |
| rs33583459 | -0.123 (0.345) | 0.722 | -0.123 (0.345) | 0.722 | -0.123 (0.345) | 0.722 | 0.655 (0.109) | <**0.001** | 1.555 (0.312) | <**0.001** | 1.174 (0.397) | **0.003** |
| rs224051056 | 0.086 (0.344) | 0.802 | 0.086 (0.344) | 0.802 | 0.086 (0.344) | 0.802 | 0.033 (0.097) | 0.734 | 0.028 (0.158) | 0.859 | 0.056 (0.201) | 0.781 |
| rs33217671 | -0.230 (0.358) | 0.521 | -0.230 (0.358) | 0.521 | -0.230 (0.358) | 0.521 | 0.241 (0.101) | 0.017 | 0.444 (0.190) | 0.020 | 0.463 (0.244) | 0.058 |
| rs38916331 | -0.045 (0.340) | 0.896 | -0.045 (0.340) | 0.896 | -0.045 (0.340) | 0.896 | 0.013 (0.107) | 0.903 | -0.014 (0.182) | 0.938 | 0.042 (0.225) | 0.851 |
| rs47869247 | -0.091 (0.354) | 0.796 | -0.091 (0.354) | 0.796 | -0.091 (0.354) | 0.796 | 0.035 (0.095) | 0.716 | 0.063 (0.157) | 0.690 | 0.056 (0.199) | 0.778 |
| rs217439518 | -0.202 (0.345) | 0.559 | -0.202 (0.345) | 0.559 | -0.202 (0.345) | 0.559 | 0.062 (0.101) | 0.540 | 0.110 (0.168) | 0.513 | 0.133 (0.210) | 0.526 |
| rs29477109 | -0.017 (0.355) | 0.962 | -0.017 (0.355) | 0.962 | -0.017 (0.355) | 0.962 | -0.458 (0.107) | <**0.001** | -0.806 (0.237) | **0.001** | -0.819 (0.299) | **0.006** |
| rs252503010 | -5.065 (1.598) | **0.002** | -5.065 (1.598) | **0.002** | -5.065 (1.598) | **0.002** | 1.479 (0.642) | 0.021 | 0.622 (0.472) | 0.188 | 0.955 (0.819) | 0.244 |
| rs265727287 | 8.857 (1.627) | <**0.001** | 8.857 (1.627) | <**0.001** | 8.857 (1.627) | <**0.001** | -1.879 (0.445) | <**0.001** | -2.686 (0.601) | <**0.001** | -2.621 (0.938) | **0.005** |
| rs246035173 | 0.698 (0.481) | 0.147 | 0.698 (0.481) | 0.147 | 0.698 (0.481) | 0.147 | -0.094 (0.123) | 0.445 | -0.170 (0.221) | 0.443 | -0.205 (0.257) | 0.426 |
| rs231489766 | -0.449 (0.497) | 0.366 | -0.449 (0.497) | 0.366 | -0.449 (0.497) | 0.366 | 0.108 (0.137) | 0.428 | 0.090 (0.226) | 0.692 | 0.140 (0.285) | 0.624 |
| rs46826545 | 0.395 (0.411) | 0.336 | 0.395 (0.411) | 0.336 | 0.395 (0.411) | 0.336 | -0.089 (0.110) | 0.419 | -0.169 (0.188) | 0.369 | -0.182 (0.234) | 0.435 |
| rs51809856 | 0.751 (0.455) | 0.099 | 0.751 (0.455) | 0.099 | 0.751 (0.455) | 0.099 | -0.099 (0.116) | 0.390 | -0.288 (0.228) | 0.207 | -0.209 (0.261) | 0.423 |
| rs6279141 | -0.141 (0.429) | 0.743 | -0.141 (0.429) | 0.743 | -0.141 (0.429) | 0.743 | 0.053 (0.106) | 0.619 | 0.064 (0.192) | 0.740 | 0.078 (0.230) | 0.736 |
| rs30535702 | 0.087 (0.357) | 0.808 | 0.087 (0.357) | 0.808 | 0.087 (0.357) | 0.808 | -0.223 (0.101) | 0.028 | -0.482 (0.194) | 0.013 | -0.414 (0.252) | 0.100 |
| rs30201629 | 0.180 (0.381) | 0.636 | 0.180 (0.381) | 0.636 | 0.180 (0.381) | 0.636 | -0.046 (0.098) | 0.641 | -0.044 (0.146) | 0.765 | -0.078 (0.183) | 0.670 |
| rs30549753 | -0.495 (0.334) | 0.138 | -0.495 (0.334) | 0.138 | -0.495 (0.334) | 0.138 | 0.129 (0.090) | 0.156 | 0.075 (0.127) | 0.552 | 0.193 (0.172) | 0.262 |
| Interaction 1 [a] | 7.178 (0.203) | <**0.001** | 7.178 (0.203) | <**0.001** | 7.178 (0.203) | <**0.001** | -1.699 (0.598) | **0.004** | -2.069 (0.382) | <**0.001** | -1.577 (0.353) | <**0.001** |
| Interaction 2 [b] | 3.249 (0.334) | <**0.001** | 3.249 (0.334) | <**0.001** | 3.249 (0.334) | <**0.001** | -0.650 (0.170) | <**0.001** | -0.934 (0.270) | **0.001** | -1.068 (0.378) | **0.005** |
| Interaction 3 [c] | 2.612 (0.316) | <**0.001** | 2.612 (0.316) | <**0.001** | 2.612 (0.316) | <**0.001** | -0.507 (0.150) | **0.001** | -0.480 (0.181) | **0.008** | -0.720 (0.298) | 0.016 |
| Interaction 4 [d] | 2.946 (0.345) | <**0.001** | 2.946 (0.345) | <**0.001** | 2.946 (0.345) | <**0.001** | -0.459 (0.128) | <**0.001** | -0.673 (0.236) | **0.004** | -0.801 (0.319) | 0.012 |

[a] rs252503010 × rs265727287 [b] rs246035173 × rs231489766 [c] rs231489766 × rs46826545 [d] rs51809856 × rs6279141

Table 4.7: Results of the sensitivity analysis for the mice SNPs data when $\sigma_e = 0.82$

| Response | Tiebia Length | | | | | | Bone Condition | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | 5% | | 10% | | 20% | | 5% | | 10% | | 20% | |
| SNPnames | $\beta$ (S.E.) | p-value | $\beta$ (S.E.) | p-value | $\beta$ (S.E.) | p-value | $\beta$ (S.E.) | p-value | $\beta$ (S.E.) | p-value | $\beta$ (S.E.) | p-value |
| rs45690064 | 0.106 (0.349) | 0.761 | 0.106 (0.349) | 0.761 | 0.106 (0.349) | 0.761 | 0.028 (0.096) | 0.772 | 0.069 (0.152) | 0.648 | 0.070 (0.204) | 0.734 |
| rs27338905 | -0.386 (0.347) | 0.266 | -0.386 (0.347) | 0.266 | -0.386 (0.347) | 0.266 | 0.095 (0.101) | 0.345 | 0.213 (0.158) | 0.178 | 0.192 (0.216) | 0.374 |
| rs32962338 | -0.547 (0.362) | 0.131 | -0.547 (0.362) | 0.131 | -0.547 (0.362) | 0.131 | 0.182 (0.095) | 0.054 | 0.258 (0.151) | 0.089 | 0.287 (0.205) | 0.161 |
| rs33583459 | -0.123 (0.345) | 0.722 | -0.123 (0.345) | 0.722 | -0.123 (0.345) | 0.722 | 0.655 (0.109) | **<0.001** | 1.236 (0.237) | **<0.001** | 1.125 (0.383) | **0.003** |
| rs224051056 | 0.086 (0.344) | 0.802 | 0.086 (0.344) | 0.802 | 0.086 (0.344) | 0.802 | 0.033 (0.097) | 0.734 | 0.040 (0.148) | 0.785 | 0.058 (0.201) | 0.772 |
| rs33217671 | -0.230 (0.358) | 0.521 | -0.230 (0.358) | 0.521 | -0.230 (0.358) | 0.521 | 0.241 (0.101) | 0.017 | 0.448 (0.174) | 0.010 | 0.460 (0.243) | 0.058 |
| rs38916331 | -0.045 (0.340) | 0.896 | -0.045 (0.340) | 0.896 | -0.045 (0.340) | 0.896 | 0.013 (0.107) | 0.903 | 0.017 (0.169) | 0.919 | 0.041 (0.224) | 0.854 |
| rs47869247 | -0.091 (0.354) | 0.796 | -0.091 (0.354) | 0.796 | -0.091 (0.354) | 0.796 | 0.035 (0.095) | 0.716 | 0.057 (0.147) | 0.699 | 0.053 (0.198) | 0.790 |
| rs217439518 | -0.202 (0.345) | 0.559 | -0.202 (0.345) | 0.559 | -0.202 (0.345) | 0.559 | 0.062 (0.101) | 0.540 | 0.108 (0.156) | 0.490 | 0.131 (0.209) | 0.531 |
| rs29477109 | -0.017 (0.355) | 0.962 | -0.017 (0.355) | 0.962 | -0.017 (0.355) | 0.962 | -0.458 (0.107) | **<0.001** | -0.816 (0.215) | **<0.001** | -0.812 (0.296) | **0.006** |
| rs252503010 | -5.065 (1.598) | **0.002** | -5.065 (1.598) | **0.002** | -5.065 (1.598) | **0.002** | 1.479 (0.642) | 0.021 | 0.743 (0.464) | 0.110 | 0.969 (0.819) | 0.237 |
| rs265727287 | 8.857 (1.627) | **<0.001** | 8.857 (1.627) | **<0.001** | 8.857 (1.627) | **<0.001** | -1.879 (0.445) | **<0.001** | -2.505 (0.549) | **<0.001** | -2.596 (0.931) | **0.005** |
| rs246035173 | 0.698 (0.481) | 0.147 | 0.698 (0.481) | 0.147 | 0.698 (0.481) | 0.147 | -0.094 (0.123) | 0.445 | -0.184 (0.201) | 0.360 | -0.198 (0.256) | 0.438 |
| rs231489766 | -0.449 (0.497) | 0.366 | -0.449 (0.497) | 0.366 | -0.449 (0.497) | 0.366 | 0.108 (0.137) | 0.428 | 0.088 (0.210) | 0.676 | 0.149 (0.287) | 0.604 |
| rs46826545 | 0.395 (0.411) | 0.336 | 0.395 (0.411) | 0.336 | 0.395 (0.411) | 0.336 | -0.089 (0.110) | 0.419 | -0.153 (0.172) | 0.375 | -0.181 (0.233) | 0.437 |
| rs51809856 | 0.751 (0.455) | 0.099 | 0.751 (0.455) | 0.099 | 0.751 (0.455) | 0.099 | -0.099 (0.116) | 0.390 | -0.258 (0.204) | 0.205 | -0.201 (0.259) | 0.436 |
| rs6279141 | -0.141 (0.429) | 0.743 | -0.141 (0.429) | 0.743 | -0.141 (0.429) | 0.743 | 0.053 (0.106) | 0.619 | 0.041 (0.174) | 0.816 | 0.082 (0.230) | 0.722 |
| rs30535702 | 0.087 (0.357) | 0.808 | 0.087 (0.357) | 0.808 | 0.087 (0.357) | 0.808 | -0.223 (0.101) | 0.028 | -0.497 (0.184) | 0.007 | -0.406 (0.250) | 0.104 |
| rs30201629 | 0.180 (0.381) | 0.636 | 0.180 (0.381) | 0.636 | 0.180 (0.381) | 0.636 | -0.046 (0.098) | 0.641 | -0.055 (0.134) | 0.682 | -0.080 (0.182) | 0.662 |
| rs30549753 | -0.495 (0.334) | 0.138 | -0.495 (0.334) | 0.138 | -0.495 (0.334) | 0.138 | 0.129 (0.090) | 0.156 | 0.100 (0.120) | 0.401 | 0.200 (0.173) | 0.247 |
| Interaction 1 [a] | 7.178 (0.203) | **<0.001** | 7.178 (0.203) | **<0.001** | 7.178 (0.203) | **<0.001** | -1.699 (0.598) | **0.004** | -1.916 (0.352) | **<0.001** | -1.563 (0.350) | **<0.001** |
| Interaction 2 [b] | 3.249 (0.334) | **<0.001** | 3.249 (0.334) | **<0.001** | 3.249 (0.334) | **<0.001** | -0.650 (0.170) | **<0.001** | -0.956 (0.256) | **<0.001** | -1.081 (0.383) | **0.005** |
| Interaction 3 [c] | 2.612 (0.316) | **<0.001** | 2.612 (0.316) | **<0.001** | 2.612 (0.316) | **<0.001** | -0.507 (0.150) | **0.001** | -0.540 (0.182) | **0.003** | -0.746 (0.308) | 0.016 |
| Interaction 4 [d] | 2.946 (0.345) | **<0.001** | 2.946 (0.345) | **<0.001** | 2.946 (0.345) | **<0.001** | -0.459 (0.128) | **<0.001** | -0.678 (0.215) | **0.002** | -0.804 (0.318) | 0.011 |

[a] rs252503010 × rs265727287 [b] rs246035173 × rs231489766 [c] rs231489766 × rs46826545 [d] rs51809856 × rs6279141

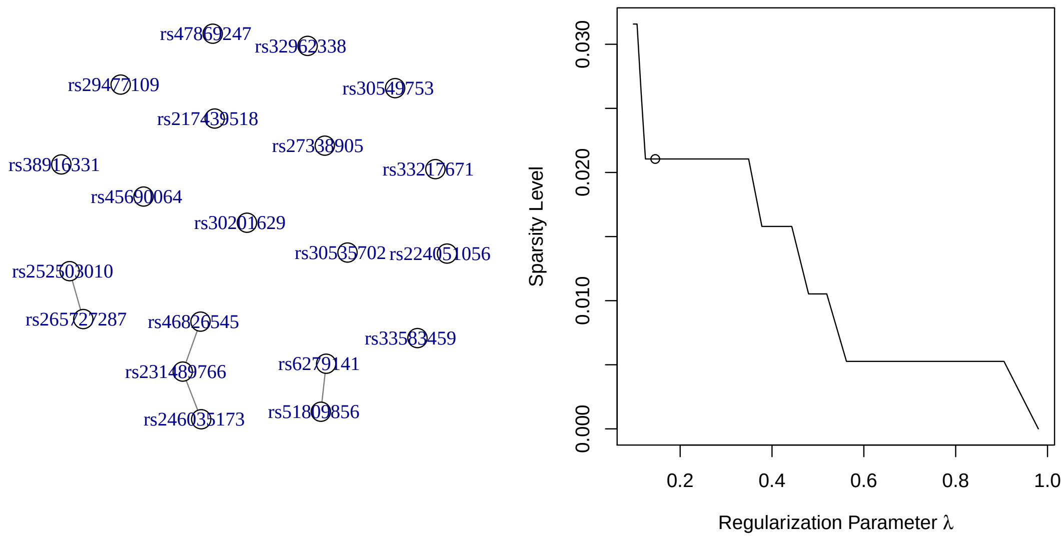


Figure 4.3: *Left panel: the diagram of the covariates structure; Right panel: the solution path sparsity levels*

# Chapter 5

# Zero-Inflated Poisson Models with Measurement Error in Response

In this chapter, we study the measurement error in the zero-inflated Poisson model. In Section 5.1, we discuss the setup of the response model as well as the measurement error model. In Section 5.2, we examine the effects of measurement error on analyzing count data and develop a method in Bayesian framework to account for the measurement error effects. In Section 5.3, we extend the method to accounting for the effects due to measurement error when validation subsamples are available. In Section 5.4, we illustrate the usage of the method by applying it to the prostate adenocarcinoma genomics data. To evaluate the performance of the method, we conduct simulation studies in Section 5.5.

## 5.1 Model Setup and Framework

### 5.1.1 Response Model

For $i = 1, \ldots, n$, let $Y_i$ denote the count outcome for subject $i$ taking a non-negative integer value and let $X_i$ denote the associated covariate vector of dimension $p_x$, where $n$ is the number of subjects in the study. We assume that $Y_i$ and $Y_{i'}$ are independent for any $i \neq i'$. The responses $Y_i$ are sampled from two sources, either from an "at-risk" group where the measurements follow a Poisson distribution, or from a "non-at-risk" group where the measurements are zero. Let $A_i$ be a latent indicator variable showing from which sources $Y_i$ is sampled, where "$A_i = 1$" represents $Y_i$ is sampled from the "at-risk" group, and

"$A_i = 0$" otherwise. For $i = 1, \ldots, n$, let $\phi_i = P(A_i = 1|X_i)$ represent the conditional probability of sampling from 'at-risk" group, given $X_i$, and let $\mu_i = E(Y_i|A_i = 1, X_i)$ denote the condition mean of $Y_i$, given being sampled from the 'at-risk" group and the covariate $X_i$, which are assumed to satisfy $0 < \phi_i < 1$, and $\mu_i > 0$. That is, $Y_i$ is sampled from the "non-at-risk" group with probability $1-\phi_i$, and sampled from the "at-risk" group with probability $\phi_i$, following a Poisson distribution with mean $\mu_i$:

$$\begin{aligned} &Y_i = 0, \text{ with probability } 1 - \phi_i, \\ &Y_i \sim \text{Poisson}(\mu_i), \text{ with probability } \phi_i. \end{aligned} \tag{5.1}$$

Therefore, the zero values of $Y_i$ may come from two sources: either from the "non-at-risk" group or from the "at-risk" group taking a zero count. Consequently, the probability mass function for response $Y_i$ is given by

$$\begin{aligned} P(Y_i = 0|X_i) &= \textstyle\sum_{k=0}^{1} P(Y_i = 0|A_i = k, X_i) P(A_i = k|X_i) \\ &= (1 - \phi_i) + \phi_i e^{-\mu_i}; \\ P(Y_i = y_i|X_i) &= \phi_i \frac{\mu_i^{y_i} e^{-\mu_i}}{y_i!} \qquad \text{for } \; y_i = 1, 2, \ldots. \end{aligned} \tag{5.2}$$

To facilitate the dependence of $\phi_i$ and $\mu_i$ on covariates $X_i$, we consider a complementary log-log regression model for $\phi_i$ and a log linear model for $\mu_i$:

$$\text{cloglog } \; \phi_i = \beta_{\phi 0} + \beta_{\phi x}^{\mathrm{T}} X_i, \tag{5.3}$$

$$\log \; \mu_i = \beta_{\mu 0} + \beta_{\mu x}^{\mathrm{T}} X_i, \tag{5.4}$$

where $(\beta_{\phi 0}, \beta_{\phi x}^{\mathrm{T}})^{\mathrm{T}}$ and $(\beta_{\mu 0}, \beta_{\mu x}^{\mathrm{T}})^{\mathrm{T}}$ are the coefficients of the binary component and the count component respectively, $\beta = (\beta_{\phi 0}, \beta_{\phi x}^{\mathrm{T}}, \beta_{\mu 0}, \beta_{\mu x}^{\mathrm{T}})^{\mathrm{T}}$, and $\text{cloglog}(t) = \log\{-\log(1-t)\}$ refers to the function of complementary log-log link. The complementary log-log link has been frequently used to model zero-inflated Poisson model in the literature (e.g. Neelon and Chung, 2017), whose interpretation is to be discussed in Section 5.2.1. We comment that although we used the same notation $X_i$ to denote the covariates for ease of notations, the covariates could be different for each component in (5.3) and (5.4) by constraining the corresponding coefficients to be zero.

### 5.1.2 Measurement Error Model

Due to the measurement error in response $Y_i$, its precise measurement is not observed for every subject $i \in \{1, \ldots, n\}$, but instead, surrogate measurement $Y_i^*$ is observed for $i = 1, \ldots, n$.

Measurement error for count data often arises from two distinct scenarios, and we call them the "*add-in*" and "*leave-out*", respectively. The add-in error generates extra counts that are not supposed to be counted when measuring $Y_i$, yielding that the surrogate $Y_i^*$ is no smaller than the true value of $Y_i$. For example, in genomics studies, we are interested in examining the count of copy number variants (CNVs). However, the mapping errors and incorrect sequencing may falsely include some insignificant CNVs, leading to the erroneous count higher than the true value (Xie and Tammi, 2009). On the contrary, the leave-out error may be caused by the loss of counts that should have been counted. In the CNV example, a significant CNV may fail to be identified due to the under-counting from the sequencing error. In the study of COVID-19, the daily reported cases number are often subject to leave-out error due to the limited test capacity and undetected asymptomatic infections as well as unreported cases with a mild symptom.

For measuring $Y_i$ with $i = 1, \ldots, n$, let $Z_{i+}$ denote the count due to the add-in error and let $Z_{i-}$ denote the count due to the leave-out error. Here we propose a measurement error model to feature the scenario where both the add-in and leave-out errors may exist; given $X_i$,

$$Y_i^* = Y_i + c_+ Z_{i+} - c_- Z_{i-}, \tag{5.5}$$

where $Z_{i+}$ is independent of $Y_i$ and follows the Poisson distribution with mean $\lambda_i$, i.e. Poisson$(\lambda_i)$; $Z_{i-}$ is independent of $Z_{i+}$ but may be dependent on $Y_i$, and the conditional distribution of $Z_{i-}$, given $Y_i = y_i$ is the Binomial distribution with the probability $\pi_i$, i.e. Binomial$(y_i, \pi_i)$. Here $c_+$ and $c_-$ are weights controlling the type of mismeasurements; they may be restricted to take values in $\{0, 1\}$ to facilitate various scenarios. For instance, if both $c_+$ and $c_-$ are zero, then $Y_i^*$ and $Y_i$ are identical, i.e., no measurement error occurs; if $c_+ = 0$ and $c_- = 1$, then only leave-out error is involved; if $c_+ = 1$ and $c_- = 0$, then only add-in error exists; if $c_+ = c_- = 1$ then both add-in and leave-out errors are equally present. In applications, the background information or researchers' experience may offer a good sense for specifying suitable values for $c+$ and $c-$.

Model (5.5) applies to count data and is a form somewhat similar to the widely used classical additive error model for featuring measurement error in continuous covariates. (Stefanski, 2000; Carroll et al., 2006; Yi, 2017, Section 2.6). But two key differences make model (5.5) unique. First, classical additive measurement error models do not differentiate error sources and use a single random variable, say $e_i$ to represent the errors; secondly, the error term $e_i$ is often assumed to be independent of true covariates. In model (5.5), however, the error term is refined by sorting out the errors of different nature. In addition, dependence of the error on the true variables is allowed. Basically, the joint distribution of $Y_i$ and $\{Z_{i+}, Z_{i-}\}$, is treated as

$$f(y_i, z_{i+}, z_{i-}|x_i) = f(y_i, z_{i-}|x_i)f(z_{i+}|x_i) \tag{5.6}$$

in model (5.5) by allowing the dependence between $Y_i$ and $Z_{i-}$, where $f(\cdot)$ represents the joint or marginal distribution for the variables indicated by the arguments.

In model (5.5), assuming a Poisson distribution for $Z_{i+}$ reflects its unboundedness yet taking a large value with a small probability. This assumption is feasible in applications where no upper limit is set for an add-in error and assuming errors beyond a certain value is not likely. On the contrary, the leave-out error cannot exceed the value of $Y_i$ itself, so assuming a Binomial distribution for the conditional distribution for $Z_{i-}$, given $Y_i$, can be reasonable.

To facilitate different degrees of measurement error, we further model $\lambda_i$ and $\pi_i$ via their dependence on predictors, say, $W_{i+}$ and $W_{i-}$, respectively, where $W_{i+}$ and $W_{i-}$ can be the same or different, and they can be part of covariates $X_i$ or identical to $X_i$. Let $W_{i+} = (W_{i1+}, \ldots, W_{ip_+})^{\mathrm{T}}$ and $W_{i-} = (W_{i1-}, \ldots, W_{ip_-})^{\mathrm{T}}$ denote the covariate vector associated with add-in and leave-out processes, respectively, where $p_+$ and $p_-$ are the dimension of $W_{i+}$ and $W_{i-}$, respectively. For ease of exposition, we assume that $W_{i+}$ and $W_{i-}$ are subsets of $X_i$; if this is not the case, we can modify our initial definition of $X_i$ to include $W_{i+}$ and $W_{i-}$ as its parts.

The mean parameter $\lambda_i$ is modeled as

$$\log \lambda_i = \alpha_{+0} + \alpha_{+w}^{\mathrm{T}} W_{i+}, \tag{5.7}$$

and the probability $\pi_i$ is postulated by a generalized linear model,

$$g(\pi_i) = \alpha_{-0} + \alpha_{-w}^{\mathrm{T}} W_{i-}, \tag{5.8}$$

where $(\alpha_{+0}, \alpha_{+w}^{\mathrm{T}})^{\mathrm{T}}$ and $(\alpha_{-0}, \alpha_{-w}^{\mathrm{T}})^{\mathrm{T}}$ are coefficient vectors and $g(\cdot)$ is a link function. Here the link function $g(\cdot)$ can be taken as the logit function $g(t) = \log \frac{t}{1-t}$, the complementary log-log link $g(t) = \log\{-\log(1-t)\}$, or the probit function $g(\cdot) = \Phi^{-1}(\cdot)$, where $\Phi(\cdot)$ is the cumulative distribution function of the standard Gaussian distribution. Let $\alpha = (\alpha_{+0}, \alpha_{+w}^{\mathrm{T}}, \alpha_{-0}, \alpha_{-w}^{\mathrm{T}})^{\mathrm{T}}$.

### 5.1.3 Impact of Naive Analysis

In the presence of measurement error in response $Y_i$, the true response $Y_i$ may not be observed. Instead, its surrogate $Y_i^*$ is available. If we naively replace the response $Y_i$ by its surrogates $Y_i^*$ in the inference procedure such as the likelihood method, the resulting estimators may not be consistent.

To see how the distribution of $Y_i^*$ is different from $Y_i$, we consider the conditional distribution of $Y_i^*$ given $X_i$:

$$
\begin{aligned}
f(y_i^*|x_i) &= \sum_{z_{i-}} \sum_{z_{i+}} f_{(Y_i^*, Z_{i+}, Z_{i-})|X_i}(y_i^*, z_{i+}, z_{i-}|x_i) \\
&= \sum_{z_{i-}} \sum_{z_{i+}} f_{(Y_i, Z_{i+}, Z_{i-})|X_i}(y_i - z_{i+} + z_{i-}, z_{i+}, z_{i-}|x_i) \\
&= \sum_{z_{i-}} \sum_{z_{i+}} f_{Y_i|X_i}(y_i - z_{i+} + z_{i-}|x_i) f(z_{i+}|x_i) f(z_{i-}|y_i, x_i), \qquad (5.9)
\end{aligned}
$$

where the second step is due to (5.5), the third step is due to the independence assumption in (5.6), $f_{Y_i|X_i}(\cdot|x_i)$ is determined by (5.2) together with (5.3) and (5.4), and $f(z_{i+}|x_i)$ and $f(z_{i-}|y_i, x_i)$ are respectively determined by (5.7) and (5.8). Expression (5.9) shows that the conditional distribution for $Y_i^*$ given $X_i$ generally differs from that for $Y_i$ given $X_i$

However, in some special cases, such as stated in Theorem 5.1 below, the conditional distribution, $f(y_i^*|x_i)$, of $Y_i^*$ given $X_i$, is closely related to the conditional distribution (5.2) of $Y_i$ given $X_i$ in the structure.

**Theorem 5.1** *Suppose $Y_i$ follows the zero-inflated Poisson distribution given by (5.2) and the measurement error model for $Y_i$ is given by (5.5).*

*(a) If $c_+ = 0$ and $c_- = 1$ in (5.5), then $Y_i^*$ also follows a zero-inflated Poisson distribution given by*

$$
\begin{aligned}
P(Y_i^* = 0|X_i) &= (1 - \phi_i^*) + \phi_i^* e^{-\mu_i^*}; \\
P(Y_i^* = y_i^*|X_i) &= \phi_i^* \frac{\mu_i^{*y_i^*} e^{-\mu_i^*}}{y_i^*!} \qquad \text{for } \; y_i^* = 1, 2, \ldots,
\end{aligned}
$$

*where $\phi_i^* = \phi_i$ and $\mu_i^* = (1 - \pi_i)\mu_i$.*

*(b) If $c_+ = 1$ and $c_- = 0$ in (5.5), then $Y_i^*$ follows a mixture distribution of two Poisson distributions, given by*

$$
P(Y_i^* = y_i^*|X_i) = (1 - \phi_i) \frac{\lambda_i^{y_i^*} e^{-\lambda_i}}{y_i^*!} + \phi_i \frac{(\mu_i + \lambda_i)^{y_i^*}}{y_i^*!} e^{-(\mu_i + \lambda_i)} \qquad \text{for } \; y_i^* = 0, 1, 2, \ldots.
$$

*(c) If $c_+ = 1$ and $c_- = 1$ in (5.5), $Y_i^*$ follows a mixture distribution of two Poisson distributions, given by*

$$
P(Y_i^* = y_i^*|X_i) = (1 - \phi_i) \frac{\lambda_i^{y_i^*} e^{-\lambda_i}}{y_i^*!} + \phi_i \frac{\mu_i^{*y_i^*}}{y_i^*!} e^{-\mu_i^*} \qquad \text{for } \; y_i^* = 0, 1, 2, \ldots,
$$

*where* $\mu_i^* = (1 - \pi_i)\mu_i + \lambda_i$.

The proof of Theorem 5.1 is presented in Appendix D.1. Theorem 5.1(a) says that if there is no add-in error in the measurement error model (5.5) and the leave-out error follows Binomial$(y_i, \pi_i)$ then the surrogate variable $Y_i^*$ assumes the same zero-inflated Poisson distribution (5.2) as the true response variable $Y_i$ except for replacing $\mu_i$ with $(1 - \pi_i)\mu_i$, where the factor $1 - \pi_i$ reflects the impact of the degree of the leave-out error. We comment that Theorem 5.1 is analogous to the well-known Poisson Process Thinning Theorem (Brown, 1979, Theorem 1) which says that if $Y_i$ follows Poisson$(\lambda_i)$, then $Y_i^*$ follows Poisson$((1 - \pi_i)\lambda_i)$.

Theorem 5.1(b) suggests that if there is no leave-out error in measurement error model (5.5) and the add-in error follows Poisson$(\lambda_i)$, then the distribution of the surrogate variable $Y_i^*$ is determined by two Poisson distributions, given by

$$\begin{aligned} Y_i^* &\sim \text{Poisson}(\lambda_i), \text{ with probability } 1 - \phi_i, \\ Y_i^* &\sim \text{Poisson}(\mu_i + \lambda_i), \text{ with probability } \phi_i. \end{aligned} \tag{5.10}$$

Theorem 5.1(c) may be viewed as a combined result from Theorem 5.1(a) and (b), saying that when both the add-in error and the leave-out error are present, the distribution of the surrogate variable $Y_i^*$ assumes the same form as (5.10) except that $\mu_i$ is replaced by $(1 - \pi_i)\mu_i$.

Next, we discuss possible biases of the naive analysis which disregards the difference between $Y_i$ and $Y_i^*$. That is, we naively assume that $Y_i^*$ follows the same distribution form as $Y_i$, then we replace $Y_i$ in (5.2) with $Y_i^*$ and let $\phi_i^*$ and $\mu_i^*$ denote the resulting quantities corresponding to $\phi_i$ and $\mu_i$ in (5.2), respectively; in addition, the same model forms as (5.3) and (5.4) are assumed for $\phi_i^*$ and $\mu_i^*$:

$$\text{cloglog } \phi_i^* = \beta_{\phi 0}^* + \beta_{\phi x}^{*\mathrm{T}} X_i, \tag{5.11}$$

$$\log \mu_i^* = \beta_{\mu 0}^* + \beta_{\mu x}^{*\mathrm{T}} X_i, \tag{5.12}$$

where $\beta^* \triangleq (\beta_{\phi 0}^*, \beta_{\phi x}^{*\mathrm{T}}, \beta_{\mu 0}^*, \beta_{\mu x}^{*\mathrm{T}})^{\mathrm{T}}$ are the associated parameters which may differ from the corresponding parameters in the models (5.3) and (5.4). Without adding any constraint on the measurement error (5.5), it is generally expected that $\beta^*$ differs from $\beta$. Even with certain conditions for the measurement error model (5.5), such as those discussed in Theorem 5.1(b)(c), $Y_i^*$ does not follow a zero-inflated Poisson distribution, and thus $\beta^* \neq \beta$. However, for the case considered in Theorem 5.1(a), the following theorem describes the relationship between $\beta^*$ and $\beta$, which shows a scenario where conducting the naive analysis can still yield consistent estimators for some parameters.

**Theorem 5.2** *If the conditions in Theorem 5.1(a) holds, then we have*

*(i)* $\beta^*_{\phi 0} = \beta_{\phi 0}$ *and* $\beta^*_{\phi x} = \beta_{\phi x}$,

*(ii)* $\beta^*_{\mu 0} = \beta_{\mu 0} + \log(1 - \pi_i)$,

*(iii)* $\beta^*_{\mu x} = \beta_{\mu x}$.

The proof of Theorem 5.2 is presented in Appendix D.2. Theorem 5.2 says that when there is only leave-out error in model (5.5), within the frequentist framework, point estimators of the parameters for the response models (5.3) and (5.4) except for the intercept in (5.4) are still consistent if using the naive method by disregarding measurement error. Furthermore, Theorem 5.2(ii) implies that the estimator for $\beta_{\mu 0}$ obtained from the naive method can be adjusted by subtracting $\log(1 - \pi_i)$ to produce a consistent estimator. On the other hand, Theorem 5.2 shows that if $\pi_i$ is unknown, nonidentifiability arises because $\beta_{\mu 0}$ and $\pi_i$ cannot be separated when using the surrogate measurements $Y^*_i$ together with the covariates $X_i$. However, this nonidentifiability issue can be circumvented if we conduct inferences in the Bayesian framework with a weakly informative prior imposed.

## 5.2 Bayesian Analysis Methodology

### 5.2.1 Bayesian Inference and Data Augmentation

Here we propose a Bayesian method for conducting inference about $\beta$ by using the surrogate measurements $Y^*_i$, together with the covariates, where the effects of measurement error are accommodated.

Let $\beta = (\beta_{\phi 0}, \beta^{\mathrm{T}}_{\phi x}, \beta_{\mu 0}, \beta^{\mathrm{T}}_{\mu x})^{\mathrm{T}}$ and let $\theta = (\beta^{\mathrm{T}}, \alpha^{\mathrm{T}})^{\mathrm{T}}$. Inference about the parameter $\theta$ is based on the posterior distribution of $\theta$ given $Y^*_i$ and $X_i$, given by

$$f(\theta|y^*_i, x_i) = \frac{f(y^*_i, \theta|x_i)}{f(y^*_i|x_i)} \propto f(y^*_i|x_i; \theta)\pi(\theta), \tag{5.13}$$

where $f(y^*_i, \theta|x_i)$ represents the joint distribution of $Y_i$ and $\theta$, $\pi(\theta)$ is the prior distribution of parameter $\theta$, $f(y^*_i|x_i; \theta)$ is given by (5.9), and $f(y^*_i|x_i) = \int f(y^*_i|x_i; \theta)\pi(\theta)d\theta$. Then, the Bayes estimator of the parameters are given by the posterior mean $\widehat{\theta} = E(\theta|Y^*_i, X_i)$.

The basic idea of implementing Bayesian estimation is to sample a sequence of parameters from their posterior distribution given by (5.13). Then the Bayes point estimator

$\widehat{\theta}$ is given by taking the sample mean of the sampled parameter sequence, and the $\gamma\%$-credibility interval is given by $(q_{1-\gamma\%}, q_{\gamma\%})$, where $0 < \gamma < 1$, $q_{\gamma\%}$ is the $\gamma\%$ quantile of the sampled parameter sequence.

To this end, one may employ a sampling algorithm such as the Gibbs sampling method to sample a sequence of values from the posteriors distribution (5.13), which, however, can be challenging due to the complex structure of the probability mass function of $Y_i$ in (5.2). To circumvent this, we consider an alternative way to express the distribution of $Y_i$ by using two latent variables, say $U_{i1}$ and $U_{i2}$, which are conditionally independent given $X_i$, and each follows a Poisson distribution. Rather than directly characterizing the distribution of $Y_i$ by using (5.2) together with (5.3) and (5.4), we separately describe (5.3) and (5.4) each using $U_{i1}$ and $U_{i2}$, respectively, to gain the flexibility in modeling of the distribution of $Y_i$.

To be specific, we assume that given $X_i$, $U_{i1}$ and $U_{i2}$ are conditionally independent, and that given $X_i$ the conditional distributions of $U_{i1}$ and $U_{i2}$ are given by

$$\begin{aligned} U_{i1}|X_i &\sim \text{Poisson}(\mu_{i1}), \\ U_{i2}|X_i &\sim \text{Poisson}(\mu_i), \end{aligned}$$

where $\mu_{i1} = \exp(\beta_{\phi 0} + \beta_{\phi x}^{\mathrm{T}} X_i)$ and $\mu_i = \exp(\beta_{\mu 0} + \beta_{\mu x}^{\mathrm{T}} X_i)$ with $\beta_{\phi 0}$, $\beta_{\phi x}$, $\beta_{\mu 0}$ and $\beta_{\mu x}$ being the parameters in (5.3) and (5.4). Then (5.3) is equivalently written as $\phi_i = 1 - \exp(\mu_{i1})$, which can be viewed as the probability $P(U_{i1} > 0|X_i)$. Therefore, the initial definition (5.1) for $Y_i$ is equivalently expressed as

$$\begin{aligned} Y_i &= 0, \text{ with probability } P(U_{i1} = 0|X_i), \\ Y_i &= U_{i2}, \text{ with probability } P(U_{i1} > 0|X_i). \end{aligned} \tag{5.14}$$

In other words, the values of $Y_i$ may be viewed by the distributions of $U_{i1}$ and $U_{i2}$ in such a way:

$$\begin{aligned} &\text{if } U_{i1} = 0, \text{ then we set } Y_i = 0; \\ &\text{if } U_{i1} > 0, \text{ then we set } Y_i = U_{i2}; \end{aligned}$$

and thus, we write $Y_i = 0 \cdot I(U_{i1} = 0) + U_{i2} \cdot I(U_{i1} > 0)$, which is

$$Y_i = U_{i2} I(U_{i1} > 0), \tag{5.15}$$

where $I(\cdot)$ is the indicator function.

Consequently, the original distribution (5.2) of $Y_i$ together with (5.3) and (5.4) can now be equivalently described by using $U_{i1}$ and $U_{i2}$ via (5.15). Thereby, using the idea of *data augmentation* (van Dyk and Meng, 2001), $U_{i1}$ and $U_{i2}$ can be used to ease sampling procedures directly based on (5.13), which is complicated to realize. In particular, rather than using $f(y_i^*|x_i;\theta)$ in (5.13) directly, we use (5.9) with $Y_i$ replaced by (5.15) in its derivation, and sampling parameter values from (5.13) can be equivalently re-expressed as follows.

To see the idea, we consider the case with $c_+ = c_- = 1$ when using (5.5). First, fixing the initial parameter $\theta$, we treat $U_{i1}$, $U_{i2}$, $Z_{i+}$ and $Z_{i-}$ as "missing data" and calculate their posterior distribution, $f(u_{i1}, u_{i2}, z_{i+}, z_{i-}|y_i^*, x_i;\theta)$, given $\{Y_i^*, X_i\}$ and $\theta$, which is given by

$$\begin{aligned}
&f(u_{i1}, u_{i2}, z_{i+}, z_{i-}|y_i^*, x_i;\theta)\\
=&P\left(U_{i1} = u_{i1}, U_{i2} = u_{i2}, Z_{i+} = y_i^* - I(u_{i1} > 0)u_{i2} + z_{i-}, Z_{i-} = z_{i-}|x_i;\theta\right)\\
=&P\left(U_{i1} = u_{i1}|x_i\right) P\left(U_{i2} = u_{i2}|x_i\right) P\left(Z_{i-} = z_{i-}|U_{i1} = u_{i1}, U_{i2} = u_{i2}, x_i\right)\\
&\times P\left(Z_{i+} = y_i^* - I(u_{i1} > 0)u_{i2} + z_{i-}|x_i\right),
\end{aligned} \tag{5.16}$$

where the first equality is due to (5.5) and (5.15), and in the second equality we use the conditional independence between $U_{i1}$ and $U_{i2}$ given $X_i$ as well as (5.6).

Next, we re-express the posterior distribution (5.13) of $\theta$ by replacing $Y_i$ with $U_{i1}$ and $U_{i2}$ and using the measurement error model (5.5):

$$\begin{aligned}
&f(\theta|y_i^*, x_i, u_{i1}, u_{i2}, z_{i+}, z_{i-})\\
\propto&f(u_{i1}, u_{i2}, z_{i+}, z_{i-}|y_i^*, x_i;\theta)\pi(\theta)\\
=&P\left(U_{i1} = u_{i1}|x_i;\beta_{\phi 0}, \beta_{\phi x}\right) P\left(U_{i2} = u_{i2}|x_i;\beta_{\mu 0}, \beta_{\mu x}\right) P\left(Z_{i-} = z_{i-}|u_{i1}, u_{i2}, x_i;\alpha_{-0}, \alpha_{-w}\right)\\
&\times P\left(Z_{i+} = y_i^* - I(u_{i1} > 0)u_{i2} + z_{i-}|x_i;\alpha_{+0}, \alpha_{+w}\right)\pi(\theta),
\end{aligned} \tag{5.17}$$

where the second step is due to (5.16) with the dependence on the parameters spelled out explicitly. The advantage of (5.17) lies in its separation of the components of $\theta$ by using distributions for different random variables, i.e., $U_{i1}$, $U_{i2}$, $Z_{i-}$ and $Z_{i+}$. For example, given the rest parameters, the posterior distribution of parameters $\beta_{\phi 0}$ and $\beta_{\phi x}$ are simplified by (5.17) as

$$\begin{aligned}
&f(\beta_{\phi 0}, \beta_{\phi x}|y_i^*, x_i, u_{i1}, u_{i2}, z_{i+}, z_{i-};\beta_{\mu 0}, \beta_{\mu x}, \alpha_{-0}, \alpha_{-w}, \alpha_{+0}, \alpha_{+w})\\
\propto&f(u_{i1}|x_i;\beta_{\phi 0}, \beta_{\phi x})\pi(\theta).
\end{aligned}$$

Therefore, sampling values of $\theta$ from (5.17) can be easily realized by sampling values for $(\beta_{\phi 0}, \beta_{\phi x}^{\mathrm{T}})^{\mathrm{T}}$, $(\beta_{\mu 0}, \beta_{\mu x}^{\mathrm{T}})^{\mathrm{T}}$, $(\alpha_{+0}, \alpha_{+w}^{\mathrm{T}})^{\mathrm{T}}$ and $(\alpha_{-0}, \alpha_{-w}^{\mathrm{T}})^{\mathrm{T}}$, separately from their posterior distribution $f(\beta_{\phi 0}, \beta_{\phi x}|u_{i1}, x_i)$, $f(\beta_{\mu 0}, \beta_{\mu x}|u_{i2}, x_i)$, $f(\alpha_{+0}, \alpha_{+w}|z_{i+}, x_i)$ and $f(\alpha_{-0}, \alpha_{-w}|z_{i-}, u_{i1}, u_{i2}, x_i)$.

### 5.2.2 Implementation: Monte Carlo Markov Chain Method with Data Augmentation

In this subsection, we describe the details of implementing the data augmentation idea described in Section 5.2.1 by using an MCMC algorithm, which is summarized as follows.

**Step 1: (Data Augmentation) Generate $U_{i1}$, $U_{i2}$, $Z_{i+}$ and $Z_{i-}$:** For $i = 1, \dots, n$ and given $Y_i^*$, $X_i$ and $\theta$, generate $U_{i1}$, $U_{i2}$, $Z_{i+}$ and $Z_{i-}$ jointly from the distribution (5.16), which can be realized by the inversion sampling algorithm described in Appendix D.3.

**Step 2: Update $\alpha_{+j}$ using (5.7):** For $j = 0, \dots, p_+$, let $\alpha_{+j}$ be the $j$th element of $\alpha_{+w}$, let $W_{i0+} = 1$ and let $W_{ij+}$ be the $j$th element of $W_{i+}$. Given $Z_{i+}$ obtained from Step 1, we generate $\alpha_{+j}$ from the posterior distribution

$$\prod_{i=1}^{n} f(\alpha_{+j}|z_{i+}) \propto \prod_{i=1}^{n} f(z_{i+}|\alpha_{+j})\pi(\alpha_+)$$
$$\propto \exp\left(\sum_{i=1}^{n} z_{i+}w_{ij+}\alpha_{+j}\right) \exp\left\{-\sum_{i=1}^{n} \exp\left(\alpha_+^{\mathrm{T}} w_{i+}\right)\right\} \pi(\alpha_+), \quad (5.18)$$

where the second step comes from (5.7), and $\pi(\alpha_+)$ is the prior distribution of $\alpha_+$ which may, for example, take a log-Gamma or a normal distribution with hyperparameters whose values are specified. When $W_{ij+}$ is binary, then we take $\pi(\alpha_+)$ to be a conjugate log-Gamma$(c, d)$ prior with $\pi(\alpha_{+j}) \propto \exp\{-c\exp(\alpha_{+j})\}\exp(d\alpha_{+j})$, so that the posterior distribution (5.18) becomes

$$\prod_{i=1}^{n} f(\alpha_{+j}|z_{i+})$$
$$\propto \exp\left(\sum_{i=1}^{n} z_{i+}w_{ij+}\alpha_{+j}\right) \exp\left\{-\sum_{i=1}^{n} \exp(\alpha_+^{\mathrm{T}} w_{i+})\right\} \exp\{-c\exp(\alpha_{+j})\}\exp(d\alpha_{+j})$$
$$= \exp\left\{\left(d + \sum_{i=1}^{n} z_{i+}w_{ij+}\right)\alpha_{+j}\right\} \exp\left[-\left\{c + \sum_{i=1}^{n} \exp\left(\sum_{j^* \neq j} w_{ij^*_+}\alpha_{+j^*}\right) w_{ij+}\right\} \exp(\alpha_{+j})\right],$$

which is the log-Gamma distribution, log-Gamma$(c + \sum_{i=1}^{n} \exp(\sum_{j^* \neq j} W_{ij^*_+}\alpha_{+j^*})W_{ij+}$, $d + \sum_{i=1}^{n} z_{i+}w_{ij+})$.

**Step 3: Update $\alpha_{-j}$:** For $j = 0, \ldots, p_-$, let $\alpha_{-j}$ be the $j$th element of $\alpha_{-w}$, let $W_{i0-} = 1$, and let that $W_{ij-}$ is the $j$th element of $W_{i-}$. We generate $\alpha_{-j}$ from the posterior distribution

$$\prod_{i=1}^{n} f(\alpha_{-j}|y_i, z_{i-}) \propto \prod_{i=1}^{n} f(y_i, z_{i-}|\alpha_{-j})\pi(\alpha_{-j})$$
$$\propto \prod_{i=1}^{n} \binom{y_i}{z_{i-}} \left\{g^{-1}(\alpha_-^{\mathrm{T}} w_{i-})\right\}^{z_{i-}} \left\{1 - g^{-1}(\alpha_-^{\mathrm{T}} w_{i-})\right\}^{y_i - z_{i-}} \pi(\alpha_{-j}),$$

where the second step comes from (5.8) and $\pi(\alpha_{-j})$ the probability density function of the prior of $\alpha_{-j}$, which can be taken as a normal distribution.

**Step 4: Update $\beta$:** Since both $U_{i1}$ and $U_{i2}$ follow a conditional Poisson distribution, given $X_i$, we update them in the same way as in Step 2. Let $\beta_{\phi j}$ and $\beta_{\mu j}$ respectively be the $j$th element of $\beta_\phi$ and $\beta_\mu$. Let $X_{i0} = 1$ and $X_{ij}$ be the $j$th element of $X_i$. For $j = 0, \ldots, p_x$, update $\beta_{\phi j}$ by sampling it from

$$\prod_{i=1}^{n} f(\beta_{\phi j}|u_{i1}) \propto \exp\left(\sum_{i=1}^{n} u_{i1} x_{ij} \beta_{\phi j}\right) \exp\left\{-\sum_{i=1}^{n} \exp(\beta_\phi^{\mathrm{T}} x_i)\right\} \pi(\beta_\phi),$$

and update $\beta_{\mu j}$ by sampling it from

$$\prod_{i=1}^{n} f(\beta_{\mu j}|u_{i2}) \propto \exp\left(\sum_{i=1}^{n} u_{i2} x_{ij} \beta_{\mu j}\right) \exp\left\{-\sum_{i=1}^{n} \exp(\beta_\mu^{\mathrm{T}} x_i)\right\} \pi(\beta_\mu).$$

where $\pi(\beta_\phi)$ and $\pi(\beta_\mu)$ are prior distributions for $\beta_\phi$ and $\beta_\mu$, respectively. For instance, if the covariates $X_i$ are binary, we may take the conjugate log-Gamma prior for $\beta_\phi$ and $\beta_\mu$.

## 5.3 Extension to the Main/Validation Studies

The Bayesian inference circumvents the traditional identifiability issue in the frequentist framework (e.g. Gelman et al., 2013, Page 412) by using weakly informative priors. In some applications, however, even weakly informative priors are not available or cannot be precisely set. In this circumstance, the study design can provide extra information regarding the measurement error process through validation data.

Let $\mathcal{M}$ denote the index set of the subjects in the main study, where $\{(y_i^*, x_i) : i \in \mathcal{M}\}$ is available. Let $\mathcal{V}$ represent the index set of the subjects in the validation data. For internal validation, the validation data contain $\{(y_i^*, y_i, x_i) : i \in \mathcal{V}\}$ with $\mathcal{V} \subset \mathcal{M}$; for external validation, the validation data contain $\{(y_i^*, y_i, w_{i+}, w_{i-}) : i \in \mathcal{V}\}$ with $\mathcal{M} \cap \mathcal{V} = \emptyset$. Let $m$ denote the size of the validation subsample $\mathcal{V}$ and let $n$ be the size of $\mathcal{M}$ as used in Sections 5.1–5.2.

### 5.3.1 Main/External Validation Study

With external validation data, we write the posterior function of $\theta$ combining the main and validation data:

$$\begin{aligned} &\left\{\prod_{i\in\mathcal{M}} f(\theta|y_i^*, x_i)\right\}\left\{\prod_{i\in\mathcal{V}} f(\theta|y_i^*, y_i, w_{i+}, w_{i-})\right\} \\ \propto\ &\pi(\theta)\left\{\prod_{i\in\mathcal{M}} f(y_i^*|x_i;\theta)\right\}\left\{\prod_{i\in\mathcal{V}} f(y_i^*|y_i, w_{i+}, w_{i-};\alpha)\right\}, \end{aligned} \tag{5.19}$$

where $f(y_i^*|x_i;\theta)$ comes from (5.9), $f(y_i^*|y_i, w_{i+}, w_{i-};\alpha)$ is modeled by (5.5), and $\pi(\theta)$ is a prior function.

Similar to the development of Section 5.2, instead of directly using (5.19) for sampling values of the parameters, we apply the following sampling procedures:

**Step 1: (Data Augmentation) Generate $U_{i1}$, $U_{i2}$, $Z_{i+}$ and $Z_{i-}$.** For $i \in \mathcal{M}$, we generate augmented data in the same way as in Section 5.2.2. For $i \in \mathcal{V}$, we generate $Z_{i+}$ and $Z_{i-}$ from their joint posterior distribution

$$f(z_{i+}, z_{i-}|Y_i^* = y_i^*, Y_i = y_i;\alpha,\beta) = f\left(Z_{i+} = y_i^* - y_i + z_{i-}, Z_{i-} = z_{i-};\alpha,\beta\right), \tag{5.20}$$

which is determined by (5.5).

**Steps 2-3: Update $\alpha_{+j}$ and $\alpha_{-j}$.** These two steps are similar to Steps 2–3 in Section 5.2.2 except for replacing the summation $\sum_{i=1}^n$ with $\sum_{i\in\mathcal{M}\cup\mathcal{V}}$. For example, we update $\alpha_{+j}$ by sampling it from the posterior

$$\prod_{i\in\mathcal{M}\cup\mathcal{V}} f(\alpha_{+j}|z_{i+}) \propto \exp\left(\sum_{i\in\mathcal{M}\cup\mathcal{V}} z_{i+}w_{ij+}\alpha_{+j}\right)\exp\left\{-\sum_{i\in\mathcal{M}\cup\mathcal{V}}\exp(\alpha_+^{\mathrm{T}}w_{i+})\right\}\pi(\alpha_+),$$

**Step 4:** This is identical to Step 4 in Section 5.2.2.

### 5.3.2 Main/Internal Validation Study

When internal validation data is available, the posterior function of parameter $\theta$ now becomes

$$\begin{aligned} &\left\{\prod_{i\in\mathcal{M}\setminus\mathcal{V}} f(\theta|y_i^*, x_i)\right\}\left\{\prod_{i\in\mathcal{V}} f(\theta|y_i^*, y_i, x_i)\right\} \\ &\propto \pi(\theta)\left\{\prod_{i\in\mathcal{M}\setminus\mathcal{V}} f(y_i^*|x_i;\theta)\right\}\left[\prod_{i\in\mathcal{V}}\{f(y_i^*|y_i;\alpha)f(y_i|x_i;\beta)\}\right], \end{aligned} \tag{5.21}$$

where $f(y_i^*|x_i;\theta)$ comes from (5.9), $f(y_i|x_i;\beta)$ is from (5.2), $f(y_i^*|y_i;\alpha)$ is from (5.5), and $\pi(\theta)$ is the prior function of parameters $\theta$. Similar to the development of Section 5.2, instead of directly using (5.21) for sampling values of the parameters, we apply the following sampling procedures:

**Step 1: (Data Augmentation) Generate $U_{i1}$, $U_{i2}$, $Z_{i+}$ and $Z_{i-}$.** For $i \in \mathcal{M} \setminus \mathcal{V}$, we generate the augmented data in a way similar to that in Section 5.2.2. For $i \in \mathcal{V}$, we generate the variables in the following steps iteratively.

1. We update the latent data augmentation variable $U_{i1}$ according to the value of $U_{i2}$ and $Y_i$, which includes three circumstances according to (5.15):
   - (i). Case 1 with $Y_i = 0$ and $U_{i2} = 0$: By (5.15), $Y_i = 0$ if and only if $U_{i1} = 0$ or $U_{i2} = 0$. That is, there is no restrictions on the $U_{i1}$ when $U_{i2} = 0$, so in this case $U_{i1}$ is generated from Poisson($\mu_{i1}$).
   - (ii). Case 2 with $Y_i = 0$ and $U_{i2} \neq 0$: (5.15) says that $U_{i1}$ is be surely be 0. Hence, we set $U_{i1} = 0$.
   - (iii). Case 3 with $Y_i > 0$: (5.15) implies that $U_{i1} > 0$. Hence, we update $U_{i1}$ by a truncated Poisson($\mu_{i1}$) at 0.
2. Given $U_{i1}$ obtained in part 1, we update the latent variable $U_{i2}$, which includes two cases:
   - (i). Case 1 with $U_{i1} > 0$: we generate $U_{i2}$ by setting it equal to $Y_{i1}$ by (5.15).
   - (ii). Case 2 with $U_{i1} = 0$: (5.15) shows that $Y_{ij} = 0$, and thus, there are no constraints on the variable $U_{i2}$ and we generate $U_{i2}$ from Poisson($\mu_i$).
3. Generate $Z_{i+}$ and $Z_{i-}$ in the same way as in Step 1 in Section 5.3.1.

**Steps 2–4:** The steps are the same as Steps 2–4 in Section 5.2.2.

## 5.4 Application to Prostate Adenocarcinoma Genomics Data

### 5.4.1 Study Background

Here we apply the proposed methods to a multi-center molecular prostate cancer study. We are interested in predicting whether or not cancer-related pathways are activated during the prostate cancer progression and how the number of genes with copy number variations (CNVs) within each pathway is associated with the risk factors. The data contain two datasets that are linked by the genes in The Cancer Genome Atlas (TCGA) data that are annotated in the Kyoto Encyclopedia of Genes and Genomes (KEGG) pathways data through website cBioPortal. The first part includes the pathway information arising from the KEGG pathways data, and the second part is the putative CNV data with 465 subjects collected from two sources with 185 subjects from Broad Institute (Banerji et al., 2012) and 280 subjects from Memorial Sloan-Kettering Cancer Center (MSKCC) (Taylor et al., 2010) for prostate adenocarcinoma.

In this analysis, similar to Neelon and Chung (2017), we consider four pathways: mitogen-activated protein kinase (MAPK) signaling, cytokine-cytokine receptor (CCR) interaction, endocytosis (EC), and P53. Genes in the MAPK pathway are related to various cellular functions, such as cell proliferation, differentiation, and migration; genes in the CCR interaction pathway are associated with inflammatory host defenses, cell growth, differentiation and death, and the restoration of homeostasis; the genes in the EC pathway are related to the mechanisms of cells transporting ligands, nutrients, proteins, and lipids from the cell surface to the cell interior; and the p53 pathway is induced by a number of stress signals, including DNA damage, oxidative stress, and activated oncogenes (Alberts et al., 2002).

In our study here, we conduct four marginal analysis separately for each pathway, where the response for each individual ($Y_i$) is defined as the count of genes with significant CNVs (with reading valued as either $-2$ or 2), which reflects the level of mutation in the individual. We implement the Vuong tests (Vuong, 1989) to assess whether or not zero-inflation exists in the response. With a p-value smaller than 0.001 for all four pathways, the test result shows a strong sign of zero inflation. We investigate two different risk factors that may be associated with the CNVs counts in two separate studies, which are reported in Sections 5.4.2 and 5.4.3, respectively. In the first study, the covariate is denoted as $X_{i1}$, which is taken as the tumor stage, which is given by an indicator variable, taking value 0 or 1, according to the T2 or T3+ tumor stage for subject $i$; and in the second study, the covariate is denoted as $X_{i2}$, which represents the cancer recurrence, with $X_i = 1$ if cancer recurrence occurs and $X_i = 0$ otherwise.

We are interested in understanding the relationship between $Y_i$ and a covariate in each study. However, due to the potential sequencing error, the CNV reading of insignificant gene can be falsely measured as significant, whereas the gene with significant CNVs can be missed to be counted, and thus, the observed count number (denoted as $Y_i^*$) may considerably differ from the true value of $Y_i$. To feature this difference, we consider the measurement error model (5.5) with $c_+ = c_- = 1$.

### 5.4.2 Association of Tumor Stage and CNVs

We conduct analysis for each pathway separately using the zero-inflated model (5.2), (5.3) and (5.4) to feature the dependence of $Y_i$ on the covariate $X_{i1}$. The dataset is combined from multiple sources, and the data quality and genetic sequencing protocols can be different. Thus, in this study, we perceive that the measurement error process is associated with the data source and use the measurement error models (5.7) and (5.8), where the covariate $W_i$ is a binary indicator for the data source, with $W_i = 1$ if the subject $i$ is from the broad institute, and 0 otherwise.

In implementing the Bayesian procedures described in Section 5.2, we consider an uninformative prior, log-Gamma$(1000, 0.001)$, for the parameters of models (5.3), (5.4) and (5.7). For the parameter $\alpha_{-0}$ in the model (5.8), we consider the prior, Normal$(-2, 10)$, where the negative mean reflects our expectation of a negative value for $\alpha_{-0}$, and a large variance shows a flat prior. We use the Gelman-Robin method (Gelman et al., 1992) to diagnose the convergence of Monte Carlo Markov chains, and the results show that MCMC series for all the parameters well converge after running 250,000 iterations of sampling steps and discarding the first 5000 as burn-in.

Our first interest lies in whether CNVs counts change as the tumor progresses for patients who have activated the pathway. We implement the proposed method described in Section 5.2.1, and for comparison, we also implement the naive method based on Neelon and Chung (2017) where the difference of $Y_i$ and $Y_i^*$ is neglected. The analysis results of parameter estimation are presented in Figure 5.1. Both the naive method and the proposed method find that the $\beta_{\mu x}$ for all pathways are not significantly different from zero, suggesting that patients in tumor stage T3+ do not have different mutations than those in T2 stage. The estimates of the intercept of the count model (5.4), $\beta_{\mu 0}$, for the MAPK, CCR and EC pathways are higher than those of $\beta_{\mu 0}$ in the P53 pathway, showing that prostate cancer patients have more mutations in the genes involved in the pathways of MAPK, CCR, and EC than the P53 pathway.

Our second interest is the probability of activation for patients as the tumor grows, which is reflected by the estimation of parameters associated with $\phi_i$. It is clear that the proposed method with the measurement error effects accounted for yields results different from the naive method. In Figure 5.2, we report credible intervals for the probability of pathway activation using (5.3) together with estimates of $\beta_{\phi 0}$ and $\beta_{\phi x}$. The proposed method indicates that the difference in the probability of pathway activation is close to zero for patients in tumor stage T2 versus the patients in T3+. However, the naive method suggests that the difference is very large.

### 5.4.3 Association of Cancer Recurrence and CNVs

We consider the endpoint to be, alternatively, the recurrence status of a prostate cancer patient after being cured. We are interested in: 1) whether the status of cancer recurrence is associated with activation of the pathway; and 2) for the subjects who have activated the pathway, whether the status of cancer recurrence is associated with the CNVs counts. We conduct an analysis for each pathway separately using (5.2), (5.3) and (5.4) to feature the dependence of $Y_i$ on the covariate $X_{i2}$. Since the covariate of cancer recurrence information is only available in the MSKCC study, we focus on the analysis of 280 subjects in the MSKCC study and consider measurement error models (5.7) and (5.8) with constant parameters $\alpha_{0+}$ and $\alpha_{0-}$ only, where $W_i$ is no longer included in the models.

In implementing the Bayesian procedures described in Section 5.2, we consider the same priors for parameters as in Section 5.4.2. We run 250,000 iterations of the sampling steps and discard the first 5000 as burn-in. The resultant Monte Carlo Markov chains converge according to the Gelman-Robin method (Gelman et al., 1992).

The results are exhibited in Figures 5.3–5.4. First, using the proposed methods in Section 5.2 in contrast to the naive method as described in Neelon and Chung (2017), we study the association between the status of recurrence of prostate cancer and the number of CNVs for patients with activated pathways. In Figure 5.3, the proposed method suggests that for all the pathways, under the significance level of 0.05, the number of CNVs is not significantly associated with a higher risk of cancer recurrence, where the estimate of $\beta_{\mu x}$ is, respectively, 0.007 and the credible interval $(-0.393, 0.401)$ for MAPK, 0.294 and the credible interval $(-0.098, 0.614)$ for CCR, 0.144 and the credible interval $(-0.346, 0.655)$ for EC, and 0.001 and the credible interval $(-1.152, 1.116)$ for P53. On the other hand, the naive method shows that the patients with higher CNVs in the CCR pathway has a higher risk of recurrence, with the estimate of $\beta_{\mu x}$ being 0.454 and the credible interval $(0.106, 0.802)$.

Secondly, we study the association between the activation of pathways and the risk of cancer recurrence. We observe that the patients with the MAPK, CCR, or EC pathway activated tend to have a higher risk of the prostate cancer recurrence, where the estimate of $\beta_{\phi x}$ is, respectively, 0.906 and the credible interval $(0.061, 1.737)$, 1.021 and the credible interval $(0.145, 1.999)$, and 1.205 and the credible interval $(0.311, 2.103)$ for each pathway. On the other hand, the naive method indicates that the activation of the pathway is associated with a lower cancer risk because the estimates of $\beta_{\phi x}$ are negative. We estimate the probability of pathway activation using (5.3) and present the results in Figure 5.4. The proposed method generally suggests that the cancer patients have low probabilities of activation of the pathway, while the naive method indicates opposite findings.

## 5.5 Simulation Studies

In this section, we conduct simulation studies to evaluate the performance of the proposed method. For the sake of comparison, we also implement the naive method where no action is taken to deal with the measurement error in response.

We conduct two simulation studies. In the first simulation study, we evaluate the performance of the proposed method under different settings of the parameters, leading to different percentages of zeros in the responses. We conduct sensitivity analyses by exploring different settings of the prior distribution of the parameters. In the second simulation study, we evaluate the performance of the methods under different degrees of measurement error.

For both simulation studies, we run 1000 simulations for each setting. The sample size is taken as $n = 5000$, and we consider model (5.3) with covariates $X_{i1}$ and (5.4) with covariates $X_{i2}$, where covariate $X_{i1}$ is generated from Binomial(0.5) and $X_{i2}$ is independently generated from Uniform$[0, 1]$. The true response $Y_i$ is generated from (5.2), where $\phi_i = 1 - \exp\{-\exp(\beta_{\phi 0} + \beta_{\phi x} X_{i1})\}$ and $\mu_i = \beta_{\mu 0} + \beta_{\mu x} X_{i2}$.

To generate surrogate measurements $Y_i^*$ of $Y_i$, we consider the measurement error models (5.7) and (5.8) each associated with a covariate $W_{i1}$ and $W_{i2}$, respectively, where $W_{i1}$ is independently generated from $U[0, 1]$, and $W_{i2}$ is independently generated from $U[0, 2]$. Furthermore, we generate $Z_{i+}$ from exponential$(\alpha_{+0} + \alpha_{+w} W_{i1})$ and $Z_{i-}$ from Binomial$(Y_i, \frac{\exp(\alpha_{-0}+\alpha_{-w} W_{i2})}{1+\exp(\alpha_{-0}+\alpha_{-w} W_{i2})})$. As a result, $Y_i^*$ is determined by $Y_i^* = Y_i + Z_{i+} - Z_{i-}$.

To summarize the simulation results, we report biases (denoted "Bias") by calculating $\frac{1}{N} \sum_{i=1}^{N} \hat{\theta}_k - \theta_k$, model-based standard errors (denoted "SEM"), empirical standard error of the point estimates (denoted "SEE"), and the coverage rate (in percent) of 95% credible intervals for a parameter, say $\theta_k$, (denoted "CR"), defined as

$$\frac{1}{N}\sum_{i=1}^{N} I(\hat{\theta}_k^{(L)} < \theta_k < \hat{\theta}_k^{(U)}),$$

where $N$ is the number of simulations, $\hat{\theta}_k^{(L)}$ and $\hat{\theta}_k^{(U)}$ are respectively the 2.5% and 97.5% quantile for the sampled parameter values.

### 5.5.1 Simulation 1: Performance of the Proposed Method with Different Zero Percentages and Hyperparameters

Two parameter settings are considered. In Setting 1, we consider $(\beta_{\phi 0}, \beta_{\phi x}, \beta_{\mu 0}, \beta_{\mu x})^{\mathrm{T}} = (-0.7, 0.7, 1, -0.5)^{\mathrm{T}}$, yielding about 60% zeros; in Setting 2, we consider $(\beta_{\phi 0}, \beta_{\phi x}, \beta_{\mu 0}, \beta_{\mu x})^{\mathrm{T}} = (-0.2, 0.7, 1, 0.5)^{\mathrm{T}}$, yielding about 30% zeros. In both settings, we consider the parameters of the measurement error model (5.5) with $\alpha_{+0} = \alpha_{-0} = 0$, and $(\alpha_{+w}, \alpha_{-w})^{\mathrm{T}} = (0.5, -2.3)^{\mathrm{T}}$. For each setting, we study the sensitivity of results with respect to different priors when implementing the proposed method and the naive method which disregards the difference in $Y_i$ and $Y_i^*$. In the first set of priors, we consider uninformative priors log-Gamma$(1000, 0.001)$ for $\beta_{\phi 0}$, $\beta_{\phi x}$, $\beta_{\mu 0}$, $\beta_{\mu x}$, $\alpha_{+w}$, and Normal$(0, 1000^2)$ for $\alpha_{-w}$. In the second set of priors, we choose log-Gamma$(1, 1)$ for $\beta_{\phi 0}$, $\beta_{\phi x}$, $\beta_{\mu 0}$, $\beta_{\mu x}$, $\alpha_{+w}$, and Normal$(-2, 2^2)$ for $\alpha_{-w}$.

Table 5.1 shows that without accounting for measurement error in response, the naive model produces biased estimates of the parameters and meaningless coverage rates of 95% credible intervals. On the other hand, the proposed method considerably reduces the biases resulting from the measurement error effects and provides reasonable standard errors. The performance of the proposed method is satisfactory for different settings, regardless of the specification of the prior distribution.

### 5.5.2 Simulation 2: Performance of Method with Different Degrees of Measurement Error

In this subsection, we evaluate how the performance of the proposed method may be affected by different degrees of measurement error resulting from different parameters in the add-in process and the leave-out process. For the add-in process, we set the parameters

$(\alpha_{+0}, \alpha_{+w})^{\mathrm{T}}$ in (5.7) to be $(-1, 0.6)^{\mathrm{T}}$ or $(2, -1.2)^{\mathrm{T}}$, leading to the mean of $Z_{i+}$ to be 0.5 (small) or 5.0 (substantial). For the leave-out process, we take the parameters $(\alpha_{-0}, \alpha_{-w})^{\mathrm{T}}$ in (5.8) to be $(-1, -1.2)^{\mathrm{T}}$ or $(-0.8, -1)^{\mathrm{T}}$, respectively, leading to 5% (low) or 10% (high) counts being neglected.

To create a dataset for a main/external validation study, we randomly choose 2500 subjects and only keep the variables of $Y_i^*$, $Y_i$, $W_{i1}$ and $W_{i2}$. For the case of a main/internal validation study, we randomly select 2500 subjects and keep their variables of $Y_i^*$, $Y_i$, $X_{i1}$, $X_{i2}$, $W_{i1}$ and $W_{i2}$ to serve as the validation data.

We implement the methods described in Sections 5.2 and 5.3, as opposed to the naive method by replacing $Y_i$ with $Y_i^*$ in the analysis. When implementing the methods, we take log-Gamma$(1, 1)$ as the prior for $\beta_{\phi 0}$, $\beta_{\phi x}$, $\beta_{\mu 0}$, $\beta_{\mu x}$, $\alpha_{+0}$, and Normal$(-2, 2^2)$ for $\alpha_{-0}$ and $\alpha_{-w}$.

The results are displayed in Table 5.2. Our proposed methods outperform the naive method, regardless of the parameter settings for the measurement error model. The proposed methods yield small finite sample biases for the point estimates and reasonable coverage rates for 95% credible intervals.

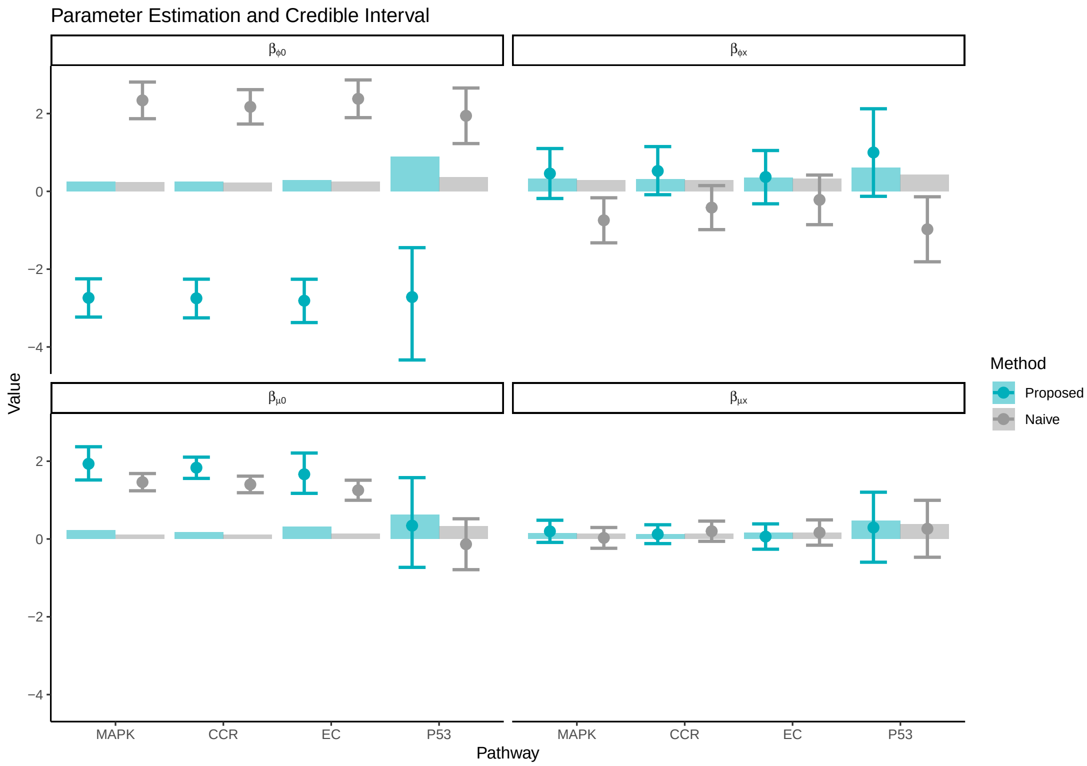


Figure 5.1: *The plot of parameter estimation of the zero-inflated Poisson models for the association between tumor stage and CNVs. The point indicates the point estimation and the line segments represent the 95% credible interval. The barplot is the associated standard error for the estimation of the corresponding parameter.*

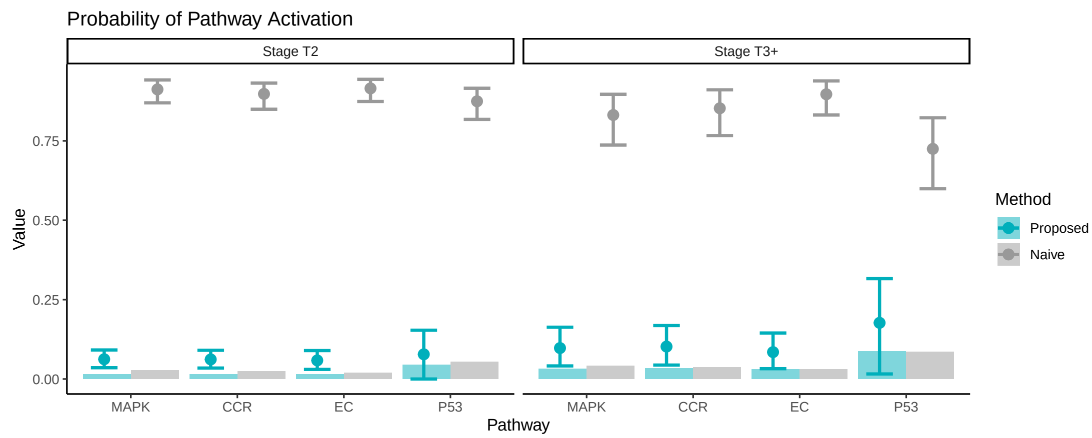


Figure 5.2: *The plot of probability of pathway activation in the study of the association between tumor stage and CNVs. The point indicates the point estimation and the line segments represent the 95% credible interval. The barplot is the associated standard error for the estimation of the corresponding parameter.*

Table 5.1: Results for Simulation 1 with different zero percentage and prior parameters

| | | Naive Mothod | | | | Proposed Method | | | |
|---|---|---|---|---|---|---|---|---|---|
| Parameter | Prior | Bias | SEE | SEM | CR% | Bias | SEE | SEM | CR% |
| | Setting 1: zero-percentage 60% | | | | | | | | |
| $\beta_{\phi 0}$ | Uninformative | 1.506 | 0.030 | 0.029 | 0.0 | 0.004 | 0.089 | 0.086 | 93.3 |
| $\beta_{\phi x}$ | | 0.340 | 0.051 | 0.049 | 0.0 | 0.004 | 0.082 | 0.079 | 94.1 |
| $\beta_{\mu 0}$ | | 0.122 | 0.020 | 0.017 | 0.0 | 0.030 | 0.069 | 0.068 | 91.9 |
| $\beta_{\mu x}$ | | 0.816 | 0.032 | 0.028 | 0.0 | 0.004 | 0.082 | 0.081 | 94.5 |
| $\alpha_{+w}$ | | - | - | - | - | 0.001 | 0.044 | 0.045 | 94.7 |
| $\alpha_{-w}$ | | - | - | - | - | 0.115 | 0.562 | 0.560 | 90.5 |
| $\beta_{\phi 0}$ | Informative | 1.506 | 0.030 | 0.029 | 0.0 | 0.007 | 0.088 | 0.085 | 93.4 |
| $\beta_{\phi x}$ | | 0.340 | 0.051 | 0.049 | 0.0 | 0.005 | 0.081 | 0.078 | 94.2 |
| $\beta_{\mu 0}$ | | 0.122 | 0.020 | 0.017 | 0.0 | 0.002 | 0.070 | 0.064 | 92.7 |
| $\beta_{\mu x}$ | | 0.817 | 0.032 | 0.028 | 0.0 | 0.004 | 0.082 | 0.080 | 94.3 |
| $\alpha_{+w}$ | | - | - | - | - | 0.005 | 0.044 | 0.045 | 95.2 |
| $\alpha_{-w}$ | | - | - | - | - | 0.243 | 0.733 | 0.648 | 91.5 |
| | Setting 2: zero-percentage 30% | | | | | | | | |
| $\beta_{\phi 0}$ | Uninformative | 1.001 | 0.030 | 0.029 | 0.0 | 0.001 | 0.049 | 0.048 | 93.9 |
| $\beta_{\phi x}$ | | 0.332 | 0.052 | 0.049 | 0.2 | 0.002 | 0.056 | 0.055 | 94.5 |
| $\beta_{\mu 0}$ | | 0.124 | 0.020 | 0.017 | 0.0 | 0.021 | 0.046 | 0.052 | 93.7 |
| $\beta_{\mu x}$ | | 0.185 | 0.033 | 0.028 | 0.0 | 0.004 | 0.047 | 0.047 | 94.5 |
| $\alpha_{+w}$ | | - | - | - | - | 0.004 | 0.053 | 0.053 | 94.7 |
| $\alpha_{-w}$ | | - | - | - | - | 0.100 | 0.370 | 0.430 | 94.0 |
| $\beta_{\phi 0}$ | Informative | 1.000 | 0.030 | 0.029 | 0.0 | 0.002 | 0.049 | 0.048 | 93.5 |
| $\beta_{\phi x}$ | | 0.332 | 0.051 | 0.049 | 0.0 | 0.003 | 0.056 | 0.055 | 95.7 |
| $\beta_{\mu 0}$ | | 0.124 | 0.020 | 0.017 | 0.0 | 0.012 | 0.043 | 0.052 | 96.6 |
| $\beta_{\mu x}$ | | 0.184 | 0.033 | 0.028 | 0.0 | 0.003 | 0.047 | 0.047 | 94.2 |
| $\alpha_{+w}$ | | - | - | - | - | 0.006 | 0.053 | 0.053 | 94.5 |
| $\alpha_{-w}$ | | - | - | - | - | 0.017 | 0.389 | 0.456 | 96.0 |

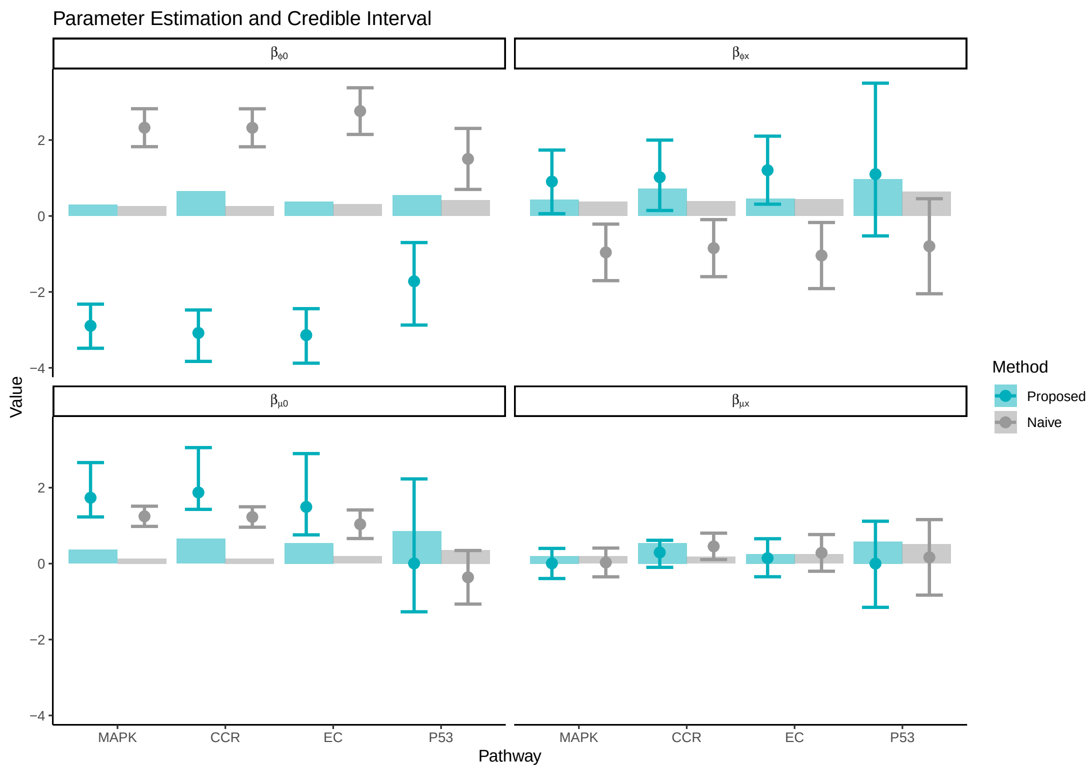


Figure 5.3: *The plot of parameter estimation of the zero-inflated Poisson models for the association between cancer recurrence and CNVs. The point indicates the point estimation and the line segments represent the 95% credible interval. The barplot is the associated standard error for the estimation of the corresponding parameter.*

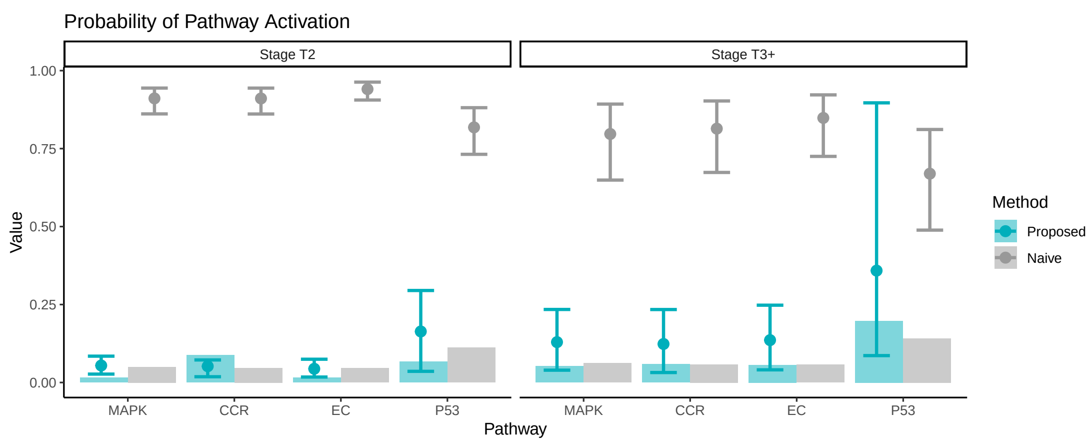


Figure 5.4: *The plot of probability of pathway activation in the study of the association between cancer recurrence and CNVs. The point indicates the point estimation and the line segments represent the 95% credible interval. The barplot is the associated standard error for the estimation of the corresponding parameter.*

Table 5.2: Results for Simulation 2 with different degrees of measurement error

| | | | | | Naive Method | | | | Proposed Method | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | | | | | | | No Validation | | | | External Validation | | | | Internal Validation | | | |
| Parameter | $\alpha_{+0}$ | $\alpha_{+w}$ | $\alpha_{-0}$ | $\alpha_{-w}$ | Bias | SEE | SEM | CR% | Bias | SEE | SEM | CR% | Bias | SEE | SEM | CR% | Bias | SEE | SEM | CR% |
| $\beta_{\phi 0}$ | -1.0 | 0.6 | -1.0 | -1.2 | 0.523 | 0.027 | 0.027 | 0.0 | 0.005 | 0.043 | 0.043 | 93.9 | 0.001 | 0.051 | 0.051 | 93.8 | 0.001 | 0.030 | 0.031 | 96.0 |
| $\beta_{\phi x}$ | | | | | 0.201 | 0.041 | 0.041 | 0.2 | 0.003 | 0.047 | 0.048 | 95.2 | 0.009 | 0.064 | 0.067 | 94.9 | 0.005 | 0.039 | 0.043 | 96.0 |
| $\beta_{\mu 0}$ | | | | | 0.054 | 0.022 | 0.020 | 23.3 | 0.028 | 0.033 | 0.041 | 94.4 | 0.003 | 0.036 | 0.037 | 95.6 | 0.000 | 0.023 | 0.023 | 94.2 |
| $\beta_{\mu x}$ | | | | | 0.081 | 0.036 | 0.032 | 30.4 | 0.005 | 0.040 | 0.041 | 95.2 | 0.010 | 0.055 | 0.057 | 95.6 | 0.002 | 0.036 | 0.036 | 94.3 |
| $\alpha_{+0}$ | | | | | - | - | - | - | 0.016 | 0.114 | 0.117 | 94.7 | 0.002 | 0.058 | 0.062 | 95.7 | 0.002 | 0.057 | 0.062 | 95.4 |
| $\alpha_{+w}$ | | | | | - | - | - | - | 0.011 | 0.157 | 0.161 | 94.5 | 0.004 | 0.092 | 0.097 | 95.9 | 0.004 | 0.093 | 0.097 | 95.7 |
| $\alpha_{-0}$ | | | | | - | - | - | - | 0.087 | 0.178 | 0.191 | 95.6 | 0.004 | 0.079 | 0.079 | 95.1 | 0.004 | 0.078 | 0.079 | 94.2 |
| $\alpha_{-w}$ | | | | | - | - | - | - | 0.466 | 0.292 | 0.481 | 94.9 | 0.005 | 0.091 | 0.092 | 95.1 | 0.005 | 0.092 | 0.091 | 95.3 |
| $\beta_{\phi 0}$ | -1.0 | 0.6 | -0.8 | -1.0 | 0.528 | 0.027 | 0.027 | 0.0 | 0.005 | 0.044 | 0.044 | 94.0 | 0.002 | 0.052 | 0.052 | 94.1 | 0.001 | 0.030 | 0.032 | 95.9 |
| $\beta_{\phi x}$ | | | | | 0.200 | 0.041 | 0.042 | 0.2 | 0.003 | 0.049 | 0.048 | 94.6 | 0.010 | 0.064 | 0.068 | 95.3 | 0.005 | 0.039 | 0.043 | 96.8 |
| $\beta_{\mu 0}$ | | | | | 0.091 | 0.022 | 0.020 | 0.5 | 0.049 | 0.036 | 0.045 | 86.9 | 0.002 | 0.037 | 0.039 | 95.3 | 0.000 | 0.023 | 0.023 | 94.7 |
| $\beta_{\mu x}$ | | | | | 0.084 | 0.037 | 0.033 | 29.2 | 0.005 | 0.042 | 0.042 | 95.0 | 0.009 | 0.058 | 0.059 | 94.4 | 0.002 | 0.037 | 0.037 | 94.1 |
| $\alpha_{+0}$ | | | | | - | - | - | - | 0.012 | 0.120 | 0.120 | 93.4 | 0.000 | 0.060 | 0.063 | 95.9 | 0.000 | 0.060 | 0.063 | 95.6 |
| $\alpha_{+w}$ | | | | | - | - | - | - | 0.010 | 0.161 | 0.161 | 94.0 | 0.002 | 0.096 | 0.098 | 95.8 | 0.002 | 0.096 | 0.099 | 96.0 |
| $\alpha_{-0}$ | | | | | - | - | - | - | 0.126 | 0.158 | 0.175 | 91.9 | 0.004 | 0.070 | 0.072 | 94.6 | 0.002 | 0.070 | 0.071 | 94.6 |
| $\alpha_{-w}$ | | | | | - | - | - | - | 0.486 | 0.259 | 0.414 | 88.8 | 0.004 | 0.074 | 0.076 | 95.3 | 0.004 | 0.074 | 0.076 | 95.1 |
| $\beta_{\phi 0}$ | 2.0 | -1.2 | -1.0 | -1.2 | 1.676 | 0.044 | 0.044 | 0.0 | 0.054 | 0.090 | 0.097 | 94.1 | 0.007 | 0.081 | 0.084 | 94.9 | 0.001 | 0.033 | 0.036 | 96.7 |
| $\beta_{\phi x}$ | | | | | 0.494 | 0.101 | 0.086 | 2.3 | 0.004 | 0.075 | 0.073 | 93.3 | 0.016 | 0.093 | 0.098 | 95.1 | 0.006 | 0.045 | 0.048 | 95.1 |
| $\beta_{\mu 0}$ | | | | | 0.785 | 0.014 | 0.011 | 0.0 | 0.053 | 0.056 | 0.072 | 97.3 | 0.001 | 0.055 | 0.059 | 95.9 | 0.001 | 0.026 | 0.026 | 95.1 |
| $\beta_{\mu x}$ | | | | | 0.334 | 0.025 | 0.019 | 0.0 | 0.001 | 0.064 | 0.068 | 96.3 | 0.021 | 0.078 | 0.083 | 95.5 | 0.003 | 0.041 | 0.041 | 95.1 |
| $\alpha_{+0}$ | | | | | - | - | - | - | 0.010 | 0.021 | 0.022 | 93.2 | 0.001 | 0.013 | 0.014 | 96.4 | 0.000 | 0.013 | 0.014 | 96.5 |
| $\alpha_{+w}$ | | | | | - | - | - | - | 0.021 | 0.044 | 0.047 | 93.7 | 0.001 | 0.028 | 0.030 | 95.6 | 0.001 | 0.028 | 0.029 | 95.1 |
| $\alpha_{-0}$ | | | | | - | - | - | - | 0.110 | 0.280 | 0.309 | 96.4 | 0.008 | 0.149 | 0.151 | 94.7 | 0.006 | 0.149 | 0.149 | 95.3 |
| $\alpha_{-w}$ | | | | | - | - | - | - | 0.699 | 0.458 | 0.740 | 96.4 | 0.046 | 0.207 | 0.208 | 93.3 | 0.038 | 0.206 | 0.205 | 94.6 |
| $\beta_{\phi 0}$ | 2.0 | -1.2 | -0.8 | -1.0 | 1.671 | 0.043 | 0.044 | 0.0 | 0.056 | 0.096 | 0.104 | 94.9 | 0.011 | 0.083 | 0.087 | 95.7 | 0.001 | 0.034 | 0.036 | 96.2 |
| $\beta_{\phi x}$ | | | | | 0.490 | 0.100 | 0.085 | 2.8 | 0.004 | 0.077 | 0.076 | 94.4 | 0.018 | 0.096 | 0.101 | 95.1 | 0.006 | 0.046 | 0.048 | 95.4 |
| $\beta_{\mu 0}$ | | | | | 0.772 | 0.014 | 0.011 | 0.0 | 0.076 | 0.058 | 0.079 | 95.9 | 0.000 | 0.056 | 0.062 | 96.3 | 0.001 | 0.026 | 0.026 | 94.6 |
| $\beta_{\mu x}$ | | | | | 0.338 | 0.025 | 0.019 | 0.0 | 0.002 | 0.067 | 0.071 | 96.2 | 0.020 | 0.082 | 0.086 | 95.0 | 0.003 | 0.042 | 0.041 | 95.2 |
| $\alpha_{+0}$ | | | | | - | - | - | - | 0.010 | 0.020 | 0.022 | 93.9 | 0.001 | 0.014 | 0.014 | 95.9 | 0.000 | 0.013 | 0.014 | 96.3 |
| $\alpha_{+w}$ | | | | | - | - | - | - | 0.020 | 0.044 | 0.048 | 93.8 | 0.000 | 0.028 | 0.030 | 95.6 | 0.001 | 0.028 | 0.030 | 95.5 |
| $\alpha_{-0}$ | | | | | - | - | - | - | 0.149 | 0.261 | 0.287 | 94.7 | 0.008 | 0.132 | 0.132 | 94.4 | 0.004 | 0.130 | 0.131 | 94.7 |
| $\alpha_{-w}$ | | | | | - | - | - | - | 0.690 | 0.367 | 0.608 | 92.8 | 0.026 | 0.155 | 0.157 | 94.5 | 0.022 | 0.155 | 0.155 | 94.2 |

# Chapter 6

# Autoregressive Models with Data Subject to Measurement Error

In this chapter, we discuss error-contaminated time series data. The notation and the setup for the autoregressive time series model and the proposed measurement error models are introduced in Section 6.1. In Section 6.2, we present the theoretical results for characterizing the impact of measurement error on the analysis of time series data. In Section 6.3, we develop an estimating equation approach to adjust for the biases due to measurement error. In Section 6.4, we implement the proposed method to analyze the COVID-19 data in four provinces in Canada.

## 6.1 Model Setup and Framework

### 6.1.1 Time Series Model

Consider a $T \times 1$ vector of time series, $X^{(T)} = (X_1, X_2, \ldots, X_T)^{\mathrm{T}}$. We are interested in modeling the dependence of $X_t$ on it previous observations $X^{(t-1)}$ and we consider it to be postulated by an autoregressive model with lag $p$

$$X_t = \phi_0 + \sum_{j=1}^{p} \phi_j X_{t-j} + \epsilon_t, \tag{6.1}$$

where $p$ is an integer smaller than $T$, $\epsilon^{(t)} = (\epsilon_1, \ldots, \epsilon_t)^{\mathrm{T}}$ is independent of $X^{(t)} = (X_1, \ldots, X_t)^{\mathrm{T}}$ with each $\epsilon_t$ having zero mean and variance $\sigma_\epsilon^2$, $\phi_0$ is a constant drift, and $\phi = (\phi_1, \ldots, \phi_p)^{\mathrm{T}}$ is the regression coefficient.

The additive form in (6.1) and the zero mean assumption of $\epsilon_t$ show that $\phi_0$ and $\phi$ are constrained by

$$\phi_0 = E(X_t) - \{E(\widetilde{X}_{t-1})\}^{\mathrm{T}}\phi, \tag{6.2}$$

where $\widetilde{X}_{t-1} = (X_{t-1}, \ldots, X_{t-p})^{\mathrm{T}}$. To make the process of $X_t$ stationary, $\phi_1, \ldots, \phi_p$ are further constrained such that all the roots of the equation in $z$

$$z^p - \phi_1 z^{p-1} - \cdots - \phi_p = 0$$

have absolute values smaller than 1 (Brockwell and Davis, 2002, Section 3.1.). For example, a stationary AR(1) process requires that $|\phi_1| < 1$, and a stationary AR(2) process needs that $(\phi_1 + \phi_2) < 1$, $(\phi_2 - \phi_1) < 1$ and $|\phi_2| < 1$. Here we are interested in the estimation of parameters, $\phi$ and $\phi_0$. Let $\mu$ denote the mean $E(X_t)$ of the time series, which equals $\frac{\phi_0}{1-\phi_1-\ldots-\phi_p}$ if $X_t$ is (weakly) stationary. When $p = 1$, the stationarity of a time series implies $\mathrm{Var}(X_t) = \frac{\sigma_\epsilon^2}{1-\phi_1^2}$ for $t = 1, \ldots, T$.

### 6.1.2 Estimation of Model Parameters

The estimation of the parameters in the AR($p$) time series model (6.1) can be carried out by the least squares method. To see this, we first focus on estimation of $\phi = (\phi_1, \ldots, \phi_p)^{\mathrm{T}}$. Let $S(\phi) = \sum_{t=p+1}^{T}\{X_t - (\phi_0 + \sum_{j=1}^{p}\phi_j X_{t-j})\}^2$ be the sum of the squared difference between $X_t$ and its linearly combined history with lag $p$. Then applying the constraint (6.2) gives $S(\phi) = \sum_{t=p+1}^{T}\left[\{X_t - E(X_t)\} - \{\widetilde{X}_{t-1} - E(\widetilde{X}_{t-1})\}^{\mathrm{T}}\phi\right]^2$.

To minimize $S(\phi)$ with respect to $\phi$, we solve $\frac{\partial S(\phi)}{\partial \phi} = 0$ for $\phi$ and obtain the solution

$$\widehat{\phi}^{(\mathrm{LS})} = \left(\sum_{t=p+1}^{T}\left\{\widetilde{X}_{t-1} - E(\widetilde{X}_{t-1})\right\}\left\{\widetilde{X}_{t-1} - E(\widetilde{X}_{t-1})\right\}^{\mathrm{T}}\right)^{-1}\sum_{t=p+1}^{T}\left\{\widetilde{X}_{t-1} - E(\widetilde{X}_{t-1})\right\}\{X_t - E(X_t)\}, \tag{6.3}$$

where for $t = 1, \ldots, T$, $E(X_t)$ can be estimated by $\frac{1}{T}\sum_{t=1}^{T} X_t$, which is denoted as $\widehat{\mu}$.

Next, by the constraint (6.2), replacing $E(X_t)$ by $\widehat{\mu}$ gives an estimator of $\phi_0$:

$$\widehat{\phi}_0^{(\mathrm{LS})} = \widehat{\mu} - \widehat{\mu} \cdot \sum_{j=1}^{p}\widehat{\phi}_j. \tag{6.4}$$

Re-expressing (6.1) as $\epsilon_t = X_t - (\phi_0 + \sum_{j=1}^{p} \phi_j X_{t-j})$ and by the definition of $S(\phi)$, we may estimate $\text{Var}(\epsilon_t) = \sigma_\epsilon^2$ by

$$\begin{aligned}
\widehat{\sigma}_\epsilon^{2(\text{LS})} &= \frac{1}{T-p} S(\widehat{\phi}) \\
&= \frac{1}{T-p} \sum_{t=p+1}^{T} \{X_t - E(X_t)\}^2 - \frac{2}{T-p} \sum_{t=p+1}^{T} \{X_t - E(X_t)\}\{\widetilde{X}_{t-1} - E(\widetilde{X}_{t-1})\}^{\mathrm{T}} \widehat{\phi} \\
&\quad + \frac{1}{T-p} \sum_{t=p+1}^{T} \widehat{\phi}^{\mathrm{T}} \{\widetilde{X}_{t-1} - E(\widetilde{X}_{t-1})\}\{\widetilde{X}_{t-1} - E(\widetilde{X}_{t-1})\}^{\mathrm{T}} \widehat{\phi} \qquad (6.5)
\end{aligned}$$

with $E(X_t)$ estimated by $\widehat{\mu}$.

Estimators (6.3)–(6.5) can be derived in an alternative way. First, by the stationarity of the $X_t$, for $k = 0, \ldots, p$ and $p \leq t$, $\text{Cov}(X_t, X_{t-k})$ is time-independent and let $\gamma_k$ denote it; it is clear that $\gamma_0$ represents $\text{Var}(X_t)$ for any $t$. Let $\Gamma$ be the autocovariance matrix

$$\Gamma = \begin{pmatrix} \gamma_0 & \cdots & \gamma_{p-1} \\ \vdots & \ddots & \vdots \\ \gamma_{p-1} & \cdots & \gamma_0 \end{pmatrix}.$$

Let $\widehat{\gamma} = (\widehat{\gamma}_1, \cdots, \widehat{\gamma}_p)^{\mathrm{T}}$ with $\widehat{\gamma}_k = \frac{1}{T-k} \sum_{t=k+1}^{T} (X_t - \widehat{\mu})(X_{t-k} - \widehat{\mu})$ being an estimator of $\gamma_k$ for $k = 0, \ldots, p$, and let $\widehat{\Gamma}$ be the estimator of $\Gamma$ with $\gamma_k$ replaced by $\widehat{\gamma}_k$ for $k = 0, \ldots, p-1$.

Next, we examine the summation terms in (6.3) and (6.5) by using the fact that as $T \to \infty$, $\frac{1}{T-p} \sum_{t=p+1}^{T} \{X_t - E(X_t)\}^2 \xrightarrow{p} \gamma_0$, $\frac{1}{T-p} \sum_{t=p+1}^{T} \{X_t - E(X_t)\}\{\widetilde{X}_{t-1} - E(\widetilde{X}_{t-1})\}^{\mathrm{T}} \xrightarrow{p} \gamma$, and $\frac{1}{T-p} \sum_{t=p+1}^{T} \{\widetilde{X}_{t-1} - E(\widetilde{X}_{t-1})\}\{\widetilde{X}_{t-1} - E(\widetilde{X}_{t-1})\}^{\mathrm{T}} \xrightarrow{p} \Gamma$. Then, (6.3)–(6.5) motivate an alternative method of finding estimators for $\phi$, $\phi_0$, and $\sigma_\epsilon^2$, by solving the estimating equations:

$$\begin{aligned}
\phi &= \widehat{\Gamma}^{-1} \widehat{\gamma}; \\
\phi_0 &= \left(1 - \sum_{i=1}^{p} \phi_i\right) \widehat{\mu}; \qquad (6.6) \\
\sigma_\epsilon^2 &= \widehat{\gamma}_0 - 2\phi^{\mathrm{T}} \widehat{\gamma} + \phi^{\mathrm{T}} \widehat{\Gamma} \phi,
\end{aligned}$$

for $\phi$, $\phi_0$, and $\sigma_\epsilon^2$. Let $\widehat{\phi}$, $\widehat{\phi}_0$ and $\widehat{\sigma}_\epsilon^2$ denote the resultant estimators of $\phi$, $\phi_0$, and $\sigma_\epsilon^2$, respectively. These estimators are asymptotically equivalent to the least squares estimators

$\widehat{\phi}^{(\text{LS})}$, $\widehat{\phi}_0^{(\text{LS})}$, and $\widehat{\sigma}_\epsilon^{2(\text{LS})}$ in a sense that $\widehat{\phi} - \widehat{\phi}^{(\text{LS})} \xrightarrow{p} 0$, $\widehat{\phi}_0 - \widehat{\phi}_0^{(\text{LS})} \xrightarrow{p} 0$ and $\widehat{\sigma}_\epsilon^2 - \widehat{\sigma}_\epsilon^{2(\text{LS})} \xrightarrow{p} 0$, as $T \to \infty$, and hence, they are consistent (Box et al., 2015, Section A.7.4).

Estimating equations (6.6) offer a unified estimation framework in its connections with not only the least squares estimation but also the maximum likelihood method under the assumption of Gaussian error as well as the Yule-Walker method. Similar to the least squares method, finding estimators using one of those approaches is asymptotically equivalent to solving (6.6) for $\phi$, $\phi_0$ and $\sigma_\epsilon^2$ (Box et al., 2015, Section A.7.4).

## 6.2 Measurement Error and Impact

### 6.2.1 Measurement Error Models

Suppose that for $t = 1, \ldots, T$, the observation of $X_t$ is subject to measurement error and the precise measurement of $X_t$ may not be observed, but its surrogate measurement $X_t^*$ is available. We consider two measurement error models.

The first measurement error model takes an additive form

$$X_t^* = \alpha_0 + \alpha_1 X_t + e_t \tag{6.7}$$

for $t = 1, \ldots, T$, where the error term $e_t$ is independent of $X_t$ with mean 0 and time-independent variance $\sigma_e^2$ and is assumed to be mutually independent for $t = 1, \ldots, T$, and $\alpha = (\alpha_0, \alpha_1)^{\text{T}}$ is the parameter vector. Here, $\alpha_0$ represents the systematic error and $\alpha_1$ represents the constant inflation (or shrinkage) due to the measurement error. For instance, if $\alpha_0 = 0$, then setting $\alpha_1 < 1$ (or $\alpha_1 > 1$) features the scenario where $X_t^*$ tends to be smaller (or larger) than $X_t$ if the noise term is ignored. This model generalizes the classical additive model considered by Staudenmayer and Buonaccorsi (2005) who considered the case with $\alpha_0 = 0$ and $\alpha_1 = 1$.

By the stationarity of the $X_t$, we note that model (6.7) yields $E(X_t^*) = \alpha_0 + \alpha_1\mu$ and

$$\text{Var}(X_t^*) = \alpha_1^2\gamma_0 + \sigma_e^2; \tag{6.8}$$

the variability of the $X_t^*$ can be greater or smaller than that of the $X_t$, depending on the value of $\alpha_1$.

The second measurement error model assumes a multiplicative form:

$$X_t^* = \beta_0 u_t X_t, \tag{6.9}$$

for $t = 1, \ldots, T$, where $\beta_0$ is a positive scaling parameter, and the $u_t$ are the error terms which are independent of each other as well as of the $X_t$, and have mean one and time-independent variance $\sigma_u^2$. Depending on the distribution of the error term $u_t$, (6.9) can feature different types of discrepancy between $X_t$ and $X_t^*$.

The stationarity of the $X_t$ together with model (6.9) implies $E(X_t^*) = \beta_0\mu$, and

$$\begin{aligned}
\mathrm{Var}(X_t^*) &= \mathrm{Var}(\beta_0 X_t u_t) \\
&= \beta_0^2 \left\{ E(X_t^2 u_t^2) - E^2(X_t u_t) \right\} \\
&= \beta_0^2 \left\{ E(X_t^2) E(u_t^2) - E^2(X_t) E^2(u_t) \right\} \\
&= \beta_0^2 \left\{ (\mathrm{Var}(X_t) + E^2(X_t))(\sigma_u^2 + 1) - E^2(X_t) \right\} \\
&= \beta_0^2 \left\{ (\sigma_u^2 + 1)\gamma_0 + \sigma_u^2 \mu^2 \right\},
\end{aligned} \tag{6.10}$$

where the third step is because of the independence of $X_t$ and $u_t$.

Since $E(X_t^*)$ is time-independent for both (6.7) and (6.9), in the following discussion, we let $\mu^*$ denote $E(X_t^*)$ for $t = 1, \ldots, T$. The modeling of the measurement error process by (6.7) or (6.9) introduces extra parameters $\{\alpha_0, \alpha_1, \sigma_e^2\}$ or $\{\beta_0, \sigma_u^2\}$, where the variance of the error term is bounded by the variability of $X_t^*$ together with others. Clearly, (6.8) shows that $\sigma_e^2 < \mathrm{Var}(X_t^*)$ and (6.10) implies that $\sigma_u^2 < \frac{\mathrm{Var}(X_t^*)}{\beta_0^2 \mu^2}$.

### 6.2.2 Naive Estimation and Bias for AR(1) Model

Estimating equations (6.6) are useful when measurements of $X_t$ are available. However, due to the measurement error, $X_t$ is not observed so (6.6) cannot be directly used for estimation of the parameters for model (6.1). As the surrogate $X_t^*$ for $X_t$ is available, one may attempt to employ the naive analysis to model (6.1) with $X_t$ replaced by $X_t^*$. Here we study the impact of measurement error on the naive analysis disregarding the difference between $X_t$ and $X_t^*$. We start with the AR(1) model, i.e., model (6.1) with $p = 1$.

If we naively replace $X_t$ in (6.1) by $X_t^*$, then the time series model (6.1) becomes

$$X_t^* = \phi_0^* + \phi_1^* X_{t-1}^* + \epsilon_t^*, \tag{6.11}$$

where $(\phi_0^*, \phi_1^*)^{\mathrm{T}}$ and $\epsilon_t^*$ show possible differences from the corresponding quantity in the model (6.1). To estimate $\phi_0^*$ and $\phi_1^*$, we may employ the ordinary least squares (OLS) method. Specifically, we minimize $S(\phi_0^*, \phi_1^*) = \sum_{t=2}^{T} (X_t^* - \phi_0^* - \phi_1^* X_{t-1}^*)^2$ with respective to $\phi_0^*$ and $\phi_1^*$, yielding the OLS estimators of $\phi_1^*$ and $\phi_0^*$:

$$\widehat{\phi}_1^* = \frac{\sum_{t=2}^{T}(X_{t-1}^* - \bar{X}_{(-1)}^*)(X_t^* - \bar{X}^*)}{\sum_{t=2}^{T}(X_{t-1}^* - \bar{X}_{(-1)}^*)^2},$$

and

$$\widehat{\phi}_0^* = \bar{X}_t^* - \widehat{\phi}_1^* \bar{X}^*, \tag{6.12}$$

where $\bar{X}_{(-1)}^* = \frac{1}{T-1}\sum_{t=2}^{T} X_{t-1}^*$ and $\bar{X}^* = \frac{1}{T-1}\sum_{t=2}^{T} X_t^*$.

**Theorem 6.1** *Let $\omega_1 = \frac{\alpha_1^2\sigma_\epsilon^2}{\alpha_1^2\sigma_\epsilon^2 + \sigma_e^2(1-\phi_1^2)}$, $\phi_1^* = \phi_1\omega_1$, and $\phi_0^* = \left(\alpha_0 + \frac{\alpha_1\phi_0}{1-\phi_1}\right)(1-\phi_1\omega_1)$. Assume the stationarity of the times series. If the measurement error process satisfies (6.7), then*

*(1) $\widehat{\phi}_1^* \xrightarrow{p} \phi_1^*$ and $\widehat{\phi}_0^* \xrightarrow{p} \phi_0^*$ as $T \to \infty$,*

*(2) $\epsilon_t^* = \alpha_0(1-\phi_1^*) + \alpha_1\phi_0 - \phi_0^* + \alpha_1(\phi_1 - \phi_1^*)X_{t-1} + (1-\phi_1^*)e_t + \alpha_1\epsilon_t$ for $t = 1, \ldots, T$, and hence $Var(\epsilon_t^*) = \phi_1^2\alpha_1^2(1-\omega_1)^2\left(\frac{\sigma_\epsilon^2}{1-\phi_1^2}\right) + (1-\omega_1\phi_1)^2\sigma_e^2 + \alpha_1^2\sigma_\epsilon^2$.*

The proof of the theorem is included in Appendix E.2. This theorem essentially implies that the naive estimator under the additive form in (6.7) is inconsistent because $\phi_1^* \neq \phi_1$ and $\phi_0^* \neq \phi_0$. The naive estimator $\widehat{\phi}_1^*$ attenuates and the attenuation factor $\omega_1$ depends on the parameters $\alpha_1$ and $\sigma_e^2$ of the measurement error model (6.7) as well as $\phi_1$ and $\sigma_\epsilon^2$ in the time series model (6.1). The coefficient $\alpha_1$ in the measurement error model (6.7) affects the estimation of the both naive estimators $\widehat{\phi}_1^*$ and $\widehat{\phi}_0^*$, while the intercept $\alpha_0$ influences the estimation of $\phi_0^*$ only, but not $\phi_1^*$ or $\mathrm{Var}(\epsilon^*)$.

**Theorem 6.2** *Let $\omega_2 = \{1 + \sigma_u^2 + \frac{(1+\phi_1)\sigma_u^2\phi_0^2}{(1-\phi_1)\sigma_\epsilon^2}\}^{-1}$, $\phi_1^* = \phi_1\omega_2$, and $\phi_0^* = \frac{\beta_0\phi_0}{1-\phi_1}(1-\omega_2\phi_1)$. If the times series is stationary and the measurement error process satisfies (6.9), then*

*(1) $\widehat{\phi}_1^* \xrightarrow{p} \phi_1^*$ and $\widehat{\phi}_0^* \xrightarrow{p} \phi_0^*$ as $T \to \infty$,*

*(2) $\epsilon_t^* = \beta_0\phi_0 u_t - \phi_0^* + \beta_0 X_{t-1}(\phi_1 u_t - \omega_2\phi_1 u_{t-1}) + \beta_0 u_t\epsilon_t$ for $t = 1, \ldots, T$, and hence $Var(\epsilon_t^*) = \beta_0^2\{\sigma_u^2\phi_0^2 + (1+\sigma_u^2)\sigma_\epsilon^2\} + \beta_0^2\phi_1^2\frac{(1+\omega_2^2)}{\omega_2}\frac{\sigma_\epsilon^2}{(1-\phi_1^2)}$.*

The proof of the theorem is included in Appendix E.3. This theorem says the attenuation effect resulting from the measurement error on estimation of $\phi_1$. The constant scaling parameter $\beta_0$ in the measurement error model (6.9) does not influence the estimation of $\phi_1$ but affects the estimation of $\phi_0$ and $\sigma_\epsilon^2$. The attenuation factor $\omega_2$ is determined by the magnitude $\sigma_u^2$ of measurement error as well as the values of $\phi_0$, $\phi_1$, and $\sigma_\epsilon^2$ of the time series model (6.1).

### 6.2.3 Naive Estimation and Bias for AR($p$) Model with $p \geq 2$

We now extend the discussion in Section 6.2.2 to the AR($p$) model with $p \geq 2$. Replacing $X_t$ with $X_t^*$ in (6.1) gives the working model

$$X_t^* = \phi_0^* + \sum_{j=1}^{p} \phi_j^* X_{t-j}^* + \epsilon_t^*, \tag{6.13}$$

where $\phi^* = (\phi_1^*, \ldots, \phi_p^*)^{\mathrm{T}}$ and $\epsilon_t^*$ may differ from the corresponding symbol in (6.1). If mimicking the procedure of using (6.6) with $X_t$ replaced by $X_t^*$ to estimate $\phi^*$, $\phi_0^*$ and $\sigma_\epsilon^{2*}$ in (6.13), then we let $\widehat{\phi}^* = (\widehat{\phi}_1^*, \ldots, \widehat{\phi}_p^*)^{\mathrm{T}}$, $\widehat{\phi}_0^*$ and $\widehat{\sigma}_\epsilon^{*2}$ denote the resultant estimators. Similar to $\widehat{\gamma}_k$ and $\widehat{\mu}$, we define $\widehat{\mu}^* = \frac{1}{T}\sum_{t=1}^{T} X_t^*$ and $\widehat{\gamma}_k^* = \frac{1}{T-k}\sum_{t=1}^{T-k}(X_t^* - \widehat{\mu}^*)(X_{t+k}^* - \widehat{\mu}^*)$ for $k = 1, \ldots, p$. Let $\widehat{\gamma}^* = (\widehat{\gamma}_1^*, \ldots, \widehat{\gamma}_p^*)^{\mathrm{T}}$ and $\widehat{\gamma}_0^* = \frac{1}{T}\sum_{t=1}^{T}(X_t^* - \widehat{\mu}^*)(X_t^* - \widehat{\mu}^*)$.

We now discuss the asymptotic results of the naive estimators under different measurement error models.

**Theorem 6.3** *Let $\mathbb{1}_p$ be the $p \times 1$ unit and let $I_p$ be the $p \times p$ identity matrix. Define $\gamma^* = \alpha_1^2\gamma$, $\gamma_0^* = \alpha_1^2\gamma_0 + \sigma_e^2$, $\phi^* = \alpha_1^2(\alpha_1^2\Gamma + \sigma_e^2 I_p)^{-1}\gamma$, $\phi_0^* = (1 - \phi^* \cdot \mathbb{1}_p)(\alpha_0 + \alpha_1\mu)$ and $\sigma_\epsilon^{2*} = \alpha_1^2\gamma_0 + \sigma_e^2 - \alpha_1^4\gamma^{T}(\alpha_1^2\Gamma + \sigma_e^2 I_p)^{-1}\gamma$. Under regularity conditions, if the time series is stationary and the measurement error process satisfies (6.7), then*

*(1)* $\widehat{\gamma}^* \xrightarrow{p} \gamma^*$ *and* $\widehat{\gamma}_0^* \xrightarrow{p} \gamma_0^*$ *as* $T \to \infty$.

*(2)* $\widehat{\phi}^* \xrightarrow{p} \phi^*$, $\widehat{\phi}_0^* \xrightarrow{p} \phi_0^*$, *and* $\widehat{\sigma}_\epsilon^{2*} \xrightarrow{p} \sigma_\epsilon^{2*}$ *as* $T \to \infty$.

*(3) Let $Q_1$ denote the $(p+1) \times (p+1)$ asymptotic covariance matrix of $\sqrt{T}\left\{(\widehat{\gamma}_0^*, \widehat{\gamma}^{*\mathrm{T}})^{\mathrm{T}} - (\gamma_0^*, \gamma^{*\mathrm{T}})^{\mathrm{T}}\right\}$ as $T \to \infty$. Then the elements of $Q_1$ are given by*

$$q_{100}^* = \alpha_1^4 q_{00} + 4\alpha_1^2\gamma_0\sigma_e^2 + E(e_t^4) - \sigma_e^4;$$
$$q_{10p}^* = \alpha_1^4 q_{0p} + 4\alpha_1^2\gamma_p\sigma_e^2;$$
$$q_{1pr}^* = \alpha_1^4 q_{pr} + 2\alpha_1^2\sigma_e^2(\gamma_{|p-r|} + \gamma_{p+r}) \text{ for } r \neq 0, r \neq p;$$
$$q_{1pp}^* = \alpha_1^4 q_{pp} + 2\alpha_1^2\sigma_e^2(\gamma_0 + \gamma_{2p}) + \sigma_e^4;$$

*for $p \geq 1$, where $q_{jk}$ is the $(j,k)$ element of the asymptotic covariance matrix of $(\widehat{\gamma}_0, \widehat{\gamma}^{\mathrm{T}})^{\mathrm{T}}$, given by (Brockwell et al., 1991, Section 7.3)*

$$q_{jk} = (\eta - 3)\gamma_j\gamma_k + \sum_{i=-\infty}^{\infty}(\gamma_i\gamma_{i-j+k} + \gamma_{i+k}\gamma_{i-j}) \tag{6.14}$$

*for $(j,k) = (0,0), (0,p), (p,p)$ and $(p,r)$ with $r \neq 0$ and $r \neq p$, with $\eta = E(\epsilon_t^4)/\sigma_\epsilon^4$.*

The proof of Theorem 6.3 is presented in the Appendix E.4. Similar to the results in Theorem 6.1, the intercept $\alpha_0$ only influences $\phi_0$ and does not influence $\phi$.

**Theorem 6.4** *Let* $\gamma^* = \beta_0^2\gamma$, $\gamma_0^* = \beta_0^2\left\{(\sigma_u^2+1)\gamma_0 + \sigma_u^2\mu^2\right\}$, $\phi^* = \left\{\Gamma + \sigma_u^2(\gamma_0+\mu^2)I_p\right\}^{-1}\gamma$, $\phi_0^* = \beta_0\left(1 - \phi^{*\mathrm{T}}\cdot \mathbb{1}_p\right)\mu$, *and* $\sigma_\epsilon^{2*} = \beta_0^2(\sigma_u^2+1)\gamma_0 + \beta_0^2\sigma_u^2\mu^2 - \beta_0^2\gamma^{\scriptscriptstyle T}\left\{\Gamma + \sigma_u^2(\gamma_0+\mu^2)I_p\right\}^{-1}\gamma$. *Under regularity conditions, if the time series are stationary and the measurement error process satisfy (6.9), then*

*(1)* $\widehat{\gamma}^* \xrightarrow{p} \gamma^*$ *and* $\widehat{\gamma}_0^* \xrightarrow{p} \gamma_0^*$ *as* $T\to\infty$.

*(2)* $\widehat{\phi}^* \xrightarrow{p} \phi^*$, $\widehat{\phi}_0^* \xrightarrow{p} \phi_0^*$, *and* $\widehat{\sigma}_\epsilon^{2*} \xrightarrow{p} \sigma_\epsilon^{2*}$ *as* $T\to\infty$.

*(3) Let* $Q_2$ *denote the* $(p+1)\times(p+1)$ *asymptotic covariance matrix of* $\sqrt{T}\left\{(\widehat{\gamma}_0^*, \widehat{\gamma}^{*\mathrm{T}})^{\mathrm{T}} - (\gamma_0^*, \gamma^{*\mathrm{T}})^{\mathrm{T}}\right\}$ *as* $T\to\infty$. *Then the elements of* $Q_2$ *are given by*

$$
\begin{aligned}
q_{200}^* =\ & \beta_0^4(\sigma_u^2+1)^2 q_{00} + \beta_0^4\{E(u_t^4) - (\sigma_u^2+1)^2\}E(X_t-\mu)^4 \\
& + 4\mu\beta_0^4\sigma_u^2(\sigma_u^2+1)v_0 + 4\mu\beta_0^4\{E(u_t^4) - E(u_t^3) - \sigma_u^2(\sigma_u^2+1)\}E(X_t-\mu)^3 \\
& + 2\mu^2\beta_0^4\left\{E(u_t^4) - 2E(u_t^3) + 1 - \sigma_u^4\right\}\gamma_0 \\
& + 4\mu^2\beta_0^4\left[\sigma_u^4\sum_{h=-\infty}^{\infty}\gamma_h + \left\{E(u_t^4) - 2E(u_t^3) + \sigma_u^2 + 1 - \sigma_u^4\right\}\gamma_0\right] + \mu^4\beta_0^4\left[E\{(u_t-1)^4\} - \sigma_u^4\right];\\
q_{20p}^* =\ & \beta_0^4 q_p(\sigma_u^2+1) + \beta_0^4\left\{E(u_t^3) - (\sigma_u^2+1)\right\}\left[E\{(X_t-\mu)^3(X_{t+p}-\mu)\} + E\{(X_t-\mu)^3(X_{t-p}-\mu)\}\right] \\
& + 2\mu\beta_0^4\sigma_u^2 v_{0p} + \mu\beta_0^4 E\{3u_t^3 - 3u_t^2 - 2\sigma_u^2\}\left[E\{(X_t-\mu)^2(X_{t-p}-\mu)\} + E\{(X_t-\mu)^2(X_{t+p}-\mu)\}\right] \\
& + 6\mu^2\beta_0^4 E(u_t-1)^3\gamma_p + 4\mu^2\beta_0^4\sigma_u^2\gamma_p;\\
q_{2pr}^* =\ & \beta_0^4 q_{pr} + \beta_0^4\sigma_u^2\left[E\{(X_t-\mu)^2(X_{t+p}-\mu)(X_{t+r}-\mu)\} + E\{(X_t-\mu)(X_{t+p}-\mu)^2(X_{t+p+r}-\mu)\}\right. \\
& \left. + E\{(X_{t-r}-\mu)(X_t-\mu)^2(X_{t+p}-\mu)\} + E\{(X_t-\mu)(X_{t+p-r}-\mu)(X_{t+p}-\mu)^2\}\right] \\
& + \mu\beta_0^4\sigma_u^2\left[E\{(X_t-\mu)(X_{t+p}-\mu)(X_{t+r}-\mu)\} + E\{(X_t-\mu)(X_{t+p}-\mu)(X_{t+p+r}-\mu)\}\right. \\
& \left. + E\{(X_{t-r}-\mu)(X_t-\mu)(X_{t+p}-\mu)\} + E\{(X_t-\mu)(X_{t+p-r}-\mu)(X_{t+p}-\mu)\}\right] \\
& + 2\mu^2\beta_0^4\sigma_u^2(\gamma_{|p-r|} + \gamma_{p+r}) \text{ for } r\neq p, r\neq 0;\\
q_{2pp}^* =\ & \beta_0^4 q_{pp} + \beta_0^4(\sigma_u^4 + 2\sigma_u^2)\,Var\{(X_t-\mu)(X_{t+p}-\mu)\} + 2\beta_0^4 E\{(X_t-\mu)(X_{t+p}-\mu)^2(X_{t+2p}-\mu)\} \\
& + \mu\beta_0^4\sigma_u^2\left[E\{(X_t-\mu)(X_{t+p}-\mu)^2\} + 2E\{(X_t-\mu)(X_{t+p}-\mu)(X_{t+2p}-\mu)\}\right. \\
& \left. + E\{(X_t-\mu)^2(X_{t+p}-\mu)\}\right] + 2\mu^2\beta_0^4\sigma_u^4\gamma_p + 2\mu^2\beta_0^4\sigma_u^2(\gamma_0+\gamma_{2p}) + \mu^4\beta_0^4\sigma_u^4,
\end{aligned}
$$

*where the* $q_{jk}$ *are given by (6.14), for* $(j,k) = (0,0), (0,p), (p,p)$ *and* $(p,r)$ *with* $r\neq 0$ *and* $r\neq p$, *and* $v_p = \lim_{T\to\infty}\frac{1}{T}\sum_{t=1}^{T}\sum_{s=1}^{T}E\{(X_t-\mu)(X_{t+p}-\mu)(X_s-\mu)\}$.

The proof of the theorem is presented in Appendix E.5. The multiplicative measurement error $u_t$ contributes to the biasedness of the parameter estimation for $\phi$, while the scaling parameter $\beta_0$ has no effects on the naive estimator $\widehat{\phi}^*$.

## 6.3 Methodology of Correcting Measurement Error Effects

### 6.3.1 Estimation of Model Parameters

In the presence of measurement error, measurements of the $X_t$ are not always available but surrogate measurements $X_t^*$ are available. It may be tempting to conduct a naive analysis by implementing (6.6) with the $X_t$ replaced by the $X_t^*$, or equivalently with $\widehat{\mu}$ and $\widehat{\gamma}_k$ replaced by $\widehat{\mu}^*$ and the $\widehat{\gamma}_k^*$, respectively, to find estimators of $\phi$, $\phi_0$ and $\sigma_\epsilon^2$. However, by Theorems 6.3–6.4, such a procedure typically yields biased estimators. In this section, we develop new estimators accounting for the measurement error effects described by either the additive model (6.7) or the multiplicative model (6.9).

Our idea is still to employ (6.6) to find consistent estimators of $\phi$, $\phi_0$ and $\sigma_\epsilon^2$, but instead of replacing $\widehat{\mu}$ and the $\widehat{\gamma}_k$ with $\widehat{\mu}^*$ and the $\widehat{\gamma}_k^*$ as in the naive analysis, we replace $\widehat{\mu}$ and the $\widehat{\gamma}_k$ in (6.6) with new functions of the $X_t^*$, denoted as $\widetilde{\mu}$ and the $\widetilde{\gamma}_k$, which adjust for the measurement error effects. Specifically, if we can find $\widetilde{\mu}$ and the $\widetilde{\gamma}_k$ such that they resemble $\widehat{\mu}$ and the $\widehat{\gamma}_k$ in the sense that as $T \to \infty$,

$$\begin{aligned} &\widetilde{\mu} \text{ and } \widehat{\mu} \text{ have the same limit in probability,} \\ \text{and} \quad &\widetilde{\gamma}_k \text{ and } \widehat{\gamma}_k \text{ have the same limit in probability for } k = 0, \ldots, p, \end{aligned} \tag{6.15}$$

then substituting $\widehat{\mu}$ and the $\widehat{\gamma}_k$ with $\widetilde{\mu}$ and the $\widetilde{\gamma}_k$ in (6.6) yields consistent estimators of $\phi$, $\phi_0$ and $\sigma_\epsilon^2$.

With the availability of the $\widetilde{\gamma}_k$ satisfying (6.15), let $\widetilde{\Gamma}$ denote $\Gamma$ with the $\gamma_k$ replaced by the $\widetilde{\gamma}_k$. Then provided regularity conditions, consistent estimators of $\phi$, $\phi_0$ and $\sigma_\epsilon^2$ can be obtained by solving the estimating equations for $\phi$, $\phi_0$, and $\sigma_\epsilon^2$:

$$\begin{aligned} \phi &= \widetilde{\Gamma}^{-1}\widetilde{\gamma}, \\ \phi_0 &= \left(1 - \sum_{i=1}^{p} \phi_i\right)\widetilde{\mu}, \\ \sigma_\epsilon^2 &= \widetilde{\gamma}_0 - 2\phi^{\mathrm{T}}\widetilde{\gamma} + \phi^{\mathrm{T}}\widetilde{\Gamma}\phi. \end{aligned} \tag{6.16}$$

It is immediate to obtain the following result.

**Theorem 6.5** *Assume regularity conditions hold and the time series are stationary. If $\widetilde{\mu}$ and the $\widetilde{\gamma}_k$ are functions of the $X_t^*$ with $t = 1, \ldots, T$ and they satisfy (6.15), and let $\widetilde{\phi}$, $\widetilde{\phi}_0$,* *and $\widetilde{\sigma}_\epsilon{}^2$ denote the estimators for $\phi$, $\phi_0$ and $\sigma_\epsilon^2$, respectively, obtained by solving (6.16). Then, as $T \to \infty$*

*(1) $\widetilde{\phi} \xrightarrow{p} \phi$, $\widetilde{\phi}_0 \xrightarrow{p} \phi_0$, and $\widetilde{\sigma}_\epsilon{}^2 \xrightarrow{p} \sigma_\epsilon^2$;*

*(2) $\sqrt{n}(\widetilde{\phi} - \phi) \xrightarrow{d} N(0, GQG^{\mathrm{T}})$,*
*where $G$ is the matrix of derivatives of $\widetilde{\phi}$ with respect to the components of $(\widehat{\gamma}_0^*, \widehat{\gamma}^{*\mathrm{T}})^{\mathrm{T}}$. Here $Q = Q_1$, the matrix in Theorem 6.3, if measurement error follows the model (6.7); and $Q = Q_2$, the matrix in Theorem 6.4, if measurement error follows the model (6.9).*

Now we discuss explicitly how to determine $\widetilde{\mu}$ and the $\widetilde{\gamma}_k$ under the measurement error model (6.7) or (6.9). With (6.7), take $\widetilde{\mu} = \frac{\widehat{\mu}^*}{\alpha_1} - \alpha_0$, $\widetilde{\gamma}_0 = \frac{1}{\alpha_1^2}(\widehat{\gamma}_0^* - \sigma_e^2)$, and $\widetilde{\gamma}_k = \frac{\widehat{\gamma}_k^*}{\alpha_1^2}$ for $k = 1, \ldots, p$. With (6.9), take $\widetilde{\mu} = \frac{\widehat{\mu}^*}{\beta_0}$, $\widetilde{\gamma}_0 = \frac{\gamma_0^*}{(1+\sigma_u^2)\beta_0^2} - \frac{\sigma_u^2 \mu^2}{\sigma_u^2 + 1}$, and $\widetilde{\gamma}_k = \frac{\widehat{\gamma}_k^*}{\beta_0^2}$ for $k = 1, \ldots, p$. By the results in Theorem 6.3(1) and Theorem 6.4(1), it can be easily verified that these $\widetilde{\mu}$ and the $\widetilde{\gamma}_k$ satisfy (6.15).

We conclude this section with a procedure of estimating the asymptotic covariance matrix for the estimator $\widetilde{\phi}$. While Theorem 6.5 presents the sandwich form of the asymptotic covariance matrix of $\widetilde{\phi}$, its evaluation involves lengthy calculations. We may alternatively employ the block bootstrap algorithm (Lahiri, 1999) to obtain variance estimates for $\widetilde{\phi}$ using the following steps. Firstly, we set a positive integer, say $N$, as the number for the bootstrap sampling; $N$ can be set as a large number such as 1000. Next, we repeat through the following five steps:

Step 1: At iteration $n \in \{1, \ldots, N\}$, we initialize a null time series $X^{(n,0)}$ of dimension 0 and specify a block length, say $b$, which is an integer between 0 and $T$. Initialize $m$=1.

Step 2: Sample an index, say $i$, from $\{0, \ldots, T - b\}$, and then define $X_{\text{add}}^{(m-1)} = \{X_{i+1}, \ldots, X_{i+b}\}$.

Step 3: Update the previous time series $X^{(n,m-1)}$ by appending $X_{\text{add}}^{(m-1)}$ to it, and let $X^{(n,m)}$ denote the new time series.

Step 4: If the dimension $X^{(n,m)}$ is smaller than $T$ then return to Steps 2 and 3; otherwise drop the elements in the time series with the index greater than $T$ to ensure the dimension of $X^{(n,m)}$ is identical to $T$ and then go to Step 5.

Step 5: Obtain an estimate $\widetilde{\phi}^{(n)}$ of parameter $\phi$ by applying the times series $X^{(n,m)}$ to (6.16). If $n < N$, then set $n$ to be $n+1$ and go back to Step 1 to repeat; otherwise stop.

Let $\widetilde{\widetilde{\phi}}^{(n)} = \frac{1}{N}\sum_{n=1}^{N}\widetilde{\widetilde{\phi}}^{(n)}$ be the sample mean. The bootstrap variance of $\widetilde{\phi}$ is then given by,

$$\mathrm{Var}_{\mathrm{boot}}(\widetilde{\phi}) = \frac{1}{N}\sum_{n=1}^{N}(\widetilde{\phi}^{(n)} - \widetilde{\widetilde{\phi}}^{(n)})^2.$$

### 6.3.2 Forecasting and Prediction Error

Forecasting is an important application of the autoregressive models. Specifically, in forecasting based on the observed time series $X_{(T)} = \{x_1, \ldots, x_T\}$, we are interested in the predictions of $\{X_{T+1}, \ldots, X_{T+H}\}$ for a positive integer $H$, which is done one by one starting from the nearest time point $T+1$ to the farthest time point $T+H$. To this end, let $h = 1, \ldots, H$, the $h$-step forecasting of $X_{T+h}$ is based on its history of lag-$p$, $\{X_{T+h-1}, \ldots, X_{T+h-p}\}$, by using the conditional expectation $E(X_{T+h}|x_{T+h-1}, \ldots, x_{T+h-p})$, denoted $\widehat{X}_{T+h}$, where for $j = T+h-1, \ldots, T+h-p$, $x_j$ is the observe value of $X_j$ if $j \leq T$; and $x_j$ is the predicted value of $X_j$, $\widehat{X}_j$, if $j > T$. This prediction minimizes the squared prediction error $E(\widehat{X}_{T+h} - X_{T+h})^2$ (e.g., Box et al., 2015, Page 131).

If no measurement error is involved, due to the zero mean of the random error term $\epsilon_t$ in the AR($p$) model (6.1), for $h = 1, \ldots, H$, the conditional expectation can be calculated by

$$\widehat{X}_{T+h} = \phi_0 + \phi_1 x_{T+h-1} + \ldots + \phi_p x_{T+h-p}. \tag{6.17}$$

When measurement error appears, the observe values $x_j$ for $j = T, \ldots, T-p+1$ in (6.17) are no longer available but their surrogates $X_j^*$ are available. We now provide a sensible estimate of $X_j$ by using the measurement error model for characterizing the relationship of $X_j$ and $X_j^*$. If measurement error follows (6.7), we "estimate" $X_j$ by

$$\widehat{X}_j = \frac{1}{\alpha_1}(X_j^* - \alpha_0) \qquad \text{for } j = t, \ldots, t-p+1; \tag{6.18}$$

if the measurement error follows (6.9), then $\widehat{X}_j$ is "estimated" by

$$\widehat{X}_j = \frac{X_j^*}{\beta_0} \qquad \text{for } j = t, \ldots, t-p+1. \tag{6.19}$$

These "estimates" are unbiased in the sense that $E(\widehat{X}_j) = X_j$ for $j = t, \dots, t-p+1$. Consequently, for $h = 1, \dots, H$, $X_{T+h}$ is predicted as

$$\widehat{X}_{T+h} = \phi_0 + \phi_1 \widehat{X}_{T+h-1} + \cdots + \phi_p \widehat{X}_{T+h-p}. \tag{6.20}$$

In contrast to the observed values $\{x_T, \dots, x_{T-p+1}\}$, also referred to as the initial values of the forecasting of $X_{T+1}, \dots, X_{T+H}$, the estimates determined by (6.18) or (6.19) introduce additional prediction error which should be characterized. Without the loss of generality, we consider $p = 1$ to illustrate the recursive calculation of the prediction error; the prediction error with a higher order of autoregressive process can be derived recursively in a similar way but with more complex expressions.

If the measurement error follows (6.7), the mean squared prediction error of the 1-step prediction is given by

$$\begin{aligned}
\mathrm{P}_{\mathrm{e}}^{(1)} &= E(\widehat{X}_{T+1} - X_{T+1})^2 \\
&= E\{(\phi_0 + \phi_1 \widehat{X}_T) - (\phi_0 + \phi_1 X_T + \epsilon_{T+1})\}^2 \\
&= E\left\{\phi_1 \left(X_t + \frac{e_T}{\alpha_1}\right) - \phi_1 X_T - \epsilon_{T+1}\right\}^2 \\
&= \frac{\phi_1^2 \sigma_e^2}{\alpha_1^2} + \sigma_\epsilon^2,
\end{aligned}$$

where the last step is due to the independence between $e_t$ and $\epsilon_{t+1}$, as well as $E(e_t^2) = \sigma_e^2$ and $E(\epsilon_t^2) = \sigma_\epsilon^2$.

Then, the $h$-step prediction error is given by

$$\begin{aligned}
\mathrm{P}_{\mathrm{e}}^{(h)} &= E(\widehat{X}_{T+h} - X_{T+h})^2 \\
&= E(\phi_0 + \phi_1 \widehat{X}_{T+h-1} - \phi_0 - \phi_1 X_{T+h-1} - \epsilon_{T+1})^2 \\
&= E\left\{\phi_1 \left(\widehat{X}_{T+h-1} - X_{T+h-1}\right) - \epsilon_{T+1}\right\}^2 \\
&= \phi_1^2 \mathrm{P}_{\mathrm{e}}^{(h-1)} + \sigma_\epsilon^2 \\
&= \frac{\phi_1^{2h} \sigma_e^2}{\alpha_1^2} + \sum_{i=0}^{h-1} \phi_1^{2i} \sigma_\epsilon^2,
\end{aligned} \tag{6.21}$$

where the last step comes from the recursive evaluation of $P_e^{(h-1)}$.

Similarly, if the measurement error follows (6.9), the mean squared prediction error is given by

$$
\begin{aligned}
\mathrm{P}_{\mathrm{e}}^{(1)} &= E(\widehat{X}_{T+1} - X_{T+1})^2 \\
&= E(\{\phi_0 + \phi_1 \widehat{X}_T\} - \{\phi_0 + \phi_1 X_T + \epsilon_{T+1}\})^2 \\
&= E\left\{\phi_1 X_t (u_T - 1) - \epsilon_{T+1}\right\}^2 \\
&= E\left\{\phi_1 X_t (u_T - 1)\right\}^2 + E\left(\epsilon_{T+1}^2\right) \\
&= \phi_1^2 E\left\{X_T^2 (u_T - 1)^2\right\} + E\left(\epsilon_{T+1}^2\right) \\
&= \phi_1^2 E\left\{X_T^2\right\} E\left(u_T^2 - 2u_T + 1\right) + E\left(\epsilon_{T+1}^2\right) \\
&= \phi_1^2 \{\mathrm{Var}(X_T) + E^2(X_T)\}\{E\left(u_T^2\right) - 2E(u_T) + 1\} + \sigma_\epsilon^2 \\
&= \phi_1^2 \{\mathrm{Var}(X_T) + E^2(X_T)\}\{\mathrm{Var}(u_T) + E^2(u_T) - 2E(u_T) + 1\} + \sigma_\epsilon^2 \\
&= \phi_1^2 \{\mathrm{Var}(X_T) + E^2(X_T)\}\mathrm{Var}(u_T) + \sigma_\epsilon^2 \\
&= \phi_1^2 \left\{\frac{\sigma_\epsilon^2}{1 - \phi_1^2} + \mu^2\right\} \sigma_u^2 + \sigma_\epsilon^2,
\end{aligned}
$$

where the fourth step is due to the independence of $\epsilon_{t+1}$, $u_t$ and $X_t$, the sixth step is due to the independence of $u_t$ and $X_t$, the second last step is due to $E(u_t) = 1$, and the last step is because $\mathrm{Var}(X_t) = \frac{\sigma_\epsilon^2}{1-\phi_1^2}$ due to the stationary AR(1) process. Hence,

$$
\begin{aligned}
\mathrm{P}_{\mathrm{e}}^{(h)} &= E(\widehat{X}_{T+h} - X_{T+h})^2 \\
&= E\{(\phi_0 + \phi_1 \widehat{X}_{T+h-1}) - (\phi_0 + \phi_1 X_{T+h-1} + \epsilon_{T+1})\}^2 \\
&= E\left\{\phi_1 \left(\widehat{X}_{T+h-1} - X_{T+h-1}\right) - \epsilon_{T+1}\right\}^2 \\
&= \phi_1^2 \mathrm{P}_{\mathrm{e}}^{(h-1)} + \sigma_\epsilon^2 \\
&= \phi_1^{2h-2} \mathrm{P}_{\mathrm{e}}^{(1)} + \sum_{i=0}^{h-2} \phi_1^{2i} \sigma_\epsilon^2 \\
&= \phi_1^{2h} \left\{\frac{\sigma_\epsilon^2}{1 - \phi_1^2} + \mu^2\right\} \sigma_u^2 + \sum_{i=0}^{h-1} \phi_1^{2i} \sigma_\epsilon^2. \qquad (6.22)
\end{aligned}
$$

The evaluation of the mean squared prediction error $\mathrm{P}_{\mathrm{e}}^{(h)}$ is carried out by replacing the parameters with their estimators. We comment that the common second term in (6.21) and (6.22), $\sum_{i=0}^{h-1} \phi_1^{2i} \sigma_\epsilon^2$, is the mean squared prediction error for the AR(1) model for error-free settings (e.g. Box et al., 2015, Page 152), which equals $\frac{1-\phi_1^{2h}}{1-\phi_1^2}\sigma_\epsilon^2$.

For an $\alpha$ with $0 < \alpha < 1$, then $h$-step $(1-\alpha)$-prediction interval is constructed as

$$\left[\widehat{X}_{T+h} - q_{\frac{\alpha}{2}} \mathrm{P}_{\mathrm{e}}^{(h)}, \widehat{X}_{T+h} + q_{\frac{\alpha}{2}} \mathrm{P}_{\mathrm{e}}^{(h)}\right],$$

where $q_{\frac{\alpha}{2}}$ the $\alpha$-level quantile of the distribution of $\widehat{X}_{T+h} - X_{T+h}$. In practice, under normal assumption of $\epsilon_t$ and $e_t$, one can take $q_{\frac{\alpha}{2}}$ to be the $\alpha$-level quantile of the standard normal distribution (Brockwell and Davis, 2002, Page 108).

## 6.4 Analysis of COVID-19 Death Rates

### 6.4.1 Study Objective

Using Canadian provincial COVID-19 data containing the daily confirmed cases and deaths from April 3, 2020 to May 4, 2020, we compare the times series of death rates for British Columbia, Ontario, Quebec, and Alberta, the four provinces in Canada which experience severe situations. The daily confirmed cases and fatalities are taken from "1Point3Acres.com" (`https://coronavirus.1point3acres.com/`).

In epidemiology, the mortality rate, defined as the proportion of cumulative deaths of the disease in the total number of people diagnosed with the disease (Kanchan et al., 2015), is often used to measure the severeness of an infectious disease. For COVID-19, determining the mortality rate is not trivial due to the difficulty in precisely determining the number of infected cases. Due to the limited test capacity, individuals with light symptoms are not being tested. Asymptomatic infections and the incubation period make it difficult to acquire an accurate number of infections. To circumvent this, we explore different definitions of death rates. *Definition 1* is from Baud et al. (2020) who estimated mortality rates by dividing the number of deaths on a given day by the number of patients with confirmed COVID-19 infection 14 days before, with the consideration of the maximum incubation time to be 14 days. On the other hand, the median time from symptom onset to intensive care unit admission is about 10 days ([3] in Baud et al., 2020), so we consider *Definition 2* which is the number of deaths of COVID-19 on day $t$ divided by the number of confirmed cases at day $(t-10)$. In comparison, we also consider *Definition 3* by calculating the death rate on day $t$ as the ratio of the number of deaths on day $t$ to the number of confirmed cases on the day $t$.

While the first two ways may help more reasonably estimate mortality rates than the third definition, these calculated rates still differ from the true mortality rates because of under-reported cases which are primarily due to limited test capacity and undetected asymptomatic infections. To reflect the discrepancy between the *reported* and the *true* mortality rates for each province, for each definition of the mortality rate, we let $X_{1,t}$, $X_{2,t}$, $X_{3,t}$, and $X_{4,t}$, represent the *true* mortality rate on day $t$ for British Columbia, Ontario, Quebec and Alberta, respectively; and let $X^*_{1,t}$, $X^*_{2,t}$, $X^*_{3,t}$ and $X^*_{4,t}$ denote the *reported* mortality rate on day $t$ in British Columbia, Ontario, Quebec and Alberta, respectively. The objective is to use the reported mortality rates $\{X^*_{it} : t = 1, \ldots, 31\}$ to infer the true mortality rates $X_{i,t}$ which are modeled by (6.1) separately for $i = 1, \ldots, 4$. In addition, we want to forecast the true mortality rate of COVID-19 for a future time period. Due to the undetected asymptomatic cases and untested cases for light symptoms, the reported mortality rates $X^*_{i,t}$ are typically overestimated (i.e., $X^*_{i,t} \geq X_{i,t}$) for $i = 1, \ldots, 4$. As there is no exact information to guide us how to characterize the relationship between $X^*_{it}$ and $X_{it}$, here we conduct sensitivity studies by considering measurement error model (6.7) or (6.9). We use the observed data $X^*_{i,t}$ from April 3, 2020 to May 4, 2020, i.e., $\{X^*_{i,t} : t = 1, ..., T_i\}$ with $T_1 = T_2 = 31$, to estimate the model parameters in (6.1) with measurement error effects accounted for, and then forecast the mortality rate of COVID-19, from May 5, 2020 to May 9, 2020, in British Columbia, Ontario, Quebec and Alberta, Canada.

### 6.4.2 Models Building

Figure 6.1 displays the trajectory of the mortality rates of COVID-19 in the four provinces that are obtained from the three definitions. To assess the stationarity of the $X^*_{it}$, we conduct the augmented Dickey–Fuller (ADF) tests (Cheung and Lai, 1995) to times series $\{X^*_{i,t} : t = 1, \ldots, T\}$, or its differencing transformation $\{X^*_{i,(t+1)} - X^*_{i,t} : t = 1, \ldots, T\}$ for $i = 1, \ldots, 4$ in each definition. Table 6.1 presents the test statistics and $p$-value of the ADF test for each time series, where "TSV" represents a test statistics value.

To determine the lag value $p$ for the autoregression model (6.1) used for the time series $\{X_{i,t} : t = 1, ..., T_i\}$ with $T_1 = T_2 = 31$ for $i = 1, \ldots, 4$, we fit the naive model (6.13) with $\epsilon^*_t$ assumed to follow a normal distribution $N(0, \sigma^{*2}_\epsilon)$, and use the AIC criterion by minimizing

$$-2 \sum_{t=p}^{T} \log f(x^*_t | x^*_{t-1}, \ldots, x^*_{t-p}) + 2p, \tag{6.23}$$

where $f(x^*_t | x^*_{t-1}, \ldots, x^*_{t-p})$ is the conditional probability of $X^*_t$ given $X^*_{t-1}, \ldots, X^*_{t-p}$. The results are summarized in Table 6.2, where no-differencing or 1-differencing is applied, the entries with "-" indicate that the corresponding model is not applicable due to the ADF test results.

We take those lag values for an AR($p$) model to feature the true mortality rate $X_{i,t}$ for each definition and $i = 1, \ldots, 4$. To be specific, for the British Columbia data, with Definition 1 we consider two models: AR(1) model for the time series with 1-order differencing and AR(2) model for the time series with no-differencing; with Definitions 2 and 3, we consider AR(2) and AR(1) models, respectively, for the time series with 1-order differencing. For the Ontario data, we consider AR(1) and AR(4) for the time series with 1-order differencing in Definitions 1 and 3, respectively, and AR(2) for Definition 2 with no transformation. For the Quebec data, we consider AR(1) and AR(2) models for the times series with 1-order differencing in Definitions 1 and 2, respectively. For Alberta data, we consider the AR(1) model for the times series with 1-order differencing for both Definitions 1 and 2.

### 6.4.3 Sensitivity Analyses

As there are no additional data available for estimating the parameters for the model (6.7) or (6.9), we conduct sensitivity analyses using the findings in the literature. Different studies showed different estimates of the asymptomatic infection rates, changing from 17.9% to 78.3% (Kimball, 2020; Day, 2020). To accommodate the heterogeneity of different studies, He et al. (2020) carried out a meta-analysis and obtained an estimate of the asymptomatic infection rate to be 46%. If under-reported confirmed cases are only caused from undetected asymptomatic cases, then $X_t = (1 - \tau_A)X_t^*$, or equivalently,

$$X_t^* = \frac{1}{1 - \tau_A} X_t, \tag{6.24}$$

where $\tau_A$ represents the rate of asymptomatic infections.

Now we use (6.24) as a starting point to conduct sensitivity analyses. In the multiplicative model (6.9), we take $\beta_0 u_t = \frac{1}{1-\tau_A}$. With $E(u_t) = 1$, we set $\beta_0 = \frac{1}{1-\tau_A}$ by setting $\tau_A = 46\%$, the value from the meta-analysis of He et al. (2020). To see different degrees of error, we consider $\sigma_u^2$ to take a small value, say $\sigma_{u1}^2$, and a large value, say, $\sigma_{u2}^2$, which is alternatively reflected by the change of the coefficient of variation, $CV = \frac{\sigma_u}{E(u_t)}$, of the error term $u_t$ from $\sigma_{u1} \times 100\%$ to $\sigma_{u2} \times 100\%$.

When using the additive model (6.7) to characterize the measurement error process, motivated by (6.24), we set $\alpha_0 = 0$ and $\alpha_1 = \frac{1}{1-46\%}$, and let $\sigma_e^2$ take a small value, say $\sigma_{e1}^2$, and a large value, say $\sigma_{e2}^2$, to feature an increasing degree of measurement error. Due to the constraints for the parameters discussed for (6.8) and (6.10), we set the values for $\sigma_{u1}$, $\sigma_{u2}$, $\sigma_{e1}$, and $\sigma_{e2}$ case by case for each definition and for each province, which are recorded in Table 6.3.

The model-fitting results are reported in Tables 6.4–6.6 for the three definitions of mortality rates, where the point estimates (EST), the associated standard errors (SE), and the p-values for the model parameters are included. Table 6.4 shows that with Definition 1, the estimates of $\phi_0$ in the absolute value from the proposed method are smaller than those of naive method, while the estimates of $\phi_1$ produced from the proposed and naive methods exhibit an opposite direction. As expected, the standard errors for the proposed method are generally larger than those of the naive method. However, both methods find no evidence to support that $\phi_0$ and $\phi_1$ are different from zero for the data of British Columbia and Ontario, suggesting that the mortality rates of these two provinces remain statistically unchanged. At the significance level 0.1, the naive method and the proposed method show different evidence for the data of Quebec and Alberta. The naive method suggests a likely downward trend with p-value 0.071 and 0.061 for testing of $\phi_0$ for Quebec and Alberta, respectively. The proposed method, on the other hand, shows that $\phi_0$ is insignificant for these two provinces.

Table 6.5 displays the results for Definition 2. For the British Columbia data, the estimates of the three parameters $\phi_1$, $\phi_2$ and $\phi_3$ produced from the proposed method are smaller than those yielded from the naive method, whereas the standard errors output from the proposed method are larger than those from the naive method. However, at the significance level 0.05, both methods find no evidence to show the significance of $\phi_0$, $\phi_1$ and $\phi_2$, suggesting that the mortality rate of British Columbia remain unchanged with time. Similar findings are revealed for the Alberta data except that the parameter estimates output from the proposed method are larger than those produced from the naive method. For the Ontario and Quebec data, the revealings from the two methods are quite different. For Ontario, both methods show that $\phi_0$ is insignificant and $\phi_1$ is significant. The evidence of $\phi_2$, however, depends on the nature of measurement error. On the contrary, the findings for Quebec do not tend to show a definite direction, and they vary with the model form or degree of the measurement error process.

Table 6.6 shows the results for Definition 3. For the British Columbia data, the estimates produced by the proposed method are smaller than those yielded from the naive method. The standard errors output from the proposed method inflate as the degree of measurement error increases. The naive and proposed methods reveal different evidence for the significance of $\phi_0$ and $\phi_1$, and the degree of measurement error affects the findings too. For the Ontario data, both methods uncover the same type of evidence for all the parameters at the significance level 0.05, except for the case with the large error under the multiplicative model.

### 6.4.4 Forecasting

With the fitted model for each time series in Section 6.4.3, we forecast the true mortality rate for the subsequent five days (May 5 – May 9) using the method described in Section 6.3.2. Specifically, since the true mortality rates are not observable, we "estimate" them using (6.18) and (6.19), respectively, for the measurement error models (6.7) and (6.9), and then we forecast the values of $X_{i,32}$, $X_{i,33}$, $X_{i,34}$, $X_{i,35}$, and $X_{i,36}$ using (6.20).

To quantify the forecasting performance, we calculate $\mathrm{P}_{\mathrm{e}}^{(h)}$ for $h = 1, \ldots, H$ for each specified model of the mortality rates $X_{i,t}$, and we report the results, together with the total $\sum_{h=1}^{H} \mathrm{P}_{\mathrm{e}}^{(h)}$ in Tables 6.7–6.9, where $H$ is set as 5. For $h = 1, \ldots, H$, we report the observed prediction error $(X_{T+h} - \widehat{X}_{T+h})^2$, and the expected prediction error defined in (6.21) and (6.22).

Forecasting results based on the three definitions of mortality rates are reported in Figures 6.2–6.8 for the four provinces, where the prediction results after May 4 are marked in blue and red for the measurement error models (6.7) and (6.9), respectively, together with prediction areas marked in shaded parts, as well as the prediction results obtained from the naive method by using (6.20) with naive estimates of $\phi$ (marked in dark yellow). In comparison, we display the reported mortality rate (in black) from Apr 3, 2020 to May 9, 2020 as well as the adjusted mortality rates obtained from (6.24) (in green); in addition, we report the fitted values using (6.17) in blue points. To compare the forecasting results in the presence of different degrees of measurement error. We report the results derived from a mild degree of measurement error in top subfigures and place those obtained from a large degree of measurement error in bottom subfigures.

The results for British Columbia are presented in Figures 6.2–6.5. With Definition 1, the methods with measurement error effects accommodated suggest that the mortality rate in the past and its forecasting values are around 4%, whereas the results obtained from the method without accounting for measurement error effects indicate that the mortality rates over time are higher than 6%. With Definition 2, the methods with or without accounting for measurement error effects reveal that the mortality rates over time are, respectively, below 3.5% and above 5%. With Definition 3, the methods with or without accounting for measurement error effects indicate that the mortality rates over time are around 3% and above 4%, respectively.

The results for Ontario are presented in Figures 6.6–6.8. With Definition 1, the methods with measurement error effects accommodated suggest that the mortality rate over time is around 7% over time, while the reported mortality rate over time is about 12.5%. With Definition 2, the methods with and without incorporating the feature of measurement error indicate the mortality rate in the past and its forecasting values are, respectively, below 6% and around 10%. With Definition 3, the mortality rate increases over time substantially. The methods with measurement error effects accommodated suggest that the mortality rate increases from 2% to above 4% whereas the reported mortality rate shows that rate increases from below 4% to above 8%.

The results for Quebec are presented in Figures 6.9–6.10. With Definition 1 the methods with measurement error effects accommodated show that the mortality rate is around 6.5% over time, whereas the method without considering measurement error indicates the mortality rate is over 10%. With Definition 2, the methods with or without addressing the measurement error effects show that the mortality rates over time are, respectively, below 6% and above 7.5%.

The results for Alberta are presented in Figures 6.11–6.12. With Definition 1 the methods with and without measurement error accommodated suggest that the mortality rates are, respectively, around 2% and 4% over time. With Definition 2, the methods with or without addressing the measurement error effects show that the historical mortality rate and its predictions are, respectively, below 2% and above 2%.

### 6.4.5 Model Assessment

The specification of lag $p$ for model (6.1) of the true mortality rates $\{X_{i,t} : t = 1, \ldots, T\}$ is based on (6.23) which is derived from the reported mortality rates $\{X^*_{i,t} : t = 1, \ldots, T\}$, but not from $\{X_{i,t} : t = 1, \ldots, T\}$ itself. This discrepancy introduces the possibility of model misspecification when featuring the series $X_{i,t}$ using (6.1). To investigate this, we conduct a sensitivity analysis by considering the AR($p$) with a different value of $p$ for the $X_{i,t}$ from Definition 1. As Table 6.2 indicates the feasibility of using AR(1) for all four provinces, here we further employ the AR(2) model to do forecasting for the period from May 5 to May 9.

In Table 6.10, we report the observed and expected prediction errors of the forecasting using AR(2) models in comparison with AR(1) models. Comparing different lag orders of the autoregressive models, we find that in terms of the observed prediction error, the selected AR(1) models have better performance than the AR(2) models for the data of Ontario and Alberta, and the results for British Columbia and Quebec are fairly similar. It is noticed that both the observed prediction error and the expected prediction error associated with the proposed method tend to become small when the degree of measurement error increases for British Columbia, Ontario, and Quebec.

Table 6.1: The results of the augmented Dickey-Fuller test

| | | British Columbia | | Ontario | | Quebec | | Alberta | |
|---|---|---|---|---|---|---|---|---|---|
| Definition | Transformation | TSV | p-value | TSV | p-value | TSV | p-value | TSV | p-value |
| Definition 1 | $X_t$ | -8.346 | <0.01 | -1.527 | 0.755 | -1.813 | 0.645 | -2.850 | 0.245 |
| | $X_{t+1} - X_t$ | -6.974 | <0.01 | -5.522 | <0.01 | -3.880 | 0.027 | -3.516 | 0.059 |
| Definition 2 | $X_t$ | -1.208 | 0.878 | -4.294 | <0.01 | -2.018 | 0.566 | -1.768 | 0.662 |
| | $X_{t+1} - X_t$ | -3.336 | 0.084 | -2.599 | 0.342 | -3.340 | 0.084 | -3.296 | 0.090 |
| Definition 3 | $X_t$ | -1.325 | 0.833 | -2.264 | 0.471 | 0.098 | 0.999 | -2.688 | 0.307 |
| | $X_{t+1} - X_t$ | -3.590 | 0.048 | -4.584 | <0.01 | -2.209 | 0.492 | -2.008 | 0.569 |

Table 6.2: The results of the augmented Dickey-Fuller test

| | British Columbia | | Ontario | | Quebec | | Alberta | |
|---|---|---|---|---|---|---|---|---|
| Definition | Differencing | lag $p$ | Differencing | lag $p$ | Differencing | lag $p$ | Differencing | lag $p$ |
| Definition 1 | 1 degree | 1 | 1 degree | 1 | 1 degree | 1 | 1 degree | 1 |
| | no differencing | 2 | - | - | - | - | - | - |
| Definition 2 | 1 degree | 2 | no differencing | 2 | 1 degree | 2 | 1 degree | 1 |
| Definition 3 | 1 degree | 1 | 1 degree | 4 | - | - | - | - |

Table 6.3: The parameter values of $\sigma_e^2$ or $\sigma_u^2$ for the measurement error model (6.7) or (6.9) that are used for sensitivity analyses

| Definition | Error Model | British Columbia | | Ontario | | Quebec | | Alberta | |
|---|---|---|---|---|---|---|---|---|---|
| Definition 1 | | AR(1) | | AR(1) | | AR(1) | | AR(1) | |
| | Additive ($\sigma_e^2$) | 0.1 | 0.2 | 0.5 | 1 | 0.5 | 1 | 0.1 | 0.3 |
| | Multiplicative ($\sigma_u^2$) | 0.3 | 0.6 | 0.5 | 1 | 0.5 | 1 | 0.4 | 0.8 |
| | | AR(2)* | | - | | - | | - | |
| | Additive ($\sigma_e^2$) | 0.1 | 0.2 | - | - | - | - | - | - |
| | Multiplicative ($\sigma_u^2$) | 0.01 | 0.02 | - | - | - | - | - | - |
| Definition 2 | | AR(2) | | AR(2)* | | AR(2) | | AR(1) | |
| | Additive ($\sigma_e^2$) | 0.05 | 0.1 | 0.05 | 0.1 | 0.1 | 0.2 | 0.05 | 0.1 |
| | Multiplicative ($\sigma_u^2$) | 0.2 | 0.5 | 0.005 | 0.01 | 0.3 | 0.6 | 0.4 | 0.8 |
| Definition 3 | | AR(2) | | AR(4) | | - | | - | |
| | Additive ($\sigma_e^2$) | 0.03 | 0.06 | 0.02 | 0.05 | - | - | - | - |
| | Multiplicative ($\sigma_u^2$) | 0.3 | 0.6 | 0.1 | 0.2 | - | - | - | - |

* The time series with no differencing

Table 6.4: Definition 1: The parameter estimation under different measurement error models: the AR(1) model with "order-1 differencing" is used to fit the data of British Columbia, Ontario, Quebec and Alberta

| | | | British Columbia | | | Ontario | | | Quebec | | | Alberta | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Method | Error Degree | Parameter | EST | SE | p-value | EST | SE | p-value | EST | SE | p-value | EST | SE | p-value |
| Naive | - | $\phi_0$ | -0.050 | 0.043 | 0.272 | -0.215 | 0.243 | 0.384 | -0.340 | 0.180 | 0.071 | -0.031 | 0.016 | 0.061 |
| | | $\phi_1$ | 0.138 | 0.214 | 0.533 | 0.215 | 0.157 | 0.183 | 0.012 | 0.124 | 0.923 | 0.052 | 0.144 | 0.721 |
| The Proposed Method with Additive Error | Small | $\phi_0$ | -0.027 | 0.025 | 0.313 | -0.113 | 0.134 | 0.406 | -0.183 | 0.111 | 0.112 | -0.017 | 0.009 | 0.088 |
| | $(\sigma_{e1}^2)$ | $\phi_1$ | 0.146 | 0.532 | 0.788 | 0.237 | 0.280 | 0.406 | 0.014 | 1.566 | 0.993 | 0.056 | 0.185 | 0.764 |
| | Large | $\phi_0$ | -0.027 | 0.025 | 0.298 | -0.097 | 0.263 | 0.715 | -0.181 | 0.100 | 0.083 | -0.014 | 0.073 | 0.845 |
| | $(\sigma_{e2}^2)$ | $\phi_1$ | 0.146 | 0.468 | 0.760 | 0.345 | 0.939 | 0.717 | 0.027 | 0.323 | 0.934 | 0.183 | 1.596 | 0.909 |
| The Proposed Method with Multiplicative Error | Small | $\phi_0$ | -0.027 | 0.024 | 0.286 | -0.107 | 0.152 | 0.488 | -0.183 | 0.099 | 0.078 | -0.017 | 0.009 | 0.080 |
| | $(\sigma_{u1}^2)$ | $\phi_1$ | 0.151 | 0.236 | 0.535 | 0.275 | 0.238 | 0.260 | 0.016 | 0.166 | 0.923 | 0.060 | 0.180 | 0.740 |
| | Large | $\phi_0$ | -0.025 | 0.024 | 0.308 | -0.078 | 1.690 | 0.964 | -0.180 | 0.127 | 0.170 | -0.016 | 0.015 | 0.299 |
| | $(\sigma_{u2}^2)$ | $\phi_1$ | 0.192 | 0.300 | 0.535 | 0.476 | 3.955 | 0.905 | 0.031 | 1.327 | 0.981 | 0.087 | 0.360 | 0.812 |

Table 6.5: Definition 2: The parameter estimation under different measurement error models: the AR(2) model with "no differencing" is used to fit the data of Ontario, the AR(1) model with "order-1 differencing" is used to fit the data of Alberta, and the AR(2) model with "order-1 differencing" is used to fit the data of British Columbia and Quebec

| | | | British Columbia | | | Ontario | | | Quebec | | | Alberta | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Method | Error Degree | Parameter | EST | SE | p-value | EST | SE | p-value | EST | SE | p-value | EST | SE | p-value |
| Naive | - | $\phi_0$ | 0.062 | 0.034 | 0.097 | 2.126 | 1.388 | 0.138 | 0.225 | 0.058 | 0.001 | -0.013 | 0.022 | 0.561 |
| | | $\phi_1$ | -0.415 | 0.186 | 0.046 | 1.167 | 0.209 | <0.001 | -0.122 | 0.136 | 0.380 | -0.124 | 0.172 | 0.477 |
| | | $\phi_2$ | -0.254 | 0.185 | 0.195 | -0.370 | 0.140 | 0.014 | -0.309 | 0.092 | 0.003 | - | - | - |
| The Proposed Method with Additive Error | Small ($\sigma_{e1}^2$) | $\phi_0$ | 0.034 | 0.020 | 0.114 | 1.146 | 0.759 | 0.144 | 0.174 | 0.042 | 0.000 | -0.007 | 0.012 | 0.567 |
| | | $\phi_1$ | -0.432 | 0.201 | 0.053 | 1.173 | 0.216 | <0.001 | 0.124 | 0.032 | 0.001 | -0.131 | 0.185 | 0.486 |
| | | $\phi_2$ | -0.268 | 0.205 | 0.215 | -0.375 | 0.141 | 0.014 | -0.130 | 0.165 | 0.435 | - | - | - |
| | Large ($\sigma_{e2}^2$) | $\phi_0$ | 0.036 | 0.024 | 0.164 | 1.138 | 0.747 | 0.140 | -0.327 | 0.096 | 0.002 | -0.007 | 0.012 | 0.554 |
| | | $\phi_1$ | -0.497 | 0.265 | 0.085 | 1.189 | 0.239 | <0.001 | 0.162 | 0.044 | 0.001 | -0.158 | 0.247 | 0.529 |
| | | $\phi_2$ | -0.320 | 0.354 | 0.384 | -0.390 | 0.172 | 0.032 | 0.132 | 0.041 | 0.004 | - | - | - |
| The Proposed Method with Multiplicative Error | Small ($\sigma_{u1}^2$) | $\phi_0$ | 0.034 | 0.020 | 0.115 | 1.139 | 0.748 | 0.141 | -0.164 | 0.229 | 0.480 | -0.007 | 0.012 | 0.564 |
| | | $\phi_1$ | -0.439 | 0.205 | 0.053 | 1.188 | 0.231 | <0.001 | -0.394 | 0.199 | 0.059 | -0.144 | 0.205 | 0.487 |
| | | $\phi_2$ | -0.273 | 0.204 | 0.205 | -0.389 | 0.162 | 0.024 | 0.128 | 0.042 | 0.006 | - | - | - |
| | Large ($\sigma_{u2}^2$) | $\phi_0$ | 0.039 | 0.032 | 0.236 | 1.112 | 0.747 | 0.149 | 0.127 | 0.036 | 0.002 | -0.008 | 0.012 | 0.546 |
| | | $\phi_1$ | -0.584 | 0.339 | 0.111 | 1.255 | 0.503 | 0.020 | -0.143 | 0.194 | 0.467 | -0.205 | 0.317 | 0.524 |
| | | $\phi_2$ | -0.393 | 0.322 | 0.245 | -0.451 | 0.510 | 0.384 | -0.353 | 0.111 | 0.004 | - | - | - |

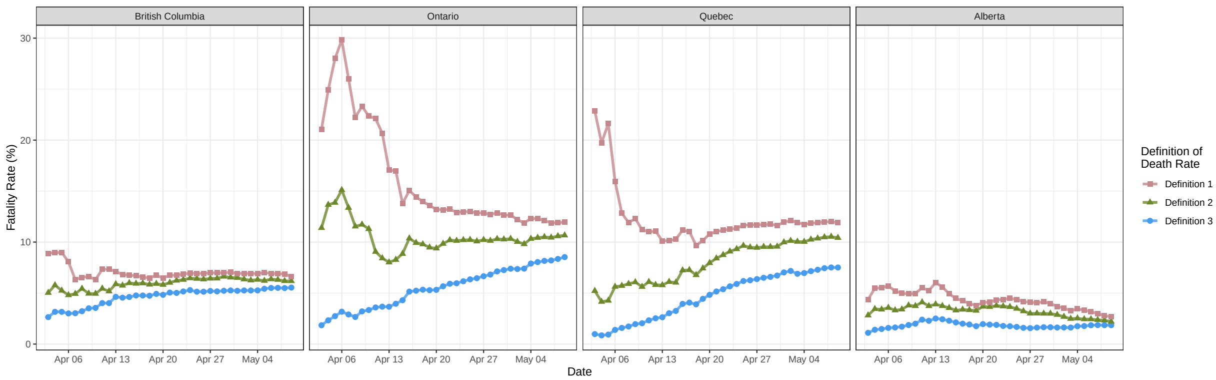


Figure 6.1: *The time series plots of the death rate with different definitions*

Table 6.6: Definition 3: The parameter estimation under different measurement error models: the AR(1) model with "order-1 differencing" is used to fit the data of British Columbia and the AR(4) model with "order-1 differencing" is used to fit the data of Ontario

| | | | British Columbia | | | Ontario | | |
|---|---|---|---|---|---|---|---|---|
| Method | Error Degree | Parameter | EST | SE | p-value | EST | SE | p-value |
| Naive | - | $\phi_0$ | 0.105 | 0.038 | 0.018 | 0.379 | 0.057 | <0.001 |
| | | $\phi_1$ | -0.207 | 0.077 | 0.020 | -0.086 | 0.099 | 0.391 |
| | | $\phi_2$ | - | - | - | -0.287 | 0.106 | 0.012 |
| | | $\phi_3$ | - | - | - | -0.301 | 0.094 | 0.004 |
| | | $\phi_4$ | - | - | - | -0.284 | 0.078 | 0.001 |
| The Proposed Method with Additive Error | Small ($\sigma^2_{e1}$) | $\phi_0$ | 0.057 | 0.021 | 0.021 | 0.206 | 0.031 | <0.001 |
| | | $\phi_1$ | -0.213 | 0.086 | 0.029 | -0.088 | 0.100 | 0.383 |
| | | $\phi_2$ | - | - | - | -0.290 | 0.109 | 0.014 |
| | | $\phi_3$ | - | - | - | -0.303 | 0.094 | 0.003 |
| | | $\phi_4$ | - | - | - | -0.287 | 0.081 | 0.002 |
| | Large ($\sigma^2_{e2}$) | $\phi_0$ | 0.058 | 0.021 | 0.017 | 0.212 | 0.036 | <0.001 |
| | | $\phi_1$ | -0.234 | 0.147 | 0.137 | -0.102 | 0.123 | 0.417 |
| | | $\phi_2$ | - | - | - | -0.306 | 0.139 | 0.037 |
| | | $\phi_3$ | - | - | - | -0.318 | 0.107 | 0.006 |
| | | $\phi_4$ | - | - | - | -0.308 | 0.093 | 0.003 |
| The Proposed Method with Multiplicative Error | Small ($\sigma^2_{u1}$) | $\phi_0$ | 0.058 | 0.023 | 0.027 | 0.210 | 0.033 | <0.001 |
| | | $\phi_1$ | -0.244 | 0.090 | 0.019 | -0.097 | 0.107 | 0.375 |
| | | $\phi_2$ | - | - | - | -0.300 | 0.117 | 0.016 |
| | | $\phi_3$ | - | - | - | -0.312 | 0.098 | 0.004 |
| | | $\phi_4$ | - | - | - | -0.300 | 0.087 | 0.002 |
| | Large ($\sigma^2_{u2}$) | $\phi_0$ | 0.066 | 0.035 | 0.087 | 0.230 | 0.058 | 0.001 |
| | | $\phi_1$ | -0.401 | 0.219 | 0.092 | -0.139 | 0.183 | 0.454 |
| | | $\phi_2$ | - | - | - | -0.347 | 0.213 | 0.116 |
| | | $\phi_3$ | - | - | - | -0.354 | 0.159 | 0.035 |
| | | $\phi_4$ | - | - | - | -0.361 | 0.149 | 0.023 |

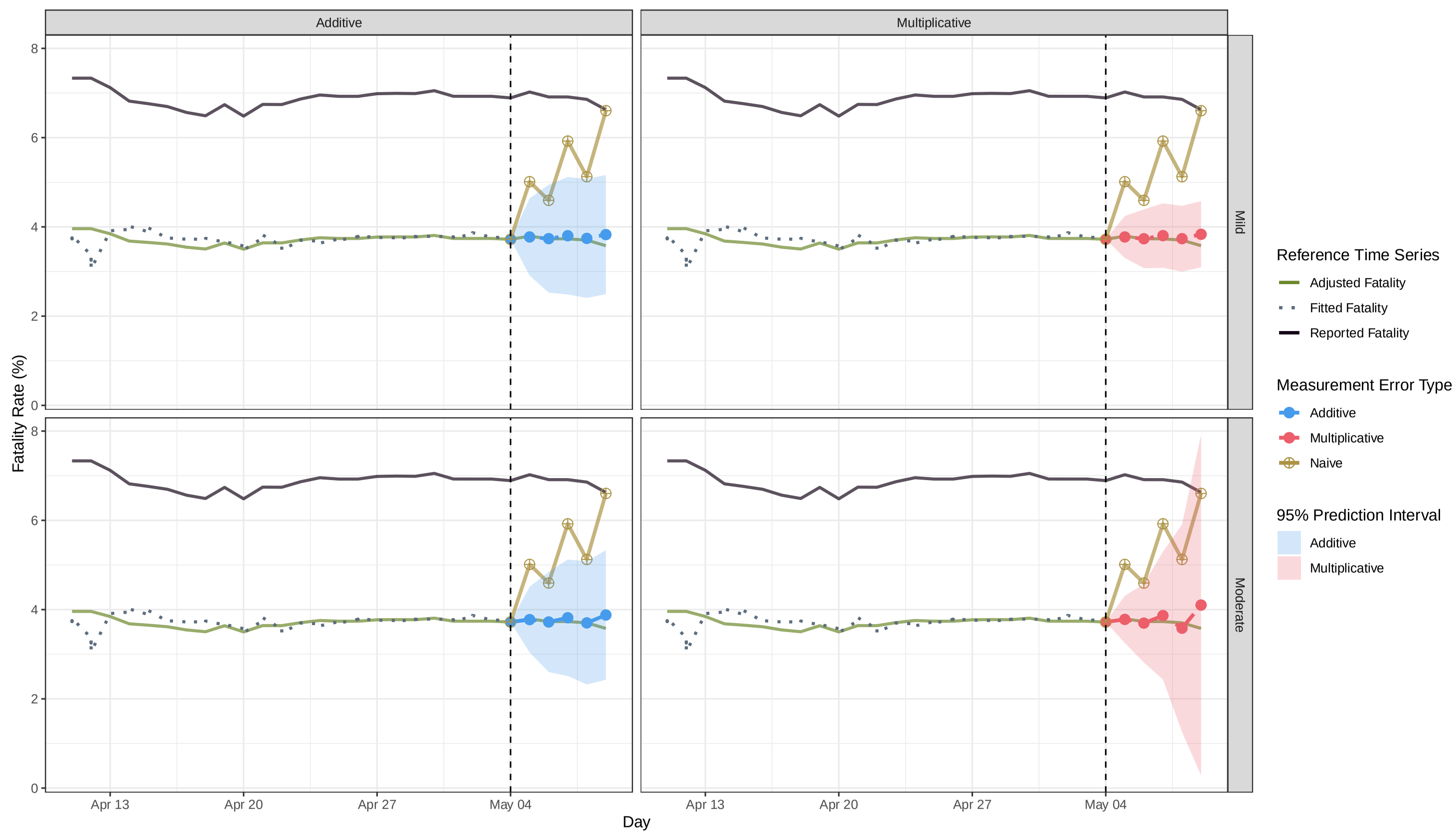


Figure 6.2: *British Columbia by Definition 1 (AR(2), no differencing): A 5-day forecasting of the true mortality rate (May 5 - May 9) based on the additive (in blue) or multiplicative (in red) versus the naive model (in dark yellow); the reported mortality rates (in black) and the adjusted true mortality rate accounting for the asymptomatic cases (in green).*

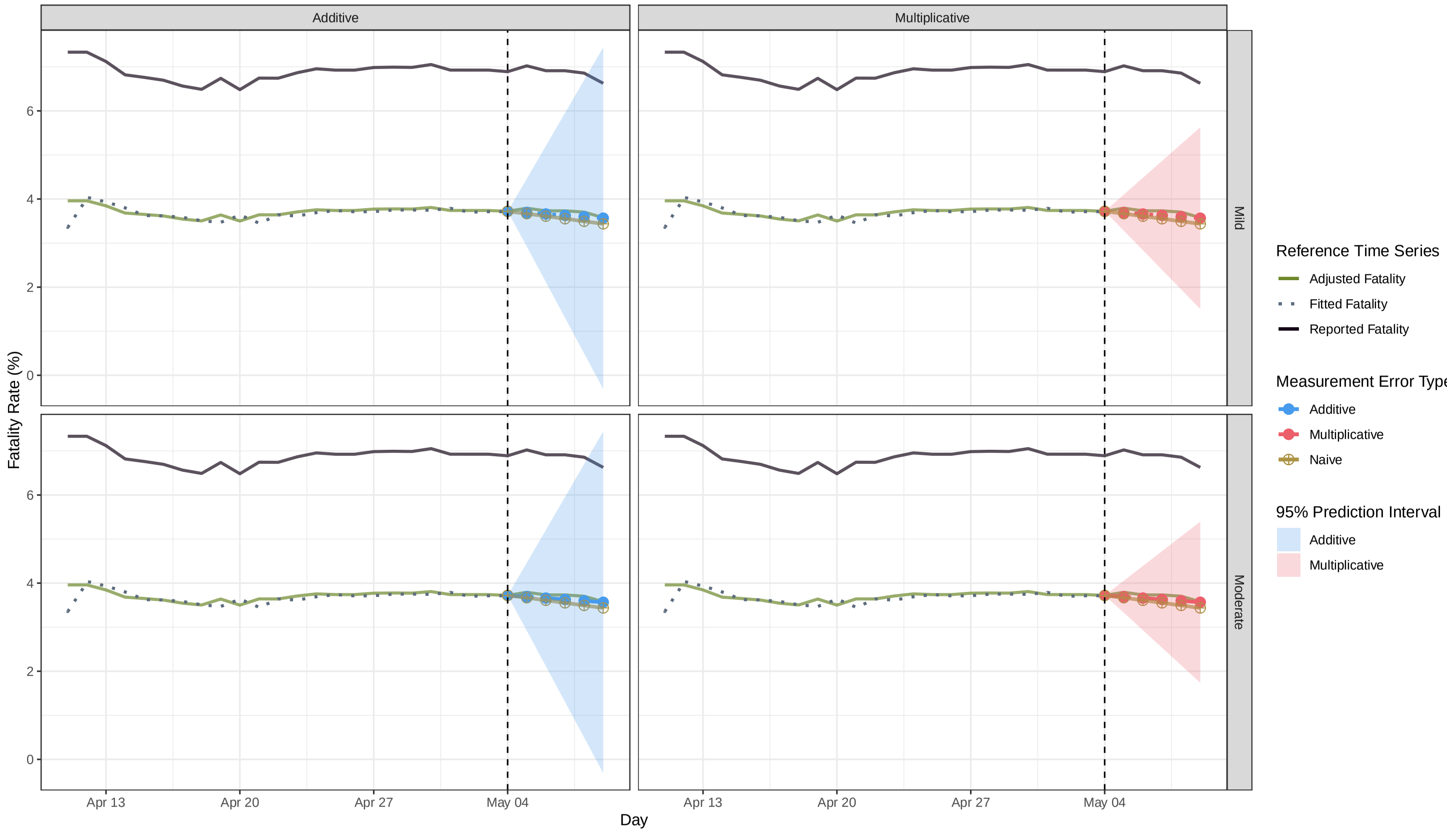


Figure 6.3: *British Columbia by Definition 1 (AR(1), order-1 differencing): A 5-day forecasting of the true mortality rate (May 5 - May 9) based on the additive (in blue) or multiplicative (in red) versus the naive model (in dark yellow); the reported mortality rates (in black) and the adjusted true mortality rate accounting for the asymptomatic cases (in green).*

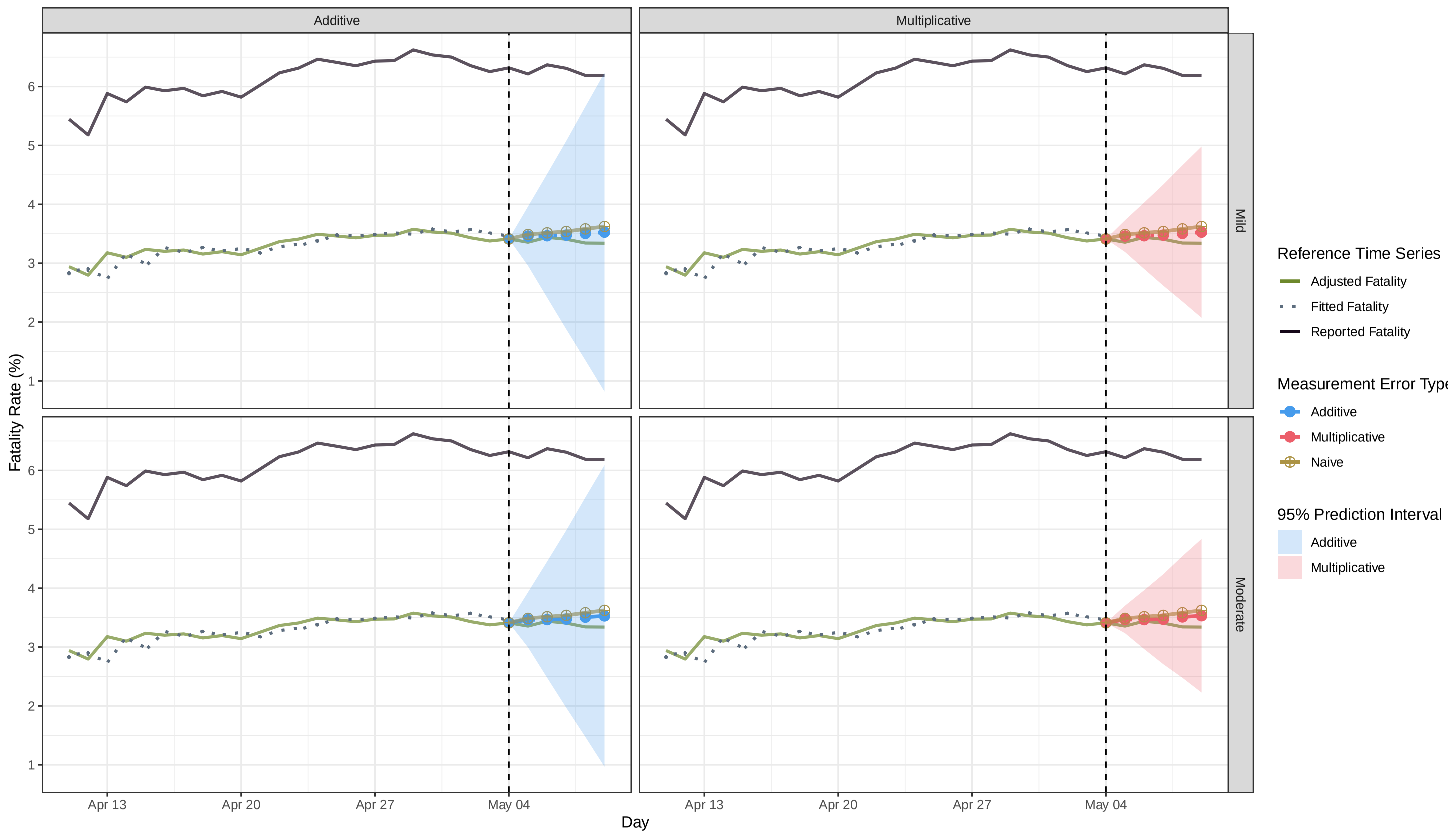


Figure 6.4: *British Columbia by Definition 2 (AR(3), order-1 differencing): A 5-day forecasting of the true mortality rate (May 5 - May 9) based on the additive (in blue) or multiplicative (in red) versus the naive model (in dark yellow); the reported mortality rates (in black) and the adjusted true mortality rate accounting for the asymptomatic cases (in green).*

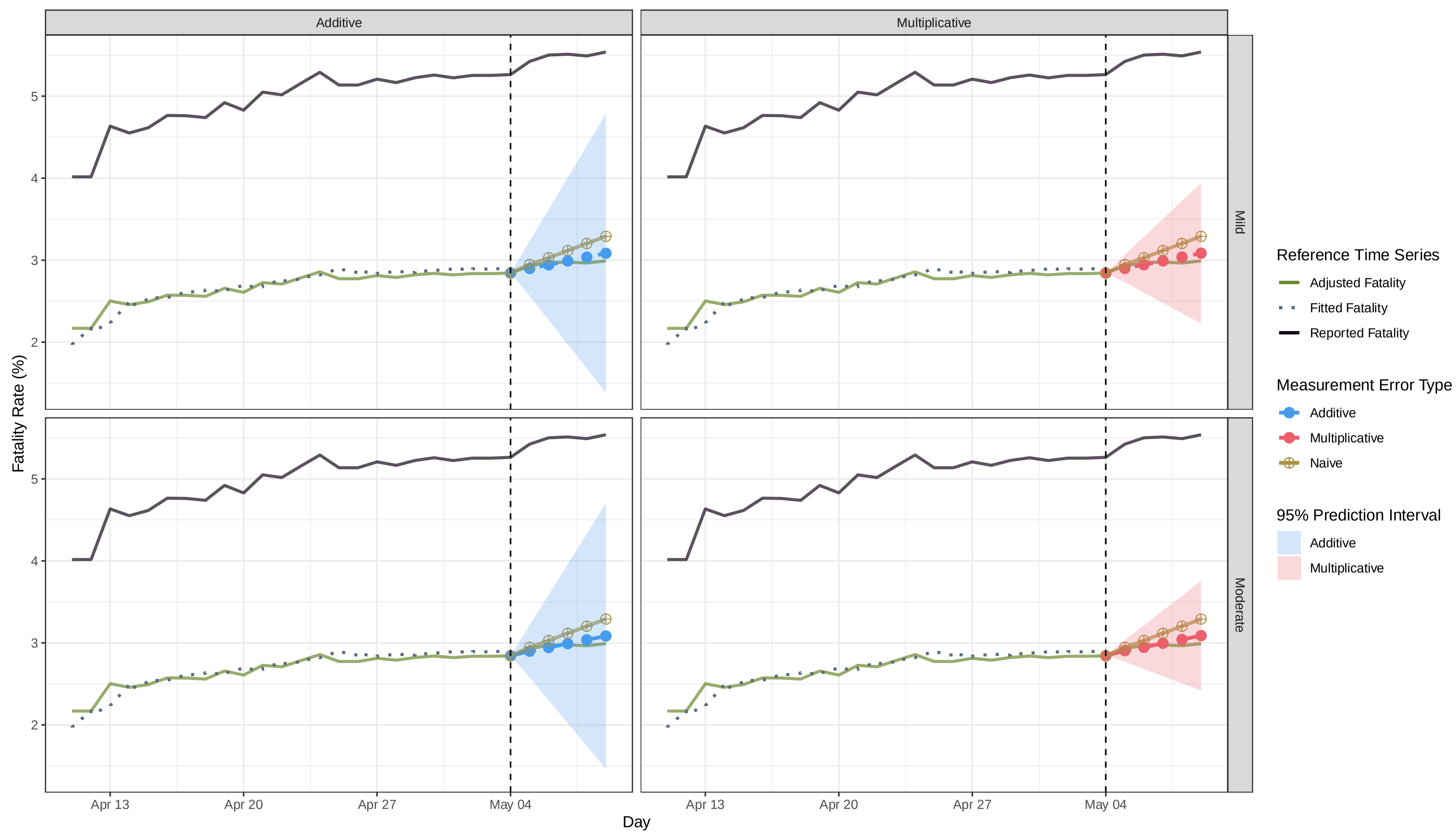


Figure 6.5: *British Columbia by Definition 3 (AR(1), order-1 differencing): A 5-day forecasting of the true mortality rate (May 5 - May 9) based on the additive (in blue) or multiplicative (in red) versus the naive model (in dark yellow); the reported mortality rates (in black) and the adjusted true mortality rate accounting for the asymptomatic cases (in green).*

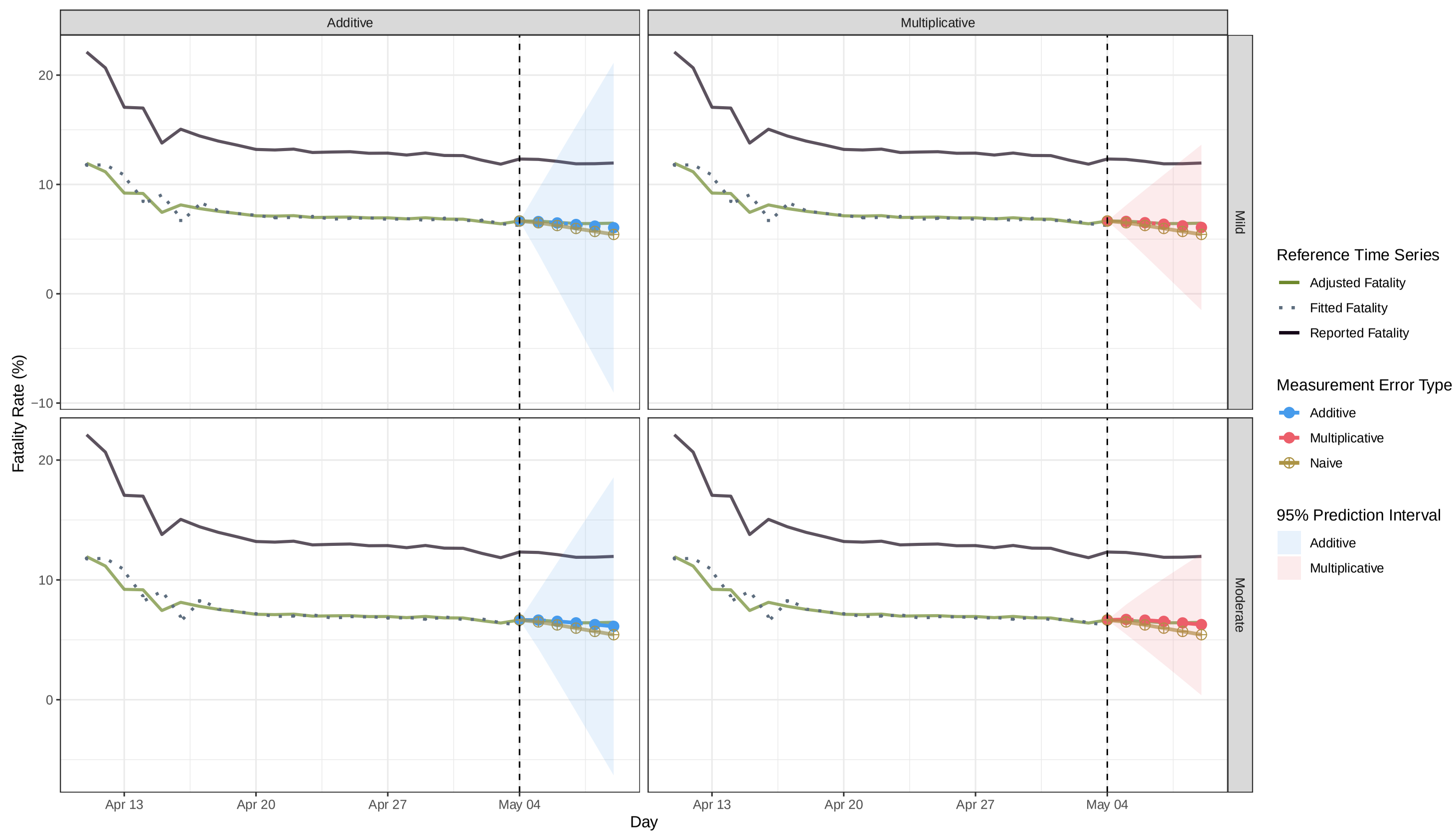


Figure 6.6: *Ontario by Definition 1 (AR(1), order-1 differencing): A 5-day forecasting of the true mortality rate (May 5 - May 9) based on the additive (in blue) or multiplicative (in red) versus the naive model (in dark yellow); the reported mortality rates (in black) and the adjusted true mortality rate accounting for the asymptomatic cases (in green).*

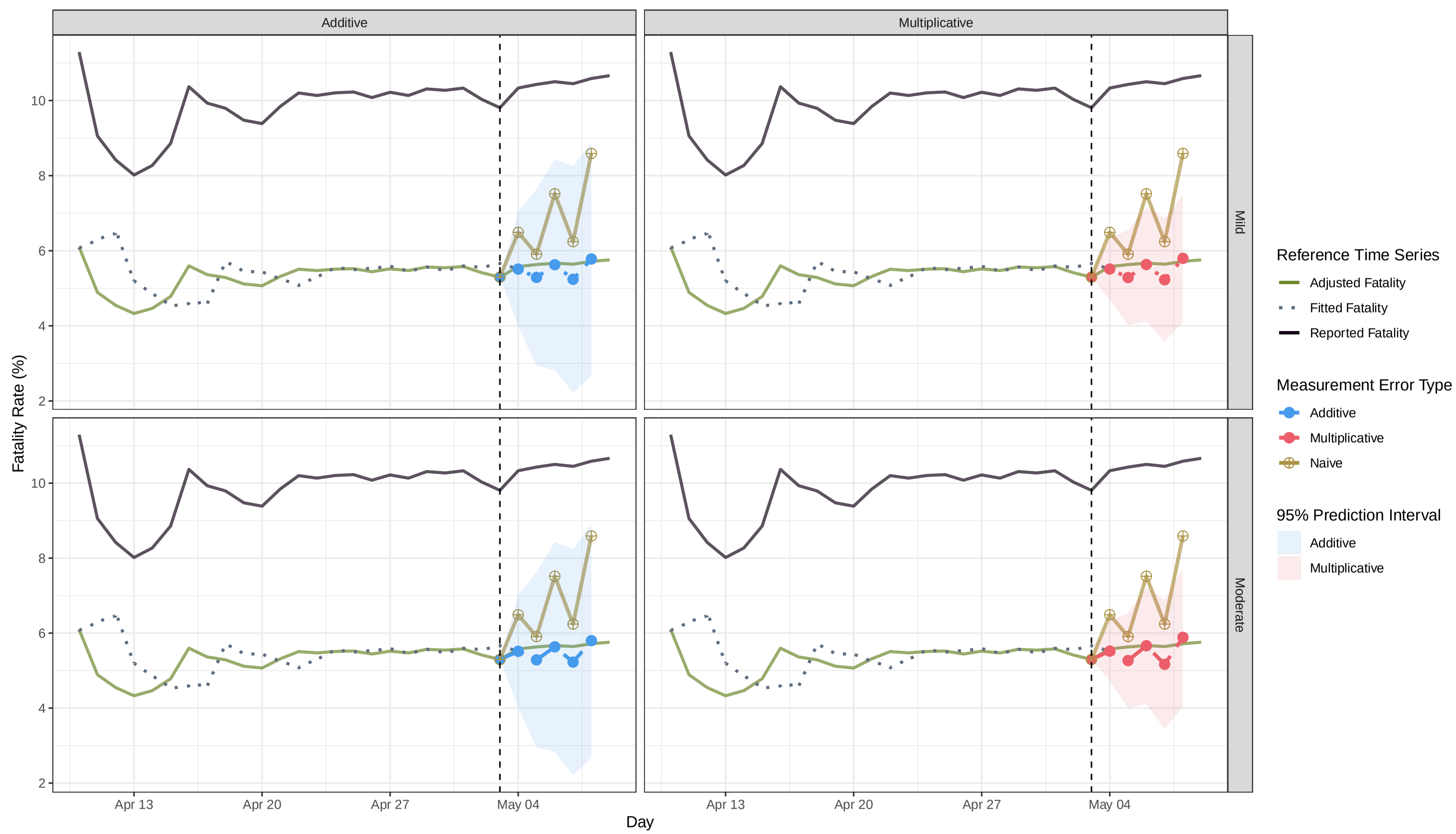


Figure 6.7: *Ontario by Definition 2 (AR(1), no differencing): A 5-day forecasting of the true mortality rate (May 5 - May 9) based on the additive (in blue) or multiplicative (in red) versus the naive model (in dark yellow); the reported mortality rates (in black) and the adjusted true mortality rate accounting for the asymptomatic cases (in green).*

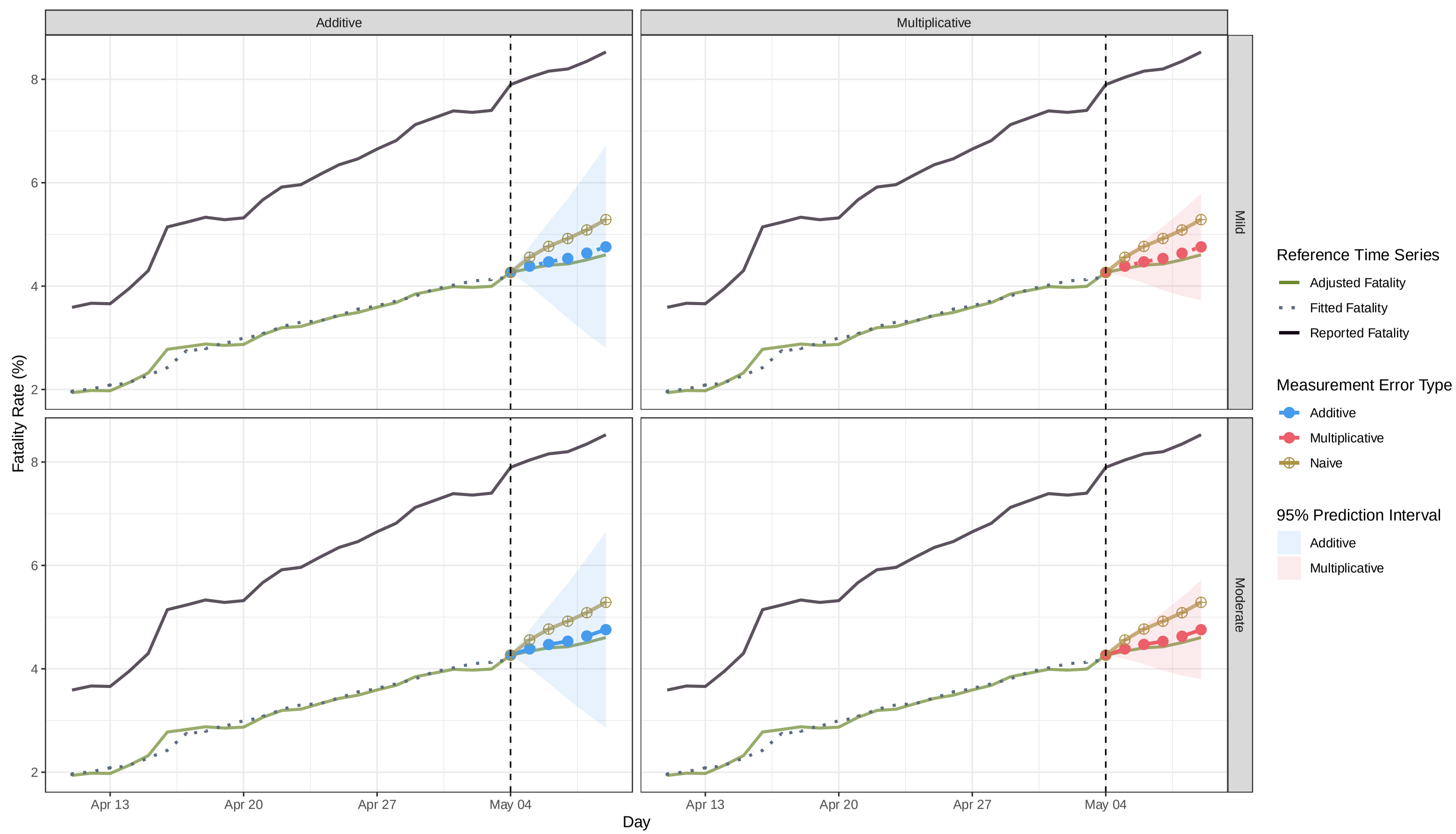


Figure 6.8: *Ontario by Definition 3 (AR(4), order-1 differencing): A 5-day forecasting of the true mortality rate (May 5 - May 9) based on the additive (in blue) or multiplicative (in red) versus the naive model (in dark yellow); the reported mortality rates (in black) and the adjusted true mortality rate accounting for the asymptomatic cases (in green).*

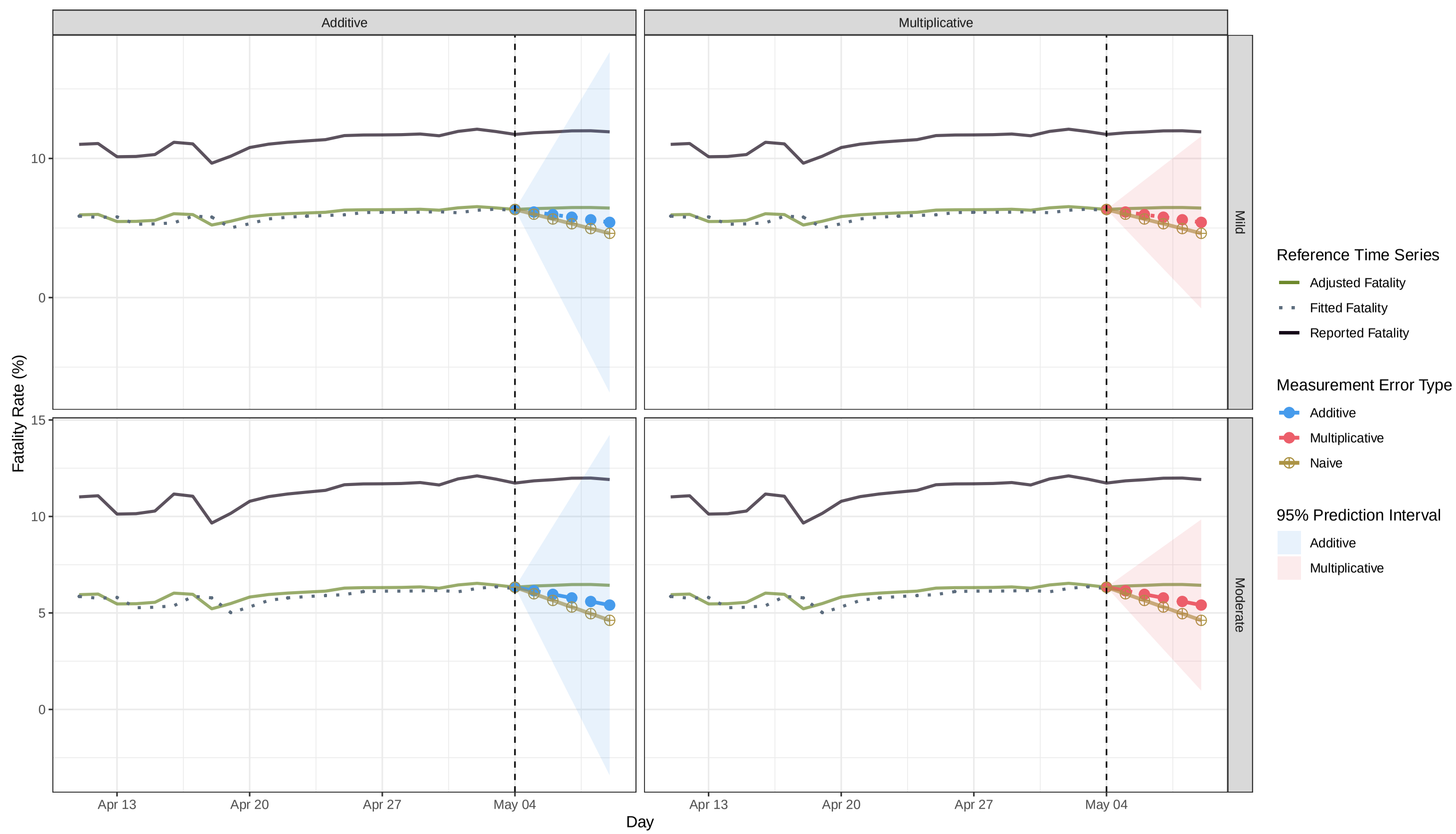


Figure 6.9: *Quebec by Definition 1 (AR(1), order-1 differencing): A 5-day forecasting of the true mortality rate (May 5 - May 9) based on the additive (in blue) or multiplicative (in red) versus the naive model (in dark yellow); the reported mortality rates (in black) and the adjusted true mortality rate accounting for the asymptomatic cases (in green).*

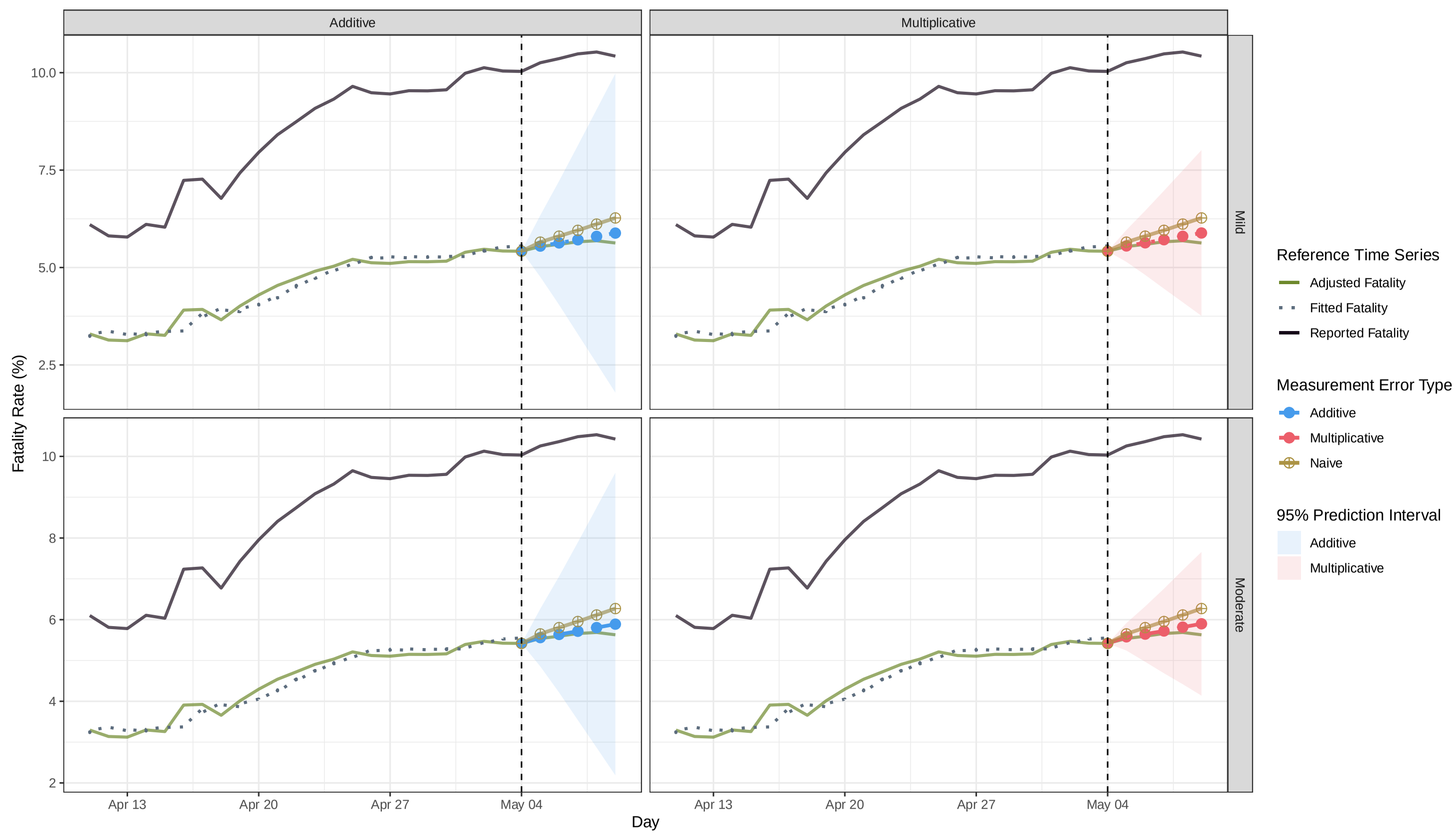


Figure 6.10: *Quebec by Definition 2 (AR(2), order-1 differencing): A 5-day forecasting of the true mortality rate (May 5 - May 9) based on the additive (in blue) or multiplicative (in red) versus the naive model (in dark yellow); the reported mortality rates (in black) and the adjusted true mortality rate accounting for the asymptomatic cases (in green).*

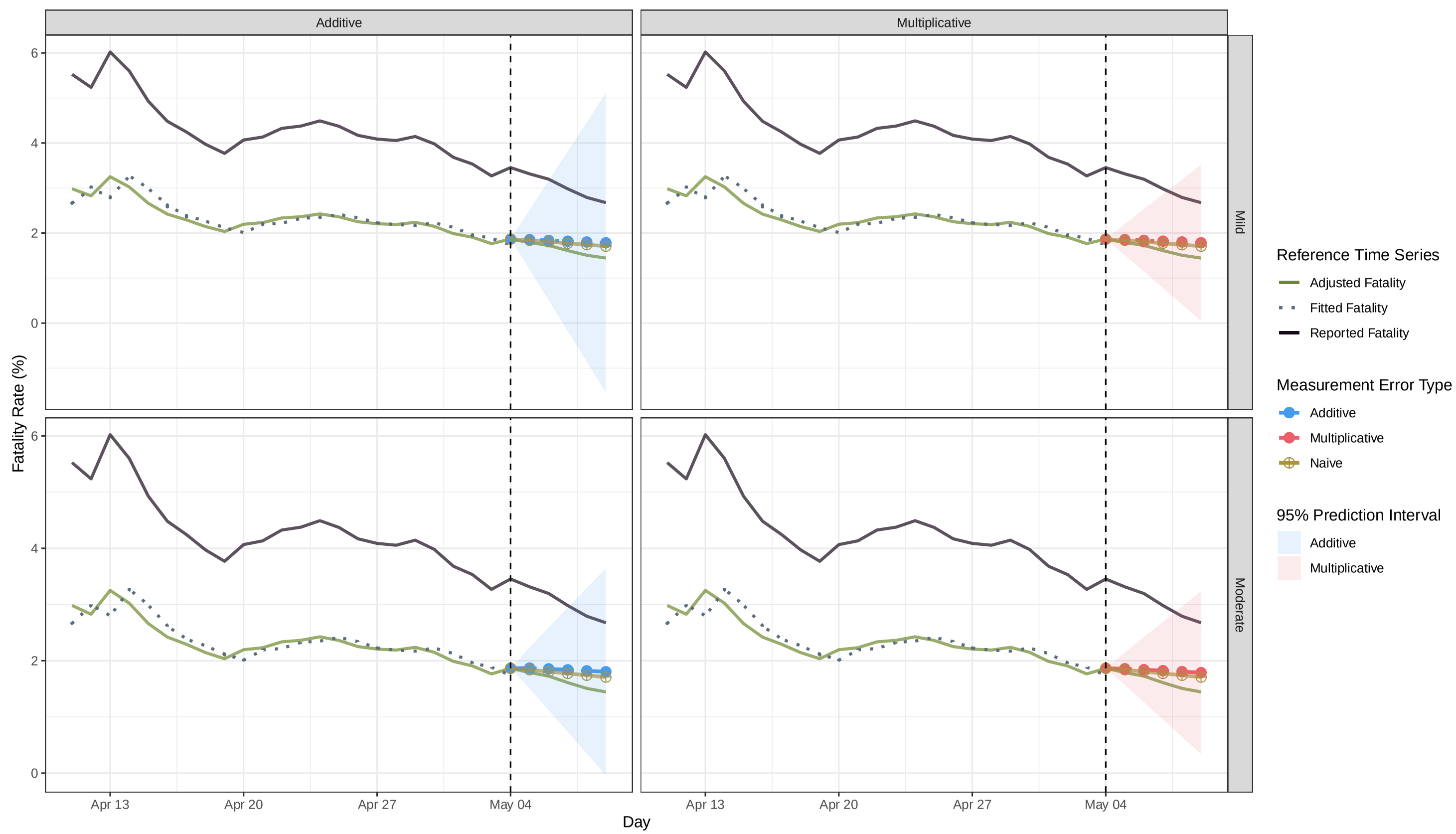


Figure 6.11: *Alberta by Definition 1 (AR(1), order-1 differencing): A 5-day forecasting of the true mortality rate (May 5 - May 9) based on the additive (in blue) or multiplicative (in red) versus the naive model (in dark yellow); the reported mortality rates (in black) and the adjusted true mortality rate accounting for the asymptomatic cases (in green).*

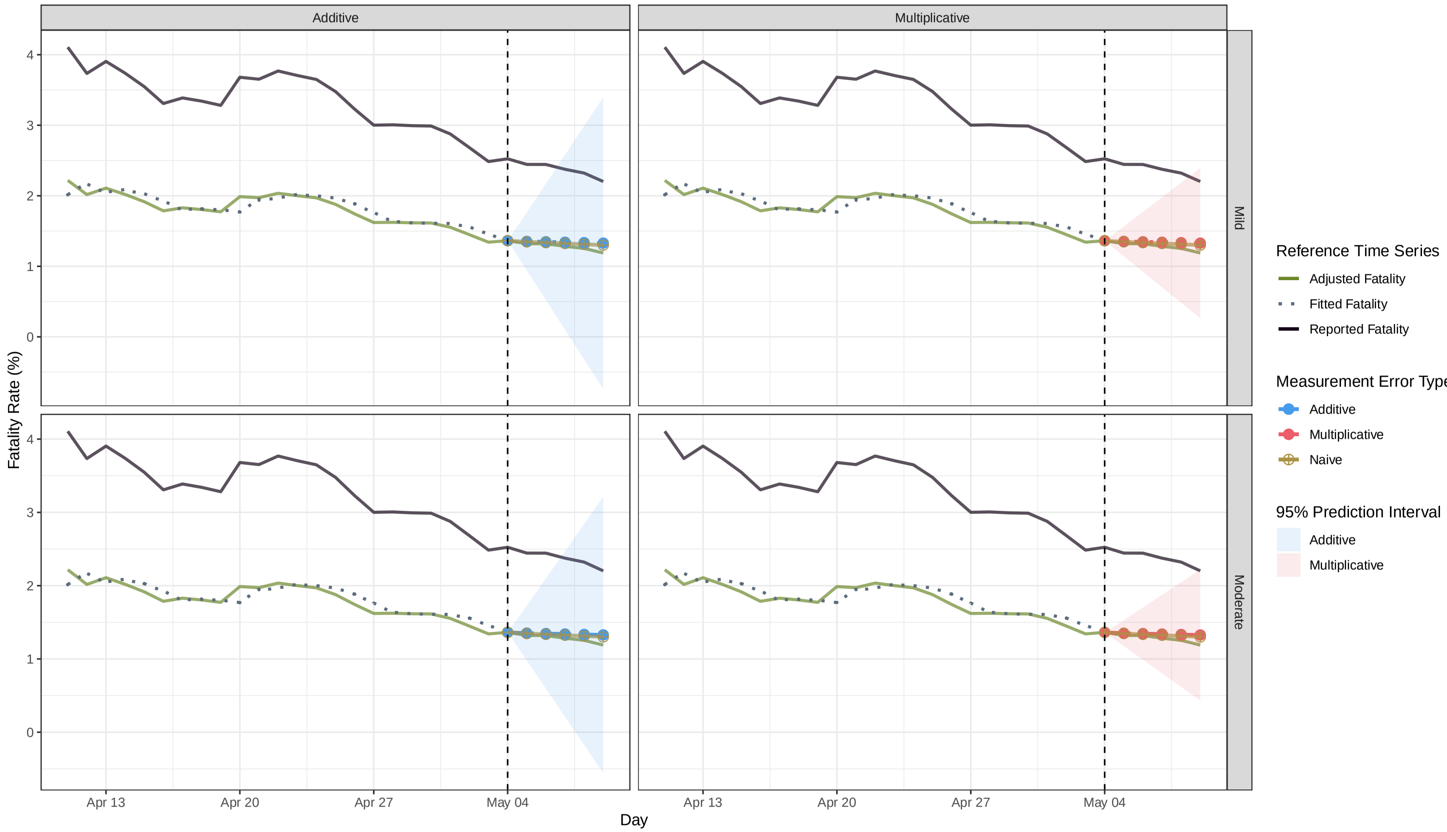


Figure 6.12: *Alberta by Definition 2 (AR(1), order-1 differencing): A 5-day forecasting of the true mortality rate (May 5 - May 9) based on the additive (in blue) or multiplicative (in red) versus the naive model (in dark yellow); the reported mortality rates (in black) and the adjusted true mortality rate accounting for the asymptomatic cases (in green).*

Table 6.7: Definition 1: The observed prediction error and expected prediction error for different definition of death rates

| | | Observed Prediction Error | | | | | | Expected Prediction Error | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Method | $\sigma_e^2$ (or $\sigma_u^2$) | Day 1 | Day 2 | Day 3 | Day 4 | Day 5 | $\sum_{h=1}^{H}$ OPE($h$) | Day 1 | Day 2 | Day 3 | Day 4 | Day 5 | $\sum_{h=1}^{H}$ EPE($h$) |
| | | Definition 1 | | | | | | | | | | | |
| | | British Columbia | | | | | | | | | | | |
| Naive | - | 0.017 | 0.006 | 0.017 | 0.058 | 0.081 | 0.178 | 0.069 | 0.081 | 0.081 | 0.082 | 0.083 | 0.396 |
| Additive | Mild | 0.011 | 0.001 | 0.005 | 0.027 | 0.035 | 0.078 | 0.066 | 0.078 | 0.078 | 0.080 | 0.080 | 0.382 |
| | Moderate | 0.012 | 0.001 | 0.005 | 0.027 | 0.036 | 0.080 | 0.057 | 0.070 | 0.070 | 0.072 | 0.073 | 0.342 |
| Multiplicative | Mild | 0.011 | 0.001 | 0.005 | 0.027 | 0.035 | 0.078 | 0.020 | 0.023 | 0.023 | 0.023 | 0.023 | 0.111 |
| | Moderate | 0.013 | 0.001 | 0.004 | 0.029 | 0.037 | 0.083 | 0.015 | 0.019 | 0.018 | 0.019 | 0.019 | 0.090 |
| | | Ontario | | | | | | | | | | | |
| Naive | - | 0.830 | 0.077 | 3.409 | 0.360 | 8.264 | 12.940 | 0.612 | 1.446 | 2.048 | 2.372 | 2.514 | 8.991 |
| Additive | Mild | 0.004 | 0.116 | 0.002 | 0.161 | 0.004 | 0.288 | 0.607 | 1.440 | 2.046 | 2.373 | 2.517 | 8.983 |
| | Moderate | 0.004 | 0.119 | 0.001 | 0.172 | 0.007 | 0.304 | 0.591 | 1.422 | 2.040 | 2.378 | 2.531 | 8.963 |
| Multiplicative | Mild | 0.004 | 0.119 | 0.001 | 0.171 | 0.007 | 0.302 | 0.176 | 0.420 | 0.602 | 0.704 | 0.754 | 2.655 |
| | Moderate | 0.003 | 0.132 | 0.000 | 0.225 | 0.029 | 0.389 | 0.169 | 0.418 | 0.630 | 0.775 | 0.888 | 2.879 |
| | | Quebec | | | | | | | | | | | |
| Naive | - | 0.163 | 0.607 | 1.357 | 2.289 | 3.294 | 7.709 | 1.811 | 1.811 | 1.811 | 1.811 | 1.811 | 9.057 |
| Additive | Mild | 0.061 | 0.216 | 0.479 | 0.778 | 1.053 | 2.587 | 1.561 | 1.561 | 1.561 | 1.561 | 1.561 | 7.807 |
| | Moderate | 0.060 | 0.215 | 0.478 | 0.776 | 1.051 | 2.580 | 0.811 | 0.811 | 0.811 | 0.811 | 0.811 | 4.057 |
| Multiplicative | Mild | 0.061 | 0.216 | 0.479 | 0.778 | 1.053 | 2.586 | 0.399 | 0.399 | 0.399 | 0.399 | 0.399 | 1.995 |
| | Moderate | 0.060 | 0.215 | 0.477 | 0.776 | 1.050 | 2.578 | 0.205 | 0.205 | 0.205 | 0.205 | 0.205 | 1.025 |
| | | Alberta | | | | | | | | | | | |
| Naive | - | 0.002 | 0.007 | 0.027 | 0.055 | 0.070 | 0.160 | 0.125 | 0.125 | 0.125 | 0.125 | 0.125 | 0.627 |
| Additive | Mild | 0.004 | 0.012 | 0.044 | 0.087 | 0.115 | 0.262 | 0.115 | 0.115 | 0.115 | 0.115 | 0.115 | 0.577 |
| | Moderate | 0.006 | 0.017 | 0.052 | 0.098 | 0.129 | 0.302 | 0.035 | 0.035 | 0.035 | 0.035 | 0.035 | 0.177 |
| Multiplicative | Mild | 0.004 | 0.012 | 0.044 | 0.087 | 0.115 | 0.263 | 0.031 | 0.031 | 0.031 | 0.031 | 0.031 | 0.157 |
| | Moderate | 0.005 | 0.013 | 0.045 | 0.089 | 0.118 | 0.270 | 0.022 | 0.022 | 0.022 | 0.022 | 0.022 | 0.109 |

Table 6.8: Definition 2: The observed prediction error and expected prediction error for different definition of death rates

| | | Observed Prediction Error | | | | | | Expected Prediction Error | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Method | $\sigma_e^2$ (or $\sigma_u^2$) | Day 1 | Day 2 | Day 3 | Day 4 | Day 5 | $\sum_{h=1}^{H}$ OPE($h$) | Day 1 | Day 2 | Day 3 | Day 4 | Day 5 | $\sum_{h=1}^{H}$ EPE($h$) |
| | | Definition 2 | | | | | | | | | | | |
| | | British Columbia | | | | | | | | | | | |
| Naive | - | 0.015 | 0.015 | 0.032 | 0.043 | 0.020 | 0.126 | 0.164 | 0.167 | 0.167 | 0.167 | 0.167 | 0.834 |
| Additive | Mild | 0.010 | 0.005 | 0.011 | 0.011 | 0.000 | 0.037 | 0.154 | 0.157 | 0.157 | 0.157 | 0.157 | 0.783 |
| | Moderate | 0.010 | 0.005 | 0.011 | 0.011 | 0.000 | 0.037 | 0.154 | 0.157 | 0.157 | 0.157 | 0.157 | 0.784 |
| Multiplicative | Mild | 0.010 | 0.005 | 0.011 | 0.011 | 0.000 | 0.037 | 0.044 | 0.044 | 0.044 | 0.044 | 0.044 | 0.222 |
| | Moderate | 0.010 | 0.005 | 0.011 | 0.011 | 0.000 | 0.037 | 0.034 | 0.035 | 0.035 | 0.035 | 0.035 | 0.174 |
| | | Ontario | | | | | | | | | | | |
| Naive | - | 0.020 | 0.087 | 0.196 | 0.521 | 1.059 | 1.884 | 2.527 | 2.643 | 2.649 | 2.649 | 2.649 | 13.117 |
| Additive | Mild | 0.001 | 0.004 | 0.007 | 0.056 | 0.175 | 0.243 | 2.264 | 2.391 | 2.399 | 2.399 | 2.399 | 11.853 |
| | Moderate | 0.000 | 0.000 | 0.000 | 0.023 | 0.110 | 0.134 | 1.453 | 1.626 | 1.646 | 1.649 | 1.649 | 8.023 |
| Multiplicative | Mild | 0.000 | 0.002 | 0.003 | 0.044 | 0.152 | 0.201 | 0.558 | 0.599 | 0.603 | 0.603 | 0.603 | 2.965 |
| | Moderate | 0.004 | 0.010 | 0.014 | 0.000 | 0.035 | 0.063 | 0.270 | 0.331 | 0.345 | 0.348 | 0.348 | 1.642 |
| | | Quebec | | | | | | | | | | | |
| Naive | - | 0.013 | 0.044 | 0.086 | 0.183 | 0.413 | 0.739 | 0.174 | 0.176 | 0.191 | 0.192 | 0.193 | 0.926 |
| Additive | Mild | 0.000 | 0.001 | 0.003 | 0.013 | 0.065 | 0.081 | 0.163 | 0.165 | 0.181 | 0.182 | 0.183 | 0.874 |
| | Moderate | 0.000 | 0.002 | 0.003 | 0.014 | 0.068 | 0.087 | 0.130 | 0.133 | 0.149 | 0.151 | 0.153 | 0.716 |
| Multiplicative | Mild | 0.000 | 0.002 | 0.003 | 0.013 | 0.066 | 0.084 | 0.044 | 0.045 | 0.049 | 0.049 | 0.050 | 0.236 |
| | Moderate | 0.002 | 0.003 | 0.004 | 0.017 | 0.073 | 0.098 | 0.030 | 0.030 | 0.033 | 0.034 | 0.034 | 0.162 |
| | | Alberta | | | | | | | | | | | |
| Naive | - | 0.001 | 0.000 | 0.002 | 0.003 | 0.012 | 0.019 | 0.047 | 0.047 | 0.047 | 0.047 | 0.047 | 0.236 |
| Additive | Mild | 0.001 | 0.001 | 0.003 | 0.007 | 0.019 | 0.031 | 0.044 | 0.045 | 0.045 | 0.045 | 0.045 | 0.223 |
| | Moderate | 0.001 | 0.001 | 0.003 | 0.006 | 0.019 | 0.031 | 0.036 | 0.037 | 0.037 | 0.037 | 0.037 | 0.185 |
| Multiplicative | Mild | 0.001 | 0.001 | 0.003 | 0.006 | 0.019 | 0.031 | 0.012 | 0.012 | 0.012 | 0.012 | 0.012 | 0.059 |
| | Moderate | 0.001 | 0.001 | 0.003 | 0.006 | 0.019 | 0.030 | 0.008 | 0.008 | 0.008 | 0.008 | 0.008 | 0.042 |

Table 6.9: Definition 3: The observed prediction error and expected prediction error for different definition of death rates

| | | Observed Prediction Error | | | | | | Expected Prediction Error | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Method | $\sigma_e^2$ (or $\sigma_u^2$) | Day 1 | Day 2 | Day 3 | Day 4 | Day 5 | $\sum_{h=1}^{H}$ OPE($h$) | Day 1 | Day 2 | Day 3 | Day 4 | Day 5 | $\sum_{h=1}^{H}$ EPE($h$) |
| | | | | | | | Definition 3 | | | | | | |
| | | British Columbia | | | | | | | | | | | |
| Naive | - | 0.000 | 0.003 | 0.020 | 0.057 | 0.090 | 0.170 | 0.030 | 0.031 | 0.031 | 0.031 | 0.031 | 0.155 |
| Additive | Mild | 0.001 | 0.001 | 0.000 | 0.005 | 0.009 | 0.016 | 0.029 | 0.030 | 0.030 | 0.030 | 0.030 | 0.151 |
| | Moderate | 0.001 | 0.001 | 0.000 | 0.005 | 0.009 | 0.016 | 0.026 | 0.028 | 0.028 | 0.028 | 0.028 | 0.137 |
| Multiplicative | Mild | 0.001 | 0.001 | 0.000 | 0.005 | 0.009 | 0.016 | 0.007 | 0.008 | 0.008 | 0.008 | 0.008 | 0.038 |
| | Moderate | 0.001 | 0.001 | 0.000 | 0.006 | 0.010 | 0.017 | 0.005 | 0.005 | 0.005 | 0.005 | 0.005 | 0.023 |
| | | Ontario | | | | | | | | | | | |
| Naive | - | 0.048 | 0.132 | 0.243 | 0.333 | 0.464 | 1.219 | 0.039 | 0.039 | 0.039 | 0.042 | 0.042 | 0.202 |
| Additive | Mild | 0.002 | 0.004 | 0.011 | 0.017 | 0.024 | 0.058 | 0.039 | 0.039 | 0.039 | 0.042 | 0.042 | 0.200 |
| | Moderate | 0.002 | 0.004 | 0.011 | 0.016 | 0.023 | 0.057 | 0.036 | 0.036 | 0.036 | 0.039 | 0.039 | 0.187 |
| Multiplicative | Mild | 0.002 | 0.004 | 0.011 | 0.016 | 0.023 | 0.057 | 0.011 | 0.011 | 0.011 | 0.012 | 0.012 | 0.056 |
| | Moderate | 0.002 | 0.004 | 0.011 | 0.015 | 0.023 | 0.055 | 0.009 | 0.009 | 0.009 | 0.010 | 0.010 | 0.048 |

Table 6.10: The observed prediction error and expected prediction error for different lag order of autoregressive models

| | | | Observed Prediction Error | | | | | | Expected Prediction Error | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Method | $\sigma_e^2$ (or $\sigma_u^2$) | Model | Day 1 | Day 2 | Day 3 | Day 4 | Day 5 | $\sum_{h=1}^{H}$ OPE($h$) | Day 1 | Day 2 | Day 3 | Day 4 | Day 5 | $\sum_{h=1}^{H}$ EPE($h$) |
| | | British Columbia | | | | | | | | | | | | |
| Naive | - | **AR(1)**[a] | 0.015 | 0.015 | 0.032 | 0.043 | 0.020 | 0.126 | 0.164 | 0.167 | 0.167 | 0.167 | 0.167 | 0.834 |
| Additive | Mild | | 0.010 | 0.005 | 0.011 | 0.011 | 0.000 | 0.037 | 0.154 | 0.157 | 0.157 | 0.157 | 0.157 | 0.783 |
| | Moderate | | 0.010 | 0.005 | 0.011 | 0.011 | 0.000 | 0.037 | 0.154 | 0.157 | 0.157 | 0.157 | 0.157 | 0.784 |
| Multiplicative | Mild | | 0.010 | 0.005 | 0.011 | 0.011 | 0.000 | 0.037 | 0.044 | 0.044 | 0.044 | 0.044 | 0.044 | 0.222 |
| | Moderate | | 0.010 | 0.005 | 0.011 | 0.011 | 0.000 | 0.037 | 0.034 | 0.035 | 0.035 | 0.035 | 0.035 | 0.174 |
| Naive | - | AR(2) | 0.016 | 0.014 | 0.031 | 0.042 | 0.019 | 0.122 | 0.161 | 0.165 | 0.167 | 0.167 | 0.167 | 0.828 |
| Additive | Mild | | 0.010 | 0.005 | 0.010 | 0.010 | 0.000 | 0.035 | 0.151 | 0.155 | 0.157 | 0.157 | 0.157 | 0.777 |
| | Moderate | | 0.010 | 0.005 | 0.010 | 0.010 | 0.000 | 0.035 | 0.151 | 0.155 | 0.157 | 0.157 | 0.157 | 0.778 |
| Multiplicative | Mild | | 0.010 | 0.005 | 0.010 | 0.010 | 0.000 | 0.035 | 0.043 | 0.044 | 0.044 | 0.044 | 0.044 | 0.220 |
| | Moderate | | 0.010 | 0.005 | 0.010 | 0.010 | 0.000 | 0.034 | 0.034 | 0.035 | 0.035 | 0.035 | 0.035 | 0.173 |
| | | Ontario | | | | | | | | | | | | |
| Naive | - | **AR(1)**[a] | 0.020 | 0.087 | 0.196 | 0.521 | 1.059 | 1.884 | 2.527 | 2.643 | 2.649 | 2.649 | 2.649 | 13.117 |
| Additive | Mild | | 0.001 | 0.004 | 0.007 | 0.056 | 0.175 | 0.243 | 2.264 | 2.391 | 2.399 | 2.399 | 2.399 | 11.853 |
| | Moderate | | 0.000 | 0.000 | 0.000 | 0.023 | 0.110 | 0.134 | 1.453 | 1.626 | 1.646 | 1.649 | 1.649 | 8.023 |
| Multiplicative | Mild | | 0.000 | 0.002 | 0.003 | 0.044 | 0.152 | 0.201 | 0.558 | 0.599 | 0.603 | 0.603 | 0.603 | 2.965 |
| | Moderate | | 0.004 | 0.010 | 0.014 | 0.000 | 0.035 | 0.063 | 0.270 | 0.331 | 0.345 | 0.348 | 0.348 | 1.642 |
| Naive | - | AR(2) | 0.073 | 0.107 | 0.240 | 0.550 | 1.111 | 2.081 | 2.517 | 2.648 | 2.648 | 2.649 | 2.649 | 13.111 |
| Additive | Mild | | 0.029 | 0.014 | 0.026 | 0.083 | 0.227 | 0.379 | 2.256 | 2.398 | 2.398 | 2.399 | 2.399 | 11.851 |
| | Moderate | | 0.045 | 0.008 | 0.031 | 0.063 | 0.221 | 0.368 | 1.470 | 1.658 | 1.646 | 1.649 | 1.649 | 8.072 |
| Multiplicative | Mild | | 0.034 | 0.012 | 0.027 | 0.076 | 0.222 | 0.370 | 0.571 | 0.606 | 0.603 | 0.603 | 0.603 | 2.986 |
| | Moderate | | 0.085 | 0.001 | 0.071 | 0.024 | 0.310 | 0.491 | 0.454 | 0.469 | 0.415 | 0.390 | 0.375 | 2.103 |
| | | Quebec | | | | | | | | | | | | |
| Naive | - | **AR(1)**[a] | 0.163 | 0.607 | 1.357 | 2.289 | 3.294 | 7.709 | 1.811 | 1.811 | 1.811 | 1.811 | 1.811 | 9.057 |
| Additive | Mild | | 0.061 | 0.216 | 0.479 | 0.778 | 1.053 | 2.587 | 1.561 | 1.561 | 1.561 | 1.561 | 1.561 | 7.807 |
| | Moderate | | 0.060 | 0.215 | 0.478 | 0.776 | 1.051 | 2.580 | 0.811 | 0.811 | 0.811 | 0.811 | 0.811 | 4.057 |
| Multiplicative | Mild | | 0.061 | 0.216 | 0.479 | 0.778 | 1.053 | 2.586 | 0.399 | 0.399 | 0.399 | 0.399 | 0.399 | 1.995 |
| | Moderate | | 0.060 | 0.215 | 0.477 | 0.776 | 1.050 | 2.578 | 0.205 | 0.205 | 0.205 | 0.205 | 0.205 | 1.025 |
| Naive | - | AR(2) | 0.129 | 0.524 | 1.226 | 2.115 | 3.085 | 7.079 | 1.746 | 1.746 | 1.809 | 1.809 | 1.811 | 8.921 |
| Additive | Mild | | 0.052 | 0.195 | 0.446 | 0.734 | 1.002 | 2.429 | 1.375 | 1.375 | 1.447 | 1.447 | 1.451 | 7.096 |
| | Moderate | | 0.032 | 0.109 | 0.247 | 0.396 | 0.519 | 1.303 | 0.413 | 0.413 | 0.407 | 0.407 | 0.402 | 2.043 |
| Multiplicative | Mild | | 0.051 | 0.190 | 0.438 | 0.723 | 0.988 | 2.390 | 0.345 | 0.345 | 0.356 | 0.356 | 0.357 | 1.760 |
| | Moderate | | 0.038 | 0.141 | 0.333 | 0.560 | 0.774 | 1.847 | 0.332 | 0.332 | 0.234 | 0.234 | 0.187 | 1.319 |
| | | Alberta | | | | | | | | | | | | |
| Naive | - | **AR(1)**[a] | 0.002 | 0.007 | 0.027 | 0.055 | 0.070 | 0.160 | 0.125 | 0.125 | 0.125 | 0.125 | 0.125 | 0.627 |
| Additive | Mild | | 0.004 | 0.012 | 0.044 | 0.087 | 0.115 | 0.262 | 0.115 | 0.115 | 0.115 | 0.115 | 0.115 | 0.577 |
| | Moderate | | 0.006 | 0.017 | 0.052 | 0.098 | 0.129 | 0.302 | 0.035 | 0.035 | 0.035 | 0.035 | 0.035 | 0.177 |
| Multiplicative | Mild | | 0.004 | 0.012 | 0.044 | 0.087 | 0.115 | 0.263 | 0.031 | 0.031 | 0.031 | 0.031 | 0.031 | 0.157 |
| | Moderate | | 0.005 | 0.013 | 0.045 | 0.089 | 0.118 | 0.270 | 0.022 | 0.022 | 0.022 | 0.022 | 0.022 | 0.109 |
| Naive | - | AR(2) | 0.003 | 0.010 | 0.033 | 0.064 | 0.081 | 0.191 | 0.122 | 0.122 | 0.125 | 0.125 | 0.125 | 0.621 |
| Additive | Mild | | 0.005 | 0.016 | 0.051 | 0.097 | 0.127 | 0.296 | 0.112 | 0.112 | 0.115 | 0.115 | 0.115 | 0.570 |
| | Moderate | | 0.006 | 0.018 | 0.056 | 0.104 | 0.136 | 0.320 | 0.081 | 0.081 | 0.085 | 0.085 | 0.085 | 0.419 |
| Multiplicative | Mild | | 0.005 | 0.016 | 0.052 | 0.099 | 0.129 | 0.301 | 0.030 | 0.031 | 0.031 | 0.031 | 0.031 | 0.155 |
| | Moderate | | 0.006 | 0.019 | 0.059 | 0.109 | 0.141 | 0.334 | 0.022 | 0.022 | 0.022 | 0.022 | 0.022 | 0.109 |

[a] The selected model

# Chapter 7

# Summary and Discussion

In this thesis, we investigate several important research problems concerning correlated responses with measurement error or misclassification. The results in this thesis have been or will be prepared as papers for dissemination. The research in Chapter 2 has been prepared as a paper, Zhang and Yi (2020b), and has been accepted by *Statistics in Medicine*; the results in Chapter 3 have been written up as a paper, Zhang and Yi (2020c), which has been invited by *Statistical Methods in Medical Research* for revision; the results in Chapter 4 have been included in the paper, Zhang and Yi (2020a), which has been submitted for publication; the results in Chapter 5 are being prepared as the paper, Zhang and Yi (2020e), which is to be submitted for publication soon; the results in Chapter 6 have already been wrapped up as the paper, Zhang and Yi (2020d), and submitted for publication. Below we present a summary for each chapter with discussions.

**Chapter 2**

When jointly modeling the mixed type of continuous and binary responses, we often encounter responses that are subject to measurement error and misclassification. To remove the bias resulting from the mismeasurement, it is necessary to address both measurement error and misclassification simultaneously. In this chapter, we develop two inference approaches to account for the effects due to mismeasurement in responses under latent variable models. The induced likelihood method can be easily implemented by R function `optim()` and the EM algorithm has the advantage of dealing with associated integrals by employing a complete likelihood formulation.

Although measurement error and misclassification is an inevitable issue in practice, such features are often ignored in genetic association studies. Even in the statistical literature, available work mainly focuses on a single type of mismeasurement in responses, either measurement error or misclassification but not both. In this chapter, we propose two valid methods to account for measurement error and misclassification in mixed continuous and discrete responses. Our methods can be applied to handle error-contaminated data arising from genomewide-association studies for which mixed responses with a continuous variable and a binary variable may be subjected to mismeasurement.

Our development is carried out for the generalized linear mixed model (2.1) where a common random effect $u_i$ is introduced to feature the mean structure of the two response components. More generally, one may use different random effects, say $u_{i1}$ and $u_{i2}$, to describe the mean structure of $Y_{i1}$ and $Y_{i2}$, respectively. That is, we write model (2.1) as

$$\begin{bmatrix} g_1(\mu_{i1}) \\ g_2(\mu_{i2}) \end{bmatrix} = \begin{bmatrix} X_i^{\mathrm{T}}\beta_1 \\ X_i^{\mathrm{T}}\beta_2 \end{bmatrix} + \begin{bmatrix} u_{i1} \\ u_{i2} \end{bmatrix}, \tag{7.1}$$

where $u_{i1}$ and $u_{i2}$ are random effects, and other symbols are defined in the same way as for (2.1). Then we set the random effects vector $u$ to be $\{(u_{i1}, u_{i2}) : i = 1, \ldots, n\}$ and modify the development accordingly.

Finally, in the development here, random effects $u$ are assumed to have the covariance structure $\sigma_g^2 R$ with a pre-specified matrix $R$ and an unknown $\sigma_g^2$. Letting $R$ be pre-specified allows us to incorporate a priori information of the study. In circumstances where $R$ is impossible to be feasibly prespecified, we write the covariance matrix of random effects to be a single matrix, say $\widetilde{R}$, which may contain multiple parameters rather than a single parameter $\sigma_g^2$ considered here. Then we carry out the inferential procedures similar to the development here by replacing $\sigma_g^2$ in the parameter vector $\theta$ in Section 2.2.1 with the parameters in $\widetilde{R}$.

**Chapter 3**

Error-contaminated mixed responses with a continuous and a binary variable present a new challenge in joint modeling and analysis of multiple responses. In this chapter, we develop a generalized estimating equation approach to incorporate the dependence among responses and develop an insertion strategy to adjust for the effects of mismeasurement in responses. We propose valid estimators that apply when either internal validation data or external validation data are available. Our methods are robust to model misspecification and produce small finite sample biases. We develop a weighted estimator to improve the efficiency of parameter estimation in the presence of internal validation data.

The generalized estimation equation is robust to model misspecification at the price of the efficiency loss. To overcome this disadvantage, in addition to the weighted estimators proposed in Section 3.3.3, other strategies may also be considered. For example, Hall and Severini (1998) proposed the extended generalized estimating equation (EGEE) based on the idea of extended Quasi-Likelihood. The EGEE has better efficiency than the original GEE approach (Prentice and Zhao, 1991) in some scenarios. The EGEE approach can be easily adapted in our method with minor modifications in the estimation part.

Measurement error and misclassification are inevitable in many cases. In this chapter, we propose several methods to address response mismeasurement in different types of study designs. We have shown that under certain regularity conditions, the proposed estimators are asymptotically normal and consistent. The methods are fast to implement and can apply for various settings.

**Chapter 4**

Identifying interactions among genetic variants is important in the analysis of gene networks. In this chapter, we develop a generalized network model to facilitate the relationship between genetic variants with a complex structure and the mixed responses via a two-step procedure. We further extend the development to handle data with measurement error and misclassification in responses. Theoretical justifications are provided to ensure the validity of the proposed method, and numerical studies demonstrate satisfactory finite sample performance of the proposed method.

In the development here, we consider continuous covariates that are featured by the Gaussian graphical model. It is interesting to generalize our method to accommodate discrete covariates or mixed covariates with both discrete and continuous components.

Our methods focus on addressing the effects due to mismeasurement in mixed bivariate responses, where covariates are assumed to be precisely measured. It is interesting to extend our work here to handle data which contain error-contaminated covariates, in addition to having mismeasured responses. In such a circumstance, adjusting the effects of measurement error in covariates is necessary for the first step for identifying the network structure for the true covariates. This research warrants exploration in depth.

**Chapter 5**

Zero-inflated Poisson models are useful in cancer genomics studies, which are, however, challenged by the presence of measurement error. While this problem is important, not much work has been available. We provide a general strategy in dealing with error-contaminated count data and proposed a flexible modeling scheme for measurement error in count data. We introduce a mixture model to facilitate an add-in process and a leave-out process for characterizing different types of measurement error associated count data.

We explore the effects of different measurement error models on the analysis. Numerical studies demonstrate satisfactory performance of the proposed method.

The development in this chapter can be modified to address the measurement error of count data in other models. For example, besides the zero-inflated model, the hurdle model (Mullahy, 1986) is also frequently used to account for excessive zeros in count data. Our Bayesian method can be adapted to suit the hurdle model. Sometimes, it is interesting to consider the overdispersion in count data. Our method can be further extended to deal with zero-inflated Negative Binomial models (Yau et al., 2003).

**Chapter 6**

We investigate the impact of measurement error on time series analysis under autoregressive models and establish analytic results under the additive and multiplicative measurement error models. We propose an estimating equation method to correct for the biases induced from the naive analysis which disregards the differences between the true measurements and their surrogate measurements. We rigorously establish the theoretical results for the proposed method. As a genuine application, we apply to the proposed method to analyze the mortality rates of COVID-19 data in four provinces, British Columbia, Ontario, Quebec, and Alberta, which have the most severe virus outbreaks in Canada. The real data analysis clearly demonstrates that incorporating measurement error in the analysis can uncover various different results.

Our method has the flexibility or robustness in that distribution assumptions are required to describe the measurement error process as well as the time series autoregressive process. While our research is motivated by the faulty nature of COVID-19 data, the proposed method can be applied to handle other problems related to error-contaminated time series. Our development here is directed to using autoregressive models to delineate time series data. The same principles can be applied to other model forms such as moving average models or autoregressive moving average models which may be used to handle error-prone time series data, where technical details can be more notationally involved.

When checking the stationarity of time series, we apply the ADF test to the observed time series $X_t^*$, which is mainly driven by the unavailability of the true values of $X_t$, as well as the fact that the weakly stationarity of observed time series implies the weakly stationarity of the true time series if measurement error is featured with (6.7) or (6.9). It is interesting to rigorously develop a formal test similar to the ADF test to handle time series subject to measurement error.

# References


Abramowitz, M. and Stegun, I. A. (1972). *Handbook of Mathematical Functions: with Formulas, Graphs, and Mathematical Tables*. Dover publications, New York.

Akavia, U. D., Litvin, O., Kim, J., Sanchez-Garcia, F., Kotliar, D., Causton, H. C., Pochanard, P., Mozes, E., Garraway, L. A., and Pe'er, D. (2010). An integrated approach to uncover drivers of cancer. *Cell*, 143(6):1005–1017.

Alberts, B., Johnson, A., Lewis, J., Raff, M., Roberts, K., and Walter, P. (2002). Transport into the cell from the plasma membrane: endocytosis. In *Molecular Biology of the Cell., 4th edition*. Garland Science, New York.

Anderson, C. A., Pettersson, F. H., Clarke, G. M., Cardon, L. R., Morris, A. P., and Zondervan, K. T. (2010). Data quality control in genetic case-control association studies. *Nature Protocols*, 5(9):1564.

Banerjee, O., Ghaoui, L. E., and d'Aspremont, A. (2008). Model selection through sparse maximum likelihood estimation for multivariate Gaussian or binary data. *Journal of Machine Learning Research*, 9(Mar):485–516.

Banerji, S., Cibulskis, K., Rangel-Escareno, C., Brown, K. K., Carter, S. L., Frederick, A. M., Lawrence, M. S., Sivachenko, A. Y., Sougnez, C., Zou, L., et al. (2012). Sequence analysis of mutations and translocations across breast cancer subtypes. *Nature*, 486(7403):405–409.

Barabási, A.-L. and Albert, R. (1999). Emergence of scaling in random networks. *Science*, 286(5439):509–512.

Baud, D., Qi, X., Nielsen-Saines, K., Musso, D., Pomar, L., and Favre, G. (2020). Real estimates of mortality following COVID-19 infection. *The Lancet Infectious Diseases*.

Benjamini, Y. and Hochberg, Y. (1995). Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society: Series B*, 57(1):289–300.

Bickel, P. J., Ritov, Y., Ryden, T., et al. (1998). Asymptotic normality of the maximum-likelihood estimator for general hidden markov models. *The Annals of Statistics*, 26(4):1614–1635.

Box, G. E., Jenkins, G. M., Reinsel, G. C., and Ljung, G. M. (2015). *Time Series Analysis: Forecasting and Control*. John Wiley & Sons, New Jersey.

Brockwell, P. J. and Davis, R. A. (2002). *Introduction to Time Series and Forecasting*. Springer-Verlag, New York.

Brockwell, P. J., Davis, R. A., and Fienberg, S. E. (1991). *Time Series: Theory and Methods*. Springer Science & Business Media, New York.

Brown, T. (1979). Position dependent and stochastic thinning of point processes. *Stochastic Processes and their Applications*, 9(2):189–193.

Buonaccorsi, J. (1991). Measurement errors, linear calibration and inferences for means. *Computational Statistics & Data Analysis*, 11(3):239–257.

Buonaccorsi, J. P. (1996). Measurement error in the response in the general linear model. *Journal of the American Statistical Association*, 91(434):633–642.

Carroll, R. J., Ruppert, D., Stefanski, L. A., and Crainiceanu, C. M. (2006). *Measurement Error in Nonlinear Models: a Modern Perspective*. CRC press, Boca Raton.

Catalano, P. J. (1997). Bivariate modelling of clustered continuous and ordered categorical outcomes. *Statistics in Medicine*, 16(8):883–900.

Chen, H., Wang, C., Conomos, M. P., Stilp, A. M., Li, Z., Sofer, T., Szpiro, A. A., Chen, W., Brehm, J. M., Celedón, J. C., et al. (2016). Control for population structure and relatedness for binary traits in genetic association studies via logistic mixed models. *The American Journal of Human Genetics*, 98(4):653–666.

Chen, L.-P., Grace, Y. Y., Zhang, Q., and He, W. (2019). Multiclass analysis and prediction with network structured covariates. *Journal of Statistical Distributions and Applications*, 6(1):6.

Chen, L.-P. and Yi, G. Y. (2020). Analysis of noisy survival data with graphical proportional hazards measurement error model. To appear in *Biometrics*. https://onlinelibrary.wiley.com/doi/abs/10.1111/biom.13331.

Chen, Z., Yi, G. Y., and Wu, C. (2011). Marginal methods for correlated binary data with misclassified responses. *Biometrika*, 98(3):647–662.

Cheng, J., Li, T., Levina, E., and Zhu, J. (2017). High-dimensional mixed graphical models. *Journal of Computational and Graphical Statistics*, 26(2):367–378.

Cheung, Y.-W. and Lai, K. S. (1995). Lag order and critical values of the augmented dickey-fuller test. *Journal of Business & Economic Statistics*, 13(3):277–280.

Day, M. (2020). COVID-19: four fifths of cases are asymptomatic, China figures indicate. *The BMJ*, 369.

de Leon, A. R. and Wu, B. (2011). Copula-based regression models for a bivariate mixed discrete and continuous outcome. *Statistics in Medicine*, 30(2):175–185.

Dedecker, J., Samson, A., and Taupin, M.-L. (2014). Estimation in autoregressive model with measurement error. *ESAIM: Probability and Statistics*, 18:277–307.

Dickey, D. A. and Fuller, W. A. (1979). Distribution of the estimators for autoregressive time series with a unit root. *Journal of the American Statistical Association*, 74(366a):427–431.

Dong, E., Du, H., and Gardner, L. (2020). An interactive web-based dashboard to track COVID-19 in real time. *The Lancet Infectious Diseases*, 20(5):533–534.

Douc, R., Matias, C., et al. (2001). Asymptotics of the maximum likelihood estimator for general hidden markov models. *Bernoulli*, 7(3):381–420.

Dunson, D. B. (2000). Bayesian latent variable models for clustered mixed outcomes. *Journal of the Royal Statistical Society: Series B*, 62(2):355–366.

Faes, C., Aerts, M., Molenberghs, G., Geys, H., Teuns, G., and Bijnens, L. (2008). A high-dimensional joint model for longitudinal outcomes of different nature. *Statistics in Medicine*, 27(22):4408–4427.

Fitzmaurice, G. M. and Laird, N. M. (1995). Regression models for a bivariate discrete and continuous outcome with clustering. *Journal of the American Statistical Association*, 90(431):845–852.

Friedman, J., Hastie, T., and Tibshirani, R. (2008). Sparse inverse covariance estimation with the graphical lasso. *Biostatistics*, 9(3):432–441.

Friedman, N. (2004). Inferring cellular networks using probabilistic graphical models. *Science*, 303(5659):799–805.

Fu, R., Wang, P., Ma, W., Taguchi, A., Wong, C.-H., Zhang, Q., Gazdar, A., Hanash, S. M., Zhou, Q., Zhong, H., and Feng, Z. (2017). A statistical method for detecting differentially expressed SNVs based on next-generation RNA-seq data. *Biometrics*, 73(1):42–51.

Furlotte, N. A. and Eskin, E. (2015). Efficient multiple-trait association and estimation of genetic correlation using the matrix-variate linear mixed model. *Genetics*, 200(1):59–68.

Gelman, A., Carlin, J. B., Stern, H. S., Dunson, D. B., Vehtari, A., and Rubin, D. B. (2013). *Bayesian Data Analysis*. CRC press, Boca Raton.

Gelman, A., Rubin, D. B., et al. (1992). Inference from iterative simulation using multiple sequences. *Statistical Science*, 7(4):457–472.

Godambe, V. P. (1991). *Estimating Functions*. Oxford University Press, Oxford.

Gueorguieva, R. V. and Agresti, A. (2001). A correlated probit model for joint modeling of clustered binary and continuous responses. *Journal of the American Statistical Association*, 96(455):1102–1112.

Hall, D. B. (2001). On the application of extended quasi-likelihood to the clustered data case. *Canadian Journal of Statistics*, 29(1):77–97.

Hall, D. B. and Severini, T. A. (1998). Extended generalized estimating equations for clustered data. *Journal of the American Statistical Association*, 93(444):1365–1375.

Hastie, T., Tibshirani, R., and Wainwright, M. (2015). *Statistical Learning with Sparsity: The Lasso and Generalizations*. CRC press, New York.

He, W., Yi, G. Y., and Zhu, Y. (2020). Estimation of the basic reproduction number, average incubation time, asymptomatic infection rate, and case fatality rate for COVID-19: Meta-analysis and sensitivity analysis. *Journal of Medical Virology*. https://onlinelibrary.wiley.com/doi/abs/10.1002/jmv.26041.

Heyde, C. C. (1997). *Quasi-likelihood and its Application: A General Approach to Optimal Parameter Estimation*. Springer Science & Business Media, New York.

Höfling, H. and Tibshirani, R. (2009). Estimation of sparse binary pairwise markov networks using pseudo-likelihoods. *Journal of Machine Learning Research*, 10(Apr):883–906.

Hossain, S., Le, N. D., Brooks-Wilson, A. R., and Spinelli, J. J. (2009). Impact of genotype misclassification on genetic association estimates and the bayesian adjustment. *American Journal of Epidemiology*, 170(8):994–1004.

James, F. (1980). Monte Carlo theory and practice. *Reports on Progress in Physics*, 43(9):1145–1189.

Jennrich, R. I. (1969). Asymptotic properties of non-linear least squares estimators. *The Annals of Mathematical Statistics*, 40(2):633–643.

Jensen, J. L., Petersen, N. V., et al. (1999). Asymptotic normality of the maximum likelihood estimator in state space models. *The Annals of Statistics*, 27(2):514–535.

Kanchan, T., Kumar, N., and Unnikrishnan, B. (2015). Mortality: Statistics. In Payne-James, J. and Byard, R. W., editors, *Encyclopedia of Forensic and Legal Medicine: Second Edition*, pages 572–577. Elsevier, Oxford.

Kimball, A. (2020). Asymptomatic and presymptomatic SARS-CoV-2 infections in residents of a long-term care skilled nursing facility—King County, Washington, March 2020. *Morbidity and Mortality Weekly Report*, 69:377–381.

Klein, N., Kneib, T., and Lang, S. (2015). Bayesian generalized additive models for location, scale, and shape for zero-inflated and overdispersed count data. *Journal of the American Statistical Association*, 110(509):405–419.

Korte, A., Vilhjálmsson, B. J., Segura, V., Platt, A., Long, Q., and Nordborg, M. (2012). A mixed-model approach for genome-wide association studies of correlated traits in structured populations. *Nature Genetics*, 44(9):1066–1071.

Kraja, A. T., Chasman, D. I., North, K. E., Reiner, A. P., Yanek, L. R., Kilpeläinen, T. O., Smith, J. A., Dehghan, A., Dupuis, J., Johnson, A. D., et al. (2014). Pleiotropic genes for metabolic syndrome and inflammation. *Molecular Genetics and Metabolism*, 112(4):317–338.

Lahiri, S. N. (1999). Theoretical comparisons of block bootstrap methods. *The Annals of Statistics*, 27(1):386–404.

Lai, C.-C., Shih, T.-P., Ko, W.-C., Tang, H.-J., and Hsueh, P.-R. (2020). Severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) and corona virus disease-2019 (COVID-19): the epidemic and the challenges. *International Journal of Antimicrobial Agents*, page 105924.

Lambert, D. (1992). Zero-inflated poisson regression, with an application to defects in manufacturing. *Technometrics*, 34(1):1–14.

Lange, C., Whittaker, J. C., and Macgregor, A. J. (2002). Generalized estimating equations: a hybrid approach for mean parameters in multivariate regression models. *Statistical Modelling*, 2(3):163–181.

Lange, K. (2003). *Mathematical and Statistical Methods for Genetic Analysis*. Springer Science & Business Media, New York.

Lee, J. D. and Hastie, T. J. (2015). Learning the structure of mixed graphical models. *Journal of Computational and Graphical Statistics*, 24(1):230–253.

Li, B., Chun, H., and Zhao, H. (2012). Sparse estimation of conditional graphical models with application to gene networks. *Journal of the American Statistical Association*, 107(497):152–167.

Li, Q., Guan, X., Wu, P., Wang, X., Zhou, L., Tong, Y., Ren, R., Leung, K. S., Lau, E. H., Wong, J. Y., et al. (2020a). Early transmission dynamics in Wuhan, China, of novel coronavirus–infected pneumonia. *New England Journal of Medicine*, 382(13):1199–1207.

Li, R., Pei, S., Chen, B., Song, Y., Zhang, T., Yang, W., and Shaman, J. (2020b). Substantial undocumented infection facilitates the rapid dissemination of novel coronavirus (SARS-CoV2). *Science*.

Liang, K.-Y. and Zeger, S. L. (1986). Longitudinal data analysis using generalized linear models. *Biometrika*, 73(1):13–22.

Lin, H., Zhou, L., Elashoff, R. M., and Li, Y. (2014). Semiparametric latent variable transformation models for multiple mixed outcomes. *Statistica Sinica*, 24(2):833–854.

Lin, L., Bandyopadhyay, D., Lipsitz, S. R., and Sinha, D. (2010). Association models for clustered data with binary and continuous responses. *Biometrics*, 66(1):287–293.

Liu, H., Roeder, K., and Wasserman, L. (2010). Stability approach to regularization selection (StARS) for high dimensional graphical models. In Lafferty, J. D., Williams,

C. K. I., Shawe-Taylor, J., Zemel, R. S., and Culotta, A., editors, *Advances in Neural Information Processing Systems 23*, volume 1, 2111–2245. Curran Associates Inc., New York.

Liu, J., Pei, Y., Papasian, C. J., and Deng, H.-W. (2009). Bivariate association analyses for the mixture of continuous and binary traits with the use of extended generalized estimating equations. *Genetic Epidemiology*, 33(3):217–227.

Louis, T. A. (1982). Finding the observed information matrix when using the EM algorithm. *Journal of the Royal Statistical Society: Series B*, 44(2):226–233.

Loukas, S. and Kemp, C. D. (1983). On computer sampling from trivariate and multivariate discrete distributions. *Journal of Statistical Computation and Simulation*, 17(2):113–123.

Lynch, J. J., Maijanen, H., and Prescher, A. (2019). Analysis of three commonly used tibia length measurement techniques. *Journal of Forensic Sciences*, 64(1):181–185.

Malécot, G. et al. (1948). *The Mathematics of Heredity.* Masson and Cie, Paris.

McCulloch, C. (2008). Joint modelling of mixed outcome types using latent variables. *Statistical Methods in Medical Research*, 17(1):53–73.

Meinshausen, N. and Bühlmann, P. (2006). High-dimensional graphs and variable selection with the LASSO. *The Annals of Statistics*, 34(3):1436–1462.

Mizumoto, K., Kagaya, K., Zarebski, A., and Chowell, G. (2020). Estimating the asymptomatic proportion of coronavirus disease 2019 (COVID-19) cases on board the diamond princess cruise ship, yokohama, japan, 2020. *Eurosurveillance*, 25(10).

Mullahy, J. (1986). Specification and testing of some modified count data models. *Journal of Econometrics*, 33(3):341–365.

Navarrete, K., Pedroso, I., De Jong, S., Stefansson, H., Steinberg, S., Stefansson, K., Ophoff, R. A., Schalkwyk, L. C., and Collier, D. A. (2013). Tcf4 (e2-2; itf2): A schizophrenia-associated gene with pleiotropic effects on human disease. *American Journal of Medical Genetics Part B: Neuropsychiatric Genetics*, 162(1):1–16.

Neelon, B. and Chung, D. (2017). The LZIP: A bayesian latent factor model for correlated zero-inflated counts. *Biometrics*, 73(1):185–196.

Network, C. G. A. et al. (2012). Comprehensive molecular portraits of human breast tumours. *Nature*, 490(7418):61.

Neuhaus, J. M. (1999). Bias and efficiency loss due to misclassified responses in binary regression. *Biometrika*, 86(4):843–855.

Neuhaus, J. M. (2002). Analysis of clustered and longitudinal binary data subject to response misclassification. *Biometrics*, 58(3):675–683.

Newey, W. K. and McFadden, D. (1994). Large sample estimation and hypothesis testing. In Engle, R. and McFadden, D. L., editors, *Handbook of Econometrics*, volume 4, 2111–2245. North-Holland, Amsterdam.

Oakes, D. (1999). Direct calculation of the information matrix via the EM. *Journal of the Royal Statistical Society: Series B*, 61(2):479–482.

Pan, W. (2001). On the robust variance estimator in generalised estimating equations. *Biometrika*, 88(3):901–906.

Parker, C., Gopalakrishnan, S., Carbonetto, P., Gonzales, N., Leung, E., Park, Y., Aryee, E., Davis, J., Blizard, D., Ackert-Bicknell, C., Lionikas, A., Pritchard, J., and Palmer, A. (2016a). Data from: Genome-wide association study of behavioral, physiological and gene expression traits in outbred CFW mice.

Parker, C. C., Gopalakrishnan, S., Carbonetto, P., Gonzales, N. M., Leung, E., Park, Y. J., Aryee, E., Davis, J., Blizard, D. A., Ackert-Bicknell, C. L., et al. (2016b). Genome-wide association study of behavioral, physiological and gene expression traits in outbred CFW mice. *Nature Genetics*, 48(8):919–926.

Paulino, C., Soares, P., and Neuhaus, J. (2003). Binomial regression with misclassification. *Biometrics*, 59(3):670–675.

Penrose, L. (1951). Measurement of pleiotropic effects in phenylketonuria. *Annals of Eugenics*, 16(1):134–141.

Pepe, M. S., Reilly, M., and Fleming, T. R. (1994). Auxiliary outcome data and the mean score method. *Journal of Statistical Planning and Inference*, 42(1-2):137–160.

Prentice, R. L. and Zhao, L. P. (1991). Estimating equations for parameters in means and covariances of multivariate discrete and continuous responses. *Biometrics*, 47(3):825–839.

Prescott, G. J. and Garthwaite, P. H. (2002). A simple bayesian analysis of misclassified binary data with a validation substudy. *Biometrics*, 58(2):454–458.

Price, A. L., Patterson, N. J., Plenge, R. M., Weinblatt, M. E., Shadick, N. A., and Reich, D. (2006). Principal components analysis corrects for stratification in genome-wide association studies. *Nature Genetics*, 38(8):904—-909.

Rajgor, D. D., Lee, M. H., Archuleta, S., Bagdasarian, N., and Quek, S. C. (2020). The many estimates of the COVID-19 case fatality rate. *The Lancet Infectious Diseases*, 20:776–777.

Ramalho, E. A. (2002). Regression models for choice-based samples with misclassification in the response variable. *Journal of Econometrics*, 106(1):171–201.

Ravikumar, P., Wainwright, M. J., and Lafferty, J. D. (2010). High-dimensional ising model selection using $\ell 1$-regularized logistic regression. *The Annals of Statistics*, 38(3):1287–1319.

Rekaya, R., Smith, S., El Hamidi Hay, N. F., and Aggrey, S. E. (2016). Analysis of binary responses with outcome-specific misclassification probability in genome-wide association studies. *The Application of Clinical Genetics*, 9(1):169–177.

Rochon, J. (1996). Analyzing bivariate repeated measures for discrete and continuous outcome variables. *Biometrics*, 52(2):740–750.

Rodrigues, J. (2003). Bayesian analysis of zero-inflated distributions. *Communications in Statistics - Theory and Methods*, 32(2):281–289.

Roeder, K., Carroll, R. J., and Lindsay, B. G. (1996). A semiparametric mixture approach to case-control studies with errors in covariables. *Journal of the American Statistical Association*, 91(434):722–732.

Rudin, W. (1976). *Principles of Mathematical Analysis*, volume 3. McGraw-hill, New York.

Sammel, M., Lin, X., and Ryan, L. (1999). Multivariate linear mixed models for multiple outcomes. *Statistics in Medicine*, 18(17-18):2479–2492.

Sammel, M. D. and Ryan, L. M. (1996). Latent variable models with fixed effects. *Biometrics*, 52(2):650–663.

Sammel, M. D., Ryan, L. M., and Legler, J. M. (1997). Latent variable models for mixed discrete and continuous outcomes. *Journal of the Royal Statistical Society: Series B*, 59(3):667–678.

Shah, S. P., Roth, A., Goya, R., Oloumi, A., Ha, G., Zhao, Y., Turashvili, G., Ding, J., Tse, K., Haffari, G., et al. (2012). The clonal and mutational evolution spectrum of primary triple-negative breast cancers. *Nature*, 486(7403):395–399.

Shim, E., Tariq, A., Choi, W., Lee, Y., and Chowell, G. (2020). Transmission potential and severity of COVID-19 in south korea. *International Journal of Infectious Diseases*, 93:339–344.

Shu, D. and Yi, G. Y. (2017). Causal inference with measurement error in outcomes: Bias analysis and estimation methods. *Statistical Methods in Medical Research*, 28(7):2049–2068.

Shukla, S., MacLennan, G. T., Hartman, D. J., Fu, P., Resnick, M. I., and Gupta, S. (2007). Activation of PI3K-Akt signaling pathway promotes prostate cancer cell invasion. *International Journal of Cancer*, 121(7):1424–1432.

Smith, S., Farhat, N., Rekaya, R., et al. (2013). Genome-wide association studies in presence of misclassified binary responses. *BMC Genetics*, 14(1):124–134.

Spiegelman, D., Rosner, B., and Logan, R. (2000). Estimation and inference for logistic regression with covariate misclassification and measurement error in main study/validation study designs. *Journal of the American Statistical Association*, 95(449):51–61.

Staudenmayer, J. and Buonaccorsi, J. P. (2005). Measurement error in linear autoregressive models. *Journal of the American Statistical Association*, 100(471):841–852.

Stefanski, L. A. (2000). Measurement error models. *Journal of the American Statistical Association*, 95(452):1353–1358.

Stefanski, L. A., Carroll, R. J., et al. (1985). Covariate measurement error in logistic regression. *The Annals of Statistics*, 13(4):1335–1351.

Stephens, P. J., Tarpey, P. S., Davies, H., Van Loo, P., Greenman, C., Wedge, D. C., Nik-Zainal, S., Martin, S., Varela, I., Bignell, G. R., et al. (2012). The landscape of cancer genes and mutational processes in breast cancer. *Nature*, 486(7403):400–404.

Tanaka, K. (2002). A unified approach to the measurement error problem in time series models. *Econometric Theory*, 18(2):278–296.

Taylor, B. S., Schultz, N., Hieronymus, H., Gopalan, A., Xiao, Y., Carver, B. S., Arora, V. K., Kaushik, P., Cerami, E., Reva, B., et al. (2010). Integrative genomic profiling of human prostate cancer. *Cancer Cell*, 18(1):11–22.

Teixeira-Pinto, A. and Normand, S.-L. T. (2009). Correlated bivariate continuous and binary outcomes: issues and applications. *Statistics in Medicine*, 28(13):1753–1773.

Tibshirani, R. (1996). Regression shrinkage and selection via the LASSO. *Journal of the Royal Statistical Society: Series B*, 58(1):267–288.

Todem, D., Kim, K., and Hsu, W.-W. (2016). Marginal mean models for zero-inflated count data. *Biometrics*, 72(3):986–994.

Trapletti, A. and Hornik, K. (2019). *tseries: Time Series Analysis and Computational Finance*. R package version 0.10-47.

Tripodis, Y. and Buonaccorsi, J. P. (2009). Prediction and forecasting in linear models with measurement error. *Journal of statistical planning and inference*, 139(12):4039–4050.

Van der Vaart, A. W. (2000). *Asymptotic Statistics*, volume 3. Cambridge University Press, Cambridge.

van Dyk, D. A. and Meng, X.-L. (2001). The art of data augmentation. *Journal of Computational and Graphical Statistics*, 10(1):1–50.

VanRaden, P. M. (2008). Efficient methods to compute genomic predictions. *Journal of Dairy Science*, 91(11):4414–4423.

Vuong, Q. H. (1989). Likelihood ratio tests for model selection and non-nested hypotheses. *Econometrica: Journal of the Econometric Society*, 57(2):307–333.

Wagner, H. and Tüchler, R. (2010). Bayesian estimation of random effects models for multivariate responses of mixed data. *Computational Statistics & Data Analysis*, 54(5):1206–1218.

Wang, L., Zhou, J., and Qu, A. (2012). Penalized generalized estimating equations for high-dimensional longitudinal data analysis. *Biometrics*, 68(2):353–360.

Wang, M. and Long, Q. (2011). Modified robust variance estimator for generalized estimating equations with improved small-sample performance. *Statistics in Medicine*, 30(11):1278–1291.

Wason, J. M. and Dudbridge, F. (2012). A general framework for two-stage analysis of genome-wide association studies and its application to case-control studies. *The American Journal of Human Genetics*, 90(5):760–773.

WHO (2020). Report of the WHO-China joint mission on coronavirus disease 2019 (COVID-19).

Wu, J. T., Leung, K., and Leung, G. M. (2020). Nowcasting and forecasting the potential domestic and international spread of the 2019-nCoV outbreak originating in wuhan, china: a modelling study. *The Lancet*, 395(10225):689–697.

Wu, Z. and McGoogan, J. M. (2020). Characteristics of and important lessons from the coronavirus disease 2019 (COVID-19) outbreak in China: summary of a report of 72314 cases from the chinese center for disease control and prevention. *Jama*, 323(13):1239–1242.

Xiang, L., Lee, A. H., Yau, K. K., and McLachlan, G. J. (2007). A score test for overdispersion in zero-inflated poisson mixed regression model. *Statistics in medicine*, 26(7):1608–1622.

Xie, C. and Tammi, M. T. (2009). CNV-seq, a new method to detect copy number variation using high-throughput sequencing. *BMC bioinformatics*, 10(1):80.

Xu, T., Demmer, R. T., and Li, G. (2020). Zero-inflated Poisson factor model with application to microbiome read counts. To appear in *Biometrics*. https://onlinelibrary.wiley.com/doi/abs/10.1111/biom.13272.

Yang, E., Ravikumar, P., Allen, G. I., and Liu, Z. (2015). Graphical models via univariate exponential family distributions. *The Journal of Machine Learning Research*, 16(1):3813–3847.

Yang, Z., Hardin, J. W., and Addy, C. L. (2010). Score tests for overdispersion in zero-inflated poisson mixed models. *Computational statistics & data analysis*, 54(5):1234–1246.

Yau, K. K., Wang, K., and Lee, A. H. (2003). Zero-inflated negative binomial mixed regression modeling of over-dispersed count data with extra zeros. *Journal of Mathematical Methods in Biosciences*, 45(4):437–452.

Yi, G. Y. (2017). *Statistical Analysis with Measurement Error or Misclassification: Strategy, Method and Application*. Springer, New York.

Yi, G. Y. and Cook, R. J. (2002). Marginal methods for incomplete longitudinal data arising in clusters. *Journal of the American Statistical Association*, 97(460):1071–1080.

Yi, G. Y., Ma, Y., Spiegelman, D., and Carroll, R. J. (2015). Functional and structural methods with mixed measurement error and misclassification in covariates. *Journal of the American Statistical Association*, 110(510):681–696.

Yi, G. Y. and Reid, N. (2010). A note on mis-specified estimating functions. *Statistica Sinica*, 20(4):1749–1769.

Yi, G. Y., Yan, Y., Liao, X., and Spiegelman, D. (2018). Parametric regression analysis with covariate misclassification in main study/validation study designs. *The International Journal of Biostatistics*, 15(1):1–24.

Yu, D., Son, W., Lim, J., and Xiao, G. (2015). Statistical completion of a partially identified graph with applications for the estimation of gene regulatory networks. *Biostatistics*, 16(4):670–685.

Yuan, M. and Lin, Y. (2007). Model selection and estimation in the Gaussian graphical model. *Biometrika*, 94(1):19–35.

Zeger, S. L. and Liang, K.-Y. (1986). Longitudinal data analysis for discrete and continuous outcomes. *Biometrics*, 42(1):121–130.

Zeger, S. L. and Qaqish, B. (1988). Markov regression models for time series: a quasi-likelihood approach. *Biometrics*, 44(4):1019–1031.

Zhang, Q. and Yi, G. Y. (2019). R package for analysis of data with mixed measurement error and misclassification in covariates: **augSIMEX**. *Journal of Statistical Computation and Simulation*, 89(12):2293–2315.

Zhang, Q. and Yi, G. Y. (2020a). Generalized network structured models with mixed responses subject to measurement error and misclassification. *Submitted for publication.*

Zhang, Q. and Yi, G. Y. (2020b). Genetic association studies with bivariate mixed responses subject to measurement error and misclassification. To appear in *Statistics in Medicine*. https://onlinelibrary.wiley.com/doi/abs/10.1002/sim.8688.

Zhang, Q. and Yi, G. Y. (2020c). Marginal analysis of bivariate mixed responses with measurement error and misclassification. *Statistical Methods in Medical Research. In Revision.*

Zhang, Q. and Yi, G. Y. (2020d). Sensitivity analysis of time series data subject to measurement error under autoregressive models with application of COVID-19 data. *Submitted for publication.*

Zhang, Q. and Yi, G. Y. (2020e). Zero-inflated poisson models with measurement error in response. *To be submitted shortly.*

Zhang, Z., Ersoz, E., Lai, C.-Q., Todhunter, R. J., Tiwari, H. K., Gore, M. A., Bradbury, P. J., Yu, J., Arnett, D. K., Ordovas, J. M., et al. (2010). Mixed linear model approach adapted for genome-wide association studies. *Nature Genetics*, 42(4):355–360.

Zhao, T., Liu, H., Roeder, K., Lafferty, J., and Wasserman, L. (2012). The **huge** package for high-dimensional undirected graph estimation in R. *Journal of Machine Learning Research*, 13(Apr):1059–1062.

Zhou, X. and Stephens, M. (2014). Efficient multivariate linear mixed model algorithms for genome-wide association studies. *Nature Methods*, 11(4):407–409.

# APPENDICES

In this part, we include supplementary materials associated with Chapters 2–6, including regularity conditions, the proofs of the theoretical results, and additional calculations or discussions.

# Appendix A

# Supplement Materials for Chapter 2

## A.1 Identifiability Issue

Model (2.8) may incur identifiability issues in some circumstances. For example, consider that $g_1(x) = x$ for (2.1) and $e_i$ follows a normal distribution for (2.8). Then, the first component of (2.1) is equivalent to

$$Y_{i1} = \beta_{10} + \beta_{11}^{\mathrm{T}} X_i + \epsilon_i, \tag{A.1}$$

where $\beta_{10}$ is the first element of $\beta_1$ in (2.1) and $\beta_{11}^{\mathrm{T}}$ is the remaining vector, and $\epsilon_i$ is independent of $X_i$ and follows normal distribution with zero mean and variance $\sigma^2$.

Plugging (A.1) into the additive measurement error model (2.8) gives

$$Y_{i1}^* = \beta_0^* + \beta_1^{*\mathrm{T}} X_i + \gamma_2 f(Y_{i2}) + \gamma_3 Z_i + e_i^*,$$

where $e_i^* = e_i + \epsilon_i$ is independent of $\{X_i, Y_{i2}, Z_i\}$ and follows $N(0, \sigma^2 + \sigma_e^2)$, and

$$\beta_0^* = \gamma_0 + \gamma_1 \beta_{10}; \quad \beta_1^{*\mathrm{T}} = \gamma_1 \beta_1^{\mathrm{T}}. \tag{A.2}$$

(A.2) shows that based on the observed data, we are not able to separate $\gamma_0$ from $\gamma_1\beta_{10}$, $\gamma_1$ from $\beta_1$, and $\sigma^2$ from $\sigma_e^2$.

To overcome model nonidentifiability, we may add extra constraints on each group of parameters as commented by Yi (2017, Page 52). For example, we may specify $\gamma_0 = -\frac{1}{2}\gamma_2$, $\gamma_1 = 1$ and $\sigma^2 = \sigma_e^2$, which is equivalent to specifying $f(t) = 2t - 1$, $\gamma_0 = 0$, $\gamma_1 = 1$ and $\sigma^2 = \sigma_e^2$.

## A.2 Gaussian Quadrature Approximation of the Expectation

In this section, we illustrate how to approximate

$$E_{u_i,Y_{i1},Y_{i2}}\{g(Y_{i1}^*,Y_{i2}^*,Y_{i1},Y_{i2},u_i;\theta)\}, \tag{A.3}$$

using Gaussian-Hermite Quadrature.

Define the notation

$$\begin{aligned}
L_{y2}(y_{i1},u_i,y_{i1^*},y_{i2^*}) &= \int_{y_{i2}} g(y_{i1}^*,y_{i2}^*,y_{i1},y_{i2},u_i;\theta)f(y_{i2}|u_i,x_i)\,\mathrm{d}y_{i2},\\
L_{y1}(u_i,y_{i1^*},y_{i2^*}) &= \int_{y_{i1}} L_{y2}(y_{i1},u_i,y_{i1^*},y_{i2^*})f(y_{i1}|u_i,x_i)\,\mathrm{d}y_{i1},\\
L_{u}(y_{i1^*},y_{i2^*}) &= \int_{u_i} L_{y1}(u_i,y_{i1^*},y_{i2^*})f(u_i|x_i)\,\mathrm{d}u_i.
\end{aligned}$$

Since $Y_{i2}$ is a binary variable following Bernoulli distribution, we can compute the exact expectation,

$$\begin{aligned}
L_{y2}(y_{i1},u_i,y_{i1^*},y_{i2^*}) = g(y_{i1}^*,y_{i2}^*,y_{i1},y_{i2}=1,u_i;\theta)f(y_{i2}=1|u_i)\\
+g(y_{i1}^*,y_{i2}^*,y_{i1},y_{i2}=0,u_i;\theta)f(y_{i2}=0|u_i).
\end{aligned}$$

Consider the case where $R$ is an diagonal matrix, using Gaussian Quadrature, we can approximate

$$\begin{aligned}
L_{y1}(u_i,y_{i1^*},y_{i2^*}) &\approx \frac{1}{\sqrt{\pi}}\sum_{j=1}^{S^{(y)}} w_j^{(y)} L_{y_2}(y_{i1}=\sqrt{2}\sigma v_j^{(y)}+\beta_1^{\mathrm{T}}x_i+u_i,u_i,y_{i1}^*,y_{i2}^*),\\
L_{u}(y_{i1^*},y_{i2^*}) &\approx \frac{1}{\sqrt{\pi}}\sum_{j=1}^{S^{(u)}} w_j^{(u)} L_{y1}(u_i=\sqrt{2R_{ii}}\sigma_g v_j^{(u)},y_{i1^*},y_{i2^*}).
\end{aligned}$$

where $v_j^{(y)}, v_j^{(u)}$ are the roots of the Hermite polynomial $H_n(x)$ for $j=1,2,...,n$ (Abramowitz and Stegun, 1972, Page 890), and $w_j^{(y)}, w_j^{(u)}$ are the associated weights given by

$$w_j^{(y)} = \frac{2^{n-1}n!\sqrt{\pi}}{n^2[H_{n-1}(v_i^{(y)})]^2} \qquad \text{and} \qquad w_j^{(u)} = \frac{2^{n-1}n!\sqrt{\pi}}{n^2[H_{n-1}(v_i^{(u)})]^2}.$$

Based on the derivation above, the expectation (A.3) can be approximated as

$$\frac{1}{\pi}\sum_{j=1}^{S^{(u)}}\sum_{k=1}^{S^{(y)}} w_j^{(u)} w_k^{(y)} \{ g(y_{i1}^*, y_{i2}^*, y_{i1}, y_{i2}=1, u_i; \theta) f(y_{i2}=1|u_i)$$

$$+ g(y_{i1}^*, y_{i2}^*, y_{i1}, y_{i2}=0, u_i; \theta) f(y_{i2}=0|u_i) \} \Big|_{\substack{y_{i1}=\sqrt{2}\sigma v_k^{(y)}+\beta_1^{\mathrm{T}} x_i + u_i, \\ u_i=\sqrt{2R_{ii}}\sigma_g v_j^{(u)}}}.$$

# Appendix B

# Proofs of the Results in Chapter 3

## B.1 Proof of $E\{U_i^{**}(\theta)|Y_{i1}, Y_{i2}, X_i\} = U_i(\theta)$

**Step 1**: First, we show that

$$E(Y_{i2}^{**}|Y_{i1}, Y_{i2}, X_i) = Y_{i2}. \tag{B.1}$$

Indeed, by the definition of $Y_{i2}^{**}$,

$$\begin{aligned}
E(Y_{i2}^{**}|Y_{i1}, Y_{i2} = j, X_i) &= E\left(\frac{Y_{i2}^{*} - \pi_{i0}}{1 - \pi_{i0} - \pi_{i1}}\Big|Y_{i1}, Y_{i2} = j, X_i\right) \\
&= \frac{E(Y_{i2}^{*}|Y_{i1}, Y_{i2} = j, X_i) - \pi_{i0}}{1 - \pi_{i0} - \pi_{i1}} \\
&= \begin{cases} \frac{1\times\pi_{i0} + 0\times(1-\pi_{i0}) - \pi_{i0}}{1-\pi_{i0}-\pi_{i1}}, & \text{if } j = 0 \\ \frac{0\times\pi_{i1} + 1\times(1-\pi_{i1}) - \pi_{i0}}{1-\pi_{i0}-\pi_{i1}}, & \text{if } j = 1 \end{cases} \\
&= \begin{cases} 0, & \text{if } j = 0 \\ 1, & \text{if } j = 1 \end{cases} \\
&= j.
\end{aligned}$$

Thus, (B.1) holds.

**Step 2**: Next, we show that

$$E(Y_{i1}^{**}|Y_{i1}, Y_{i2}, X_i) = Y_{i1}. \tag{B.2}$$

By the definition of $Y_{i1}^{**}$, we obtain that

$$
\begin{aligned}
E(Y_{i1}^{**}|Y_{i1},Y_{i2},X_i) &= E\left(\frac{Y_{i1}^{*}-\gamma_0-\gamma_2 Y_{i2}^{**}-\gamma_3^{\mathrm{T}}X_i}{\gamma_1}\middle|Y_{i1},Y_{i2},X_i\right)\\
&= E\left(\frac{Y_{i1}^{*}-\gamma_0-\gamma_3^{\mathrm{T}}X_i}{\gamma_1}\middle|Y_{i1},Y_{i2},X_i\right)-E\left(\frac{\gamma_2}{\gamma_1}Y_{i2}^{**}\middle|Y_{i1},Y_{i2},X_i\right)\\
&= \frac{1}{\gamma_1}E\left(Y_{i1}^{*}|Y_{i1},Y_{i2},X_i\right)-\frac{\gamma_0}{\gamma_1}-\frac{\gamma_2}{\gamma_1}Y_{i2}-\frac{\gamma_3^{\mathrm{T}}X_i}{\gamma_1}\\
&= Y_{i1},
\end{aligned}
$$

where the third step is due to (B.1), and the last step comes from measurement error model (3.9) together with $E(e_i|Y_{i1},Y_{i2},X_i)=0$.

**Step 3**: Since $e_i$ in (3.9) is independent of $Y_{i1}$, $Y_{i2}$ and $Y_{i2}^{**}$, we have that

$$E\left(Y_{i1}e_i|Y_{i1},Y_{i2},X_i\right)=0 \tag{B.3}$$

and

$$E\left\{e_i(Y_{i2}-Y_{i2}^{**})|Y_{i1},Y_{i2},X_i\right\}=0. \tag{B.4}$$

Now we prove that

$$E\left(\Delta_i|Y_{i1},Y_{i2},X_i\right)=E\left\{(Y_{i2}-Y_{i2}^{**})^2|Y_{i1},Y_{i2},X_i\right\}. \tag{B.5}$$

Indeed, by the definition of $\Delta_i$,

$$
\begin{aligned}
E(\Delta_i|Y_{i1},Y_{i1},Y_{i2}=j,X_i) &= E\left(\frac{\Delta_{i0}^{1-Y_{i2}^{*}}\Delta_{i1}^{Y_{i2}^{*}}-\Delta_{i0}\pi_{i1}-\Delta_{i1}\pi_{i0}}{1-\pi_{i1}-\pi_{i0}}\middle|Y_{i1},Y_{i2}=j,X_i\right)\\
&= \begin{cases}\pi_{i0}\left(\frac{\Delta_{i1}-\Delta_{i0}\pi_{i1}-\Delta_{i1}\pi_{i0}}{1-\pi_{i1}-\pi_{i0}}\right)+(1-\pi_{i0})\left(\frac{\Delta_{i0}-\Delta_{i0}\pi_{i1}-\Delta_{i1}\pi_{i0}}{1-\pi_{i1}-\pi_{i0}}\right), & \text{if } j=0\\ (1-\pi_{i1})\left(\frac{\Delta_{i1}-\Delta_{i0}\pi_{i1}-\Delta_{i1}\pi_{i0}}{1-\pi_{i1}-\pi_{i0}}\right)+\pi_{i1}\left(\frac{\Delta_{i0}-\Delta_{i0}\pi_{i1}-\Delta_{i1}\pi_{i0}}{1-\pi_{i1}-\pi_{i0}}\right), & \text{if } j=1\end{cases}\\
&= \begin{cases}\Delta_{i0}, & \text{if } j=0\\ \Delta_{i1}, & \text{if } j=1\end{cases}\\
&= \Delta_{ij}.
\end{aligned}
$$

On the other hand,

$$
\begin{aligned}
E\left\{(Y_{i2}-Y_{i2}^{**})^2\,|Y_{i1},Y_{i2}=j,X_i\right\} &= E\left\{\left(Y_{i2}-\frac{Y_{i2}^{*}-\pi_{i0}}{1-\pi_{i0}-\pi_{i1}}\right)^2\middle|Y_{i1},Y_{i2}=j,X_i\right\} \\
&= \begin{cases} \pi_{i0}\left(-\frac{1-\pi_{i0}}{1-\pi_{i0}-\pi_{i1}}\right)^2+(1-\pi_{i0})\left(\frac{\pi_{i0}}{1-\pi_{i0}-\pi_{i1}}\right)^2, & \text{if } j=0 \\ (1-\pi_{i1})\left(1-\frac{1-\pi_{i0}}{1-\pi_{i0}-\pi_{i1}}\right)^2+\pi_{i1}\left(1+\frac{\pi_{i0}}{1-\pi_{i0}-\pi_{i1}}\right)^2, & \text{if } j=1 \end{cases} \\
&= \begin{cases} \Delta_{i0}, & \text{if } j=0 \\ \Delta_{i1}, & \text{if } j=1 \end{cases} \\
&= \Delta_{ij}.
\end{aligned}
$$

Hence, (B.5) is proved.

**Step 4**: We show that

$$E(Y_{i11}^{**}|Y_{i1},Y_{i2},X_i)=Y_{i1}^2. \tag{B.6}$$

By the definition of $Y_{i11}^{**}$,

$$
\begin{aligned}
E(Y_{i11}^{**}|Y_{i1},Y_{i2},X_i) =& E\left(Y_{i1}^{**2}-\frac{\sigma_e^2}{\gamma_1^2}-\frac{\gamma_2^2}{\gamma_1^2}\Delta_i\middle|Y_{i1},Y_{i2},X_i\right) \\
=& E\left[\left\{\frac{\gamma_1 Y_{i1}+e_i+\gamma_2(Y_{i2}-Y_{i2}^{**})}{\gamma_1}\right\}^2-\frac{\sigma_e^2}{\gamma_1^2}-\frac{\gamma_2^2}{\gamma_1^2}\Delta_i\middle|Y_{i1},Y_{i2},X_i\right] \\
=& E\left\{Y_{i1}^2+\frac{\gamma_2^2}{\gamma_1^2}(Y_{i2}-Y_{i2}^{**})^2-\frac{\gamma_2^2}{\gamma_1^2}\Delta_i\right. \\
& \left.+\frac{2Y_{i1}e_i}{\gamma_1}+\frac{2\gamma_2 e_i(Y_{i2}-Y_{i2}^{**})}{\gamma_1^2}+\frac{2\gamma_2 Y_{i1}(Y_{i2}-Y_{i2}^{**})}{\gamma_1}\middle|Y_{i1},Y_{i2},X_i\right\} \\
=& Y_{i1}^2+\frac{\gamma_2^2}{\gamma_1^2}E\left\{(Y_{i2}-Y_{i2}^{**})^2-\Delta_i\middle|Y_{i1},Y_{i2},X_i\right\} \\
=& Y_{i1}^2,
\end{aligned}
$$

where the second step comes from the definition of $Y_{i1}^{**}$ and the model (3.9), the fourth step comes from (B.3), (B.4) and $E\left\{Y_{i1}(Y_{i2}-Y_{i2}^{**})|Y_{i1},Y_{i2},X_i\right\}=0$ which is due to (B.1), and the last step is due to (B.5).

**Step 5**: We show that

$$E(Y_{i12}^{**}|Y_{i1}, Y_{i2}, X_i) = Y_{i1}Y_{i2}. \tag{B.7}$$

Similarly, by definition, we obtain that

$$\begin{aligned}
&E(Y_{i12}^{**}|Y_{i1}, Y_{i2}, X_i)\\
=&E\left(Y_{i1}^{**}Y_{i2}^{**} + \frac{\gamma_2}{\gamma_1}\Delta_i\middle|Y_{i1}, Y_{i2}, X_i\right)\\
=&E\left(\frac{Y_{i1}^{*} - \gamma_0 - \gamma_2 Y_{i2}^{**} - \gamma_3^{\mathrm{T}} X_i}{\gamma_1}Y_{i2}^{**} + \frac{\gamma_2}{\gamma_1}\Delta_i\middle|Y_{i1}, Y_{i2}, X_i\right)\\
=&E\left\{\frac{Y_{i1}^{*} - \gamma_0 - \gamma_2 Y_{i2} - \gamma_3^{\mathrm{T}} X_i + \gamma_2(Y_{i2} - Y_{i2}^{**})}{\gamma_1}Y_{i2}^{**} + \frac{\gamma_2}{\gamma_1}\Delta_i\middle|Y_{i1}, Y_{i2}, X_i\right\}\\
=&E\left(\frac{Y_{i1}^{*} - \gamma_0 - \gamma_2 Y_{i2} - \gamma_3^{\mathrm{T}} X_i}{\gamma_1}Y_{i2}^{**}\middle|Y_{i1}, Y_{i2}, X_i\right) + \frac{\gamma_2}{\gamma_1}E\left\{(Y_{i2} - Y_{i2}^{**})Y_{i2}^{**} + \Delta_i\middle|Y_{i1}, Y_{i2}, X_i\right\}\\
=&E\left(\frac{Y_{i1}^{*} - \gamma_0 - \gamma_2 Y_{i2} - \gamma_3^{\mathrm{T}} X_i}{\gamma_1}\middle|Y_{i1}, Y_{i2}, X_i\right)E\left(Y_{i2}^{**}\middle|Y_{i1}, Y_{i2}, X_i\right)\\
&+\frac{\gamma_2}{\gamma_1}E\left\{(Y_{i2} - Y_{i2}^{**})Y_{i2}^{**}\middle|Y_{i1}, Y_{i2}, X_i\right\} + \frac{\gamma_2}{\gamma_1}E\left(\Delta_i\middle|Y_{i1}, Y_{i2}, X_i\right)\\
=&E\left(Y_{i1} + \frac{e_i}{\gamma_1}\middle|Y_{i1}, Y_{i2}, X_i\right)Y_{i2} + \frac{\gamma_2}{\gamma_1}E\left\{(Y_{i2} - Y_{i2}^{**})Y_{i2}^{**}\middle|Y_{i1}, Y_{i2}, X_i\right\} + \frac{\gamma_2}{\gamma_1}E\left(\Delta_i\middle|Y_{i1}, Y_{i2}, X_i\right)\\
=&Y_{i1}Y_{i2} + \frac{\gamma_2}{\gamma_1}E\left\{(Y_{i2} - Y_{i2}^{**})Y_{i2}^{**}\middle|Y_{i1}, Y_{i2}, X_i\right\} + \frac{\gamma_2}{\gamma_1}E\left(\Delta_i\middle|Y_{i1}, Y_{i2}, X_i\right),
\end{aligned} \tag{B.8}$$

where the fifth step is due to the conditional independence assumption for $Y_{i1}^*$ and $Y_{i2}^*$ given by (2.4), and the sixth step comes from (3.9) and (B.1), and last step comes from $E(e_i|Y_{i1}, Y_{i2}, X_i) = 0$.

By (B.5), we obtain that

$$\begin{aligned}
&E(\Delta_i|Y_{i1}, Y_{i2}, X_i) = E\left\{(Y_{i2} - Y_{i2}^{**})^2|Y_{i1}, Y_{i2}, X_i\right\}\\
=&E\left\{(Y_{i2} - Y_{i2}^{**})Y_{i2}|Y_{i1}, Y_{i2}, X_i\right\} - E\left\{(Y_{i2} - Y_{i2}^{**})Y_{i2}^{**}|Y_{i1}, Y_{i2}, X_i\right\}\\
=&-E\left\{(Y_{i2} - Y_{i2}^{**})Y_{i2}^{**}|Y_{i1}, Y_{i2}, X_i\right\},
\end{aligned}$$

where the last step is due to (B.1). Consequently, (B.8) gives (B.7).

**Step 6**: In (3.10) we replace $Y_{i1}$, $Y_{i1}^2$ and $Y_{i1}Y_{i2}$, respectively, with $Y_{i1}^{**}$, $Y_{i11}^{**}$, and $Y_{i12}^{**}$, and then we obtain $U_i^{**}(\theta) = (U_{i1}^{**\mathrm{T}}(\theta), U_{i2}^{**\mathrm{T}}(\theta))^{\mathrm{T}}$ where

$$\begin{aligned} U_{i1}^{**}(\theta) &= D_{1i}^{\mathrm{T}} V_{i1}^{-1} \begin{pmatrix} Y_{i1}^{**} - \mu_{i1} \\ Y_{i2}^{**} - \mu_{i2} \end{pmatrix}, \\ U_{i2}^{**}(\theta) &= D_{2i}^{\mathrm{T}} V_{i2}^{-1} \begin{pmatrix} Y_{i11}^{**} - 2\mu_{i1} Y_{i1}^{**} + \mu_{i1}^2 - \xi_{i1} \\ Y_{i12}^{**} - Y_{i1}^{**}\mu_{i2} - Y_{i2}^{**}\mu_{i1} + \mu_{i1}\mu_{i2} - \xi_{i2} \\ Y_{i2}^{**} - 2\mu_{i2} Y_{i2}^{**} + \mu_{i2}^2 - \xi_{i3} \end{pmatrix}. \end{aligned}$$

Then applying (B.1), (B.2), (B.6), and (B.7) gives that

$$E\{U_i^{**}(\theta) | Y_{i1}, Y_{i2}, X_i\} = U_i(\theta).$$

## B.2 The Consistency and Normality of the Proposed Estimator with External Validation Data

Assume that subjects are randomly assigned to the validation or nonvalidation sample. Let $\delta_i = I(i \in \mathcal{V})$, where $I(\cdot)$ is the indicator function. Define $H_{i1}(\theta, \eta) = (1 - \delta_i) U_{i1}^{**}(\theta, \eta)$, $H_{i2}(\theta, \eta) = (1 - \delta_i) U_{i2}^{**}(\theta, \eta)$, $H_{i3}(\eta) = \delta_i S_i(\eta)$, and $H_i(\theta, \eta) = \{H_{i1}^{\mathrm{T}}(\theta, \eta), H_{i2}^{\mathrm{T}}(\theta, \eta), H_{i3}^{\mathrm{T}}(\eta)\}^{\mathrm{T}}$. Then, (3.13) is equivalent to

$$U^{(E)}(\theta, \eta) = \sum_{i \in \mathcal{M} \bigcup \mathcal{V}} H_i(\theta, \eta).$$

Since $H_i(\theta, \eta)$ is an unbiased estimating function, i.e., $E\{H_i(\theta, \eta)\} = 0$, then by estimating function theory (Godambe, 1991; Newey and McFadden, 1994; Heyde, 1997, Ch.12) we conclude that under regularity conditions, solving $\sum_{i \in \mathcal{M} \bigcup \mathcal{V}} H_i(\theta, \eta) = 0$ gives a consistent estimator, $(\widehat{\theta}_E^{\mathrm{T}}, \widehat{\eta}_E^{\mathrm{T}})^{\mathrm{T}}$, of $(\theta^{\mathrm{T}}, \eta^{\mathrm{T}})^{\mathrm{T}}$.

Applying the Taylor series expansion to $\sum_{i \in \mathcal{M} \bigcup \mathcal{V}} H_i(\widehat{\theta}_E, \widehat{\eta}_E) = 0$, we obtain

$$\sum_{i \in \mathcal{M} \bigcup \mathcal{V}} H_i(\theta, \eta) + \sum_{i \in \mathcal{M} \bigcup \mathcal{V}} \begin{pmatrix} \frac{\partial H_i(\theta, \eta)}{\partial \theta^{\mathrm{T}}} & \frac{\partial H_i(\theta, \eta)}{\partial \eta^{\mathrm{T}}} \end{pmatrix} \left\{ \begin{pmatrix} \widehat{\theta}_E \\ \widehat{\eta}_E \end{pmatrix} - \begin{pmatrix} \theta \\ \eta \end{pmatrix} \right\} + o_p(1) = 0,$$

which leads to

$$\begin{aligned} &\sqrt{1 + \frac{m}{n}} \left\{ -\frac{1}{n+m} \sum_{i \in \mathcal{M} \bigcup \mathcal{V}} \begin{pmatrix} \frac{\partial H_i(\theta, \eta)}{\partial \theta^{\mathrm{T}}} & \frac{\partial H_i(\theta, \eta)}{\partial \eta^{\mathrm{T}}} \end{pmatrix} \right\} \sqrt{n} \left\{ \begin{pmatrix} \widehat{\theta}_E \\ \widehat{\eta}_E \end{pmatrix} - \begin{pmatrix} \theta \\ \eta \end{pmatrix} \right\} \\ &= \frac{1}{\sqrt{n+m}} \sum_{i \in \mathcal{M} \bigcup \mathcal{V}} H_i(\theta, \eta) + o_p(1). \end{aligned} \tag{B.9}$$

Let

$$\Gamma_E = \lim_{n\to\infty} \left\{ -\frac{1}{n+m} \sum_{i\in\mathcal{M}\bigcup\mathcal{V}} E\begin{pmatrix} \frac{\partial H_i(\theta,\eta)}{\partial\theta^{\mathrm{T}}} & \frac{\partial H_i(\theta,\eta)}{\partial\eta^{\mathrm{T}}} \end{pmatrix} \right\}$$

and

$$\Sigma_E = \lim_{n\to\infty} \mathrm{Var}\left( \frac{1}{\sqrt{n+m}} \sum_{i\in\mathcal{M}\bigcup\mathcal{V}} H_i(\theta,\eta) \right).$$

Then applying the central limit theorem to (B.9) gives that

$$\sqrt{n}\left\{ (\hat{\theta}_E^{\mathrm{T}}, \hat{\eta}_E^{\mathrm{T}})^{\mathrm{T}} - (\theta^{\mathrm{T}}, \eta^{\mathrm{T}})^{\mathrm{T}} \right\} \xrightarrow{d} N\left(0, \frac{1}{1+\rho}\Gamma_E^{-1}\Sigma_E(\Gamma_E^{-1})^{\mathrm{T}}\right) \qquad \text{as} \qquad n\to\infty.$$

Now it remains to show that $\Gamma_E$ and $\Sigma_E$ are identical to (3.14). By the definitions of $H_i(\theta,\eta)$, we derive that

$$\begin{aligned}
\Gamma_E &= \lim_{n\to\infty} \left\{ -\frac{1}{n+m} E \begin{pmatrix} \sum\limits_{i\in\mathcal{M}} \frac{\partial U_{i1}^{**}(\theta,\eta)}{\partial\theta^{\mathrm{T}}} & \sum\limits_{i\in\mathcal{M}} \frac{\partial U_{i1}^{**}(\theta,\eta)}{\partial\eta^{\mathrm{T}}} \\ \sum\limits_{i\in\mathcal{M}} \frac{\partial U_{i2}^{**}(\theta,\eta)}{\partial\theta^{\mathrm{T}}} & \sum\limits_{i\in\mathcal{M}} \frac{\partial U_{i2}^{**}(\theta,\eta)}{\partial\eta^{\mathrm{T}}} \\ 0 & 0 \end{pmatrix} - \frac{1}{n+m} E \begin{pmatrix} 0 & 0 \\ 0 & 0 \\ 0 & \sum_{i\in\mathcal{V}} \frac{\partial S_i(\eta)}{\partial\eta^{\mathrm{T}}} \end{pmatrix} \right\} \\
&= \lim_{n\to\infty} \left\{ -\frac{n}{n+m} E \begin{pmatrix} \frac{\partial U_{i1}^{**}(\theta,\eta)}{\partial\theta^{\mathrm{T}}} & \frac{\partial U_{i1}^{**}(\theta,\eta)}{\partial\eta^{\mathrm{T}}} \\ \frac{\partial U_{i2}^{**}(\theta,\eta)}{\partial\theta^{\mathrm{T}}} & \frac{\partial U_{i2}^{**}(\theta,\eta)}{\partial\eta^{\mathrm{T}}} \\ 0 & 0 \end{pmatrix} - \frac{m}{n+m} E \begin{pmatrix} 0 & 0 \\ 0 & 0 \\ 0 & \frac{\partial S_i}{\partial\eta^{\mathrm{T}}} \end{pmatrix} \right\} \\
&= -\frac{1}{1+\rho} \begin{pmatrix} E\left(\frac{\partial U_{i1}^{**}(\theta,\eta)}{\partial\theta^{\mathrm{T}}}\right) & E\left(\frac{\partial U_{i1}^{**}(\theta,\eta)}{\partial\eta^{\mathrm{T}}}\right) \\ E\left(\frac{\partial U_{i2}^{**}(\theta,\eta)}{\partial\theta^{\mathrm{T}}}\right) & E\left(\frac{\partial U_{i2}^{**}(\theta,\eta)}{\partial\eta^{\mathrm{T}}}\right) \\ 0 & 0 \end{pmatrix} - \frac{\rho}{1+\rho} \begin{pmatrix} 0 & 0 \\ 0 & 0 \\ 0 & E\left(\frac{\partial S_i(\eta)}{\partial\eta^{\mathrm{T}}}\right) \end{pmatrix}.
\end{aligned}$$

Similarly, we have that

$$
\begin{aligned}
\Sigma_E &= \lim_{n\to\infty} \frac{1}{n+m} \sum_{i\in\mathcal{M}\bigcup\mathcal{V}} \mathrm{Var}\{H_i(\theta,\eta)\} \\
&= \lim_{n\to\infty} \frac{1}{n+m} \sum_{i\in\mathcal{M}\bigcup\mathcal{V}} E\{H_i(\theta,\eta)H_i^{\mathrm{T}}(\theta,\eta)\} \\
&= \lim_{n\to\infty} \frac{1}{n+m} \sum_{i\in\mathcal{M}\bigcup\mathcal{V}} E \begin{pmatrix} (1-\delta_i)^2 U_{i1}^{**}(\theta,\eta)U_{i1}^{**\mathrm{T}}(\theta,\eta) & (1-\delta_i)^2 U_{i1}^{**}(\theta,\eta)U_{i2}^{**\mathrm{T}}(\theta,\eta) & 0 \\ (1-\delta_i)^2 U_{i2}^{**}(\theta,\eta)U_{i1}^{**\mathrm{T}}(\theta,\eta) & (1-\delta_i)^2 U_{i2}^{**}(\theta,\eta)U_{i2}^{**\mathrm{T}}(\theta,\eta) & 0 \\ 0 & 0 & \delta_i^2 S_i(\eta)S_i^{\mathrm{T}}(\eta) \end{pmatrix} \\
&= \lim_{n\to\infty} \frac{n}{n+m}\frac{1}{n} \sum_{i\in\mathcal{M}} \begin{pmatrix} E\{U_{i1}^{**}(\theta,\eta)U_{i1}^{**\mathrm{T}}(\theta,\eta)\} & E\{U_{i1}^{**}(\theta,\eta)U_{i2}^{**\mathrm{T}}(\theta,\eta)\} & 0 \\ E\{U_{i2}^{**}(\theta,\eta)U_{i1}^{**\mathrm{T}}(\theta,\eta)\} & E\{U_{i2}^{**}(\theta,\eta)U_{i2}^{**\mathrm{T}}(\theta,\eta)\} & 0 \\ 0 & 0 & 0 \end{pmatrix} \\
&\quad + \lim_{n\to\infty} \frac{m}{n+m}\frac{1}{m} \sum_{i\in\mathcal{V}} \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & E\{S_i(\eta)S_i^{\mathrm{T}}(\eta)\} \end{pmatrix} \\
&= \frac{1}{1+\rho} \begin{pmatrix} E\{U_{i1}^{**}(\theta,\eta)U_{i1}^{**\mathrm{T}}(\theta,\eta)\} & E\{U_{i1}^{**}(\theta,\eta)U_{i2}^{**\mathrm{T}}(\theta,\eta)\} & 0 \\ E\{U_{i2}^{**}(\theta,\eta)U_{i1}^{**\mathrm{T}}(\theta,\eta)\} & E\{U_{i2}^{**}(\theta,\eta)U_{i2}^{**\mathrm{T}}(\theta,\eta)\} & 0 \\ 0 & 0 & 0 \end{pmatrix} + \frac{\rho}{1+\rho} \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & E\{S_i(\eta)S_i^{\mathrm{T}}(\eta)\} \end{pmatrix},
\end{aligned}
$$

where the third step is due to $(1-\delta_i)^2 = (1-\delta_i)$ and $\delta_i^2 = \delta_i$.

## B.3 The Consistency and Normality of the Proposed Estimator with Internal Validation Data

Similar to Section B.2, we assume that subjects are randomly assigned to the validation or nonvalidation sample. We define $H_{i1}(\theta,\eta) = (1-\delta_i)U_{i1}^{**}(\theta,\eta) + \delta_i U_{i1}(\theta,\eta)$, $H_{i2}(\theta,\eta) = (1-\delta_i)U_{i2}^{**}(\theta,\eta) + \delta_i U_{i2}(\theta,\eta)$, $H_{i3}(\theta,\eta) = \delta_i S_i(\eta)$, and $H_i(\theta,\eta) = \{H_{i1}^{\mathrm{T}}(\theta,\eta), H_{i2}^{\mathrm{T}}(\theta,\eta), H_{i3}^{\mathrm{T}}(\eta)\}^{\mathrm{T}}$. Then, (3.15) is equivalent to

$$U^{(I)}(\theta,\eta) = \sum_{i\in\mathcal{M}} H_i(\theta,\eta).$$

Similar to Section B.2, we conclude that under regularity conditions, solving $\sum_{i\in\mathcal{M}} H_i(\theta,\eta) = 0$ gives a consistent estimator, $(\hat{\theta}_I^{\mathrm{T}}, \hat{\eta}_I^{\mathrm{T}})^{\mathrm{T}}$, of $(\theta^{\mathrm{T}}, \eta^{\mathrm{T}})^{\mathrm{T}}$.

Applying the Taylor series expansion to $\sum_{i\in\mathcal{M}} H_i(\hat{\theta}_I^{\mathrm{T}}, \hat{\eta}_I^{\mathrm{T}}) = 0$, we obtain

$$\sum_{i\in\mathcal{M}} H_i(\theta,\eta) + \sum_{i\in\mathcal{M}} \begin{pmatrix} \frac{\partial H_i(\theta,\eta)}{\partial\theta^{\mathrm{T}}} & \frac{\partial H_i(\theta,\eta)}{\partial\eta^{\mathrm{T}}} \end{pmatrix} \left\{ \begin{pmatrix} \widehat{\theta}_I \\ \widehat{\eta}_I \end{pmatrix} - \begin{pmatrix} \theta \\ \eta \end{pmatrix} \right\} + o_p(1) = 0,$$

which leads to

$$\left\{-\frac{1}{n}\sum_{i\in\mathcal{M}}\begin{pmatrix}\frac{\partial H_i(\theta,\eta)}{\partial\theta^{\mathrm{T}}} & \frac{\partial H_i(\theta,\eta)}{\partial\eta^{\mathrm{T}}}\end{pmatrix}\right\}\sqrt{n}\left\{\begin{pmatrix}\widehat{\theta}_I\\ \widehat{\eta}_I\end{pmatrix}-\begin{pmatrix}\theta\\ \eta\end{pmatrix}\right\}=\frac{1}{\sqrt{n}}\sum_{i\in\mathcal{M}}H_i(\theta,\eta)+o_p(1). \quad \text{(B.10)}$$

Let

$$\Gamma_{\mathrm{I}}=\lim_{n\to\infty}\left\{-\frac{1}{n}\sum_{i\in\mathcal{M}}E\begin{pmatrix}\frac{\partial H_i(\theta,\eta)}{\partial\theta^{\mathrm{T}}} & \frac{\partial H_i(\theta,\eta)}{\partial\eta^{\mathrm{T}}}\end{pmatrix}\right\}$$

and

$$\Sigma_{\mathrm{I}}=\lim_{n\to\infty}\mathrm{Var}\left(\frac{1}{\sqrt{n}}\sum_{i\in\mathcal{M}}H_i(\theta,\eta)\right).$$

Then applying the central limit theorem to (B.10) gives that

$$\sqrt{n}\left\{(\hat{\theta}_{\mathrm{I}}^{\mathrm{T}},\hat{\eta}_{\mathrm{I}}^{\mathrm{T}})^{\mathrm{T}}-(\theta^{\mathrm{T}},\eta^{\mathrm{T}})^{\mathrm{T}}\right\}\xrightarrow{d}N\left(0,\Gamma_{\mathrm{I}}^{-1}\Sigma_{\mathrm{I}}(\Gamma_{\mathrm{I}}^{-1})^{\mathrm{T}}\right)\qquad\text{as}\qquad n\to\infty.$$

Now it remains to show that $\Gamma_{\mathrm{I}}$ and $\Sigma_{\mathrm{I}}$ are identical to (3.17). By definition of $H_i(\theta,\eta)$, $\Gamma_{\mathrm{I}}$ equals

$$\begin{aligned}
&\lim_{n\to\infty}\left\{-\frac{1}{n}E\begin{pmatrix}\sum_{i\in\mathcal{M}\setminus\mathcal{V}}\frac{\partial U_{i1}^{**}(\theta,\eta)}{\partial\theta^{\mathrm{T}}} & \sum_{i\in\mathcal{M}\setminus\mathcal{V}}\frac{\partial U_{i1}^{**}(\theta,\eta)}{\partial\eta^{\mathrm{T}}}\\ \sum_{i\in\mathcal{M}\setminus\mathcal{V}}\frac{\partial U_{i2}^{**}(\theta,\eta)}{\partial\theta^{\mathrm{T}}} & \sum_{i\in\mathcal{M}\setminus\mathcal{V}}\frac{\partial U_{i2}^{**}(\theta,\eta)}{\partial\eta^{\mathrm{T}}}\\ 0 & 0\end{pmatrix}-\frac{1}{n}E\begin{pmatrix}\sum_{i\in\mathcal{V}}\frac{\partial U_{i1}(\theta,\eta)}{\partial\theta^{\mathrm{T}}} & 0\\ \sum_{i\in\mathcal{V}}\frac{\partial U_{i2}(\theta,\eta)}{\partial\theta^{\mathrm{T}}} & 0\\ 0 & \sum_{i\in\mathcal{V}}\frac{\partial S_i(\eta)}{\partial\eta^{\mathrm{T}}}\end{pmatrix}\right\}\\
&=\lim_{n\to\infty}\left\{-\frac{n-m}{n}E\begin{pmatrix}\frac{\partial U_{i1}^{**}(\theta,\eta)}{\partial\theta^{\mathrm{T}}} & \frac{\partial U_{i1}^{**}(\theta,\eta)}{\partial\eta^{\mathrm{T}}}\\ \frac{\partial U_{i2}^{**}(\theta,\eta)}{\partial\theta^{\mathrm{T}}} & \frac{\partial U_{i2}^{**}(\theta,\eta)}{\partial\eta^{\mathrm{T}}}\\ 0 & 0\end{pmatrix}-\frac{m}{n}E\begin{pmatrix}\frac{\partial U_{i1}(\theta,\eta)}{\partial\theta^{\mathrm{T}}} & 0\\ \frac{\partial U_{i2}(\theta,\eta)}{\partial\theta^{\mathrm{T}}} & 0\\ 0 & \frac{\partial S_i(\eta)}{\partial\eta^{\mathrm{T}}}\end{pmatrix}\right\}\\
&=-(1-\rho)\begin{pmatrix}E\left(\frac{\partial U_{i1}^{**}(\theta,\eta)}{\partial\theta^{\mathrm{T}}}\right) & E\left(\frac{\partial U_{i1}^{**}(\theta,\eta)}{\partial\eta^{\mathrm{T}}}\right)\\ E\left(\frac{\partial U_{i2}^{**}(\theta,\eta)}{\partial\theta^{\mathrm{T}}}\right) & E\left(\frac{\partial U_{i2}^{**}(\theta,\eta)}{\partial\theta^{\mathrm{T}}}\right)\\ 0 & 0\end{pmatrix}-\rho\begin{pmatrix}E\left(\frac{\partial U_{i1}(\theta,\eta)}{\partial\theta^{\mathrm{T}}}\right) & 0\\ E\left(\frac{\partial U_{i2}(\theta,\eta)}{\partial\theta^{\mathrm{T}}}\right) & 0\\ 0 & E\left(\frac{\partial S_i(\eta)}{\partial\eta^{\mathrm{T}}}\right)\end{pmatrix}.
\end{aligned}$$

Similarly, we have that

$$
\begin{aligned}
\Sigma_{\mathrm{I}} &= \lim_{n\to\infty}\frac{1}{n}\sum_{i\in\mathcal{M}}\mathrm{Var}\left\{H_i(\theta,\eta)\right\}\\
&= \lim_{n\to\infty}\frac{1}{n}\sum_{i\in\mathcal{M}}E\{H_i(\theta,\eta)H_i^{\mathrm{T}}(\theta,\eta)\}\\
&= \lim_{n\to\infty}\frac{1}{n}\sum_{i\in\mathcal{M}}E\begin{pmatrix}(1-\delta_i)^2U_{i1}^{**}(\theta,\eta)U_{i1}^{**\mathrm{T}}(\theta,\eta) & (1-\delta_i)^2U_{i1}^{**}(\theta,\eta)U_{i2}^{**\mathrm{T}}(\theta,\eta) & 0\\ (1-\delta_i)^2U_{i2}^{**}(\theta,\eta)U_{i1}^{**\mathrm{T}}(\theta,\eta) & (1-\delta_i)^2U_{i2}^{**}(\theta,\eta)U_{i2}^{**\mathrm{T}}(\theta,\eta) & 0\\ 0 & 0 & 0\end{pmatrix}\\
&\quad+\lim_{n\to\infty}\frac{1}{n}\sum_{i\in\mathcal{M}}E\begin{pmatrix}\delta_i^2U_{i1}(\theta,\eta)U_{i1}^{\mathrm{T}}(\theta,\eta) & \delta_i^2U_{i1}(\theta,\eta)U_{i2}^{\mathrm{T}}(\theta,\eta) & \delta_i^2U_{i1}(\theta,\eta)S_i^{\mathrm{T}}(\eta)\\ \delta_i^2U_{i2}(\theta,\eta)U_{i1}^{\mathrm{T}}(\theta,\eta) & \delta_i^2U_{i1}(\theta,\eta)U_{i1}^{\mathrm{T}}(\theta,\eta) & \delta_i^2U_{i1}(\theta,\eta)S_i^{\mathrm{T}}(\eta)\\ \delta_i^2U_{i1}(\theta,\eta)S_i^{\mathrm{T}}(\eta) & \delta_i^2U_{i2}(\theta,\eta)S_i^{\mathrm{T}}(\eta) & \delta_i^2S_i(\eta)S_i^{\mathrm{T}}(\eta)\end{pmatrix}\\
&= \lim_{n\to\infty}\frac{n-m}{n}\frac{1}{n-m}\sum_{i\in\mathcal{M}\setminus\mathcal{V}}\begin{pmatrix}E\{U_{i1}^{**}(\theta,\eta)U_{i1}^{**\mathrm{T}}(\theta,\eta)\} & E\{U_{i1}^{**}(\theta,\eta)U_{i2}^{**\mathrm{T}}(\theta,\eta)\} & 0\\ E\{U_{i2}^{**}(\theta,\eta)U_{i1}^{**\mathrm{T}}(\theta,\eta)\} & E\{U_{i2}^{**}(\theta,\eta)U_{i2}^{**\mathrm{T}}(\theta,\eta)\} & 0\\ 0 & 0 & 0\end{pmatrix}\\
&\quad+\lim_{n\to\infty}\frac{m}{n}\frac{1}{m}\sum_{i\in\mathcal{V}}\begin{pmatrix}E\{U_{i1}(\theta,\eta)U_{i1}^{\mathrm{T}}(\theta,\eta)\} & E\{U_{i1}(\theta,\eta)U_{i2}^{\mathrm{T}}(\theta,\eta)\} & E\{U_{i1}(\theta,\eta)S_i^{\mathrm{T}}(\theta,\eta)\}\\ E\{U_{i2}(\theta,\eta)U_{i1}^{\mathrm{T}}(\theta,\eta)\} & E\{U_{i1}(\theta,\eta)U_{i1}^{\mathrm{T}}(\theta,\eta)\} & E\{U_{i1}(\theta,\eta)S_i^{\mathrm{T}}(\theta,\eta)\}\\ E\{U_{i1}(\theta,\eta)S_i^{\mathrm{T}}(\theta,\eta)\} & E\{U_{i2}(\theta,\eta)S_i^{\mathrm{T}}(\theta,\eta)\} & E\{S_i(\theta,\eta)S_i^{\mathrm{T}}(\theta,\eta)\}\end{pmatrix}\\
&=(1-\rho)\begin{pmatrix}E\{U_{i1}^{**}(\theta,\eta)U_{i1}^{**\mathrm{T}}(\theta,\eta)\} & E\{U_{i1}^{**}(\theta,\eta)U_{i2}^{**\mathrm{T}}(\theta,\eta)\} & 0\\ E\{U_{i2}^{**}(\theta,\eta)U_{i1}^{**\mathrm{T}}(\theta,\eta)\} & E\{U_{i2}^{**}(\theta,\eta)U_{i2}^{**\mathrm{T}}(\theta,\eta)\} & 0\\ 0 & 0 & 0\end{pmatrix}\\
&\quad+\rho\begin{pmatrix}E\{U_{i1}(\theta,\eta)U_{i1}^{\mathrm{T}}(\theta,\eta)\} & E\{U_{i1}(\theta,\eta)U_{i2}^{\mathrm{T}}(\theta,\eta)\} & E\{U_{i1}(\theta,\eta)S_i^{\mathrm{T}}(\eta)\}\\ E(U_{i2}(\theta,\eta)U_{i1}^{\mathrm{T}}(\theta,\eta)\} & E(U_{i1}(\theta,\eta)U_{i1}^{\mathrm{T}}(\theta,\eta)\} & E\{U_{i1}(\theta,\eta)S_i^{\mathrm{T}}(\eta)\}\\ E\{U_{i1}(\theta,\eta)S_i^{\mathrm{T}}(\eta)\} & E\{U_{i2}(\theta,\eta)S_i^{\mathrm{T}}(\eta)\} & E\{S_i(\eta)S_i^{\mathrm{T}}(\eta)\}\end{pmatrix},
\end{aligned}
$$

where the third step is due to $\delta_i(1-\delta_i)=0$ and the fourth step comes from $(1-\delta_i)^2=(1-\delta_i)$ and $\delta_i^2=\delta_i$.

## B.4 Determination the values of $\widehat{\textbf{Var}}(\widehat{\theta}_{Ir}^{(1)})$, $\widehat{\textbf{Var}}(\widehat{\theta}_{Ir}^{(0)})$ and $\widehat{\textbf{Cov}}(\widehat{\theta}_{Ir}^{(0)}, \widehat{\theta}_{Ir}^{(1)})$

To study the covariance between $\widehat{\theta}_{Ir}^{(0)}$ and $\widehat{\theta}_{Ir}^{(1)}$, we jointly combine the estimating procedure by stacking the estimating functions from validation data and nonvalidation data. However, this procedure makes the resulting dimension of estimating functions be greater than the dimension of parameter $\theta$. To overcome this problem, following the spirit of (Shu and Yi, 2017) , we enlarge the original parameter space by using different symbols, say, $\theta^{(0)}$ and $\theta^{(1)}$, respectively, to represent the parameter $\theta$ in the estimating function of $\widehat{\theta}_{I}^{(0)}$ and $\widehat{\theta}_{I}^{(1)}$, where the true value of $\theta^{(0)}$ and $\theta^{(1)}$ are identical to that of $\theta$. Specifically, consider the estimating functions

$$\Psi_i(\theta^{(0)\mathrm{T}}, \eta^{\mathrm{T}}, \theta^{(1)\mathrm{T}}) = \begin{pmatrix} I(i \in \mathcal{V}) \cdot U_{i1}(\theta^{(0)}, \eta) \\ I(i \in \mathcal{V}) \cdot U_{i2}(\theta^{(0)}, \eta) \\ I(i \in \mathcal{V}) \cdot S_i(\eta) \\ I(i \in \mathcal{M}\backslash\mathcal{V}) \cdot U_{i1}^{**}(\theta^{(1)}, \eta) \\ I(i \in \mathcal{M}\backslash\mathcal{V}) \cdot U_{i2}^{**}(\theta^{(1)}, \eta) \end{pmatrix}. \tag{B.11}$$

Solving the estimating function $\sum_{i=1}^{n} \Psi_i(\theta^{(0)\mathrm{T}}, \eta^{\mathrm{T}}, \theta^{(1)\mathrm{T}}) = 0$, we obtain an estimator, $(\widehat{\theta}^{(0)\mathrm{T}}, \widehat{\eta}^{\mathrm{T}}, \widehat{\theta}^{(1)\mathrm{T}})^{\mathrm{T}}$, of $(\theta^{(0)\mathrm{T}}, \eta^{\mathrm{T}}, \theta^{(1)\mathrm{T}})^{\mathrm{T}}$. By estimating function theory (e.g. Godambe, 1991; Newey and McFadden, 1994; Yi, 2017, Section 1.3.2), the variance of $(\widehat{\theta}^{(0)\mathrm{T}}, \widehat{\eta}^{\mathrm{T}}, \widehat{\theta}^{(1)\mathrm{T}})^{\mathrm{T}}$ can be estimated by the empirical sandwich estimator

$$\widehat{\mathrm{Var}}\left\{(\widehat{\theta}^{(0)\mathrm{T}}, \widehat{\eta}^{\mathrm{T}}, \widehat{\theta}^{(1)\mathrm{T}})^{\mathrm{T}}\right\} = \frac{1}{n}\Gamma_{\Psi}^{-1}\Sigma_{\Psi}\Gamma_{\Psi}^{-1\mathrm{T}}, \tag{B.12}$$

where $\Gamma_{\Psi} = \frac{1}{n}\sum_{i=1}^{n}\left\{-\frac{\partial}{\partial(\theta^{(0)\mathrm{T}}, \eta^{\mathrm{T}}, \theta^{(1)\mathrm{T}})^{\mathrm{T}}}\Psi_i(\widehat{\theta}^{(0)}, \widehat{\eta}, \widehat{\theta}^{(1)})\right\}$ and
$\Sigma_{\Psi} = \frac{1}{n}\sum_{i=1}^{n}\left\{\Psi_i(\widehat{\theta}^{(0)}, \widehat{\eta}, \widehat{\theta}^{(1)})\Psi_i^{\mathrm{T}}(\widehat{\theta}^{(0)}, \widehat{\eta}, \widehat{\theta}^{(1)})\right\}$.

Therefore, $\widehat{\mathrm{Var}}(\widehat{\theta}_{Ir}^{(0)})$, $\widehat{\mathrm{Var}}(\widehat{\theta}_{Ir}^{(1)})$ and $\widehat{\mathrm{Cov}}(\widehat{\theta}_{Ir}^{(0)}, \widehat{\theta}_{Ir}^{(1)})$ are, respectively, the covariance matrix, $\widehat{\mathrm{Var}}\left\{(\widehat{\theta}^{(0)\mathrm{T}}, \widehat{\eta}^{\mathrm{T}}, \widehat{\theta}^{(1)\mathrm{T}})^{\mathrm{T}}\right\}$, corresponding to elements $(r, r)$, $(r + q, r + q)$ and $(r, r + q)$ where $q = p_{\theta} + p_{\eta}$.

# Appendix C

# Conditions and Proofs of the Results in Chapter 4

## C.1 Regularity Conditions

(R1) The dimension $p$ of the covariates is of a polynomial order of the sample size $n$. That is, $p = O(n^{\gamma})$ for a constant $\gamma > 0$.

(R2) There exists $0 \le \kappa < 1$ so that

$$\max_{s \in V} |\mathcal{N}(s)| = O(n^{\kappa}).$$

(R3) There exists some $m < \infty$ so that

$$\max_{s \in V,\, t \in \mathcal{N}(s)} |\mathcal{N}(s) \bigcap \mathcal{N}(t)| \le m.$$

(R4) There exists a constant $\delta > 0$ and $\xi$ with $\kappa < \xi \le 1$ for $\kappa$ in condition (R2), such that for every edge $(s,t) \in E$,

$$|\pi_{st}| \ge \delta n^{-\frac{1-\xi}{2}},$$

where $\pi_{st}$ is the partial correlation between $X_{is}$ and $X_{it}$ after eliminating the linear effects from all remaining variables $\{X_{ik} : k \in V \setminus \{s,t\}\}$.

(R5) The covariance matrix of $X_i$ is non-singular.

(R6) The parameter space $\mathcal{B}$ of $(\beta_{\mathrm{M}}^{\mathrm{T}}, \beta_{\mathrm{I}}^{\mathrm{T}}, \phi^{\mathrm{T}})^{\mathrm{T}}$ is compact.

(R7) Given the data, the estimating functions $\sum_{i=1}^{n} U_i(\beta_{\mathrm{M}}, \beta_{\mathrm{I}}, \phi)$ are continuous in $(\beta_{\mathrm{M}}^{\mathrm{T}}, \beta_{\mathrm{I}}^{\mathrm{T}}, \phi^{\mathrm{T}})^{\mathrm{T}}$ everywhere, and satisfy the condition (12.5) in Theorem 12.1 of Heyde (1997).

(R8) Given the data, the estimating functions $\sum_{i=1}^{n} U_i(\beta_{\mathrm{M}}, \beta_{\mathrm{I}}, \phi)$ are continuously differentiable in a neighbourhood of $(\beta_{\mathrm{M}}^{\mathrm{T}}, \beta_{\mathrm{I}}^{\mathrm{T}}, \phi^{\mathrm{T}})^{\mathrm{T}}$.

(R9) For $U_i(w; \beta_{\mathrm{M}}, \beta_{\mathrm{I}}, \phi)$ defined in (4.8) with $w = (y, x^{\mathrm{T}})^{\mathrm{T}}$, there exists a function $h(w)$ with $E\{h(W_i)\} < \infty$, such that $|U_i(w, \beta_{\mathrm{M}}, \beta_{\mathrm{I}}, \phi)| < h(w)$ for all $\beta_{\mathrm{M}}$, $\beta_{\mathrm{I}}$, and $\phi$, where $W_i = (Y_i, X_i^{\mathrm{T}})^{\mathrm{T}}$.

(R10) The equation $E_{\mathrm{Y|X}}\{U_i(\beta_{\mathrm{M}(0)}, \beta_{\mathrm{I}(0)}, \phi_0)\} = 0$ has an unique solution.

Conditions (R1)–(R5) are the regularity conditions discussed by Meinshausen and Bühlmann (2006). Condition (R1) allows for a high dimension of the covariates and regulates the dimension of covariates on a scale relative to the sample size. Conditions (R2) and (R3) basically regulate the sparsity of the graph and the maximum possible growth rate of the size of neighborhoods. Condition (R4) provides a lower bound of the magnitude of partial correlations to ensure the consistency of variable selection of edge set $E$. Condition (R5) requires the existence of the precision matrix. Conditions (R6)-(R8) are the regularity conditions for estimating functions discussed by Heyde (1997). Condition (R9) is the condition for Theorem 2 in Jennrich (1969). Condition (R10) is used to show the consistency in Theorem 2, which was also assumed by Yi and Reid (2010), among others.

## C.2 Proof of Theorem 4.1

For $\kappa$ and $\xi$ defined in Conditions (R3) and (R4), consider a tuning parameter $\lambda$ in (4.6) satisfying $\lambda \sim dn^{-(1-\epsilon)/2}$ with $\kappa < \epsilon < \xi$ and a scaling constant $d > 0$. Then according to Meinshausen and Bühlmann (2006, Page 1445), with regularity conditions assumed, there exists $c > 0$ such that

$$P(\widehat{E} = E) = 1 - O(\exp(-cn^{\epsilon})), \tag{C.1}$$

yielding, by the definition of $\widehat{\beta}_{\mathrm{II}}$, that

$$P(\widehat{\beta}_{\mathrm{II}} = \beta_{\mathrm{II}}) \geq 1 - O(\exp(-cn^{\epsilon})).$$

Thus, the conclusion follows from that $0 < \epsilon < \xi \leq 1$.

## C.3 Proof of Theorem 4.2

**Proof of Theorem 4.2** (i)

**Step 1**: We introduce basic notation first

Let $(\beta_{\mathrm{M}}, \beta_{\mathrm{I}}, \beta_{\mathrm{II}}, \phi)$ be the generic symbol of parameters from parameter space $\mathcal{B}$, and let $(\beta_0^{\mathrm{T}}, \phi_0^{\mathrm{T}})^{\mathrm{T}} = (\beta_{\mathrm{M}(0)}^{\mathrm{T}}, \beta_{\mathrm{I}(0)}^{\mathrm{T}}, \beta_{\mathrm{II}(0)}^{\mathrm{T}}, \phi_0^{\mathrm{T}})^{\mathrm{T}}$ denote the true value of the parameters $\beta = (\beta_{\mathrm{M}}^{\mathrm{T}}, \beta_{\mathrm{I}}^{\mathrm{T}}, \beta_{\mathrm{II}}^{\mathrm{T}}, \phi^{\mathrm{T}})^{\mathrm{T}}$. Let $\mathcal{E} = \{E^a : E^a \subseteq \widetilde{E}\}$ denote the collection of all possible subsets of $\widetilde{E}$. For an estimated edge set $\widehat{E}$ for $E$, let $\beta^{*(\widehat{E})} = \{\beta_{st} : (s,t) \in \widehat{E}\}$ denote the subvector of $(\beta_{\mathrm{I}}^{\mathrm{T}}, \beta_{\mathrm{II}}^{\mathrm{T}})^{\mathrm{T}}$ with the indexes included in $\widehat{E}$, and let $\beta^{**(\widehat{E})} = \{\beta_{st} : (s,t) \notin \widehat{E}\}$ be the complement of $\beta^{*(\widehat{E})}$, i.e., the subvector of $(\beta_{\mathrm{I}}^{\mathrm{T}}, \beta_{\mathrm{II}}^{\mathrm{T}})^{\mathrm{T}}$ with the indexes not included in $\widehat{E}$. The introduction of $\beta^{*(\widehat{E})}$ and $\beta^{**(\widehat{E})}$ offers a new way to partition the vector $(\beta_{\mathrm{I}}^{\mathrm{T}}, \beta_{\mathrm{II}}^{\mathrm{T}})^{\mathrm{T}}$, or $E$, according to the estimated set $\widehat{E}$.

For $U_i(\beta_{\mathrm{M}}, \beta_{\mathrm{I}}, \phi)$ in (4.8) and a generic element $E^a$ in $\mathcal{E}$, let $U_i^{\dagger}(\beta_{\mathrm{M}}, \beta^{*(E^a)}, \phi)$ denote the estimating function $U_i(\beta_{\mathrm{M}}, \beta_{\mathrm{I}}, \phi)$ with $\beta_{\mathrm{I}}$ replaced by $\beta^{*(E^a)}$, and define

$$H(\beta_{\mathrm{M}}, \beta_{\mathrm{I}}, \beta_{\mathrm{II}}, \phi) = \begin{pmatrix} E_{\mathrm{Y|X}}\{U_i(\beta_{\mathrm{M}}, \beta_{\mathrm{I}}, \phi)\} \\ \beta_{\mathrm{II}} \end{pmatrix},$$

$$H_n^{\dagger}(\beta_{\mathrm{M}}, \beta_{\mathrm{I}}, \beta_{\mathrm{II}}, \phi) = \sum_{E^a \in \mathcal{E}} \begin{pmatrix} \frac{1}{n}\sum_{i=1}^n U_i^{\dagger}(\beta_{\mathrm{M}}, \beta^{*(E^a)}, \phi) \\ \beta^{**(E^a)} \end{pmatrix} \cdot I(E^a = \widehat{E}), \tag{C.2}$$

and

$$H^{\dagger}(\beta_{\mathrm{M}}, \beta_{\mathrm{I}}, \beta_{\mathrm{II}}, \phi) = E_{\mathrm{Y|X}}\{H_n^{\dagger}(\beta_{\mathrm{M}}, \beta_{\mathrm{I}}, \beta_{\mathrm{II}}, \phi)\}, \tag{C.3}$$

the expectation is taken with respect to $Y_i$ given $X_i$, and $H^{\dagger}(\beta_{\mathrm{M}}, \beta_{\mathrm{I}}, \beta_{\mathrm{II}}, \phi)$ can be written as $\sum_{E^a \in \mathcal{E}} \begin{pmatrix} E_{\mathrm{Y|X}}\{U_i^{\dagger}(\beta_{\mathrm{M}}, \beta^{*(E^a)}, \phi)\} \\ \beta^{**(E^a)} \end{pmatrix} \cdot I(E^a = \widehat{E})$.

**Step 2**: To show the consistency of $(\widehat{\beta}_{\mathrm{M}}^{\mathrm{T}}, \widehat{\beta}_{\mathrm{I}}^{\mathrm{T}}, \widehat{\phi}^{\mathrm{T}})^{\mathrm{T}}$, we apply Theorem 5.9 of Van der Vaart (2000, Page 46) by varying the required conditions. That is, it suffices to show that

Claim 1: $\inf_{(\beta^{\mathrm{T}}, \phi^{\mathrm{T}})^{\mathrm{T}} \in \mathcal{B}(\eta)} \|H(\beta, \phi)\| > 0 = \|H(\beta_0, \phi_0)\|$.

Claim 2: $\sup_{\beta,\phi} \left\|H_n^{\dagger}(\beta, \phi) - H(\beta, \phi)\right\| \xrightarrow{p} 0$ as $n \to \infty$.

**Step 3**: Show Claim 1.

By Condition (R10) and the definition of $\beta_{\text{II}}$, $(\beta_{\text{M}(0)}^{\text{T}}, \beta_{\text{I}(0)}^{\text{T}}, \beta_{\text{II}(0)}^{\text{T}}, \phi_0^{\text{T}})^{\text{T}}$ is the unique solution to the equation $H(\beta_{\text{M}}, \beta_{\text{I}}, \beta_{\text{II}}, \phi) = 0$.

Given the data, by Conditions (R6) and (R7), applying by the Heine-Cantor Theorem (Rudin, 1976, Theorem 4.19), $\left(U_i^{\text{T}}(\beta_{\text{M}}, \beta_{\text{I}}, \phi), \beta_{\text{II}}^{\text{T}}\right)^{\text{T}}$ is uniformly continuous in $\beta$ and $\phi$, implying that $\|H(\beta, \phi)\|$ is continuous, where $\|\cdot\|$ is the Euclidean norm. Since the set $\mathcal{B}(\eta) = \left\{(\beta^{\text{T}}, \phi^{\text{T}})^{\text{T}} \in \mathcal{B} : \left\|(\beta^{\text{T}}, \phi^{\text{T}})^{\text{T}} - (\beta_0^{\text{T}}, \phi_0^{\text{T}})^{\text{T}}\right\| \geq \eta\right\}$ is a compact subset of $\mathcal{B}$ for any $\eta > 0$, there exists $(\beta_1^{\text{T}}, \phi_1^{\text{T}})^{\text{T}} \in \mathcal{B}(\eta)$ such that

$$\inf_{(\beta^{\text{T}}, \phi^{\text{T}})^{\text{T}} \in \mathcal{B}(\eta)} \|H(\beta, \phi)\| = \|H(\beta_1, \phi_1)\|.$$

Since $(\beta_0^{\text{T}}, \phi_0^{\text{T}})^{\text{T}}$ is the unique solution of $H(\beta, \phi) = 0$, then for any $(\beta_1^{\text{T}}, \phi_1^{\text{T}})^{\text{T}} \neq (\beta_0^{\text{T}}, \phi_0^{\text{T}})^{\text{T}}$, we have that $\|H(\beta_1, \phi_1)\| > 0$. That is,

$$\inf_{(\beta^{\text{T}}, \phi^{\text{T}})^{\text{T}} \in \mathcal{B}(\eta)} \|H(\beta, \phi)\| > 0 = \|H(\beta_0, \phi_0)\|.$$

**Step 4**: Show Claim 2.

Noting that

$$\sup_{\beta,\phi} \left\|H_n^{\dagger}(\beta, \phi) - H(\beta, \phi)\right\| \leq \sup_{\beta,\phi} \left\|H_n^{\dagger}(\beta, \phi) - H^{\dagger}(\beta, \phi)\right\| + \sup_{\beta,\phi} \left\|H^{\dagger}(\beta, \phi) - H(\beta, \phi)\right\|, \tag{C.4}$$

we examine the two terms on the right-hand side of (C.4) separately.

1°. For the first term on the right-hand side of (C.4), by (C.2) and (C.3),

$$\begin{aligned}
&\sup_{\beta,\phi} \left\|H_n^{\dagger}(\beta, \phi) - H^{\dagger}(\beta, \phi)\right\| \\
=&\sup_{\beta,\phi} \left\| \sum_{E^a \in \mathcal{E}} \begin{pmatrix} \frac{1}{n}\sum_{i=1}^n U_i^{\dagger}(\beta_{\text{M}}, \beta^{*(E^a)}, \phi) - E\{U_i^{\dagger}(\beta_{\text{M}}, \beta^{*(E^a)}, \phi)\} \\ \beta^{**(E^a)} - \beta^{**(E^a)} \end{pmatrix} \cdot I(E^a = \widehat{E}) \right\|, \\
=&\sup_{\beta,\phi} \left\| \sum_{E^a \in \mathcal{E}} \left[ \frac{1}{n}\sum_{i=1}^n U_i^{\dagger}(\beta_{\text{M}}, \beta^{*(E^a)}, \phi) - E\{U_i^{\dagger}(\beta_{\text{M}}, \beta^{*(E^a)}, \phi)\} \right] \cdot I(E^a = \widehat{E}) \right\|, \\
=&\sup_{\beta,\phi} \left\| \frac{1}{n}\sum_{i=1}^n U_i^{\dagger}(\beta_{\text{M}}, \beta^{*(\widehat{E})}, \phi) - E\{U_i^{\dagger}(\beta_{\text{M}}, \beta^{*(\widehat{E})}, \phi)\} \right\|.
\end{aligned} \tag{C.5}$$

Now we show the convergence of $\frac{1}{n}\sum_{i=1}^{n} U_i^{\dagger}(\beta_{\text{M}}, \beta^{*(\widehat{E})}, \phi)$ for a given estimated graph $\widehat{E}$. By Condition (R9) and the uniform weak law of large numbers (Newey and McFadden, 1994, Lemma 2.4), we have that

$$\sup_{\beta,\phi} \left\| \frac{1}{n}\sum_{i=1}^{n} U_i^{\dagger}(\beta_{\text{M}}, \beta^{*(\widehat{E})}, \phi) - E\{U_i^{\dagger}(\beta_{\text{M}}, \beta^{*(\widehat{E})}, \phi)\} \right\| \xrightarrow{p} 0,$$

and thus by (C.5),

$$\sup_{\beta,\phi} \left\| H_n^{\dagger}(\beta,\phi) - H^{\dagger}(\beta,\phi) \right\| \xrightarrow{p} 0 \qquad \text{as } n \to \infty. \tag{C.6}$$

2°. Next, for the second term on the right-hand side of (C.4), we have that

$$\begin{aligned}
&\lim_{n\to\infty} P\left( \sup_{\beta,\phi} \left\| H^{\dagger}(\beta,\phi) - H(\beta,\phi) \right\| > \epsilon \right) \\
&= \lim_{n\to\infty} P\left( \sup_{\beta,\phi} \left\| H^{\dagger}(\beta,\phi) - H(\beta,\phi) \right\| > \epsilon \middle| \widehat{E} = E \right) P(\widehat{E} = E) \\
&\qquad + \lim_{n\to\infty} P\left( \sup_{\beta,\phi} \left\| H^{\dagger}(\beta,\phi) - H(\beta,\phi) \right\| > \epsilon \middle| \widehat{E} \neq E \right) P(\widehat{E} \neq E) \\
&= \lim_{n\to\infty} P\left( \sup_{\beta,\phi} \left\| H^{\dagger}(\beta,\phi) - H(\beta,\phi) \right\| > \epsilon \middle| \widehat{E} = E \right) \\
&= \lim_{n\to\infty} P\left( \sup_{\beta,\phi} \left\| \sum_{E^a \in \mathcal{E}} \begin{pmatrix} E\{U_i^{\dagger}(\beta_{\text{M}}, \beta^{*(E^a)}, \phi)\} \\ \beta^{**(E^a)} \end{pmatrix} \cdot I(E^a = \widehat{E}) - H(\beta,\phi) \right\| > \epsilon \middle| \widehat{E} = E \right) \\
&= \lim_{n\to\infty} P\left( \sup_{\beta,\phi} \left\| \sum_{E^a \in \mathcal{E}} \begin{pmatrix} E\{U_i^{\dagger}(\beta_{\text{M}}, \beta^{*(E^a)}, \phi)\} \\ \beta^{**(E^a)} \end{pmatrix} \cdot I(E^a = E) - H(\beta,\phi) \right\| > \epsilon \right) \\
&= \lim_{n\to\infty} P\left( \sup_{\beta,\phi} \left\| \begin{pmatrix} E\{U_i^{\dagger}(\beta_{\text{M}}, \beta^{*(E)}, \phi)\} \\ \beta^{**(E)} \end{pmatrix} - H(\beta,\phi) \right\| > \epsilon \right) = 0,
\end{aligned}$$

where the second step is because of $\lim_{n\to\infty} P(\widehat{E} \neq E) = 0$, the third step is because of (C.3) and (C.2), and last step is by the definition $H(\beta,\phi)$.

Therefore,

$$\sup_{\beta,\phi} \left\| H^{\dagger}(\beta,\phi) - H(\beta,\phi) \right\| \xrightarrow{p} 0 \text{ as } n \to \infty. \tag{C.7}$$

Combining (C.4), (C.6) and (C.7) shows Claim 2.

**Proof of Theorem 4.2** (ii)

By (C.1),

$$P(\widehat{E} = E) \longrightarrow 0 \text{ as } n \to \infty,$$

which implies that as $n \to \infty$

$$P\left\{\sum_{i=1}^{n} U_i^{\dagger}(\beta_{\mathrm{M}}, \beta^{*(\widehat{E})}, \phi) = \sum_{i=1}^{n} U_i(\beta_{\mathrm{M}}, \beta_{\mathrm{I}}, \phi)\right\} \longrightarrow 0,$$

and

$$P\left\{\sum_{i=1}^{n}\begin{pmatrix}\frac{\partial U_i^{\dagger}(\beta_{\mathrm{M}},\beta^{*(\widehat{E})},\phi)}{\partial\beta_{\mathrm{M}}^{\mathrm{T}}} & \frac{\partial U_i(\beta_{\mathrm{M}},\beta^{*(\widehat{E})},\phi)}{\partial\beta_{\mathrm{I}}^{\mathrm{T}}} & \frac{\partial U_i(\beta_{\mathrm{M}},\beta^{*(\widehat{E})},\phi)}{\partial\phi^{\mathrm{T}}}\end{pmatrix} = \sum_{i=1}^{n}\begin{pmatrix}\frac{\partial U_i(\beta_{\mathrm{M}},\beta_{\mathrm{I}},\phi)}{\partial\beta_{\mathrm{M}}^{\mathrm{T}}} & \frac{\partial U_i(\beta_{\mathrm{M}},\beta_{\mathrm{I}},\phi)}{\partial\beta_{\mathrm{I}}^{\mathrm{T}}} & \frac{\partial U_i(\beta_{\mathrm{M}},\beta_{\mathrm{I}},\phi)}{\partial\phi^{\mathrm{T}}}\end{pmatrix}\right\} \longrightarrow 0,$$

because by definition of $U_i^{\dagger}(\beta_{\mathrm{M}}, \beta^{*(\widehat{E})}, \phi)$, we have that $U_i^{\dagger}(\beta_{\mathrm{M}}, \beta^{*(\widehat{E})}, \phi) = U_i(\beta_{\mathrm{M}}, \beta_{\mathrm{I}}, \phi)$ if $\widehat{E} = E$.

Hence, we have that

$$\sum_{i=1}^{n} U_i^{\dagger}(\beta_{\mathrm{M}}, \beta^{*(\widehat{E})}, \phi) = \sum_{i=1}^{n} U_i(\beta_{\mathrm{M}}, \beta_{\mathrm{I}}, \phi) + o_p(1) \tag{C.8}$$

and

$$\begin{aligned}&\sum_{i=1}^{n}\begin{pmatrix}\frac{\partial U_i^{\dagger}(\beta_{\mathrm{M}},\beta^{*(\widehat{E})},\phi)}{\partial\beta_{\mathrm{M}}^{\mathrm{T}}} & \frac{\partial U_i(\beta_{\mathrm{M}},\beta^{*(\widehat{E})},\phi)}{\partial\beta_{\mathrm{I}}^{\mathrm{T}}} & \frac{\partial U_i(\beta_{\mathrm{M}},\beta^{*(\widehat{E})},\phi)}{\partial\phi^{\mathrm{T}}}\end{pmatrix}\\ &=\sum_{i=1}^{n}\begin{pmatrix}\frac{\partial U_i(\beta_{\mathrm{M}},\beta_{\mathrm{I}},\phi)}{\partial\beta_{\mathrm{M}}^{\mathrm{T}}} & \frac{\partial U_i(\beta_{\mathrm{M}},\beta_{\mathrm{I}},\phi)}{\partial\beta_{\mathrm{I}}^{\mathrm{T}}} & \frac{\partial U_i(\beta_{\mathrm{M}},\beta_{\mathrm{I}},\phi)}{\partial\phi^{\mathrm{T}}}\end{pmatrix} + o_p(1).\end{aligned}$$

Applying the Taylor series expansion to $\sum_{i=1}^{n} U_i^{\dagger}(\widehat{\beta}_{\mathrm{M}}, \widehat{\beta}^{*(\widehat{E})}, \widehat{\phi}) = 0$ around $(\beta_{\mathrm{M}(0)}^{\mathrm{T}}, \beta_{\mathrm{I}(0)}^{\mathrm{T}}, \phi_0^{\mathrm{T}})^{\mathrm{T}}$, we obtain

$$\begin{aligned}&\sum_{i=1}^{n} U_i(\beta_{\mathrm{M}(0)}, \beta_{\mathrm{I}(0)}, \phi) + \sum_{i=1}^{n}\left.\begin{pmatrix}\frac{\partial U_i(\beta_{\mathrm{M}},\beta_{\mathrm{I}},\phi)}{\partial\beta_{\mathrm{M}}^{\mathrm{T}}} & \frac{\partial U_i(\beta_{\mathrm{M}},\beta_{\mathrm{I}},\phi)}{\partial\beta_{\mathrm{I}}^{\mathrm{T}}} & \frac{\partial U_i(\beta_{\mathrm{M}},\beta_{\mathrm{I}},\phi)}{\partial\phi^{\mathrm{T}}}\end{pmatrix}\right|_{\substack{\beta_{\mathrm{M}}=\beta_{\mathrm{M}(0)}\\ \beta_{\mathrm{I}}=\beta_{\mathrm{I}(0)}\\ \phi=\phi_0}}\\ &\qquad\times\left\{\begin{pmatrix}\widehat{\beta}_{\mathrm{M}}\\ \widehat{\beta}_{\mathrm{I}}\\ \widehat{\phi}\end{pmatrix} - \begin{pmatrix}\beta_{\mathrm{M}(0)}\\ \beta_{\mathrm{I}(0)}\\ \phi_0\end{pmatrix}\right\} + o_p(1) = 0,\end{aligned}$$

yielding

$$\left\{-\frac{1}{n}\sum_{i=1}^{n}\left(\frac{\partial U_i(\beta_{\rm M},\beta_{\rm I},\phi)}{\partial\beta_{\rm M}^{\rm T}} \quad \frac{\partial U_i(\beta_{\rm M},\beta_{\rm I},\phi)}{\partial\beta_{\rm I}^{\rm T}} \quad \frac{\partial U_i(\beta_{\rm M},\beta_{\rm I},\phi)}{\partial\phi^{\rm T}}\right)\right\}\Bigg|_{\substack{\beta_{\rm M}=\beta_{\rm M(0)}\\ \beta_{\rm I}=\beta_{\rm I(0)}\\ \phi=\phi_0}} \times \sqrt{n}\left\{\begin{pmatrix}\widehat{\beta}_{\rm M}\\ \widehat{\beta}_{\rm I}\\ \widehat{\phi}\end{pmatrix} - \begin{pmatrix}\beta_{\rm M(0)}\\ \beta_{\rm I(0)}\\ \phi_0\end{pmatrix}\right\}$$

$$=\frac{1}{\sqrt{n}}\sum_{i=1}^{n}U_i(\beta_{\rm M(0)},\beta_{\rm I(0)},\phi_0)+o_p(1). \tag{C.9}$$

Let

$$\Gamma_0=\lim_{n\to\infty}E_{(X,Y)}\left\{\frac{1}{n}\sum_{i=1}^{n}\left(\frac{\partial U_i(\beta_{\rm M},\beta_{\rm I},\phi)}{\partial\beta_{\rm M}^{\rm T}} \quad \frac{\partial U_i(\beta_{\rm M},\beta_{\rm I},\phi)}{\partial\beta_{\rm I}^{\rm T}} \quad \frac{\partial U_i(\beta_{\rm M},\beta_{\rm I},\phi)}{\partial\phi^{\rm T}}\right)\right\}\Bigg|_{\substack{\beta_{\rm M}=\beta_{\rm M(0)}\\ \beta_{\rm I}=\beta_{\rm I(0)}\\ \phi=\phi_0}}$$

and

$$\Sigma_0=\lim_{n\to\infty}{\rm Var}\left\{\frac{1}{\sqrt{n}}\sum_{i=1}^{n}U_i(\beta_{\rm M(0)},\beta_{\rm I(0)},\phi_0)\right\}.$$

Then applying the central limit theorem to (C.9) gives that

$$\sqrt{n}\left\{(\widehat{\beta}_{\rm M}^{\rm T},\widehat{\beta}_{\rm I}^{\rm T},\widehat{\phi}^{\rm T})^{\rm T}-(\beta_{\rm M(0)}^{\rm T},\beta_{\rm I(0)}^{\rm T},\phi_0^{\rm T})^{\rm T}\right\}\xrightarrow{d}N\left(0,\Gamma_0^{-1}\Sigma_0(\Gamma_0^{-1})^{\rm T}\right) \qquad \text{as} \qquad n\to\infty,$$

where the expectation and variance are taken with respect to the joint distribution of $Y_i$ and $X_i$ with the model parameters taken as their true value.

Now it remains to show that $\Gamma_0$ and $\Sigma_0$ are identical to those specified as in (4.9). By the assumption that the variables are independent and identically are distributed, it is immediate that

$$\Gamma_0=\left[E\left\{\frac{\partial U_i(\beta_{\rm M},\beta_{\rm I},\phi)}{\partial\beta_{\rm M}^{\rm T}}\right\} \quad E\left\{\frac{\partial U_i(\beta_{\rm M},\beta_{\rm I},\phi)}{\partial\beta_{\rm I}^{\rm T}}\right\} \quad E\left\{\frac{\partial U_i(\beta_{\rm M},\beta_{\rm I},\phi)}{\partial\phi^{\rm T}}\right\}\right]\Bigg|_{\substack{\beta_{\rm M}=\beta_{\rm M(0)}\\ \beta_{\rm I}=\beta_{\rm I(0)}\\ \phi=\phi_0}}$$

and

$$\begin{aligned}\Sigma_0&=\lim_{n\to\infty}{\rm Var}\left\{\frac{1}{\sqrt{n}}\sum_{i=1}^{n}U_i(\beta_{\rm M(0)},\beta_{\rm I(0)},\phi_0)\right\}\\ &=\lim_{n\to\infty}\frac{1}{n}\sum_{i=1}^{n}{\rm Var}\left\{U_i(\beta_{\rm M(0)},\beta_{\rm I(0)},\phi_0)\right\}\\ &=E\left\{U_i^{\rm T}(\beta_{\rm M(0)},\beta_{\rm I(0)},\phi_0)U_i(\beta_{\rm M(0)},\beta_{\rm I(0)},\phi_0)\right\}-E^{\rm T}\left\{U_i(\beta_{\rm M(0)},\beta_{\rm I(0)},\phi_0)\right\}E\left\{U_i(\beta_{\rm M(0)},\beta_{\rm I(0)},\phi_0\right\}\\ &=E\left\{U_i^{\rm T}(\beta_{\rm M(0)},\beta_{\rm I(0)},\phi_0)U_i(\beta_{\rm M(0)},\beta_{\rm I(0)},\phi_0)\right\}.\end{aligned}$$

## C.4 Proof of Theorem 4.4

To reflect the randomness introduced by the selection of the edge set, we use $U_i^{*\dagger}(\beta_{\mathrm{M}}, \beta^{*(\widehat{E})}, \phi, \eta)$ to denote the estimating function $U_i^*(\beta_{\mathrm{M}}, \beta_{\mathrm{I}}, \phi, \eta)$ in (4.14) with the edge set $E$, estimate by $\widehat{E}$, where $\beta^{*(\widehat{E})} = \{\beta_{st} : (s,t) \in \widehat{E}\}$ represents the subvector of $(\beta_{\mathrm{I}}^{\mathrm{T}}, \beta_{\mathrm{II}}^{\mathrm{T}})^{\mathrm{T}}$ with the indexes included in $\widehat{E}$.

Assume the external validation sample is randomly formed. To simplify the notation, let $\beta = (\beta_{\mathrm{M}}^{\mathrm{T}}, \beta_{\mathrm{I}}^{\mathrm{T}}, \phi^{\mathrm{T}})^{\mathrm{T}}$ and $\beta^\dagger = (\beta_{\mathrm{M}}^{\mathrm{T}}, \beta^{*(\widehat{E})}, \phi^{\mathrm{T}})^{\mathrm{T}}$. Let $\delta_i = I(i \in \mathcal{V})$, where $I(\cdot)$ is the indicator function. Define $F_{i1}(\beta, \eta) = (1-\delta_i)U_i^*(\beta, \eta)$, $F_{i2}(\eta) = \delta_i S_i(\eta)$, and $F_i(\beta, \eta) = \{F_{i1}^{\mathrm{T}}(\beta, \eta), F_{i2}^{\mathrm{T}}(\eta)\}^{\mathrm{T}}$, and define $F_{i1}^\dagger(\beta^\dagger, \eta) = (1-\delta_i)U_i^{*\dagger}(\beta^\dagger, \eta)$, and $F_i^\dagger(\beta^\dagger, \eta) = \{F_{i1}^{\dagger\mathrm{T}}(\beta^\dagger, \eta), F_{i2}^{\mathrm{T}}(\eta)\}^{\mathrm{T}}$. Then, (4.14) is equivalent to solving

$$U^{(\mathrm{EV})}(\beta, \eta) = \sum_{i \in \mathcal{M} \bigcup \mathcal{V}} F_i^\dagger(\beta^\dagger, \eta).$$

Similar to the proof of Theorem 4.2(i), with $U_i(\beta_{\mathrm{M}}, \beta_{\mathrm{I}}, \phi)$ replaced by $F_i(\beta, \eta)$ and $U_i^\dagger(\beta_{\mathrm{M}}, \beta^{*(\widehat{E})}, \phi)$ replaced by $F_i^\dagger(\beta^\dagger, \eta)$, we can show that solving $\sum_{i \in \mathcal{M} \bigcup \mathcal{V}} F_i^\dagger(\beta^\dagger, \eta) = 0$ gives a consistent estimator, $(\widehat{\beta}^{(\mathrm{EV})\mathrm{T}}, \widehat{\eta}^{(\mathrm{EV})\mathrm{T}})^{\mathrm{T}}$, of $(\beta^{\mathrm{T}}, \eta^{\mathrm{T}})^{\mathrm{T}}$.

Similar to the derivation of (C.8) in the proof of Theorem 4.2(ii), we have that

$$\sum_{i=1}^n U_i^{*\dagger}(\beta_{\mathrm{M}}, \beta^{*(E)}, \phi, \eta) = \sum_{i=1}^n U_i^*(\beta_{\mathrm{M}}, \beta_{\mathrm{I}}, \phi, \eta) + o_p(1)$$

and hence

$$\sum_{i \in \mathcal{M} \bigcup \mathcal{V}} F_i^\dagger(\beta^\dagger, \eta) = \sum_{i \in \mathcal{M} \bigcup \mathcal{V}} F_i(\beta, \eta) + o_p(1).$$

Applying the Taylor series expansion to $\sum_{i \in \mathcal{M} \bigcup \mathcal{V}} F_i^\dagger(\widehat{\beta}^{(\mathrm{EV})}, \widehat{\eta}^{(\mathrm{EV})}) = 0$, we obtain

$$\sum_{i \in \mathcal{M} \bigcup \mathcal{V}} F_i(\beta_0, \eta_0) + \sum_{i \in \mathcal{M} \bigcup \mathcal{V}} \begin{pmatrix} \frac{\partial F_i(\beta,\eta)}{\partial \beta^{\mathrm{T}}} & \frac{\partial F_i(\beta,\eta)}{\partial \eta^{\mathrm{T}}} \end{pmatrix} \Bigg|_{\substack{\beta=\beta_0 \\ \eta=\eta_0}} \times \left\{ \begin{pmatrix} \widehat{\beta} \\ \widehat{\eta} \end{pmatrix} - \begin{pmatrix} \beta_0 \\ \eta_0 \end{pmatrix} \right\} + o_p(1) = 0,$$

which leads to

$$\begin{aligned} &\sqrt{1 + \frac{m}{n}} \left\{ \frac{1}{n+m} \sum_{i \in \mathcal{M} \bigcup \mathcal{V}} \begin{pmatrix} \frac{\partial F_i(\beta,\eta)}{\partial \beta^{\mathrm{T}}} & \frac{\partial F_i(\beta,\eta)}{\partial \eta^{\mathrm{T}}} \end{pmatrix} \right\} \Bigg|_{\substack{\beta=\beta_0 \\ \eta=\eta_0}} \times \sqrt{n} \left\{ \begin{pmatrix} \widehat{\beta} \\ \widehat{\eta} \end{pmatrix} - \begin{pmatrix} \beta_0 \\ \eta_0 \end{pmatrix} \right\} \\ =& \frac{1}{\sqrt{n+m}} \sum_{i \in \mathcal{M} \bigcup \mathcal{V}} F_i(\beta_0, \eta_0) + o_p(1). \end{aligned} \tag{C.10}$$

Let

$$\Gamma_{\text{(EV)}} = \lim_{n\to\infty}\left\{\frac{1}{n+m}\sum_{i\in\mathcal{M}\bigcup\mathcal{V}} E\begin{pmatrix}\frac{\partial F_i(\beta,\eta)}{\partial\beta^{\mathrm{T}}} & \frac{\partial F_i(\beta,\eta)}{\partial\eta^{\mathrm{T}}}\end{pmatrix}\right\}\Bigg|_{\substack{\beta=\beta_0\\ \eta=\eta_0}}$$

and

$$\Sigma_{\text{(EV)}} = \lim_{n\to\infty}\operatorname{Var}\left(\frac{1}{\sqrt{n+m}}\sum_{i\in\mathcal{M}\bigcup\mathcal{V}} F_i(\beta_0,\eta_0)\right).$$

Then applying the central limit theorem to (C.10) gives that

$$\sqrt{n}\left\{(\widehat{\beta}^{\text{(EV)}\mathrm{T}}, \widehat{\eta}^{\text{(EV)}\mathrm{T}})^{\mathrm{T}} - (\beta_0^{\mathrm{T}}, \eta_0^{\mathrm{T}})^{\mathrm{T}}\right\} \xrightarrow{d} N\left(0, \frac{1}{1+\rho}\Gamma_{\text{(EV)}}^{-1}\Sigma_{\text{(EV)}}(\Gamma_{\text{(EV)}}^{-1})^{\mathrm{T}}\right) \qquad \text{as} \qquad n\to\infty.$$

Now it remains to show that $\Gamma_{\text{(EV)}}$ and $\Sigma_{\text{(EV)}}$ are identical to (4.15). By the definitions of $F_i(\beta,\eta)$, we derive that

$$\begin{aligned}
\Gamma_{\text{(EV)}} &= \lim_{n\to\infty}\left\{\frac{1}{n+m}E\begin{pmatrix}\sum_{i\in\mathcal{M}}\frac{\partial U_i^*(\beta,\eta)}{\partial\beta^{\mathrm{T}}} & \sum_{i\in\mathcal{M}}\frac{\partial U_i^*(\beta,\eta)}{\partial\eta^{\mathrm{T}}}\\ 0 & 0\end{pmatrix} + \frac{1}{n+m}E\begin{pmatrix}0 & 0\\ 0 & \sum_{i\in\mathcal{V}}\frac{\partial S_i(\eta)}{\partial\eta^{\mathrm{T}}}\end{pmatrix}\right\}\\
&= \lim_{n\to\infty}\left\{\frac{n}{n+m}E\begin{pmatrix}\frac{\partial U_i^*(\beta,\eta)}{\partial\beta^{\mathrm{T}}} & \frac{\partial U_i^*(\beta,\eta)}{\partial\eta^{\mathrm{T}}}\\ 0 & 0\end{pmatrix} + \frac{m}{n+m}E\begin{pmatrix}0 & 0\\ 0 & \frac{\partial S_i(\eta)}{\partial\eta^{\mathrm{T}}}\end{pmatrix}\right\}\\
&= \frac{1}{1+\rho}\begin{pmatrix}E\left(\frac{\partial U_i^*(\beta,\eta)}{\partial\beta^{\mathrm{T}}}\right) & E\left(\frac{\partial U_i^*(\beta,\eta)}{\partial\eta^{\mathrm{T}}}\right)\\ 0 & 0\end{pmatrix} + \frac{\rho}{1+\rho}\begin{pmatrix}0 & 0\\ 0 & E\left(\frac{\partial S_i(\eta)}{\partial\eta^{\mathrm{T}}}\right)\end{pmatrix}.
\end{aligned}$$

For $\Sigma_{\text{(EV)}}$, we have that

$$
\begin{aligned}
\Sigma_{\text{(EV)}} &= \lim_{n\to\infty} \frac{1}{n+m} \sum_{i\in\mathcal{M}\bigcup\mathcal{V}} \text{Var}\{F_i(\beta,\eta)\} \\
&= \lim_{n\to\infty} \frac{1}{n+m} \sum_{i\in\mathcal{M}\bigcup\mathcal{V}} E\{F_i(\beta,\eta)F_i^{\mathrm{T}}(\beta,\eta)\} \\
&= \lim_{n\to\infty} \frac{1}{n+m} \sum_{i\in\mathcal{M}\bigcup\mathcal{V}} E\begin{pmatrix} (1-\delta_i)^2 U_i^*(\beta,\eta)U_i^{*\mathrm{T}}(\beta,\eta) & 0 \\ 0 & \delta_i^2 S_i(\eta)S_i^{\mathrm{T}}(\eta) \end{pmatrix} \\
&= \lim_{n\to\infty} \frac{n}{n+m}\frac{1}{n} \sum_{i\in\mathcal{M}} \begin{pmatrix} E\{U_i^*(\beta,\eta)U_i^{*\mathrm{T}}(\beta,\eta)\} & 0 \\ 0 & 0 \end{pmatrix} \\
&\quad + \lim_{n\to\infty} \frac{m}{n+m}\frac{1}{m} \sum_{i\in\mathcal{V}} \begin{pmatrix} 0 & 0 \\ 0 & E\{S_i(\eta)S_i^{\mathrm{T}}(\eta)\} \end{pmatrix} \\
&= \frac{1}{1+\rho} \begin{pmatrix} E\{U_i^*(\beta,\eta)U_i^{*\mathrm{T}}(\beta,\eta)\} & 0 \\ 0 & 0 \end{pmatrix} + \frac{\rho}{1+\rho} \begin{pmatrix} 0 & 0 \\ 0 & E\{S_i(\eta)S_i^{\mathrm{T}}(\eta)\} \end{pmatrix},
\end{aligned}
$$

where the fourth step is due to $(1-\delta_i)^2 = (1-\delta_i)$ and $\delta_i^2 = \delta_i$.

## C.5 Proof of Theorem 4.5

Similar to Section C.4, we assume the internal validation subsample is randomly formed. We define $F_{i1}(\beta,\eta) = (1-\delta_i)U_{i1}^*(\beta,\eta) + \delta_i U_{i1}(\beta,\eta)$, $F_{i2}(\beta,\eta) = \delta_i S_i(\eta)$, and $F_i(\beta,\eta) = \{F_{i1}^{\mathrm{T}}(\beta,\eta), F_{i2}^{\mathrm{T}}(\eta)\}^{\mathrm{T}}$, and define $F_{i1}^{\dagger}(\beta^{\dagger},\eta) = (1-\delta_i)U_{i1}^*(\beta^{\dagger},\eta) + \delta_i U_{i1}(\beta^{\dagger},\eta)$, and $F_i^{\dagger}(\beta^{\dagger},\eta) = \{F_{i1}^{\dagger\mathrm{T}}(\beta^{\dagger},\eta), F_{i2}^{\mathrm{T}}(\eta)\}^{\mathrm{T}}$. Then, (4.17) is equivalent to

$$
U^{\text{(IV)}}(\beta,\eta) = \sum_{i\in\mathcal{M}} F_i^{\dagger}(\beta^{\dagger},\eta).
$$

Similar to the proof of Theorem 4.2(i), with $U_i(\beta_{\text{M}},\beta_{\text{I}},\phi)$ replaced by $F_i(\beta,\eta)$ and $U_i^{\dagger}(\beta_{\text{M}},\beta^{*(\widehat{E})},\phi)$ replaced by $F_i^{\dagger}(\beta^{\dagger},\eta)$, we conclude that under regularity conditions, solving $\sum_{i\in\mathcal{M}} F_i^{\dagger}(\beta^{\dagger},\eta) = 0$ gives a consistent estimator, $(\widehat{\beta}^{\text{(IV)}\mathrm{T}}, \widehat{\eta}^{\text{(IV)}\mathrm{T}})^{\mathrm{T}}$, of $(\beta_0,\eta_0)^{\mathrm{T}}$.

Applying the Taylor series expansion to $\sum_{i\in\mathcal{M}} F_i(\widehat{\beta}^{\text{(IV)}\mathrm{T}}, \widehat{\eta}^{\text{(IV)}\mathrm{T}}) = 0$, we obtain

$$
\sum_{i\in\mathcal{M}} F_i(\beta_0,\eta_0) + \sum_{i\in\mathcal{M}} \begin{pmatrix} \frac{\partial F_i(\beta,\eta)}{\partial\beta^{\mathrm{T}}} & \frac{\partial F_i(\beta,\eta)}{\partial\eta^{\mathrm{T}}} \end{pmatrix}\Bigg|_{\substack{\beta=\beta_0\\ \eta=\eta_0}} \left\{ \begin{pmatrix} \widehat{\beta} \\ \widehat{\eta} \end{pmatrix} - \begin{pmatrix} \beta_0 \\ \eta_0 \end{pmatrix} \right\} + o_p(1) = 0,
$$

which leads to

$$\left\{-\frac{1}{n}\sum_{i\in\mathcal{M}}\begin{pmatrix}\frac{\partial F_i(\beta,\eta)}{\partial\beta^{\mathrm{T}}} & \frac{\partial F_i(\beta,\eta)}{\partial\eta^{\mathrm{T}}}\end{pmatrix}\right\}\Bigg|_{\substack{\beta=\beta_0\\ \eta=\eta_0}} \times\sqrt{n}\left\{\begin{pmatrix}\widehat{\beta}\\ \widehat{\eta}\end{pmatrix}-\begin{pmatrix}\beta_0\\ \eta_0\end{pmatrix}\right\}=\frac{1}{\sqrt{n}}\sum_{i\in\mathcal{M}}F_i(\beta_0,\eta_0)+o_p(1). \tag{C.11}$$

Let

$$\Gamma_{\text{(IV)}}=\lim_{n\to\infty}\left\{\frac{1}{n}\sum_{i\in\mathcal{M}}E\begin{pmatrix}\frac{\partial F_i(\beta,\eta)}{\partial\beta^{\mathrm{T}}} & \frac{\partial F_i(\beta,\eta)}{\partial\eta^{\mathrm{T}}}\end{pmatrix}\right\}\Bigg|_{\substack{\beta=\beta_0\\ \eta=\eta_0}}$$

and

$$\Sigma_{\text{(IV)}}=\lim_{n\to\infty}\mathrm{Var}\left(\frac{1}{\sqrt{n}}\sum_{i\in\mathcal{M}}F_i(\beta_0,\eta_0)\right).$$

Then applying the central limit theorem to (C.11) gives that

$$\sqrt{n}\left\{(\widehat{\beta}^{\text{(IV)}\mathrm{T}},\widehat{\eta}^{\text{(IV)}\mathrm{T}})^{\mathrm{T}}-(\beta_0^{\mathrm{T}},\eta_0^{\mathrm{T}})^{\mathrm{T}}\right\}\xrightarrow{d}N\left(0,\Gamma_{\text{(IV)}}^{-1}\Sigma_{\text{(IV)}}(\Gamma_{\text{(IV)}}^{-1})^{\mathrm{T}}\right)\qquad\text{as}\qquad n\to\infty.$$

Now it remains to show that $\Gamma_{\text{(IV)}}$ and $\Sigma_{\text{(IV)}}$ are identical to (4.19). By definition of $F_i(\beta,\eta)$, $\Gamma_{\text{(IV)}}$ equals

$$\begin{aligned}
&\lim_{n\to\infty}\left\{\frac{1}{n}E\begin{pmatrix}\sum_{i\in\mathcal{M}\backslash\mathcal{V}}\frac{\partial U_i^*(\beta,\eta)}{\partial\beta^{\mathrm{T}}} & \sum_{i\in\mathcal{M}\backslash\mathcal{V}}\frac{\partial U_i^*(\beta,\eta)}{\partial\eta^{\mathrm{T}}}\\ 0 & 0\end{pmatrix}+\frac{1}{n}E\begin{pmatrix}\sum_{i\in\mathcal{V}}\frac{\partial U_i(\beta,\eta)}{\partial\beta^{\mathrm{T}}} & 0\\ 0 & \sum_{i\in\mathcal{V}}\frac{\partial S_i(\eta)}{\partial\eta^{\mathrm{T}}}\end{pmatrix}\right\}\\
=&\lim_{n\to\infty}\left\{\frac{n-m}{n}E\begin{pmatrix}\frac{\partial U_i^*(\beta,\eta)}{\partial\beta^{\mathrm{T}}} & \frac{\partial U_i^*(\beta,\eta)}{\partial\eta^{\mathrm{T}}}\\ 0 & 0\end{pmatrix}+\frac{m}{n}E\begin{pmatrix}\frac{\partial U_i(\beta,\eta)}{\partial\beta^{\mathrm{T}}} & 0\\ 0 & \frac{\partial S_i(\eta)}{\partial\eta^{\mathrm{T}}}\end{pmatrix}\right\}\\
=&(1-\rho)\begin{pmatrix}E\left(\frac{\partial U_i^*(\beta,\eta)}{\partial\beta^{\mathrm{T}}}\right) & E\left(\frac{\partial U_i^*(\beta,\eta)}{\partial\eta^{\mathrm{T}}}\right)\\ 0 & 0\end{pmatrix}+\rho\begin{pmatrix}E\left(\frac{\partial U_i(\beta,\eta)}{\partial\beta^{\mathrm{T}}}\right) & 0\\ 0 & E\left(\frac{\partial S_i(\eta)}{\partial\eta^{\mathrm{T}}}\right)\end{pmatrix}.
\end{aligned}$$

Similarly, we have that

$$
\begin{aligned}
\Sigma_{\text{(IV)}} &= \lim_{n\to\infty} \frac{1}{n} \sum_{i\in\mathcal{M}} \operatorname{Var}\{F_i(\beta,\eta)\} \\
&= \lim_{n\to\infty} \frac{1}{n} \sum_{i\in\mathcal{M}} E\{F_i(\beta,\eta)F_i^{\mathrm{T}}(\beta,\eta)\} \\
&= \lim_{n\to\infty} \frac{1}{n} \sum_{i\in\mathcal{M}} E \begin{pmatrix} (1-\delta_i)^2 U_i^*(\beta,\eta)U_i^{*\mathrm{T}}(\beta,\eta) & 0 \\ 0 & 0 \end{pmatrix} \\
&\quad + \lim_{n\to\infty} \frac{1}{n} \sum_{i\in\mathcal{M}} E \begin{pmatrix} \delta_i^2 U_i(\beta,\eta)U_i^{\mathrm{T}}(\beta,\eta) & \delta_i^2 U_i(\beta,\eta)S_i^{\mathrm{T}}(\eta) \\ \delta_i^2 U_i(\beta,\eta)S_i^{\mathrm{T}}(\eta) & \delta_i^2 S_i(\eta)S_i^{\mathrm{T}}(\eta) \end{pmatrix} \\
&= \lim_{n\to\infty} \frac{n-m}{n} \frac{1}{n-m} \sum_{i\in\mathcal{M}\backslash\mathcal{V}} \begin{pmatrix} E\{U_i^*(\beta,\eta)U_i^{*\mathrm{T}}(\beta,\eta)\} & 0 \\ 0 & 0 \end{pmatrix} \\
&\quad + \lim_{n\to\infty} \frac{m}{n} \frac{1}{m} \sum_{i\in\mathcal{V}} \begin{pmatrix} E\{U_i(\beta,\eta)U_i^{\mathrm{T}}(\beta,\eta)\} & E\{U_i(\beta,\eta)S_i^{\mathrm{T}}(\beta,\eta)\} \\ E\{U_i(\beta,\eta)S_i^{\mathrm{T}}(\beta,\eta)\} & E\{S_i(\beta,\eta)S_i^{\mathrm{T}}(\beta,\eta)\} \end{pmatrix} \\
&= (1-\rho) \begin{pmatrix} E\{U_i^*(\beta,\eta)U_i^{*\mathrm{T}}(\beta,\eta)\} & 0 \\ 0 & 0 \end{pmatrix} \\
&\quad + \rho \begin{pmatrix} E\{U_i(\beta,\eta)U_i^{\mathrm{T}}(\beta,\eta)\} & E\{U_i(\beta,\eta)S_i^{\mathrm{T}}(\eta)\} \\ E\{U_i(\beta,\eta)S_i^{\mathrm{T}}(\eta)\} & E\{S_i(\eta)S_i^{\mathrm{T}}(\eta)\} \end{pmatrix},
\end{aligned}
$$

where the third step is due to $\delta_i(1-\delta_i) = 0$ and the fourth step comes from $(1-\delta_i)^2 = (1-\delta_i)$ and $\delta_i^2 = \delta_i$.

## C.6 Proof of Theorem 4.6

By (4.12) and (4.20), showing (4.23) is equivalent to showing

$$
\{(1-\rho)\Gamma + \rho\Gamma_0\}^{-1}\{(1-\rho)\Sigma + \rho\Sigma_0 - \rho\Delta\}\{(1-\rho)\Gamma + \rho\Gamma_0\}^{-1\mathrm{T}} \leq \Gamma^{-1}\Sigma\Gamma^{-1\mathrm{T}}. \quad \text{(C.12)}
$$

Left multiplying $(1-\rho)\Gamma + \rho\Gamma_0$ and right multiplying $\{(1-\rho)\Gamma + \rho\Gamma_0\}^{\mathrm{T}}$ on both sides of (C.12), we obtain

$$
(1-\rho)\Sigma + \rho\Sigma_0 - \rho\Delta \leq (1-\rho)^2\Sigma + \rho(1-\rho)\Gamma_0\Gamma^{-1}\Sigma + (1-\rho)\rho\Sigma\Gamma^{-1\mathrm{T}}\Gamma_0^{\mathrm{T}} + \rho^2\Gamma_0\Gamma^{-1}\Sigma\Gamma^{-1\mathrm{T}}\Gamma_0^{\mathrm{T}}. \quad \text{(C.13)}
$$

Left multiplying $\Gamma_0^{-1}$ and right multiplying $\Gamma_0^{-1\mathrm{T}}$ on the both sides of (C.13) gives

$$\begin{aligned}&(1-\rho)\Gamma_0^{-1}\Sigma\Gamma_0^{-1\mathrm{T}}+\rho\Gamma_0^{-1}\Sigma_0\Gamma_0^{-1\mathrm{T}}-\rho\Gamma_0^{-1}\Delta\Gamma_0^{-1\mathrm{T}}\\ \leq&(1-\rho)^2\Gamma_0^{-1}\Sigma\Gamma_0^{-1\mathrm{T}}+\rho(1-\rho)\Gamma^{-1}\Sigma\Gamma_0^{-1\mathrm{T}}+(1-\rho)\rho\Gamma_0^{-1}\Sigma\Gamma^{-1\mathrm{T}}+\rho^2\Gamma^{-1}\Sigma\Gamma^{-1\mathrm{T}}.\end{aligned}$$

Then combining the terms with $\Gamma_0^{-1}\Sigma\Gamma_0^{-1\mathrm{T}}$ and dividing the both sides by $\rho$, we get

$$\begin{aligned}(1-\rho)\Gamma_0^{-1}\Sigma\Gamma_0^{-1\mathrm{T}}+\Gamma_0^{-1}\Sigma_0\Gamma_0^{-1\mathrm{T}}-\Gamma_0^{-1}\Delta\Gamma_0^{-1\mathrm{T}}\leq(1-\rho)\Gamma^{-1}\Sigma\Gamma_0^{-1\mathrm{T}}+(1-\rho)\Gamma_0^{-1}\Sigma\Gamma^{-1\mathrm{T}}\\ +\rho\Gamma^{-1}\Sigma\Gamma^{-1\mathrm{T}}.\end{aligned}\tag{C.14}$$

It now suffices to show that (C.14) is true when the conditions (4.21) and (4.22) are satisfied. For the case with $\rho=1$, the inequality (C.14) is equivalent to

$$\Gamma_0^{-1}\Sigma_0\Gamma_0^{-1\mathrm{T}}-\Gamma_0^{-1}\Delta\Gamma_0^{-1\mathrm{T}}\leq\Gamma^{-1}\Sigma\Gamma^{-1\mathrm{T}},$$

which is true by the condition (4.21) together with the fact that $\Gamma_0^{-1}\Delta\Gamma_0^{-1\mathrm{T}}$ is non-negative definite.

For the case with $\rho<1$, dividing $(1-\rho)$ on both sides, (C.14) is equivalent to

$$\Gamma_0^{-1}\Sigma\Gamma_0^{-1\mathrm{T}}+\frac{1}{1-\rho}\Gamma_0^{-1}\Sigma_0\Gamma_0^{-1\mathrm{T}}-\frac{1}{1-\rho}\Gamma_0^{-1}\Delta\Gamma_0^{-1\mathrm{T}}\leq\Gamma^{-1}\Sigma\Gamma_0^{-1\mathrm{T}}+\Gamma_0^{-1}\Sigma\Gamma^{-1\mathrm{T}}+\frac{\rho}{1-\rho}\Gamma^{-1}\Sigma\Gamma^{-1\mathrm{T}},\tag{C.15}$$

which is true, because

$$\begin{aligned}&\text{the left hand side of (C.15)}\\ \leq&\Gamma_0^{-1}\Sigma\Gamma_0^{-1\mathrm{T}}+\frac{1}{1-\rho}\Gamma_0^{-1}\Sigma_0\Gamma_0^{-1\mathrm{T}}-\Gamma_0^{-1}\Delta\Gamma_0^{-1\mathrm{T}}\\ =&\Gamma_0^{-1}\Sigma\Gamma_0^{-1\mathrm{T}}+\Gamma_0^{-1}\Sigma_0\Gamma_0^{-1\mathrm{T}}+\frac{\rho}{1-\rho}\Gamma_0^{-1}\Sigma_0\Gamma_0^{-1\mathrm{T}}-\Gamma_0^{-1}\Delta\Gamma_0^{-1\mathrm{T}}\\ \leq&\Gamma_0^{-1}\Sigma\Gamma_0^{-1\mathrm{T}}+\Gamma_0^{-1}\Sigma_0\Gamma_0^{-1\mathrm{T}}+\frac{\rho}{1-\rho}\Gamma^{-1}\Sigma\Gamma^{-1\mathrm{T}}-\Gamma_0^{-1}\Delta\Gamma_0^{-1\mathrm{T}}\\ \leq&\Gamma^{-1}\Sigma\Gamma_0^{-1\mathrm{T}}+\Gamma_0^{-1}\Sigma\Gamma^{-1\mathrm{T}}+\frac{\rho}{1-\rho}\Gamma^{-1}\Sigma\Gamma^{-1\mathrm{T}}\\ =&\text{ the right hand side of (C.15)},\end{aligned}$$

where the first step is because $\rho\in(0,1)$, the third step is due to the condition (4.21) that $\Gamma_0^{-1}\Sigma_0\Gamma_0^{-1\mathrm{T}}\leq\Gamma^{-1}\Sigma\Gamma^{-1\mathrm{T}}$, and the fourth step is due to the condition (4.22).

# Appendix D

# Proofs of the Results in Chapter 5

## D.1 Proof of Theorem 5.1

**Proof of Theorem 5.1(a):**

First, consider

$$\begin{aligned}
P(Y_i^* = 0|X_i) &= \sum_{k=0}^{\infty} P(Y_i^* = 0|Y_i = k, X_i)P(Y_i = k|X_i) \\
&= \sum_{k=0}^{\infty} P(Z_{i-} = k|Y_i = k, X_i)P(Y_i = k|X_i) \\
&= P(Y_i = 0|X_i)P(Z_{i-} = 0|Y_i = 0, X_i) \\
&\quad + \sum_{k=1}^{\infty} P(Z_{i-} = k|Y_i = k, X_i)P(Y_i = k|X_i) \\
&= P(Y_i = 0|X_i) + \sum_{k=1}^{\infty} \binom{k}{k} \pi_i^k (1-\pi_i)^0 \times \frac{\mu_i^k}{k!} e^{-\mu_i} \\
&= P(Y_i = 0|X_i) + \phi_i \left( \sum_{k=1}^{\infty} \frac{(\pi_i \mu_i)^k}{k!} e^{-\mu_i \pi_i} \right) e^{-\mu_i + \mu_i \pi_i} \\
&= P(Y_i = 0|X_i) + \phi_i (1 - e^{-\mu_i \pi_i}) e^{-\mu_i(1-\pi_i)} \\
&= 1 - \phi_i + \phi_i e^{-\mu_i(1-\pi_i)},
\end{aligned} \tag{D.1}$$

where the second step comes from (5.5) with $c_+ = 0$ and $c_- = 1$, the fourth step comes from (5.2) and the distributional assumption of $Z_{i-}$, the fifth step is because $\sum_{k=0}^{\infty} \frac{(\pi_i \mu_i)^k}{k!} e^{-\mu_i \pi_i} = 1$, and the last step is due to (5.2).

Next, for $Y_i^* = k \geq 1$, we calculate

$$\begin{aligned}
P(Y_i^* = k|X_i) &= \sum_{r=0}^{\infty} P(Z_{i-} = r|Y = k+r, X_i) P(Y = k+r|X_i) \\
&= \phi_i \sum_{r=0}^{\infty} \binom{k+r}{r} \pi_i^r (1-\pi_i)^k \times \frac{\mu_i^{k+r}}{(k+r)!} e^{-\mu_i} \\
&= \phi_i \sum_{r=0}^{\infty} \frac{1}{r!k!} (1-\pi_i)^k \mu_i^k (\mu_i \pi_i)^r e^{-\mu_i} \\
&= \phi_i \frac{\{\mu_i(1-\pi_i)\}^k}{k!} e^{-\mu_i(1-\pi_i)},
\end{aligned} \tag{D.2}$$

where the first step is due to (5.5) with $c_+ = 0$ and $c_- = 1$ and the second step is because of (5.2) and the distributional assumption of $Z_{i-}$.

Therefore, comparing (D.1) and (D.2) to (5.2), we conclude that conditional on $X_i$, $Y_i^*$ follows a zero-inflated Poisson distribution with mean $\mu_i(1-\pi_i)$ and the probability $\phi_i$.

**Proof of Theorem 5.1(b)**:

First, we consider the case with $Y_i^* = 0$. Under (5.5) with $c_+ = 1$ and $c_- = 0$, we note that $Y_i^* = 0$ if and only if $Y_i = 0$ and $Z_{i+} = 0$. Then,

$$\begin{aligned}
P(Y_i^* = 0|X_i) &= P(Y_i = 0, Z_{i+} = 0|X_i) \\
&= (1 - \phi_i + \phi_i e^{-\mu_i}) e^{-\lambda_i} \\
&= (1-\phi_i) e^{-\lambda_i} + \phi_i e^{-(\mu_i + \lambda_i)},
\end{aligned}$$

where the second step is due to the conditional independence assumption between $Y_i$ and $Z_{i+}$ given $X_i$, as well as (5.2) and the distributional assumption of $Z_{i+}$.

Next, for $k \geq 1$, we have that

$$\begin{aligned}
f(Y_i^* = k|X_i) &= P(Y_i = 0, Z_{i+} = k|X_i) + \sum_{t=1}^{k} P(Y_i = t, Z_{i+} = k-t|X_i) \\
&= (1 - \phi_i + \phi_i e^{-\mu_i}) \frac{\lambda_i^k e^{-\lambda_i}}{k!} + \sum_{t=1}^{k} \phi_i \frac{\mu_i^t e^{-\mu_i}}{t!} \times \frac{\lambda_i^{k-t}}{(k-t)!} e^{-\lambda_i} \\
&= (1 - \phi_i) \frac{\lambda_i^k e^{-\lambda_i}}{k!} + \sum_{t=0}^{k} \phi_i \frac{\mu_i^t e^{-\mu_i}}{t!} \times \frac{\lambda^{k-t}}{(k-t)!} e^{-\lambda_i} \\
&= (1 - \phi_i) \frac{\lambda_i^k e^{-\lambda_i}}{k!} + \phi_i e^{-\mu_i} e^{-\lambda_i} \sum_{t=0}^{k} \frac{\mu_i^t \lambda^{k-t}}{t!(k-t)!} \\
&= (1 - \phi_i) \frac{\lambda_i^k e^{-\lambda_i}}{k!} + \phi_i \frac{(\mu_i + \lambda_i)^k}{k!} e^{-(\mu_i + \lambda_i)} \sum_{t=0}^{k} \frac{k!}{t!(k-t)!} \left( \frac{\mu_i}{\mu_i + \lambda_i} \right)^t \left( \frac{\lambda_i}{\mu_i + \lambda_i} \right)^{k-t} \\
&= (1 - \phi_i) \frac{\lambda_i^k e^{-\lambda_i}}{k!} + \phi_i \frac{(\mu_i + \lambda_i)^k}{k!} e^{-(\mu_i + \lambda_i)},
\end{aligned}$$

where the second step is due to the conditional independence assumption between $Y_i$ and $Z_{i+}$ given $X_i$, as well as (5.2) and the distributional assumption of $Z_{i+}$, and the last step is due to the Binomial theorem. Thus, the conclusion follows.

**Proof of Theorem 5.1(c):**

Model (5.5) with $c_+ = c_- = 1$ can viewed as $Y_i^* = (Y_i - Z_{i-}) + Z_{i+}$, where by Theorem 5.1(a), the first term $(Y_i - Z_{i-})$ follows a zero-inflated Poisson distribution with parameters $\phi_i$ and $\mu_i^* = (1 - \pi_i)\mu_i$. Then applying Theorem 5.1(b) to $Y_i^*$, we conclude that

$$P(Y_i^* = y_i^*|X_i) = (1 - \phi_i) \frac{\lambda_i^{y_i^*} e^{-\lambda_i}}{y_i^*!} + \phi_i \frac{\mu_i^{*y_i^*}}{y_i^*!} e^{-\mu_i^*} \qquad \text{for } \; y_i^* = 0, 1, 2, \ldots,$$

where $\mu_i^* = (1 - \pi_i)\mu_i + \lambda_i$.

## D.2 Proof of Theorem 5.2

*Proof:*

By Theorem 1(a), we have that if $c_+ = 0$ and $c_- = 1$, $Y_i^*$ follows zero-inflated Poisson distribution with parameter $\phi_i^*$ and $\mu_i^*$, where $\phi_i^* = \phi_i$ and $\mu_i^* = \mu_i(1-\pi)$. Thus, by (5.4),

$$\begin{aligned} \log\ \mu_i^* &= \log\ (1-\pi_i) + \log\ \ \mu_i \\ &= \log(1-\pi_i) + \beta_{\mu 0} + \beta_{\mu x}^{\mathrm{T}} X_i \\ &= \beta_{\mu 0}^* + \beta_{\mu x}^{*\mathrm{T}} X_i. \end{aligned} \tag{D.3}$$

Comparing (D.3) to (5.12), we conclude that $\beta_{\mu 0}^* = \beta_{\mu 0} + \log(1-\pi_i)$ and $\beta_{\mu x}^* = \beta_{\mu x}$.

## D.3 Inverse Sampling of Multivariate Discrete Variables

To execute inverse sampling for multivariate discrete variables (Loukas and Kemp, 1983), we evaluate the joint distribution of $U_{i1}$, $U_{i2}$, $Z_{i-}$ and $Z_{i+}$, expressed in (5.16). Noting that although $U_{i1}$ and $U_{i2}$ are unbounded (and $Z_{i-}$ is bounded by $U_{i2}$), the probability (5.16) is extremely small for sufficiently large values. We focus only those values of $U_{i1}$ and $U_{i2}$ bounded by sufficiently large values, say, $J$ and $K$, respectively, and hence $Z_{i-}$ is also bounded by $K$. Specifically, for $j = 0, \ldots, J$, $k = 0, \ldots, K$, and $l = 0, \ldots, K$, we evaluate (5.16)

$$\begin{aligned} p_{jkl} &= P(U_{i1} = j, U_{i2} = k, Z_{i-} = l, Z_{i+} = y_i^* - I(j>0)k + l|x_i) \\ &= P(Z_{i+} = y_i^* - I(j>0)k + l|x_i)P(Z_{i-} = l|U_{i1} = j, U_{i2} = k|x_i)P(U_{i1} = j|x_i)P(U_{i2} = k|x_i), \end{aligned}$$

where each probability is computed based on the distribution assumed for the associated random variables.

To ensure legitimate probabilities induced from the imposition of bounds to $U_{i1}$ and $U_{i2}$, we normalize the $p_{ijk}$ by calculating $p_{jkl}^* = \frac{p_{jkl}}{\sum_{j',k',l'} p_{j'k'l'}}$. Let $\Phi$ be the $(J+1)(K+1)^2$-dimensional column vector consisting of the $p_{jkl}^*$ with $j = 0, \ldots, J; k = 0, \ldots, K$; and $l = 0, \ldots, K$, and let $\Phi_t$ denote the $t$th element of $\Phi$ for $t = 1, \ldots, (J+1)(K+1)^2$.

Generate a random value $V$ from Uniform$[0,1]$ and find the smallest $x$ such that $\sum_{t=1}^{x} \Phi_t \geq V$. Examining $\Phi_x$, we identify $j_0$, $k_0$ and $l_0$ such that $\Phi_x = p_{j_0 k_0 l_0}^*$. Then we set $j_0$, $k_0$ and $l_0$ to be the values for $U_{i1}$, $U_{i2}$, and $Z_{i-}$, respectively, and take $Z_{i+}$ as $Y_i^* - I(U_{i1} > 0)U_{i2} + Z_{i-} = y_i^* - I(j>0)k + l$.

# Appendix E

# Conditions and Proofs of the Results in Chapter 6

## E.1 Regularity Conditions

(R1) The time series $\{X_t : t = 1 \ldots, T\}$ is stationary.

(R2) The observed error-prone time series $\{X_t^* : t = 1 \ldots, T\}$ is stationary.

(R3) For any $t \in \{1, \ldots, T\}$, $\frac{1}{T}\sum_{s=1}^{T} \gamma_{|s-t|} \to 0$ as $T \to \infty$.

(R4) For any $p$, $\frac{1}{T}\sum_{t=1}^{T}\sum_{s=1}^{T} E\{(X_t - \mu)(X_{t+p} - \mu)(X_s - \mu)\} < \infty$.

While the two process $\{X_t : t = 1, \ldots, T\}$ and $\{X_t^* : t = 1, \ldots, T\}$ are constrained by the measurement error model (6.7) or (6.9), they can both be assumed to be stationary without inducing conflicting requirements on the associated processes. Obviously, the weak stationarity of $\{X_t : t = 1, \ldots, T\}$ implies the weak stationarity of $\{X_t^* : t = 1, \ldots, T\}$ if they are linked by (6.7) or (6.9). Condition (R3) says that as the time series goes long enough, the average of the covariances between any paired variables is is negligible. Condition (R4) requires the summation of the third moment of $X_t$ is $O(T)$, which is needed in Theorem 6.4 when $\phi_0 \neq 0$; this condition can be satisfied if $E(\epsilon_t^3) = 0$, for example.

## E.2 The proof of Theorem 6.1

Applying the weak law of large numbers to $\widehat{\phi}_1^*$ given by (6.12), we obtain that the estimator $\widehat{\phi}_1^*$ converges in probability to $\frac{\mathrm{Cov}(X_t^*, X_{t-1}^*)}{\mathrm{Var}(X_{t-1}^*)}$, which is denoted as $\phi_1^*$. Now we further examine $\phi_1^*$ by using the AR(1) model (6.1) and the measurement error model (6.7):

$$\begin{aligned}\phi_1^* &= \frac{\mathrm{Cov}(X_t^*, X_{t-1}^*)}{\mathrm{Var}(X_{t-1}^*)}\\ &= \frac{\mathrm{Cov}(\alpha_0 + \alpha_1 X_t + e_t, \alpha_0 + \alpha_1 X_{t-1} + e_{t-1})}{\mathrm{Var}(\alpha_0 + \alpha_1 X_t + e_t)}\\ &= \frac{\alpha_1^2 \mathrm{Cov}(X_t, X_{t-1})}{\alpha_1^2 \mathrm{Var}(X_t) + \mathrm{Var}(e_t)}\\ &= \frac{\alpha_1^2 \mathrm{Cov}(\phi_0 + \phi_1 X_{t-1} + \epsilon_t, X_{t-1})}{\alpha_1^2 \mathrm{Var}(X_t) + \mathrm{Var}(e_t)}\\ &= \phi_1 \cdot \frac{\alpha_1^2 \mathrm{Var}(X_{t-1})}{\alpha_1^2 \mathrm{Var}(X_t) + Var(e_t)},\end{aligned}$$

where the second step is due to (6.7), the third step is because of the independence among the $X_t$ and the $e_t$, and the fourth step is because of (6.1). Since the time series $\{X_t\}$ is stationary, it follows that $\mathrm{Var}(X_t) = \mathrm{Var}(X_{t-1}) = \frac{\sigma_\epsilon^2}{1-\phi_1^2}$, and hence

$$\phi_1^* = \phi_1 \cdot \frac{\alpha_1^2 \sigma_\epsilon^2}{\alpha_1^2 \sigma_\epsilon^2 + \sigma_e^2 (1 - \phi_1^2)} = \phi_1 \omega_1. \tag{E.1}$$

Next, applying the Slutsky's theorem to (6.12), we have that as $T \to \infty$,

$$\widehat{\phi}_0^* \xrightarrow{p} E(X_t^*) - \phi_1^* E(X_t^*),$$

where the limit equals $\left(\alpha_0 + \frac{\alpha_1 \phi_0}{1-\phi_1}\right)(1 - \phi_1 \omega_1)$ by (E.1) and the fact that $E(X_t^*) = \alpha_0 + \frac{\alpha_1 \phi_0}{1-\phi_1}$.

Finally, plugging the AR(1) model (6.1) into the measurement error model (6.11), we obtain that

$$X_t^* = \alpha_0 + \alpha_1(\phi_0 + \phi_1 X_{t-1} + \epsilon_t) + e_t. \tag{E.2}$$

On the other hand, plugging the measurement error model (6.7) into the working model (6.11), we obtain that

$$X_t^* = \phi_0^* + \phi_1^*(\alpha_0 + \alpha_1 X_{t-1} + e_t) + \epsilon_t^*. \tag{E.3}$$

Then equating (E.2) and (E.3) that

$$\epsilon^* = \alpha_0(1-\phi_1^*) + \alpha_1\phi_0 - \phi_0^* + \alpha_1(\phi_1 - \phi_1^*)X_{t-1} + (1-\phi_1^*)e_t + \alpha_1\epsilon_t.$$

Consequently, by the independence assumption for $X_{t-1}$, $e_t$ and $\epsilon_t$, we obtain that

$$\begin{aligned} Var(\epsilon_t^*) &= \phi_1^2\alpha_1^2(1-\omega_1)^2\mathrm{Var}(X_{t-1}) + (1-\omega_1\phi_1)^2\mathrm{Var}(e_t) + \alpha_1^2\mathrm{Var}(\epsilon_t) \\ &= \phi_1^2\alpha_1^2(1-\omega_1)^2\left(\frac{\sigma_\epsilon^2}{1-\phi_1^2}\right) + (1-\omega_1\phi_1)^2\sigma_e^2 + \alpha_1^2\sigma_\epsilon^2. \end{aligned}$$

## E.3 The proof of Theorem 6.2

As noted in the beginning of E.2, as $T \to \infty$, $\widehat{\phi}_1^* \xrightarrow{p} \phi_1^*$ where

$$\widehat{\phi}_1^* = \frac{\mathrm{Cov}(X_t^*, X_{t-1}^*)}{\mathrm{Var}(X_{t-1}^*)}.$$

Now we further examine $\phi_1^*$ by using the AR(1) model (6.1) and the measurement error model (6.9):

$$\begin{aligned} \phi_1^* &= \frac{\mathrm{Cov}(X_t^*, X_{t-1}^*)}{\mathrm{Var}(X_{t-1}^*)} \\ &= \frac{\mathrm{Cov}(\beta_0 u_t X_t, \beta_0 u_{t-1} X_{t-1})}{\mathrm{Var}(\beta_0 u_{t-1} X_{t-1})} \\ &= \frac{\beta_0^2 \mathrm{Cov}(u_t X_t, u_{t-1} X_{t-1})}{\beta_0^2 \mathrm{Var}(u_{t-1} X_{t-1})} \\ &= \frac{\mathrm{Cov}\{u_t(\phi_0 + \phi_1 X_{t-1} + \epsilon_t), u_{t-1} X_{t-1}\}}{\mathrm{Var}(X_{t-1} u_{t-1})} \\ &= \phi_1 \frac{\mathrm{Cov}(u_t X_{t-1}, u_{t-1} X_{t-1})}{\mathrm{Var}(u_{t-1} X_{t-1})} \\ &= \phi_1 \frac{E(u_t u_{t-1} X_{t-1}^2) - E(u_t X_{t-1}) E(u_{t-1} X_{t-1})}{E(u_{t-1}^2 X_{t-1}^2) - E^2(u_{t-1} X_{t-1})} \\ &= \phi_1 \frac{E(u_t) E(u_{t-1}) E(X_{t-1}^2) - E(u_t) E(u_{t-1}) E^2(X_{t-1})}{E(u_{t-1}^2) E(X_{t-1}^2) - E^2(u_{t-1} X_{t-1})}, \end{aligned}$$

where the second step is due to measurement error model (6.9).

Then, because $u_t$, $u_{t-1}$ and $X_{t-1}$ are mutually independent, we further have that

$$\begin{aligned}
\phi_1^* &= \phi_1 \frac{E(u_t)E(u_{t-1})\mathrm{Var}(X_{t-1})}{\{\mathrm{Var}(u_{t-1}) + E^2(u_{t-1})\}\{\mathrm{Var}(X_{t-1}) + E^2(X_{t-1})\} - E^2(u_{t-1})E^2(X_{t-1})} \\
&= \phi_1 \frac{\mathrm{Var}(X_{t-1})}{\{\mathrm{Var}(u_{t-1}) + 1\}\{\mathrm{Var}(X_{t-1}) + E^2(X_{t-1})\} - E^2(X_{t-1})} \\
&= \phi_1 \frac{\mathrm{Var}(X_{t-1})}{\mathrm{Var}(u_{t-1})\mathrm{Var}(X_{t-1}) + \mathrm{Var}(u_{t-1})E^2(X_{t-1}) + \mathrm{Var}(X_{t-1})}, \qquad \text{(E.4)}
\end{aligned}$$

where the second last step is due to $E(u_t) = 1$. Since the time series $\{X_t\}$ is stationary, it follows that $E(X_t) = E(X_{t-1}) = \frac{\phi_0}{1-\phi_1}$ and $\mathrm{Var}(X_t) = \mathrm{Var}(X_{t-1}) = \frac{\sigma_\epsilon^2}{1-\phi_1^2}$. Hence (E.4) becomes

$$\begin{aligned}
\phi_1^* &= \phi_1 \frac{\mathrm{Var}(X_{t-1})}{\mathrm{Var}(u_{t-1})\mathrm{Var}(X_{t-1}) + \mathrm{Var}(u_{t-1})E^2(X_{t-1}) + \mathrm{Var}(X_{t-1})} \\
&= \phi_1 \frac{\frac{\sigma_\epsilon^2}{1-\phi_1^2}}{\sigma_u^2 \frac{\sigma_\epsilon^2}{1-\phi_1^2} + \sigma_u^2 \left(\frac{\phi_0}{1-\phi_1}\right)^2 + \frac{\sigma_\epsilon^2}{1-\phi_1^2}} \\
&= \phi_1 \frac{\sigma_\epsilon^2}{\sigma_\epsilon^2\sigma_u^2 + \sigma_\epsilon^2 + \sigma_u^2\phi_0^2\frac{1+\phi_1}{1-\phi_1}} = \phi_1\omega_2. \qquad \text{(E.5)}
\end{aligned}$$

Next, applying the Slustky's Theorem to (6.12) gives that as $T \to \infty$,

$$\widehat{\phi}_0^* \xrightarrow{p} \left(\frac{\beta_0\phi_0}{1-\phi_1}\right)(1 - \phi_1\omega_2)$$

by (E.5) as well as $E(X_t^*) = \frac{\beta_0\phi_0}{1-\phi_1}$.

Finally plugging the AR(1) model (6.1) into the measurement error model (6.9), we obtain that

$$X_t^* = \beta_0(\phi_0 + \phi_1 X_{t-1} + \epsilon_t)u_t. \qquad \text{(E.6)}$$

On the other hand, plugging the measurement error model (6.9) into the working model (6.11), we obtain that

$$X_t^* = \phi_0^* + \phi_1^*(\beta_0 X_{t-1}u_{t-1}) + \epsilon_t^*. \qquad \text{(E.7)}$$

Then equating (E.6) and (E.7) gives that

$$\epsilon^* = \beta_0\phi_0 u_t - \phi_0^* + \beta_0 X_{t-1}(\phi_1 u_t - \omega_2\phi_1 u_{t-1}) + \beta_0 u_t \epsilon_t.$$

yielding that

$$
\begin{aligned}
Var(\epsilon_t^*) &= \phi_0^2\beta_0^2\text{Var}(u_t) + \beta_0^2\phi_1^2\text{Var}(X_{t-1}u_t) + \beta_0^2\omega_2^2\phi_1^2\text{Var}(X_{t-1}u_{t-1}) + \beta_0^2 Var(u_t\epsilon_t) \\
&= \phi_0^2\beta_0^2\sigma_u^2 + (\beta_0^2\phi_1^2 + \beta_0^2\omega_2^2\phi_1^2)\{E(X_{t-1}^2u_{t-1}^2) - E^2(X_t)E^2(u_{t-1})\} \\
&\quad + \beta_0^2\{E(u_t^2)E(\epsilon_t^2) - E^2(u_t)E^2(\epsilon_t)\} \\
&= \phi_0^2\beta_0^2\sigma_u^2 + (\beta_0^2\phi_1^2 + \beta_0^2\omega_2^2\phi_1^2)\{E(X_{t-1}^2)E(u_{t-1}^2) - E^2(X_t)E^2(u_{t-1})\} + \beta_0^2(\sigma_u^2+1)\sigma_\epsilon^2 \\
&= \beta_0^2\{\sigma_u^2\phi_0^2 + (1+\sigma_u^2)\sigma_\epsilon^2\} \\
&\quad + \beta_0^2\phi_1^2(1+\omega_2^2)\left[\{\text{Var}(u_{t-1}) + E^2(u_{t-1})\}\{\text{Var}(X_{t-1}) + E^2(X_{t-1})\} - E^2(X_{t-1})\right] \\
&= \beta_0^2\{\sigma_u^2\phi_0^2 + (1+\sigma_u^2)\sigma_\epsilon^2\} \\
&\quad + \beta_0^2\phi_1^2(1+\omega_2^2)\left[\{\text{Var}(u_{t-1}) + 1\}\{\text{Var}(X_{t-1}) + E^2(X_{t-1})\} - E^2(X_{t-1})\right] \\
&= \beta_0^2\{\sigma_u^2\phi_0^2 + (1+\sigma_u^2)\sigma_\epsilon^2\} \\
&\quad + \beta_0^2\phi_1^2(1+\omega_2^2)\left\{\text{Var}(u_{t-1})\text{Var}(X_{t-1}) + \text{Var}(u_{t-1})E^2(X_{t-1}) + \text{Var}(X_{t-1})\right\} \\
&= \beta_0^2\{\sigma_u^2\phi_0^2 + (1+\sigma_u^2)\sigma_\epsilon^2\} + \beta_0^2\phi_1^2(1+\omega_2^2)\frac{Var(X_{t-1})}{\omega_2} \\
&= \beta_0^2\{\sigma_u^2\phi_0^2 + (1+\sigma_u^2)\sigma_\epsilon^2\} + \beta_0^2\phi_1^2\frac{1+\omega_2^2}{\omega_2}\frac{\sigma_\epsilon^2}{1-\phi_1^2},
\end{aligned}
$$

where the second step is because of the independence assumption as well as $E(u_{t-1}^2) = E(u_t^2)$ and $E(u_{t-1}) = E(u_t)$ such that $\text{Var}(X_{t-1}u_t) = \text{Var}(X_{t-1}u_{t-1})$, and the second last step is due to $\omega_2 = \frac{\text{Var}(X_{t-1})}{\text{Var}(u_{t-1})\text{Var}(X_{t-1}) + \text{Var}(u_{t-1})E^2(X_{t-1}) + \text{Var}(X_{t-1})}$ in (E.5).

## E.4 The proof of Theorem 6.3

**Proof of Theorem 6.3(1):**

For $k = 1, \ldots, p$, applying the weak law of large numbers to $\widehat{\gamma}_k^*$, we obtain that as $T \to \infty$, the estimator $\widehat{\gamma}_k^*$ converges in probability to $\text{Cov}(X_t^*, X_{t-k}^*)$, denoted $\gamma_k^*$.

Next, we examine $\gamma_k$. By the form of measurement error model (6.7), we have that for $0 < k < t$,

$$
\begin{aligned}
\text{Cov}(X_t^*, X_{t-k}^*) &= \text{Cov}(\alpha_0 + \alpha_1 X_t + e_t, \alpha_0 + \alpha_1 X_{t-k} + e_{t-k}) \\
&= \alpha_1^2\text{Cov}(X_t, X_{t-k}) = \alpha_1^2\gamma_k,
\end{aligned}
$$

and by (6.8), $\text{Var}(X_t^*) = \alpha_1^2\gamma_0 + \sigma_e^2$, which is denoted as $\gamma_0^*$.

Thus, Theorem 6.3(1) follows.

**Proof of Theorem 6.3(2)**:

First, by Theorem 6.3(1), we write

$$\widehat{\gamma}^* = \alpha_1^2\gamma + o_p(1) \tag{E.8}$$

and

$$\widehat{\Gamma}^* = \alpha_1^2\Gamma + \sigma_e^2 I_p + o_p(1),$$

where $\widehat{\Gamma}^* = \begin{pmatrix} \widehat{\gamma}_0^* & \cdots & \widehat{\gamma}_{p-1}^* \\ \vdots & \ddots & \vdots \\ \widehat{\gamma}_{p-1}^* & \cdots & \widehat{\gamma}_0^* \end{pmatrix}$. Then the naive estimator $\widehat{\phi}^*$ is obtained by replacing $\widehat{\gamma}_k$ in (6.6) with $\widehat{\gamma}_k^*$,

$$\widehat{\phi}^* = \left\{\alpha_1^2\Gamma + \sigma_e^2 I_p + o_p(1)\right\}^{-1}\left\{\alpha_1^2\gamma + o_p(1)\right\} = \alpha_1^2\left(\alpha_1^2\Gamma + \sigma_e^2 I_p\right)^{-1}\gamma + o_p(1), \tag{E.9}$$

and hence $\phi^* = \alpha_1^2\left(\alpha_1^2\Gamma + \sigma_e^2 I_p\right)^{-1}\gamma$ such that $\widehat{\phi}^* \xrightarrow{p} \phi^*$ as $T \to \infty$.

Again, replacing $\widehat{\gamma}_k$ in (6.6) with $\widehat{\gamma}_k^*$ gives the naive estimator $\widehat{\phi}_0^*$

$$\begin{aligned}
\widehat{\phi}_0^* &= \frac{1}{T-p}\sum_{t=p}^{T} X_t^* - \left(\sum_{k=1}^{p}\widehat{\phi}_k^*\right)\left(\frac{1}{T-p}\sum_{t=p}^{T} X_{t-k}^*\right) \\
&= E(X_t^*) - E(X_t^*)\sum_{k=1}^{p}\widehat{\phi}_k^* + o_p(1) \\
&= \alpha_0 + \alpha_1 E(X_t) - \{\alpha_0 + \alpha_1 E(X_t)\}\sum_{k=1}^{p}\{\phi_k^* + o_p(1)\} + o_p(1) \\
&= \left(1 - \phi^{*\mathrm{T}} \cdot \mathbb{1}_p\right)(\alpha_0 + \alpha_1\mu) + o_p(1),
\end{aligned}$$

where $\widehat{\phi}_k$ and $\phi_k$ are respectively the $k$th element of $\widehat{\phi}$ and $\phi$, the third step is because $\widehat{\phi}_k = \phi_k + o_p(1)$ by (E.9) as well as the model form (6.7), and the last step is due to the stationarity of the time series $\{X_t\}$ such that $E(X_t) = \mu$.

Finally, noting that the native estimator $\widehat{\sigma}_\epsilon^{2*}$ is given by $\widehat{\sigma}_\epsilon^{2*} = \widehat{\gamma}_0^* - 2\widehat{\phi}^{*\mathrm{T}}\widehat{\gamma}^* + \widehat{\phi}^{*\mathrm{T}}\widehat{\Gamma}^*\widehat{\phi}^*$ by applying a version similar to (6.6), we obtain that

$$\begin{aligned}
\widehat{\sigma}_\epsilon^{2*} &= \widehat{\gamma}_0^* - 2\widehat{\phi}^{*\mathrm{T}}\widehat{\gamma}^* + \widehat{\phi}^{*\mathrm{T}}\widehat{\Gamma}^*\widehat{\phi}^* \\
&= (\alpha_1^2\gamma_0^2 + \sigma_e^2) - 2\alpha_1^4\gamma^{\mathrm{T}}(\alpha_1^2\Gamma + \sigma_e^2 I_p)^{-1}\gamma \\
&\qquad + \alpha_1^4\gamma^{\mathrm{T}}(\alpha_1^2\Gamma + \sigma_e^2 I_p)^{-1}(\alpha_1^2\Gamma + \sigma_e^2 I_p)(\alpha_1^2\Gamma + \sigma_e^2 I_p)^{-1}\gamma + o_p(1) \\
&= \alpha_1^2\gamma_0 + \sigma_e^2 - \alpha_1^4\gamma^{\mathrm{T}}(\alpha_1^2\Gamma + \sigma_e^2 I_p)^{-1}\gamma + o_p(1),
\end{aligned}$$

where the second step is due to (6.8), (E.8) and (E.9).

**Proof of Theorem 6.3(3)**:

**Step 1**: We show certain identities before proving Theorem 6.3(3).

1. By model (6.7), we have that

$$\begin{aligned}
X_t^* - \widehat{\mu}^* &= \alpha_0 + \alpha_1 X_t + e_t - \frac{1}{T}\sum_{t=1}^{T}(\alpha_0 + \alpha_1 X_t + e_t) \\
&= \alpha_1\left(X_t - \frac{1}{T}\sum_{t=1}^{T} X_t\right) + \left(e_t - \frac{1}{T}\sum_{t=1}^{T} e_t\right) \\
&= \alpha_1(X_t - \widehat{\mu}) + (e_t - \bar{e}), \qquad \text{(E.10)}
\end{aligned}$$

where the first step is because $\widehat{\mu}^* = \frac{1}{T}\sum_{t=1}^{T} X_t^*$ and in the last step $\bar{e} = \frac{1}{T}\sum_{t=1}^{T} e_t$.

2. For any $t$ and $s$, we have that

$$\begin{aligned}
&\mathrm{Cov}\left\{(X_t - \widehat{\mu})^2, (X_s - \widehat{\mu})(e_s - \bar{e})\right\} \\
=&E\{(X_t - \widehat{\mu})^2(X_s - \widehat{\mu})(e_s - \bar{e})\} - \{E(X_t - \widehat{\mu})^2\}E\{(X_s - \widehat{\mu})(e_s - \bar{e})\} \\
=&E\{(X_t - \widehat{\mu})^2(X_s - \widehat{\mu})\}E(e_s - \bar{e}) - \{E(X_t - \widehat{\mu})^2\}E(X_s - \widehat{\mu})E(e_s - \bar{e}) \\
=&0, \qquad \text{(E.11)}
\end{aligned}$$

where the second step is due to the independence of $e_t$ and $X_t$, and the last step is by $E(e_s - \bar{e}) = 0$.

3. By the independence of $e_t$ and $e_s$ for $t \neq s$, we have that

$$
\begin{aligned}
&\operatorname{Cov}\{(X_t - \widehat{\mu})(e_t - \bar{e}), (X_s - \widehat{\mu})(e_s - \bar{e})\} \\
=& E\{(X_t - \widehat{\mu})(e_t - \bar{e})(X_s - \widehat{\mu})(e_s - \bar{e})\} - E\{(X_t - \widehat{\mu})(e_t - \bar{e})\}E\{(X_s - \widehat{\mu})(e_s - \bar{e})\} \\
=& E\{(X_t - \widehat{\mu})(X_s - \widehat{\mu})\}E\{(e_t - \bar{e})\}E\{(e_s - \bar{e})\} \\
&\qquad - E\{(X_t - \widehat{\mu})\}E\{(e_t - \bar{e})\}E\{(X_s - \widehat{\mu})\}E\{(e_s - \bar{e})\} \\
=& 0,
\end{aligned} \tag{E.12}
$$

where the second step is due to the independence of the $e_t$ and the $X_t$, and the last step is by $E(e_s - \bar{e}) = 0$.

4. For any $t$, we have that

$$
\begin{aligned}
&\operatorname{Var}\{(X_t - \widehat{\mu})(e_t - \bar{e})\} \\
=& E\{(X_t - \widehat{\mu})^2(e_t - \bar{e})^2\} - E^2\{(X_t - \widehat{\mu})(e_t - \bar{e})\} \\
=& E\{(X_t - \widehat{\mu})^2\}E\{(e_t - \bar{e})^2\} - E^2\{(X_t - \widehat{\mu})\}E^2\{(e_t - \bar{e})\} \\
=& E\{(X_t - \widehat{\mu})^2\}E\{(e_t - \bar{e})^2\}.
\end{aligned} \tag{E.13}
$$

5. For any $t$, we have

$$
\begin{aligned}
&\lim_{T\to\infty} E\{(X_t - \widehat{\mu})^2\} \\
=& \lim_{T\to\infty} E\{(X_t - \mu)^2 + (\mu - \widehat{\mu})^2 + 2(X_t - \mu)(\mu - \widehat{\mu})\} \\
=& \gamma_0 + \lim_{T\to\infty} E\{(\widehat{\mu} - \mu)^2\} + 2\lim_{T\to\infty} E\{(X_t - \mu)(\mu - \widehat{\mu})\} \\
=& \gamma_0 + \lim_{T\to\infty} E\{(\widehat{\mu} - \mu)^2\} - 2\lim_{T\to\infty} E\left[(X_t - \mu)\{\frac{1}{T}\sum_{s=1}^{T}(X_s - \mu)\}\right] \\
=& \gamma_0 + \lim_{T\to\infty} Var(\widehat{\mu}) - 2\lim_{T\to\infty} \frac{1}{T}\sum_{s=1}^{T} E\{(X_t - \mu)(X_s - \mu)\} \\
=& \gamma_0 + 0 - 2\lim_{T\to\infty} \frac{1}{T}\sum_{s=1}^{T} \gamma_{|s-t|} \\
=& \gamma_0,
\end{aligned} \tag{E.14}
$$

where the third step is due to $\widehat{\mu} - \mu = \frac{1}{T}\sum_{s=1}(X_s - \mu)$, and the fourth step is because $E(\widehat{\mu} - \mu) = 0$ by stationarity of the time series, the second last step is due to $\lim_{T\to\infty} Var(\widehat{\mu}) = 0$ (Brockwell et al., 1991, Theorem 7.1.1.), and the last step due to Condition (R3).

6. Similar to (E.14), we have that

$$\begin{aligned}
&\lim_{T\to\infty} E\{(X_t-\widehat{\mu})(X_{t-p}-\widehat{\mu})\}\\
=&\lim_{T\to\infty} E\{(X_t-\mu+\mu-\widehat{\mu})(X_{t-p}-\mu+\mu-\widehat{\mu})\}\\
=&\lim_{T\to\infty} [E\{(X_t-\mu)(X_{t-p}-\mu)\}+E\{(\mu-\widehat{\mu})(X_{t-p}-\mu)\} \qquad \text{(E.15)}\\
&\quad +E\{(\mu-\widehat{\mu})(X_t-\mu)\}+E\{(\mu-\widehat{\mu})(\mu-\widehat{\mu})\}]\\
=&\gamma_p+\lim_{T\to\infty}\frac{1}{T}\sum_{s=1}^{T}(\gamma_{|t-s|}+\gamma_{|t-s-p|})+\lim_{T\to\infty}\mathrm{Var}(\widehat{\mu})\\
=&\gamma_p, \qquad \text{(E.16)}
\end{aligned}$$

where the last step is due to Condition (R3) and $\lim_{T\to\infty} Var(\widehat{\mu})=0$ (Brockwell et al., 1991, Theorem 7.1.1).

7. For any $t$, we have

$$\begin{aligned}
&E\{(e_t-\bar{e})^2\}\\
=&E\{e_t^2-2e_t\bar{e}+\bar{e}^2\}\\
=&\left\{E(e_t^2)-\frac{2}{T}\sum_{s=1}^{T}E(e_te_s)+\frac{1}{T^2}\sum_{t=1}^{T}\sum_{s=1}^{T}E(e_te_s)\right\}\\
=&E(e_t^2)+\left\{-\frac{2}{T}E(e_te_t)+\frac{1}{T^2}\sum_{t=1}^{T}E(e_t^2)\right\}\\
=&\frac{T-1}{T}E(e_t^2)=\frac{T-1}{T}\sigma_e^2, \qquad \text{(E.17)}
\end{aligned}$$

so $\lim_{T\to\infty} E\{(e_t-\bar{e})^2\}=\sigma_e^2$.

8. By the independence of $e_t$ and $X_t$, for any $s$ and $t$, we have that

$$\begin{aligned}
&\mathrm{Cov}\{(X_t-\widehat{\mu})(e_t-\bar{e}),(e_s-\bar{e})^2\}\\
=&E\{(X_t-\widehat{\mu})(e_t-\bar{e})(e_s-\bar{e})^2\}-E\{(X_t-\widehat{\mu})(e_t-\bar{e})\}E\{(e_s-\bar{e})^2\}\\
=&E(X_t-\widehat{\mu})E\{(e_t-\bar{e})(e_s-\bar{e})^2\}-E(X_t-\widehat{\mu})E(e_t-\bar{e})E(e_s-\bar{e})^2\\
=&0, \qquad \text{(E.18)}
\end{aligned}$$

where the last step is due to $E(X_t-\widehat{\mu})=0$ and $E(e_t-\bar{e})=0$.

9. For any $t \neq s$, $\text{Cov}\{(e_t - \bar{e})^2, (e_s - \bar{e})^2\} = 0$; and for $t = s$,

$$
\begin{aligned}
&\text{Var}\{(e_t - \bar{e})^2\} \\
=&E\{(e_t - \bar{e})^4\} - E^2\{(e_t - \bar{e})^2\} \\
=&E(e_t^4) - 4E(e_t^3\bar{e}) + 6E(e_t^2\bar{e}^2) - 4E(e_t\bar{e}^3) + E(\bar{e}_t^4) - \{E(e_t^2) - 2E(e_t\bar{e}) + E(\bar{e}^2)\}^2 \\
=&E(e_t^4) - \frac{4}{T}E(e_t^4) + \left[\frac{6(T-1)}{T^2}\{E(e_t^2)\}^2 + \frac{6}{T^2}E(e_t^4)\right] - \frac{4}{T^3}E(e_t^4) \\
&+ \left[\frac{1}{T^3}E(e_t^4) + \frac{3(T-1)}{T^3}\{E(e_t^2)\}^2\right] - \left\{E(e_t^2) - \frac{2}{T}E(e_t^2) + \frac{1}{T}E(e_t^2)\right\}^2, \qquad \text{(E.19)}
\end{aligned}
$$

so $\lim_{T\to\infty} \text{Var}\{(e_t - \bar{e})^2\} = E(e_t^4) - \{E(e_t^2)\}^2 = E(e_t^4) - \sigma_e^4$.

10. Similar to the derivation in (E.19), we can show $\text{Cov}\{(e_t - \bar{e})^2, (e_s - \bar{e})(e_{s+p} - \bar{e})\} = 0$ for $s \neq t$ and $s \neq t - p$. For a given $t$,

$$
\begin{aligned}
&\text{Cov}\{(e_t - \bar{e})^2, (e_t - \bar{e})(e_{t+p} - \bar{e})\} \\
=&E\{(e_t - \bar{e})^3(e_{t+p} - \bar{e})\} - E\{(e_t - \bar{e})^2\}E\{(e_t - \bar{e})(e_{t+p} - \bar{e})\}, \qquad \text{(E.20)}
\end{aligned}
$$

which can be derived analogously to the (E.19) that $\lim_{T\to\infty} E\{(e_t - \bar{e})^3(e_{t+p} - \bar{e})\} - E\{(e_t - \bar{e})^2\}E\{(e_t - \bar{e})(e_{t+p} - \bar{e})\} = E\{e_t^3 e_{t+p}\} - E\{e_t^2\}E\{e_t e_{t+p}\} = 0$ and similarly $\lim_{T\to\infty} \text{Cov}\{(e_t - \bar{e})^2, (e_{t-p} - \bar{e})(e_t - \bar{e})\} = 0$.

11. For any $t$,

$$
\begin{aligned}
&\text{Cov}\left\{(X_t - \widehat{\mu})(e_{t+p} - \bar{e}), (X_{t+p-r} - \widehat{\mu})(e_{t+p} - \bar{e})\right\} \\
&= \left[E\left\{(X_t - \widehat{\mu})(X_{t+p-r} - \widehat{\mu})(e_{t+p} - \bar{e})^2\right\} - E(X_t - \widehat{\mu})E(X_{t+p-r} - \widehat{\mu})E^2(e_{t+p} - \bar{e})\right] \\
&= E\left\{(X_t - \widehat{\mu})(X_{t+p-r} - \widehat{\mu})\right\} E\left\{(e_{t+p} - \bar{e})^2\right\} \\
&= \gamma_{|p-r|}\left(\frac{T-1}{T}\right)\sigma_e^2, \qquad \text{(E.21)}
\end{aligned}
$$

where the second step is because of $E(X_t - \widehat{\mu}) = 0$ and the independence of $X_t$ and $e_t$, the third step is due to (E.17) and (E.15). Hence,

$$\lim_{T\to\infty} \text{Cov}\left\{(X_t - \widehat{\mu})(e_{t+p} - \bar{e}), (X_{t+p-r} - \widehat{\mu})(e_{t+p} - \bar{e})\right\} = \gamma_{|p-r|}\sigma_e^2.$$

Similarly,

$$\lim_{T\to\infty} \text{Cov}\left\{(X_{t+p} - \widehat{\mu})(e_t - \bar{e}), (X_{t-r} - \widehat{\mu})(e_t - \bar{e})\right\} = \gamma_{|p-r|}\sigma_e^2.$$

Similarly,

$$
\begin{aligned}
&\text{Cov}\left\{(X_t - \widehat{\mu})(e_{t+p} - \bar{e}), (X_{t+p+r} - \widehat{\mu})(e_{t+p} - \bar{e})\right\} \\
&= \Big[E\left\{(X_t - \widehat{\mu})(X_{t+p+r} - \widehat{\mu})(e_{t+p} - \bar{e})^2\right\} \\
&\qquad - E(X_t - \widehat{\mu})E(X_{t+p-r} - \widehat{\mu})E^2(e_{t+p} - \bar{e})\Big] \\
&= E\left\{(X_t - \widehat{\mu})(X_{t+p+r} - \widehat{\mu})\right\} E\left\{(e_t - \bar{e})^2\right\} \\
&= \gamma_{p+r}\left(\frac{T-1}{T}\right)\sigma_e^2,
\end{aligned}
\tag{E.22}
$$

and hence

$$
\lim_{T\to\infty} \text{Cov}\left\{(X_t - \widehat{\mu})(e_{t+p} - \bar{e}), (X_{t+p+r} - \widehat{\mu})(e_{t+p} - \bar{e})\right\} = \gamma_{p+r}\sigma_e^2.
$$

Similarly,

$$
\lim_{T\to\infty} \text{Cov}\left\{(X_{t+p} - \widehat{\mu})(e_t - \bar{e}), (X_{t+r} - \widehat{\mu})(e_t - \bar{e})\right\} = \gamma_{p+r}\sigma_e^2.
$$

12. By independence assumption between $\{e_t\}$, if $t \neq s$ or $p \neq r$, we have that

$$
\text{Cov}\left\{(e_t - \bar{e})(e_{t+p} - \bar{e}), (e_s - \bar{e})(e_{s+r} - \bar{e})\right\} = 0. \tag{E.23}
$$

In addition, by (E.17), we have that

$$
\begin{aligned}
&\text{Var}\left\{(e_t - \bar{e})(e_{t+p} - \bar{e})\right\} \\
&= E\left\{(e_t - \bar{e})^2(e_{t+p} - \bar{e})^2\right\} \\
&= E\left\{(e_t - \bar{e})^2\right\} E\left\{(e_{t+p} - \bar{e})^2\right\}, \\
&= \left(\frac{T-1}{T}\right)^2 \sigma_e^4,
\end{aligned}
\tag{E.24}
$$

so $\lim_{T\to\infty} \text{Var}\left\{(e_t - \bar{e})(e_{t+p} - \bar{e})\right\} = \sigma_e^4$.

**Step 2**: Now we prove the results in (3).

$1^\circ$. We first show the derivation of $q^*_{100}$ as follows:

$$
\begin{aligned}
q^*_{100} &= \lim_{T\to\infty} T\mathrm{Cov}\left\{\frac{1}{T}\sum_{t=1}^{T}(X^*_t-\widehat{\mu}^*)^2, \frac{1}{T}\sum_{s=1}^{T}(X^*_s-\widehat{\mu}^*)^2\right\} \\
&= \lim_{T\to\infty} T\mathrm{Cov}\left[\frac{1}{T}\sum_{t=1}^{T}\left\{\alpha_1^2(X_t-\widehat{\mu})^2+2\alpha_1(X_t-\widehat{\mu})(e_t-\bar{e})+(e_t-\bar{e})^2\right\},\right. \\
&\qquad\qquad \left.\frac{1}{T}\sum_{s=1}^{T}\alpha_1^2(X_s-\widehat{\mu})^2+2\alpha_1(X_s-\widehat{\mu})(e_s-\bar{e})+(e_s-\bar{e})^2\right] \\
&= \alpha_1^4 \lim_{T\to\infty} T\mathrm{Cov}\left\{\frac{1}{T}\sum_{t=1}^{T}(X_t-\widehat{\mu})^2, \frac{1}{T}\sum_{s=1}^{T}(X_s-\widehat{\mu})^2\right\} \\
&\qquad + \lim_{T\to\infty} T\mathrm{Cov}\left\{\frac{1}{T}\sum_{t=1}^{T}2\alpha_1(X_t-\widehat{\mu})(e_t-\bar{e}), \frac{1}{T}\sum_{s=1}^{T}2\alpha_1(X_s-\widehat{\mu})(e_s-\bar{e})\right\} \\
&\qquad + \lim_{T\to\infty} T\mathrm{Cov}\left\{\frac{1}{T}\sum_{t=1}^{T}(e_t-\bar{e})^2, \frac{1}{T}\sum_{s=1}^{T}(e_s-\bar{e})^2\right\} \\
&= \alpha_1^4 q_{00} + \lim_{T\to\infty}\frac{4\alpha_1^2}{T}\sum_{t=1}^{T}\sum_{s=1}^{T}\mathrm{Cov}\left\{(X_t-\widehat{\mu})(e_t-\bar{e}), (X_s-\widehat{\mu})(e_s-\bar{e})\right\} \\
&\qquad + \lim_{T\to\infty}\frac{1}{T}\sum_{t=1}^{T}\sum_{s=1}^{T}\mathrm{Cov}\left\{(e_t-\bar{e})^2, (e_s-\bar{e})^2\right\} \\
&= \alpha_1^4 q_{00} + \lim_{T\to\infty}\frac{4\alpha_1^2}{T}\sum_{t=1}^{T}\mathrm{Cov}\left\{(X_t-\widehat{\mu})(e_t-\bar{e}), (X_t-\widehat{\mu})(e_t-\bar{e})\right\} \\
&\qquad + \lim_{T\to\infty}\frac{1}{T}\sum_{t=1}^{T}\mathrm{Cov}\left\{(e_t-\bar{e})^2, (e_t-\bar{e})^2\right\} \\
&= \alpha_1^4 q_{00} + 4\alpha_1^2 E\left\{(X_t-\widehat{\mu})^2(e_t-\bar{e})^2\right\} + E(e_t^4) - \left\{E(e_t^2)\right\}^2 \\
&= \alpha_1^4 q_{00} + 4\alpha_1^2\gamma_0\sigma_e^2 + E(e_t^4) - \sigma_e^4,
\end{aligned}
$$

where the second step is due to (E.10), the third step is because of (E.11), (E.18), and the definition $q_{00} = \lim_{T\to\infty} T\mathrm{Cov}\left\{\frac{1}{T}\sum_{t=1}^{T}(X_t-\widehat{\mu})^2, \frac{1}{T}\sum_{s=1}^{T}(X_s-\widehat{\mu})^2\right\}$, the fifth step is due to (E.12) and (E.19), and the sixth step is because (E.13) and (E.19), and the last step is because (E.17) and (E.18).

$2^\circ$. We derive the value of $q^*_{10p}$:

$$
\begin{aligned}
q^*_{10p} &= \lim_{T\to\infty} T\mathrm{Cov}\left\{\frac{1}{T}\sum_{t=1}^{T}(X^*_t-\widehat{\mu}^*)^2, \frac{1}{T-p}\sum_{s=1}^{T-p}(X^*_s-\widehat{\mu}^*)(X^*_{s+p}-\widehat{\mu}^*)\right\}\\
&= \lim_{T\to\infty} T\mathrm{Cov}\left[\frac{1}{T}\sum_{t=1}^{T}\left\{\alpha_1^2(X_t-\widehat{\mu})^2+2\alpha_1(X_t-\widehat{\mu})(e_t-\bar{e})+(e_t-\bar{e})^2\right\},\right.\\
&\qquad\qquad \frac{1}{T-p}\sum_{s=1}^{T-p}\alpha_1^2(X_s-\widehat{\mu})(X_{s+p}-\widehat{\mu})+\alpha_1(X_s-\widehat{\mu})(e_{s+p}-\bar{e})\\
&\qquad\qquad \left.+\alpha_1(X_{s+p}-\widehat{\mu})(e_s-\bar{e})+(e_s-\bar{e})(e_{s+p}-\bar{e})\right]\\
&= \alpha_1^4\lim_{T\to\infty} T\mathrm{Cov}\left\{\frac{1}{T}\sum_{t=1}^{T}(X_t-\widehat{\mu})^2, \frac{1}{T-p}\sum_{s=1}^{T-p}(X_s-\widehat{\mu})(X_{s+p}-\widehat{\mu})\right\}\\
&\qquad + \lim_{T\to\infty} T\mathrm{Cov}\left\{\frac{1}{T}\sum_{t=1}^{T}2\alpha_1(X_t-\widehat{\mu})(e_t-\bar{e}), \frac{1}{T-p}\sum_{s=1}^{T-p}\alpha_1(X_s-\widehat{\mu})(e_{s+p}-\bar{e})\right\}\\
&\qquad + \lim_{T\to\infty} T\mathrm{Cov}\left\{\frac{1}{T}\sum_{t=1}^{T}2\alpha_1(X_t-\widehat{\mu})(e_t-\bar{e}), \frac{1}{T-p}\sum_{s=1}^{T-p}\alpha_1(X_{s+p}-\widehat{\mu})(e_s-\bar{e})\right\}\\
&\qquad + \lim_{T\to\infty} T\mathrm{Cov}\left\{\frac{1}{T}\sum_{t=1}^{T}(e_t-\bar{e})^2, \frac{1}{T-p}\sum_{s=1}^{T-p}(e_s-\bar{e})(e_{s+p}-\bar{e})\right\}\\
&= \alpha_1^4 q_{0p} + \lim_{T\to\infty}\frac{2\alpha_1^2}{T-p}\sum_{t=1}^{T}\sum_{s=1}^{T-p}\mathrm{Cov}\left\{(X_t-\widehat{\mu})(e_t-\bar{e}),(X_s-\widehat{\mu})(e_{s+p}-\bar{e_s})\right\}\\
&\qquad + \lim_{T\to\infty}\frac{2\alpha_1^2}{T-p}\sum_{t=1}^{T}\sum_{s=1}^{T-p}\mathrm{Cov}\left\{(X_t-\widehat{\mu})(e_t-\bar{e}),(X_{s+p}-\widehat{\mu})(e_s-\bar{e})\right\}\\
&= \alpha_1^4 q_{0p} + \lim_{T\to\infty}\frac{2\alpha_1^2}{T-p}\sum_{\substack{t=p\\(s=t-p)}}^{T}\mathrm{Cov}\left\{(X_t-\widehat{\mu})(e_t-\bar{e}),(X_{t-p}-\widehat{\mu})(e_t-\bar{e})\right\}\\
&\qquad + \lim_{T\to\infty}\frac{2\alpha_1^2}{T-p}\sum_{\substack{t=1\\(s=t)}}^{T-p}\mathrm{Cov}\left\{(X_t-\widehat{\mu})(e_t-\bar{e}),(X_{t+p}-\widehat{\mu})(e_t-\bar{e})\right\}\\
&= \alpha_1^4 q_{0p} + 2\alpha_1^2 E\left\{(X_t-\widehat{\mu})(X_{t-p}-\widehat{\mu})(e_t-\bar{e})^2\right\} + 2\alpha_1^2 E\left\{(X_t-\widehat{\mu})(X_{t+p}-\widehat{\mu})(e_t-\bar{e})^2\right\}\\
&= \alpha_1^4 q_{0p} + 4\alpha_1^2\gamma_p\sigma_e^2,
\end{aligned}
$$

where the second step is due to (E.10), the third step is because of (E.11) and (E.18), the fourth step is by definition that $q_{0p}=\lim_{T\to\infty}T\mathrm{Cov}\left\{\frac{1}{T}\sum_{t=1}^{T}(X_t-\widehat{\mu})^2, \frac{1}{T-p}\sum_{s=1}^{T-p}(X_s-\widehat{\mu})(X_{s+p}-\widehat{\mu})\right\}$ and (E.20), the fifth step is due to (E.12), and the last step is result from (E.17) and (E.15).

3°. We derive $q^*_{1pr}$ for $p>0$, $r>0$ and $p\neq r$:

$$
\begin{aligned}
q^*_{1pr} &= \lim_{T\to\infty} T\mathrm{Cov}\left\{\frac{1}{T-p}\sum_{t=1}^{T-p}(X^*_t-\widehat{\mu}^*)(X^*_{t+p}-\widehat{\mu}^*),\frac{1}{T-r}\sum_{s=1}^{T-r}(X^*_s-\widehat{\mu}^*)(X^*_{s+r}-\widehat{\mu}^*)\right\}\\
&= \lim_{T\to\infty} T\mathrm{Cov}\Bigg[\frac{1}{T-p}\sum_{t=1}^{T-p}\left\{\alpha_1^2(X_t-\widehat{\mu})(X_{t+p}-\widehat{\mu})+\alpha_1(X_t-\widehat{\mu})(e_{t+p}-\bar{e})\right.\\
&\qquad \left.+\alpha_1(X_{t+p}-\widehat{\mu})(e_t-\bar{e})+(e_t-\bar{e})(e_{t+p}-\bar{e})\right\},\\
&\qquad \frac{1}{T-r}\sum_{s=1}^{T-r}\alpha_1^2(X_s-\widehat{\mu})(X_{s+r}-\widehat{\mu})+\alpha_1(X_s-\widehat{\mu})(e_{s+r}-\bar{e})\\
&\qquad +\alpha_1(X_{s+r}-\widehat{\mu})(e_s-\bar{e})+(e_s-\bar{e})(e_{s+r}-\bar{e})\Bigg]\\
&= \alpha_1^4\lim_{T\to\infty} T\mathrm{Cov}\left\{\frac{1}{T-p}\sum_{t=1}^{T-p}(X_t-\widehat{\mu})(X_{t+p}-\widehat{\mu}),\frac{1}{T-r}\sum_{s=1}^{T-r}(X_s-\widehat{\mu})(X_{s+r}-\widehat{\mu})\right\}\\
&\qquad +\lim_{T\to\infty} T\mathrm{Cov}\left\{\frac{1}{T-p}\sum_{t=1}^{T-p}\alpha_1(X_t-\widehat{\mu})(e_{t+p}-\bar{e}),\frac{1}{T-r}\sum_{s=1}^{T-r}\alpha_1(X_s-\widehat{\mu})(e_{s+r}-\bar{e})\right\}\\
&\qquad +\lim_{T\to\infty} T\mathrm{Cov}\left\{\frac{1}{T-p}\sum_{t=1}^{T-p}\alpha_1(X_t-\widehat{\mu})(e_{t+p}-\bar{e}),\frac{1}{T-r}\sum_{s=1}^{T-r}\alpha_1(X_{s+r}-\widehat{\mu})(e_s-\bar{e})\right\}\\
&\qquad +\lim_{T\to\infty} T\mathrm{Cov}\left\{\frac{1}{T-p}\sum_{t=1}^{T-p}\alpha_1(X_{t+p}-\widehat{\mu})(e_t-\bar{e}),\frac{1}{T-r}\sum_{s=1}^{T-r}\alpha_1(X_s-\widehat{\mu})(e_{s+r}-\bar{e})\right\}\\
&\qquad +\lim_{T\to\infty} T\mathrm{Cov}\left\{\frac{1}{T-p}\sum_{t=1}^{T-p}\alpha_1(X_{t+p}-\widehat{\mu})(e_t-\bar{e}),\frac{1}{T-r}\sum_{s=1}^{T-r}\alpha_1(X_{s+r}-\widehat{\mu})(e_s-\bar{e})\right\}\\
&= \alpha_1^4 q_{pr}+\alpha_1^2\lim_{T\to\infty}\frac{T}{(T-p)(T-r)}\sum_{\substack{t=\max(1,r-p+1)\\(s=t+p-r)}}^{T-p}\mathrm{Cov}\left\{(X_t-\widehat{\mu})(e_{t+p}-\bar{e}),(X_{t+p-r}-\widehat{\mu})(e_{t+p}-\bar{e})\right\}\\
&\quad +\alpha_1^2\lim_{T\to\infty}\frac{T}{(T-p)(T-r)}\sum_{\substack{t=1\\(s=t+p)}}^{T-p-r}\mathrm{Cov}\left\{(X_t-\widehat{\mu})(e_{t+p}-\bar{e}),(X_{t+p+r}-\widehat{\mu})(e_{t+p}-\bar{e})\right\}\\
&\quad +\alpha_1^2\lim_{T\to\infty}\frac{T}{(T-p)(T-r)}\sum_{\substack{t=r+1\\(s=t-r)}}^{T-p}\mathrm{Cov}\left\{(X_{t+p}-\widehat{\mu})(e_t-\bar{e}),(X_{t-r}-\widehat{\mu})(e_t-\bar{e})\right\}\\
&\quad +\alpha_1^2\lim_{T\to\infty}\frac{T}{(T-p)(T-r)}\sum_{\substack{t=1\\(s=t)}}^{T-\max(p,r)}\mathrm{Cov}\left\{(X_{t+p}-\widehat{\mu})(e_t-\bar{e}),(X_{t+r}-\widehat{\mu})(e_t-\bar{e})\right\}\\
&= \alpha_1^4 q_{pr}+2\alpha_1^2\sigma_e^2(\gamma_{|p-r|}+\gamma_{p+r}). \qquad \text{(E.25)}
\end{aligned}
$$

where the second step is due to (E.10), the third step is because of (E.11) and a similar version to (E.18), the fourth step is because (E.23) and by the definition that $q_{pr} = \lim_{T\to\infty} T\text{Cov}\left\{\frac{1}{T-p}\sum_{t=1}^{T-p}(X_t-\widehat{\mu})(X_{t+p}-\widehat{\mu}), \frac{1}{T-r}\sum_{s=1}^{T-r}(X_s-\widehat{\mu})(X_{s+r}-\widehat{\mu})\right\}$, and the last step is from (E.21) and (E.22).

4°. Finally, we present the derivation of $q^*_{1pp}$ for $p \neq 0$,

$$
\begin{aligned}
q^*_{1pp} &= \lim_{T\to\infty} T\text{Cov}\left\{\frac{1}{T-p}\sum_{t=1}^{T-p}(X^*_t-\widehat{\mu}^*)(X^*_{t+p}-\widehat{\mu}^*), \frac{1}{T-p}\sum_{s=1}^{T-p}(X^*_s-\widehat{\mu}^*)(X^*_{s+p}-\widehat{\mu}^*)\right\} \\
&= \lim_{T\to\infty} T\text{Cov}\Bigg[\frac{1}{T-p}\sum_{t=1}^{T-p}\{\alpha_1^2(X_t-\widehat{\mu})(X_{t+p}-\widehat{\mu}) + \alpha_1(X_t-\widehat{\mu})(e_{t+p}-\bar{e}) \\
&\qquad +\alpha_1(X_{t+p}-\widehat{\mu})(e_t-\bar{e}) + (e_t-\bar{e})(e_{t+p}-\bar{e})\}, \\
&\qquad \frac{1}{T-p}\sum_{s=1}^{T-p}\alpha_1^2(X_s-\widehat{\mu})(X_{s+p}-\widehat{\mu}) + \alpha_1(X_s-\widehat{\mu})(e_{s+p}-\bar{e}) \\
&\qquad + \alpha_1(X_{s+p}-\widehat{\mu})(e_s-\bar{e}) + (e_s-\bar{e})(e_{s+p}-\bar{e})\Bigg] \\
&= \alpha_1^4 \lim_{T\to\infty} T\text{Cov}\left\{\frac{1}{T-p}\sum_{t=1}^{T-p}(X_t-\widehat{\mu})(X_{t+p}-\widehat{\mu}), \frac{1}{T-p}\sum_{s=1}^{T-p}(X_s-\widehat{\mu})(X_{s+p}-\widehat{\mu})\right\} \\
&\qquad + \lim_{T\to\infty} T\text{Cov}\left\{\frac{1}{T-p}\sum_{t=1}^{T-p}\alpha_1(X_t-\widehat{\mu})(e_{t+p}-\bar{e}), \frac{1}{T-p}\sum_{s=1}^{T-p}\alpha_1(X_s-\widehat{\mu})(e_{s+p}-\bar{e})\right\} \\
&\qquad + \lim_{T\to\infty} T\text{Cov}\left\{\frac{1}{T-p}\sum_{t=1}^{T-p}\alpha_1(X_t-\widehat{\mu})(e_{t+p}-\bar{e}), \frac{1}{T-p}\sum_{s=1}^{T-p}\alpha_1(X_{s+p}-\widehat{\mu})(e_s-\bar{e})\right\} \\
&\qquad + \lim_{T\to\infty} T\text{Cov}\left\{\frac{1}{T-p}\sum_{t=1}^{T-p}\alpha_1(X_{t+p}-\widehat{\mu})(e_t-\bar{e}), \frac{1}{T-p}\sum_{s=1}^{T-p}\alpha_1(X_s-\widehat{\mu})(e_{s+p}-\bar{e})\right\} \\
&\qquad + \lim_{T\to\infty} T\text{Cov}\left\{\frac{1}{T-p}\sum_{t=1}^{T-p}\alpha_1(X_{t+p}-\widehat{\mu})(e_t-\bar{e}), \frac{1}{T-p}\sum_{s=1}^{T-p}\alpha_1(X_{s+p}-\widehat{\mu})(e_s-\bar{e})\right\} \\
&\qquad + \lim_{T\to\infty} T\text{Cov}\left\{\frac{1}{T-p}\sum_{t=1}^{T-p}(e_t-\bar{e})(e_{t+p}-\bar{e}), \frac{1}{T-p}\sum_{s=1}^{T-p}(e_s-\bar{e})(e_{s+p}-\bar{e})\right\},
\end{aligned}
$$

where the second step is due to (E.10), the third step is because of (E.11) and a similar version to (E.18),

Because (E.23) and by the definition that

$$q_{pp} = \lim_{T\to\infty} T\text{Cov}\left\{\frac{1}{T-p}\sum_{t=1}^{T-p}(X_t-\widehat{\mu})(X_{t+p}-\widehat{\mu}), \frac{1}{T-p}\sum_{s=1}^{T-p}(X_s-\widehat{\mu})(X_{s+p}-\widehat{\mu})\right\},$$

we have that

$$
\begin{aligned}
q_{1pp}^* &= \alpha_1^4 q_{pp} + \alpha_1^2 \lim_{T\to\infty} \frac{T}{(T-p)^2} \sum_{\substack{t=1\\s=t}}^{T-p} \mathrm{Cov}\left\{(X_t-\widehat{\mu})(e_{t+p}-\bar{e}),(X_t-\widehat{\mu})(e_{t+p}-\bar{e})\right\} \\
&\quad + \alpha_1^2 \lim_{T\to\infty} \frac{T}{(T-p)^2} \sum_{\substack{t=1\\s=t+p}}^{T-2p} \mathrm{Cov}\left\{(X_t-\widehat{\mu})(e_{t+p}-\bar{e}),(X_{t+2p}-\widehat{\mu})(e_{t+p}-\bar{e})\right\} \\
&\quad + \alpha_1^2 \lim_{T\to\infty} \frac{T}{(T-p)^2} \sum_{\substack{t=1+p\\s=t-p}}^{T-p} \mathrm{Cov}\left\{(X_{t+p}-\widehat{\mu})(e_t-\bar{e}),(X_{t-p}-\widehat{\mu})(e_t-\bar{e})\right\} \\
&\quad + \alpha_1^2 \lim_{T\to\infty} \frac{T}{(T-p)^2} \sum_{\substack{t=1\\s=t}}^{T} \mathrm{Cov}\left\{(X_{t+p}-\widehat{\mu})(e_t-\bar{e}),(X_{t+p}-\widehat{\mu})(e_t-\bar{e})\right\} \\
&\quad + \alpha_1^2 \lim_{T\to\infty} \frac{T}{(T-p)^2} \mathrm{Var}\left\{(e_t-\bar{e})(e_{t+p}-\bar{e})\right\} \\
&= \alpha_1^4 q_{pp} + 2\alpha_1^2\sigma_e^2(\gamma_0+\gamma_{2p}) + \sigma_e^4, \qquad \text{(E.26)}
\end{aligned}
$$

where the last step is because of (E.24), and (E.21) and (E.22) with $q = p$.

## E.5 The proof of Theorem 6.4

**Proof of Theorem 6.4(1):**

For $k = 1, \ldots, p$, applying the weak law of large numbers to $\widehat{\gamma}_k^*$, we obtain that as $T \to \infty$, the estimator $\widehat{\gamma}_k^*$ converges in probability to $\mathrm{Cov}(X_t^*, X_{t-k}^*)$, which is denoted as $\gamma_k^*$.

Next, we examine $\gamma_k$. By the form of measurement error model (6.9), we have that for $0 < k < t$,

$$
\begin{aligned}
&\mathrm{Cov}(X_t^*, X_{t-k}^*) \\
&= \mathrm{Cov}(\beta_0 X_t u_t, \beta_0 X_{t-k} u_{t-k}) \\
&= \beta_0^2\{E(X_t u_t X_{t-k} u_{t-k}) - E(X_t u_t)E(X_{t-k} u_{t-k})\} \\
&= \beta_0^2\{E(u_t)E(u_{t-k})\mathrm{Cov}(X_t, X_{t-k})\} \\
&= \beta_0^2\{\mathrm{Cov}(X_t, X_{t-k})\} = \beta_0^2\gamma_k,
\end{aligned}
$$

and by (6.10), $\mathrm{Var}(X_t^*) = \beta_0^2 \{(\sigma_u^2+1)\gamma_0 + \sigma_u^2\mu^2\}$, which is denoted as $\gamma_0^*$. Thus, Theorem 6.4(1) follows.

**Proof of Theorem 6.4(2)**:

First, by Theorem 6.4(1), we write

$$\widehat{\gamma}^* = \beta_0^2\gamma + o_p(1)$$

and

$$\begin{aligned}\widehat{\Gamma}^* &= \begin{pmatrix} \beta_0^2(\sigma_u^2+1)\gamma_0 + \beta_0\sigma_u^2\mu^2 & \beta_0^2\gamma_1 & \cdots & \beta_0^2\gamma_{p-1} \\ \vdots & & \ddots & \vdots \\ \beta_0^2\gamma_{p-1} & \beta_0^2\gamma_{p-2} & \cdots & \beta_0^2(\sigma_u^2+1)\gamma_0 + \beta_0\sigma_u^2\mu^2 \end{pmatrix} + o_p(1) \\ &= \beta_0^2\left\{\Gamma + \sigma_u^2(\gamma_0+\mu^2)I_p\right\} + o_p(1),\end{aligned}$$

where $\widehat{\Gamma}^* = \begin{pmatrix} \widehat{\gamma}_0^* & \cdots & \widehat{\gamma}_{p-1}^* \\ \vdots & \ddots & \vdots \\ \widehat{\gamma}_{p-1}^* & \cdots & \widehat{\gamma}_0^* \end{pmatrix}$. Then the naive estimator $\widehat{\phi}^*$ is obtained by replacing $\widehat{\gamma}_k$ in (6.6) with $\widehat{\gamma}_k^*$,

$$\widehat{\phi}^* = [\beta_0^2\left\{\Gamma + \sigma_u^2(\gamma_0+\mu^2)I_p\right\} + o_p(1)]^{-1}\{\beta_0^2\gamma + o_p(1)\} = \left\{\Gamma + \sigma_u^2(\gamma_0+\mu^2)I_p\right\}^{-1}\gamma + o_p(1), \tag{E.27}$$

and hence $\phi^* = \{\Gamma + \sigma_u^2(\gamma_0+\mu^2)I_p\}^{-1}\gamma$ such that $\widehat{\phi}^* \xrightarrow{p} \phi^*$ as $T \to \infty$.

Again, by replacing $\widehat{\gamma}_k$ in (6.6) with $\widehat{\gamma}_k^*$ gives the naive estimator $\widehat{\phi}_0^*$

$$\begin{aligned}\widehat{\phi}_0^* &= \frac{1}{T-p}\sum_{t=p}^{T} X_t^* - \left(\sum_{k=1}^{p}\widehat{\phi}_k^*\right)\left(\frac{1}{T-p}\sum_{t=p}^{T} X_{t-k}^*\right) \\ &= E(X_t^*) - E(X_t^*)\sum_{k=1}^{p}\widehat{\phi}_k^* + o_p(1) \\ &= \beta_0 E(X_t) - \beta_0 E(X_t)\sum_{k=1}^{p}\{\phi_k^* + o_p(1)\} + o_p(1), \\ &= \beta_0(1 - \phi^{*\mathrm{T}}\cdot \mathbb{1}_p)\mu + o_p(1),\end{aligned}$$

where $\widehat{\phi}_k$ and $\phi_k$ are respectively the $k$th element of $\widehat{\phi}$ and $\phi$, the third step is because $\widehat{\phi}_k = \phi_k + o_p(1)$ by (E.27) as well as the model form (6.9), and the last step is due to the stationarity of the time series $\{X_t\}$ such that $E(X_t) = \mu$.

Finally, noting that the native estimator $\widehat{\sigma}_\epsilon^{*2}$ is given by $\widehat{\sigma}_\epsilon^{*2} = \widehat{\gamma}_0^* - 2\widehat{\phi}^{*\mathrm{T}}\widehat{\gamma}^* + \widehat{\phi}^{*\mathrm{T}}\widehat{\Gamma}^*\widehat{\phi}^*$ by applying a version similar to (6.6), we obtain that

$$
\begin{aligned}
\widehat{\sigma}_\epsilon^{*2} =& \widehat{\gamma}_0^* - 2\widehat{\phi}^{*\mathrm{T}}\widehat{\gamma}^* + \widehat{\phi}^{*\mathrm{T}}\widehat{\Gamma}^*\widehat{\phi}^* \\
=& \beta_0^2\left\{(\sigma_u^2+1)\gamma_0 + \sigma_u^2\mu^2\right\} - 2\beta_0^2\gamma^{\mathrm{T}}\{\Gamma + \sigma_u^2(\gamma_0+\mu^2)I\}^{-1}\gamma \\
&+ \beta_0^2\gamma^{\mathrm{T}}\{\Gamma + \sigma_u^2(\gamma_0+\mu^2)I\}^{-1}\{\Gamma + \sigma_u^2(\gamma_0+\mu^2)I\}\{\Gamma + \sigma_u^2(\gamma_0+\mu^2)I\}^{-1}\gamma + o_p(1) \\
=& \beta_0^2\left\{(\sigma_u^2+1)\gamma_0 + \sigma_u^2\mu^2\right\} - \beta_0^2\gamma^{\mathrm{T}}\{\Gamma + \sigma_u^2(\gamma_0+\mu^2)I\}^{-1}\gamma + o_p(1).
\end{aligned}
$$

**Proof of Theorem 6.4(3)**:

**Step 1**: We show that as $T \to \infty$,

$$
\sqrt{T}\left(\frac{1}{T}\sum_{t=1}^{T}(X_t^*-\mu^*)(X_{t+p}^*-\mu^*) - \frac{1}{T-p}\sum_{t=1}^{T-p}(X_t^*-\widehat{\mu}^*)(X_{t+p}^*-\widehat{\mu}^*)\right) = o_p(1). \quad \text{(E.28)}
$$

With some simple algebra,

$$
\begin{aligned}
&\sqrt{T}\left\{\frac{1}{T}\sum_{t=1}^{T}(X_t^*-\mu^*)(X_{t+p}^*-\mu^*) - \frac{1}{T-p}\sum_{t=1}^{T-p}(X_t^*-\widehat{\mu}^*)(X_{t+p}^*-\widehat{\mu}^*)\right\} \\
&= \sqrt{T}\left\{\frac{1}{T}\sum_{t=1}^{T}(X_t^*-\mu^*)(X_{t+p}^*-\mu^*) - \frac{1}{T-p}\sum_{t=1}^{T-p}(X_t^*-\mu^*+\mu^*-\widehat{\mu}^*)(X_{t+p}^*-\mu^*+\mu^*-\widehat{\mu}^*)\right\} \\
&= \sqrt{T}\left\{\frac{1}{T}\sum_{t=1}^{T}(X_t^*-\mu^*)(X_{t+p}^*-\mu^*) - \frac{1}{T-p}\sum_{t=1}^{T-p}(X_t^*-\mu^*)(X_{t+p}^*-\mu^*)\right. \\
&\qquad \left. -\frac{1}{T-p}\sum_{t=1}^{T-p}(X_t^*-\mu^*)(\mu^*-\widehat{\mu}^*) - \frac{1}{T-p}\sum_{t=1}^{T-p}(X_{t+p}^*-\mu^*)(\mu^*-\widehat{\mu}^*) - \frac{1}{T-p}\sum_{t=1}^{T-p}(\mu^*-\widehat{\mu}^*)^2\right\} \\
&= \sqrt{T}\left(\frac{T-p}{T}-1\right)\frac{1}{T-p}\sum_{t=1}^{T-p}(X_t^*-\mu^*)(X_{t+p}^*-\mu^*) + \frac{1}{\sqrt{T}}\sum_{t=T-p+1}^{T}(X_t^*-\mu^*)(X_{t+p}^*-\mu^*) \\
&\qquad + \sqrt{T}(\widehat{\mu}^*-\mu^*)\left(\frac{1}{T-p}\sum_{t=1}^{T-p}X_t^* + \frac{1}{T-p}\sum_{t=1}^{T-p}X_{t+p}^* - \widehat{\mu}^* - \mu^*\right) \qquad \text{(E.29)} \\
&\triangleq I_1 + I_2 + I_3.
\end{aligned}
$$

Now we examine each term in (E.29) as $T \to \infty$ separately. First,

$$\begin{aligned} I_1 &= -\frac{p}{\sqrt{T}} \frac{1}{T-p} \sum_{t=1}^{T-p} (X_t^* - \mu^*)(X_{t+p}^* - \mu^*) \\ &= -\frac{p}{\sqrt{T}} \{\gamma_p^* + o_p(1)\} = o_p(1) \qquad \text{as} \quad T \to \infty. \end{aligned} \tag{E.30}$$

Next, we examine the second term $I_2$ in (E.29). Since $T^{-\frac{1}{2}} E[\sum_{t=T-p+1}^{T} (X_t^* - \mu^*)(X_{t+p}^* - \mu^*)] \le T^{-\frac{1}{2}} p \mathrm{Var}(X_t)$ (Brockwell et al., 1991, Page 230) and $T^{-\frac{1}{2}} p \mathrm{Var}(X_t) \to 0$ as $T \to \infty$, we have that

$$I_2 = \frac{1}{\sqrt{T}} \sum_{t=T-p+1}^{T} (X_t^* - \mu^*)(X_{t+p}^* - \mu^*) = o_p(1). \tag{E.31}$$

Finally, we examine $I_3$ in (E.29).

$$\begin{aligned} &\frac{1}{T-p} \sum_{t=1}^{T-p} X_{t+p}^* - \widehat{\mu}^* \\ =&\frac{1}{T-p} \sum_{t=1}^{T-p} X_{t+p}^* - \frac{1}{T} \sum_{t=1}^{p} X_t^* - \frac{1}{T} \sum_{t=p+1}^{T} X_t^* \\ =&\frac{1}{T-p} \sum_{t=1}^{T-p} X_{t+p}^* - \frac{1}{T} \sum_{t=1}^{p} X_t^* - \frac{1}{T} \sum_{t=1}^{T-p} X_{t+p}^* \\ =&(\frac{1}{T-p} - \frac{1}{T}) \sum_{t=1}^{T-p} X_{t+p}^* - \frac{1}{T} \sum_{t=1}^{p} X_t^* \\ =&o_p(1) \qquad \text{as} \quad T \to \infty, \end{aligned} \tag{E.32}$$

where $\widehat{\mu}^* = \frac{1}{T} \sum_{t=1}^{T} X_t^*$, and $\frac{1}{T} \sum_{t=1}^{p} X_t^* = o_p(1)$ because $E(\frac{1}{T} \sum_{t=1}^{p} X_t^*) = \frac{1}{T} p E(X_t) \to 0$ as $T \to \infty$. In addition, by the weak law of large numbers,

$$\frac{1}{T-p} \sum_{t=1}^{T-p} X_t^* - \mu^* \xrightarrow{p} 0 \qquad \text{as} \quad T \to \infty. \tag{E.33}$$

By condition (R2) and the central limit theorem for strictly stationary $p$-dependent sequences (Brockwell et al., 1991, Theorem 6.4.2), we have

$$\sqrt{T}(\widehat{\mu}^* - \mu^*) = O_p(1). \tag{E.34}$$

Therefore, applying (E.30), (E.31), (E.32), (E.33) and (E.34) yields (E.28).

**Step 2**: We now show that as $T \to \infty$, the asymptotic covariance matrix of $\sqrt{T}\left\{(\widehat{\gamma}_0^*, \widehat{\gamma}^{*\mathrm{T}})^{\mathrm{T}} - (\gamma_0^*, \gamma^{*\mathrm{T}})^{\mathrm{T}}\right\}$ equals

$$\lim_{T\to\infty} \mathrm{Cov}\left\{\frac{1}{\sqrt{T}}\sum_{t=1}^{T}(X_t^* - \mu^*)(X_{t+r}^* - \mu^*), \frac{1}{\sqrt{T}}\sum_{s=1}^{T}(X_s^* - \mu^*)(X_{s+q}^* - \mu^*)\right\}.$$

For $k \leq p$

$$\begin{aligned}
&\sqrt{T}(\widehat{\gamma}_k - \gamma_k)\\
=&\sqrt{T}\left\{\frac{1}{T-k}\sum_{t=1}^{T-k}(X_t^* - \widehat{\mu}^*)(X_{t+k}^* - \widehat{\mu}^*) - \gamma_k\right\}\\
=&\sqrt{T}\left\{\frac{1}{T}\sum_{t=1}^{T}(X_t^* - \mu^*)(X_{t+k}^* - \mu^*) - \gamma_k\right\}\\
&+\sqrt{T}\left\{\frac{1}{T-k}\sum_{t=1}^{T-k}(X_t^* - \widehat{\mu}^*)(X_{t+k}^* - \widehat{\mu}^*) - \frac{1}{T}\sum_{t=1}^{T}(X_t^* - \mu^*)(X_{t+k}^* - \mu^*)\right\}\\
=&\left\{\frac{1}{\sqrt{T}}\sum_{t=1}^{T}(X_t^* - \mu^*)(X_{t+k}^* - \mu^*) - \gamma_k\right\} + o_p(1),
\end{aligned}$$

where the last step is due to (E.28).

Hence, the $(r, q)$ element of matrix $\lim_{T\to\infty} \mathrm{Var}\left(\sqrt{T}\left\{(\widehat{\gamma}_0^*, \widehat{\gamma}^{*\mathrm{T}})^{\mathrm{T}} - (\gamma_0^*, \gamma^{*\mathrm{T}})^{\mathrm{T}}\right\}\right)$ is given by

$$\lim_{T\to\infty} \mathrm{Cov}\left\{\frac{1}{\sqrt{T}}\sum_{t=1}^{T}(X_t^* - \mu^*)(X_{t+r}^* - \mu^*), \frac{1}{\sqrt{T}}\sum_{s=1}^{T}(X_s^* - \mu^*)(X_{s+q}^* - \mu^*)\right\}.$$

**Step 3**: We show certain identities to be used for proving Theorem 6.4(3):

1. By model (6.9), we have that

$$\begin{aligned}
X_t^* - \mu^* &= \beta_0 X_t u_t - \beta_0 \mu\\
&= \beta_0 X_t u_t - \beta_0 u_t \mu + \beta_0 u_t \mu - \beta_0 \mu\\
&= \beta_0\{u_t(X_t - \mu) + \mu(u_t - 1)\}, \qquad \text{(E.35)}
\end{aligned}$$

where the first step is because $\mu^* = E(\beta_0 X_t u_t) = \beta_0 E(X_t)E(u_t) = \beta_0\mu$.

2. We have that

$$\begin{aligned}
&\lim_{T\to\infty}\frac{1}{T}\sum_{t=1}^{T}\sum_{s=1}^{T}\mathrm{Cov}\left\{u_t^2(X_t-\mu)^2, u_s^2(X_s-\mu)^2\right\}\\
=&\lim_{T\to\infty}\frac{1}{T}\sum_{t=1}^{T}\sum_{s=1}^{T}\left[E\{u_t^2u_s^2(X_t-\mu)^2(X_s-\mu)^2\} - E(u_t^2)E(u_s^2)E\{(X_t-\mu)^2\}E\{(X_s-\mu)^2\}\right],\\
=&\lim_{T\to\infty}\frac{1}{T}\sum_{t=1}^{T}\sum_{\substack{s=1\\ s\neq t}}^{T}\left[E(u_t^2u_s^2)E\{(X_t-\mu)^2(X_s-\mu)^2\} - E(u_t^2)E(u_s^2)E\{(X_t-\mu)^2\}E\{(X_s-\mu)^2\}\right]\\
&\quad+\lim_{T\to\infty}\frac{1}{T}\sum_{\substack{t=1\\ s=t}}^{T}\left[E(u_t^4)E\{(X_t-\mu)^4\} - E^2(u_t^2)E^2\{(X_t-\mu)^2\}\right],\\
=&\lim_{T\to\infty}\frac{1}{T}\sum_{t=1}^{T}\sum_{\substack{s=1\\ s\neq t}}^{T}\left[E(u_t^2)E(u_s^2)\mathrm{Cov}\{(X_t-\mu)^2,(X_s-\mu)^2\}\right] + \lim_{T\to\infty}\frac{1}{T}\sum_{\substack{t=1\\ s=t}}^{T}E^2(u_t^2)\mathrm{Var}\{(X_t-\mu)^2\}\\
&\quad+\lim_{T\to\infty}\frac{1}{T}\sum_{\substack{t=1\\ s=t}}^{T}\left\{E(u_t^4)-E^2(u_t^2)\right\}E\{(X_t-\mu)^4\}\\
=&\lim_{T\to\infty}\frac{1}{T}\sum_{t=1}^{T}\sum_{s=1}^{T}\left[E(u_t^2)E(u_s^2)\mathrm{Cov}\{(X_t-\mu)^2,(X_s-\mu)^2\}\right] \qquad \text{(E.36)}\\
&\quad+\lim_{T\to\infty}\frac{1}{T}\sum_{t=1}^{T}\left\{E(u_t^4)-E^2(u_t^2)\right\}E\{(X_t-\mu)^4\}\\
=&(\sigma_u^2+1)^2 q_{00} + \{E(u_t^4) - (\sigma_u^2+1)^2\}E\{(X_t-\mu)^4\}, \qquad \text{(E.37)}
\end{aligned}$$

where the second and third step is due to the independence between $u_t$ and $X_t$. In the last step, we use the definition $q_{00} = \lim_{T\to\infty}\frac{1}{T}\sum_{t=1}^{T}\sum_{s=1}^{T}\mathrm{Cov}\left\{(X_t-\mu)^2,(X_s-\mu)^2\right\}$, $E(u_t^2) = \sigma_u^2+1$, and the fact that $E(u_t^4)$ and $E\{(X_t-\mu)^4\}$ are time-independent which are derived from Conditions (R1) and (R2) together with independence between $u_t$ and $X_t$.

3. Similar to the derivation in (E.36), now we derive the summation of $\text{Cov}\{\beta_0^2 u_t^2 (X_t - \mu)^2, \beta_0^2 u_s u_{s+p}(X_s - \mu)(X_{s+p} - \mu)\}$ for $p > 0$,

$$
\begin{aligned}
&\lim_{T\to\infty} \frac{1}{T} \sum_{t=1}^{T} \sum_{s=1}^{T} \text{Cov}\{\beta_0^2 u_t^2 (X_t - \mu)^2, \beta_0^2 u_s u_{s+p}(X_s - \mu)(X_{s+p} - \mu)\} \\
&= \lim_{T\to\infty} \frac{\beta_0^4}{T} \sum_{t=1}^{T} \sum_{s=1}^{T} \left[ E(u_t^2 u_s u_{s+p}) E\{(X_t - \mu)^2 (X_s - \mu)(X_{s+p} - \mu)\} \right. \\
&\left. - E(u_t^2) E(u_s) E(u_{s+p}) E(X_t - \mu)^2 E\{(X_s - \mu)(X_{s+p} - \mu)\} \right] \\
&= \lim_{T\to\infty} \frac{\beta_0^4}{T} \sum_{t=1}^{T} \sum_{s=1}^{T} E(u_t^2) E(u_s) E(u_{s+p}) \text{Cov}\{(X_t - \mu)^2, (X_s - \mu)(X_{s+p} - \mu)\} \\
&\quad + \lim_{T\to\infty} \frac{\beta_0^4}{T} \sum_{\substack{t=1 \\ s=t}}^{T} \left\{ E(u_t^3) E(u_{t+p}) - E(u_t^2) E(u_t) E(u_{t+p}) \right\} E\{(X_t - \mu)^3 (X_{t+p} - \mu)\} \\
&\quad + \lim_{T\to\infty} \frac{\beta_0^4}{T} \sum_{\substack{t=1 \\ s=t-p}}^{T} \left\{ E(u_t^3) E(u_{t-p}) - E(u_t^2) E(u_t) E(u_{t-p}) \right\} E\{(X_t - \mu)^3 (X_{t-p} - \mu)\} \\
&= \lim_{T\to\infty} \frac{\beta_0^4}{T} \sum_{t=1}^{T} \sum_{s=1}^{T} (\sigma_u^2 + 1) \text{Cov}\{(X_t - \mu)^2, (X_s - \mu)(X_{s+p} - \mu)\} \\
&\quad + \beta_0^4 \left\{ E(u_t^3) - E(u_t^2) \right\} E\{(X_t - \mu)^3 (X_{t+p} - \mu)\} \\
&\quad + \beta_0^4 \left\{ E(u_t^3) - E(u_t^2) \right\} E\{(X_t - \mu)^3 (X_{t-p} - \mu)\}, \\
&= \beta_0^4 \left\{ E(u_t^3) - (\sigma_u^2 + 1) \right\} \left[ E\{(X_t - \mu)^3 (X_{t+p} - \mu)\} + E\{(X_t - \mu)^3 (X_{t-p} - \mu)\} \right] \\
&\quad + \beta_0^4 q_{0p} (\sigma_u^2 + 1),
\end{aligned} \tag{E.38}
$$

where the first step is because $X_t$ and $u_t$ are independent, and the second last step is due to $E(u_t^2) = Var(u_t) + E(u_t^2) = \sigma_u^2 + 1$ and is derived similar to the second and third step in (E.36), and the last step is because of the definition that $q_{0p} = \lim_{T\to\infty} \frac{1}{T} \sum_{t=1}^{T} \sum_{s=1}^{T} \text{Cov}\{(X_t - \mu)^2, (X_s - \mu)(X_{s+p} - \mu)\}$ and the fact that $E\{(X_t - \mu)^3 (X_{t+p} - \mu)\}$, $E\{(X_t - \mu)^3 (X_{t-p} - \mu)\}$ and $E(u_t^3)$ are time-independent, derived from Conditions (R1) and (R2) together with the independence between $u_t$ and $X_t$.

4. Analogous to the derivation in (E.36) and (E.38), we derive the summation of $\text{Cov}\{u_t u_{t+p}(X_t - \mu)(X_{t+p} - \mu), u_s u_{s+r}(X_s - \mu)(X_{s+r} - \mu)\}$ for $p > 0$, $r > 0$ and $p \neq r$,

$$
\begin{aligned}
&\beta_0^4 \lim_{T\to\infty} \frac{1}{T} \sum_{t=1}^{T} \sum_{s=1}^{T} \text{Cov}\{u_t u_{t+p}(X_t - \mu)(X_{t+p} - \mu), u_s u_{s+r}(X_s - \mu)(X_{s+r} - \mu)\} \\
=&\beta_0^4 \lim_{T\to\infty} \frac{1}{T} \sum_{t=1}^{T} \sum_{s=1}^{T} E(u_t u_{t+p} u_s u_{s+r}) \text{Cov}\{(X_t - \mu)(X_{t+p} - \mu), (X_s - \mu)(X_{s+r} - \mu)\} \\
&\quad + \beta_0^4 \lim_{T\to\infty} \frac{1}{T} \sum_{\substack{t=1 \\ s=t}}^{T} \left\{E(u_t^2)E(u_{t+p})E(u_{t+r}) - 1\right\} E\{(X_t - \mu)^2 (X_{t+p} - \mu)(X_{t+r} - \mu)\} \\
&\quad + \beta_0^4 \lim_{T\to\infty} \frac{1}{T} \sum_{\substack{t=1 \\ s=t+p}}^{T} \left\{E(u_{t+p}^2)E(u_t)E(u_{t+p+r}) - 1\right\} E\{(X_t - \mu)(X_{t+p} - \mu)^2 (X_{t+p+r} - \mu)\} \\
&\quad + \beta_0^4 \lim_{T\to\infty} \frac{1}{T} \sum_{\substack{t=1 \\ s=t-r}}^{T} \left\{E(u_t^2)E(u_{t+p})E(u_{t-r}) - 1\right\} E\{(X_{t-r} - \mu)(X_t - \mu)^2 (X_{t+p} - \mu)\} \\
&\quad + \beta_0^4 \lim_{T\to\infty} \frac{1}{T} \sum_{\substack{t=1 \\ s=t+p-r}}^{T} \left\{E(u_{t+p}^2)E(u_t)E(u_{t+p-r}) - 1\right\} E\{(X_t - \mu)(X_{t+p-r} - \mu)X_{t+p} - \mu)^2\} \\
=&\beta_0^4 q_{pr} + \beta_0^4 \sigma_u^2 E\{(X_t - \mu)^2 (X_{t+p} - \mu)(X_{t+r} - \mu)\} \\
&\quad + \beta_0^4 \sigma_u^2 E\{(X_t - \mu)(X_{t+p} - \mu)^2 (X_{t+p+r} - \mu)\} \\
&\quad + \beta_0^4 \sigma_u^2 E\{(X_{t-r} - \mu)(X_t - \mu)^2 (X_{t+p} - \mu)\} \\
&\quad + \beta_0^4 \sigma_u^2 E\{(X_t - \mu)(X_{t+p-r} - \mu)(X_{t+p} - \mu)^2\},
\end{aligned}
\tag{E.39}
$$

where the third step is derived analogously to the second step of (E.38), and $E(u_t u_{t+p} u_s u_{s+r}) = 1$, and the last step is due to the definition $q_{pr} = \lim_{T\to\infty} \frac{1}{T} \sum_{t=1}^{T} \sum_{s=1}^{T} \text{Cov}\{(X_t - \mu)(X_{t+p} - \mu), (X_s - \mu)(X_{s+r} - \mu)\}$ and the fact that $E\{(X_t - \mu)^2 (X_{t+p} - \mu)(X_{t+r} - \mu)\}$, $E\{(X_t - \mu)(X_{t+p} - \mu)^2 (X_{t+p+r} - \mu)\}$, $E\{(X_{t-r} - \mu)(X_t - \mu)^2 (X_{t+p} - \mu)\}$, and $E\{(X_t - \mu)(X_{t+p} - \mu)^2 (X_{t+2p} - \mu)\}$ are time-independent derived from Conditions (R1) and (R2).

5. Similar to the derivation in (E.36), (E.38), and (E.39), we derive the summation of $\mathrm{Cov}\{u_t u_{t+p}(X_t-\mu)(X_{t+p}-\mu), u_s u_{s+p}(X_s-\mu)(X_{s+p}-\mu)\}$ for $p>0$,

$$
\begin{aligned}
&\beta_0^4 \lim_{T\to\infty}\frac{1}{T}\sum_{t=1}^{T}\sum_{s=1}^{T}\mathrm{Cov}\{u_t u_{t+p}(X_t-\mu)(X_{t+p}-\mu), u_s u_{s+p}(X_s-\mu)(X_{s+p}-\mu)\}\\
&=\beta_0^4 \lim_{T\to\infty}\frac{1}{T}\sum_{t=1}^{T}\sum_{s=1}^{T}E(u_t)E(u_{t+p})E(u_s)E(u_{s+p})\mathrm{Cov}\{(X_t-\mu)(X_{t+p}-\mu),(X_s-\mu)(X_{s+p}-\mu)\\
&\quad+\beta_0^4 \lim_{T\to\infty}\frac{1}{T}\sum_{\substack{t=1\\ s=t}}^{T}\left\{E(u_t^2)E(u_{t+p}^2)-1\right\}\mathrm{Var}\{(X_t-\mu)(X_{t+p}-\mu)\}\\
&\quad+\beta_0^4 \lim_{T\to\infty}\frac{1}{T}\sum_{\substack{t=1\\ s=t+p}}^{T}\left\{E(u_{t+p}^2)E(u_t)E(u_{t+2p})-1\right\}E\{(X_t-\mu)(X_{t+p}-\mu)^2(X_{t+2p}-\mu)\}\\
&\quad+\beta_0^4 \lim_{T\to\infty}\frac{1}{T}\sum_{\substack{t=1\\ s=t-p}}^{T}\left\{E(u_t^2)E(u_{t-p})E(u_{t+p})-1\right\}E\{(X_{t-p}-\mu)(X_t-\mu)^2(X_{t+p}-\mu)\}\\
&=\beta_0^4 q_{pp}+\beta_0^4(\sigma_u^4+2\sigma_u^2)\mathrm{Var}\{(X_t-\mu)(X_{t+p}-\mu)\}+2\beta_0^4 E\{(X_t-\mu)(X_{t+p}-\mu)^2(X_{t+2p}-\mu)\},
\end{aligned}
\tag{E.40}
$$

where the last step is by the definition $q_{pp}=\lim_{T\to\infty}\frac{1}{T}\sum_{t=1}^{T}\sum_{s=1}^{T}\mathrm{Cov}\{(X_t-\mu)(X_{t+p}-\mu),(X_s-\mu)(X_{s+p}-\mu)\}$ and $E\{(X_t-\mu)(X_{t+p}-\mu)^2(X_{t+2p}-\mu)\}=E\{(X_{t-p}-\mu)(X_t-\mu)^2(X_{t+p}-\mu)\}$ due to the stationarity of the time series and the fact that $\mathrm{Var}\{(X_t-\mu)(X_{t+p}-\mu)\}$ and $E\{(X_t-\mu)(X_{t+p}-\mu)^2(X_{t+2p}-\mu)\}$ are time-independent, resulted from the Conditions (R1) and (R2).

6. For any $t$, $s$ and $p$, we have that

$$
\begin{aligned}
&\mathrm{Cov}\{(X_t-\mu)(X_{t-p}-\mu),(X_s-\mu)\}\\
=&E\{(X_t-\mu)(X_{t-p}-\mu)(X_s-\mu)\}-E\{(X_t-\mu)(X_{t-p}-\mu)\}E(X_s-\mu)\\
=&E\{(X_t-\mu)(X_{t-p}-\mu)(X_s-\mu)\},
\end{aligned}
\tag{E.41}
$$

where the last step is because $E(X_s-\mu)=0$.

7. For any $t$ and $s$, we have that

$$
\begin{aligned}
&\mathrm{Cov}\{u_t(u_t-1)(X_t-\mu), u_s(u_s-1)(X_s-\mu)\} \\
=&E\{u_t(u_t-1)(X_t-\mu)u_s(u_s-1)(X_s-\mu)\} - E\{u_t(u_t-1)(X_t-\mu)\}E\{u_s(u_s-1)(X_s-\mu)\} \\
=&E\{u_t(u_t-1)(X_t-\mu)u_s(u_s-1)(X_s-\mu)\} \\
=&E\{u_t(u_t-1)u_s(u_s-1)\}E\{(X_t-\mu)(X_s-\mu)\}, \qquad \text{(E.42)}
\end{aligned}
$$

where the second step is because of the independence between $u_t$ and $X_t$ and that $E(X_t - \mu) = 0$. Then, $E\{u_t(u_t-1)u_s(u_s-1)\} = \sigma_u^4$ for $t \neq s$ and $E\{u_t^2(u_t-1)^2\} = E(u_t^4) - 2E(u_t^3) + \sigma_u^2 + 1$ for any $t$.

By (E.42), we have that

$$
\begin{aligned}
&\lim_{T\to\infty} \frac{1}{T} \sum_{t=1}^{T} \sum_{s=1}^{T} \mathrm{Cov}\left\{u_t(u_t-1)(X_t-\mu), u_s(u_s-1)(X_s-\mu)\right\} \\
=&\lim_{T\to\infty} \frac{1}{T} \sum_{t=1}^{T} \sum_{s=1}^{T} E\{u_t(u_t-1)u_s(u_s-1)\}E\{(X_t-\mu)(X_s-\mu)\} \\
=&\lim_{T\to\infty} \frac{1}{T} \sum_{t=1}^{T} \sum_{s=1}^{T} \sigma_u^4 E\{(X_t-\mu)(X_s-\mu)\} \\
&\quad + \lim_{T\to\infty} \frac{1}{T} \sum_{\substack{t=1 \\ s=t}}^{T} \left\{E(u_t^4) - 2E(u_t^3) + \sigma_u^2 + 1 - \sigma_u^4\right\} E\{(X_t-\mu)^2\} \\
=&\sigma_u^4 \sum_{h=-\infty}^{\infty} \gamma_h + \left\{E(u_t^4) - 2E(u_t^3) + \sigma_u^2 + 1 - \sigma_u^4\right\} \gamma_0, \qquad \text{(E.43)}
\end{aligned}
$$

where the last is because $\lim_{T\to\infty} \frac{1}{T} \sum_{t=1}^{T} \sum_{s=1}^{T} E\{(X_t-\mu)(X_s-\mu)\} = \sum_{h=-\infty}^{\infty} \gamma_h$ (Brockwell et al., 1991, Theorem 7.1.1).

8. For any $t$, $s$ and $p > 0$, we have that

$$
\begin{aligned}
&\mathrm{Cov}\{u_t(u_t-1)(X_t-\mu), u_{s+p}(u_s-1)(X_{s+p}-\mu)\} \\
=&E\{u_t(u_t-1)(X_t-\mu)u_{s+p}(u_s-1)(X_{s+p}-\mu)\} \\
&\quad - E\{u_t(u_t-1)(X_t-\mu)\}E\{u_{s+p}(u_s-1)(X_{s+p}-\mu)\} \\
=&E\{u_t(u_t-1)u_{s+p}(u_s-1)\}E\{(X_t-\mu)(X_{s+p}-\mu)\} \\
=&E\{u_t(u_t-1)u_{s+p}(u_s-1)\}\gamma_{|s+p-t|}, \qquad \text{(E.44)}
\end{aligned}
$$

where the second step is because of the independence between $u_t$ and $X_t$ and that $E(X_t - \mu) = 0$. Then, $E\{u_t(u_t-1)u_{s+p}(u_s-1)\} = 0$ for $t \neq s$ and $E\{u_t(u_t-1)^2 u_{t+p}\} = E\{u_t(u_t-1)^2\} = E\{(u_t-1)^3\} + \sigma_u^2$ for any $s = t$.

9. By independence of $u_t$ and $u_s$, for $t \neq s$, we have that

$$\mathrm{Cov}\{u_t^2(X_t-\mu)^2, (u_s-1)^2\} = 0, \tag{E.45}$$

and for any $t$,

$$\begin{aligned}
&\mathrm{Cov}\{u_t^2(X_t-\mu)^2, (u_t-1)^2\} \\
=&E\{u_t^2(u_t-1)^2(X_t-\mu)^2\} - E\{u_t^2(X_t-\mu)^2\}E\{(u_t-1)^2\} \\
=&\left[E\{u_t^2(u_t-1)^2\} - E(u_t^2)E(u_t-1)^2\right] E\{(X_t-\mu)^2\} \\
=&\left\{E(u_t^4) - 2E(u_t^3) + \sigma_u^2 + 1 - \sigma_u^4 - \sigma_u^2\right\}\gamma_0 \\
=&\left\{E(u_t^4) - 2E(u_t^3) + 1 - \sigma_u^4\right\}\gamma_0.
\end{aligned} \tag{E.46}$$

10. By independence of $u_t$ and $u_s$, for $s \neq t$, $s \neq t+p$ and any $p$, we have that

$$\mathrm{Cov}\{u_t u_{t+p}(X_t-\mu)(X_{t+p}-\mu), (u_s-1)^2\} = 0. \tag{E.47}$$

For any $t$ and $p > 0$,

$$\begin{aligned}
&\mathrm{Cov}\{u_t u_{t+p}(X_t-\mu)(X_{t+p}-\mu), (u_t-1)^2\} \\
=&E\{u_t u_{t+p}(u_t-1)^2(X_t-\mu)(X_{t+p}-\mu)\} - E\{u_t u_{t+p}(X_t-\mu)(X_{t+p}-\mu)\}E\{(u_t-1)^2\}, \\
=&\left[E\{u_t u_{t+p}(u_t-1)^2\} - E(u_t u_{t+p})E\{(u_t-1)^2\}\right] E\{(X_t-\mu)(X_{t+p}-\mu)\} \\
=&E\left\{(u_t-1)^3\right\}\gamma_p,
\end{aligned} \tag{E.48}$$

and

$$\mathrm{Cov}\{u_t u_{t-p}(X_t-\mu)(X_{t-p}-\mu), (u_t-1)^2\} = E\left\{(u_t-1)^3\right\}\gamma_p.$$

11. For any $t$ and $s$, and $r \neq p$ and $r > 0$, we have that

$$\mathrm{Cov}\{u_t u_{t+p}(X_t-\mu)(X_{t+p}-\mu), (u_s-1)(u_{s+r}-1)\} = 0. \tag{E.49}$$

By independence of $u_t$ and $u_s$, for $t \neq s$ and any $p$, we have that

$$\mathrm{Cov}\{u_t u_{t+p}(X_t-\mu)(X_{t+p}-\mu), (u_s-1)(u_{s+p}-1)\} = 0, \tag{E.50}$$

and for any $t$ and $p > 0$,

$$\begin{aligned}
&\text{Cov}\{u_t u_{t+p}(X_t-\mu)(X_{t+p}-\mu),(u_t-1)(u_{t+p}-1)\} \\
=&E\{u_t u_{t+p}(u_t-1)(t_{t+p}-1)(X_t-\mu)(X_{t+p}-\mu)\} \qquad \text{(E.51)}\\
&-E\{u_t u_{t+p}(X_t-\mu)(X_{t+p}-\mu)\}E\{(u_t-1)(u_{t+p}-1)\} \\
=&E\{u_t(u_t-1)\}E\{u_{t+p}(u_{t+p}-1)\}E\{(X_t-\mu)(X_{t+p}-\mu)\} \\
=&\sigma_u^4\gamma_p. \qquad \text{(E.52)}
\end{aligned}$$

12. For any $t$, we have that

$$\begin{aligned}
&\text{Cov}\{u_t(u_t-1)(X_t-\mu),(u_s-1)^2\} \\
&=E\{u_t(u_t-1)(X_t-\mu)(u_s-1)^2\}-E\{u_t(u_t-1)(X_t-\mu)\}E\{(u_s-1)^2\} \\
&=\left[E\{u_t(u_t-1)(u_s-1)^2\}-E\{u_t(u_t-1)\}E\{(u_s-1)^2\}\right]E(X_t-\mu)=0, \qquad \text{(E.53)}
\end{aligned}$$

where the last step is because $E(X_t-\mu)=0$.

13. By independence assumption between $\{u_t\}$, if $t \neq s$ or $p \neq r$, we have that

$$\text{Cov}\left\{(u_t-1)(u_{t+p}-1),(u_s-1)(u_{s+r}-1)\right\}=0. \qquad \text{(E.54)}$$

In addition, for any $t$ and $p$ we have that

$$\begin{aligned}
&\text{Var}\left\{(u_t-1)(u_{t+p}-1)\right\} \\
&=E\left\{(u_t-1)^2(u_{t+p}-1)^2\right\} \\
&=E\left\{(u_t-1)^2\right\}E\left\{(u_{t+p}-1)^2\right\} \\
&=\sigma_u^4, \qquad \text{(E.55)}
\end{aligned}$$

and for any $t$, we have that

$$\begin{aligned}
&\text{Var}(u_t-1)^2 \\
=&E\{(u_t-1)^4\}-E^2\{(u_t-1)^2\} \\
=&E\{(u_t-1)^4\}-\sigma_u^4. \qquad \text{(E.56)}
\end{aligned}$$

**Step 4**: Now we prove the results in (3).

1°. We first show the derivation of $q^*_{200}$ as follows:

$$
\begin{aligned}
q^*_{200} &= \lim_{T\to\infty} T\mathrm{Cov}\left\{\frac{1}{T}\sum_{t=1}^{T}(X^*_t-\mu^*)^2, \frac{1}{T}\sum_{s=1}^{T}(X^*_s-\mu^*)^2\right\} \\
&= \lim_{T\to\infty}\frac{\beta_0^4}{T}\sum_{t=1}^{T}\sum_{s=1}^{T}\mathrm{Cov}\left\{u_t^2(X_t-\mu)^2+2\mu u_t(u_t-1)(X_t-\mu)+\mu^2(u_t-1)^2,\right. \\
&\quad \left. u_s^2(X_s-\mu)^2+2\mu u_s(u_s-1)(X_s-\mu)+\mu^2(u_s-1)^2\right\} \\
&= \lim_{T\to\infty}\frac{\beta_0^4}{T}\sum_{t=1}^{T}\sum_{s=1}^{T}\mathrm{Cov}\left\{u_t^2(X_t-\mu)^2, u_s^2(X_s-\mu)^2\right\} \\
&\quad + \lim_{T\to\infty}\frac{4\mu\beta_0^4}{T}\sum_{t=1}^{T}\sum_{s=1}^{T}\mathrm{Cov}\left\{u_t^2(X_t-\mu)^2, u_s(u_s-1)(X_s-\mu)\right\} \\
&\quad + \lim_{T\to\infty}\frac{2\mu^2\beta_0^4}{T}\sum_{t=1}^{T}\sum_{s=1}^{T}\mathrm{Cov}\left\{u_t^2(X_t-\mu)^2, (u_s-1)^2\right\} \\
&\quad + \lim_{T\to\infty}\frac{4\mu^2\beta_0^4}{T}\sum_{t=1}^{T}\sum_{s=1}^{T}\mathrm{Cov}\left\{u_t(u_t-1)(X_t-\mu), u_s(u_s-1)(X_s-\mu)\right\} \\
&\quad + \lim_{T\to\infty}\frac{\mu^4\beta_0^4}{T}\sum_{t=1}^{T}\sum_{s=1}^{T}\mathrm{Cov}\left\{(u_t-1)^2, (u_s-1)^2\right\}, \\
&= \beta_0^4(\sigma_u^2+1)^2q_0+\beta_0^4\{E(u_t^4)-(\sigma_u^2+1)^2\}E\{(X_t-\mu)^4\} \\
&\quad + 4\mu\beta_0^4\sigma_u^2(\sigma_u^2+1)v_{00}+4\mu\beta_0^4\{E(u_t^4)-E(u_t^3)-\sigma_u^2(\sigma_u^2+1)\}E\{(X_t-\mu)^3\} \\
&\quad + 2\mu^2\beta_0^4\left\{E(u_t^4)-2E(u_t^3)+1-\sigma_u^4\right\}\gamma_0 \\
&\quad + 4\mu^2\beta_0^4\left[\sigma_u^4\sum_{h=-\infty}^{\infty}\gamma_h+\left\{E(u_t^4)-2E(u_t^3)+\sigma_u^2+1-\sigma_u^4\right\}\gamma_0\right] \\
&\quad + \mu^4\beta_0^4\left[E\{(u_t-1)^4\}-\sigma_u^4\right],
\end{aligned}
$$

where the second step is due to (E.35), the third step is because of (E.53), the last step is by (E.36), (E.41), (E.43), (E.45), (E.46), and (E.56).

$2^\circ$. Then we derive the value of $q^*_{20p}$:

$$
\begin{aligned}
q^*_{20p} &= \lim_{T\to\infty} T\mathrm{Cov}\left\{\frac{1}{T}\sum_{t=1}^{T}(X^*_t-\mu^*)^2, \frac{1}{T}\sum_{s=1}^{T}(X^*_s-\mu^*)(X^*_{s+p}-\mu^*)\right\} \\
&= \lim_{T\to\infty}\frac{\beta_0^4}{T}\sum_{t=1}^{T}\sum_{s=1}^{T}\mathrm{Cov}\left\{u_t^2(X_t-\mu)^2+2\mu u_t(u_t-1)(X_t-\mu)+\mu^2(u_t-1)^2,\right. \\
&\quad u_s u_{s+p}(X_s-\mu)(X_{s+p}-\mu)+\mu u_s(u_{s+p}-1)(X_s-\mu) \\
&\quad \left.+\mu u_{s+p}(u_s-1)(X_{s+p}-\mu)+\mu^2(u_s-1)(u_{s+p}-1)\right\} \\
&= \lim_{T\to\infty}\frac{\beta_0^4}{T}\sum_{t=1}^{T}\sum_{s=1}^{T}\mathrm{Cov}\left\{u_t^2(X_t-\mu)^2, u_s u_{s+p}(X_s-\mu)(X_{s+p}-\mu)\right\} \\
&\quad + \lim_{T\to\infty}\frac{\mu\beta_0^4}{T}\sum_{t=1}^{T}\sum_{s=1}^{T}\mathrm{Cov}\left\{u_t^2(X_t-\mu)^2, u_s(u_{s+p}-1)(X_s-\mu)+u_{s+p}(u_s-1)(X_{s+p}-\mu)\right\} \\
&\quad + \lim_{T\to\infty}\frac{2\mu\beta_0^4}{T}\sum_{t=1}^{T}\sum_{s=1}^{T}\mathrm{Cov}\left\{u_s u_{s+p}(X_s-\mu)(X_{s+p}-\mu), u_t(u_t-1)(X_t-\mu)\right\} \\
&\quad + \lim_{T\to\infty}\frac{\mu^2\beta_0^4}{T}\sum_{t=1}^{T}\sum_{s=1}^{T}\left[\mathrm{Cov}\left\{u_t^2(X_t-\mu)^2, (u_s-1)(u_{s+p}-1)\right\}\right. \\
&\quad \left.+\ \mathrm{Cov}\left\{(u_t-1)^2, u_s u_{s+p}(X_s-\mu)(X_{s+p}-\mu)\right\}\right] \\
&\quad + \lim_{T\to\infty}\frac{2\mu^2\beta_0^4}{T}\sum_{t=1}^{T}\sum_{s=1}^{T}\mathrm{Cov}\left\{u_t(u_t-1)(X_t-\mu), u_s(u_{s+p}-1)(X_s-\mu)\right\} \\
&\quad + \lim_{T\to\infty}\frac{2\mu^2\beta_0^4}{T}\sum_{t=1}^{T}\sum_{s=1}^{T}\mathrm{Cov}\left\{u_t(u_t-1)(X_t-\mu), u_{s+p}(u_s-1)(X_{s+p}-\mu)\right\} \\
&\quad + \lim_{T\to\infty}\frac{\mu^4\beta_0^4}{T}\sum_{t=1}^{T}\sum_{s=1}^{T}\mathrm{Cov}\left\{(u_t-1)^2, (u_s-1)(u_{s+p}-1)\right\}, \\
&= \beta_0^4 q_{0p}(\sigma_u^2+1)+\beta_0^4\left\{E(u_t^3)-(\sigma_u^2+1)\right\}\left[E\{(X_t-\mu)^3(X_{t+p}-\mu)\}+E\{(X_t-\mu)^3(X_{t-p}-\mu)\}\right] \\
&\quad + \mu\beta_0^4 E\{u_t^3-u_t^2\}\left[E\{(X_t-\mu)^2(X_{t-p}-\mu)\}+E\{(X_t-\mu)^2(X_{t+p}-\mu)\}\right] \\
&\quad + 2\mu\beta_0^4\sigma_u^2 v_{0p}+2\mu\beta_0^4 E\{u_t^3-u_t^2-\sigma_u^2\}\left[E\{(X_t-\mu)^2(X_{t-p}-\mu)\}+E\{(X_t-\mu)^2(X_{t+p}-\mu)\}\right] \\
&\quad + 2\mu^2\beta_0^4 E(u_t-1)^3\gamma_p+4\mu^2\beta_0^4\left\{E(u_t-1)^3+\sigma_u^2\right\}\gamma_p+\mu^4\beta_0^4\sigma_u^4 \\
&= \beta_0^4 q_p(\sigma_u^2+1)+\beta_0^4\left\{E(u_t^3)-(\sigma_u^2+1)\right\}\left[E\{(X_t-\mu)^3(X_{t+p}-\mu)\}+E\{(X_t-\mu)^3(X_{t-p}-\mu)\}\right] \\
&\quad + 2\mu\beta_0^4\sigma_u^2 v_p+\mu\beta_0^4 E\{3u_t^3-3u_t^2-2\sigma_u^2\}\left[E\{(X_t-\mu)^2(X_{t-p}-\mu)\}+E\{(X_t-\mu)^2(X_{t+p}-\mu)\}\right] \\
&\quad + 6\mu^2\beta_0^4 E(u_t-1)^3\gamma_p+4\mu^2\beta_0^4\sigma_u^2\gamma_p,
\end{aligned}
$$

where the second step is by (E.35), the third step is because (E.41) and (E.53), and the second last step is because (E.38), (E.49), (E.48), (E.44), and (E.54).

3°. Then we derive the value of $q^*_{2pr}$ for $r \neq p$

$$
\begin{aligned}
q^*_{2pr} &= \lim_{T\to\infty} T\mathrm{Cov}\left\{\frac{1}{T}\sum_{t=1}^{T}(X^*_t-\mu^*)(X^*_{t+p}-\mu^*), \frac{1}{T}\sum_{s=1}^{T}(X^*_s-\mu^*)(X^*_{s+r}-\mu^*)\right\} \\
&= \lim_{T\to\infty}\frac{\beta_0^4}{T}\sum_{t=1}^{T}\sum_{s=1}^{T}\mathrm{Cov}\left\{u_tu_{t+p}(X_t-\mu)(X_{t+p}-\mu)+\mu u_t(u_{t+p}-1)(X_t-\mu)\right. \\
&\quad +\mu u_{t+p}(u_t-1)(X_{t+p}-\mu)+\mu^2(u_t-1)(u_{t+p}-1), \\
&\quad \left. u_su_{s+r}(X_s-\mu)(X_{s+r}-\mu)+\mu u_s(u_{s+r}-1)(X_s-\mu)+\mu u_{s+r}(u_s-1)(X_{s+r}-\mu)+\mu^2(u_s-1)(u_{s+r}-1)\right\} \\
&= \lim_{T\to\infty}\frac{\beta_0^4}{T}\sum_{t=1}^{T}\sum_{s=1}^{T}\mathrm{Cov}\left\{u_tu_{t+p}(X_t-\mu)(X_{t+p}-\mu), u_su_{s+r}(X_s-\mu)(X_{s+r}-\mu)\right\} \\
&\quad +\lim_{T\to\infty}\frac{\mu\beta_0^4}{T}\sum_{t=1}^{T}\sum_{s=1}^{T}\mathrm{Cov}\left\{u_su_{s+r}(X_s-\mu)(X_{s+r}-\mu), u_t(u_{t+p}-1)(X_t-\mu)\right\} \\
&\quad +\lim_{T\to\infty}\frac{\mu\beta_0^4}{T}\sum_{t=1}^{T}\sum_{s=1}^{T}\mathrm{Cov}\left\{u_su_{s+r}(X_s-\mu)(X_{s+r}-\mu), u_{t+p}(u_t-1)(X_{t+p}-\mu)\right\} \\
&\quad +\lim_{T\to\infty}\frac{\mu\beta_0^4}{T}\sum_{t=1}^{T}\sum_{s=1}^{T}\mathrm{Cov}\left\{u_tu_{t+p}(X_t-\mu)(X_{t+p}-\mu), u_s(u_{s+r}-1)(X_s-\mu)\right\} \\
&\quad +\lim_{T\to\infty}\frac{\mu\beta_0^4}{T}\sum_{t=1}^{T}\sum_{s=1}^{T}\mathrm{Cov}\left\{u_tu_{t+p}(X_t-\mu)(X_{t+p}-\mu), u_{s+r}(u_s-1)(X_{s+r}-\mu)\right\} \\
&\quad +\lim_{T\to\infty}\frac{2\mu^2\beta_0^4}{T}\sum_{t=1}^{T}\sum_{s=1}^{T}\mathrm{Cov}\left\{u_tu_{t+p}(X_t-\mu)(X_{t+p}-\mu), (u_s-1)(u_{s+r}-1)\right\} \\
&\quad +\lim_{T\to\infty}\frac{\mu^2\beta_0^4}{T}\sum_{t=1}^{T}\sum_{s=1}^{T}\mathrm{Cov}\left\{u_t(u_{t+p}-1)(X_t-\mu), u_s(u_{s+r}-1)(X_s-\mu)\right\} \\
&\quad +\lim_{T\to\infty}\frac{\mu^2\beta_0^4}{T}\sum_{t=1}^{T}\sum_{s=1}^{T}\mathrm{Cov}\left\{u_t(u_{t+p}-1)(X_t-\mu), u_{s+r}(u_s-1)(X_{s+r}-\mu)\right\} \\
&\quad +\lim_{T\to\infty}\frac{\mu^2\beta_0^4}{T}\sum_{t=1}^{T}\sum_{s=1}^{T}\mathrm{Cov}\left\{u_{t+p}(u_t-1)(X_{t+p}-\mu), u_s(u_{s+r}-1)(X_s-\mu)\right\} \\
&\quad +\lim_{T\to\infty}\frac{\mu^2\beta_0^4}{T}\sum_{t=1}^{T}\sum_{s=1}^{T}\mathrm{Cov}\left\{u_{t+p}(u_t-1)(X_{t+p}-\mu), u_{s+r}(u_s-1)(X_{s+r}-\mu)\right\} \\
&\quad +\lim_{T\to\infty}\frac{\mu^4\beta_0^4}{T}\sum_{t=1}^{T}\sum_{s=1}^{T}\mathrm{Cov}\left\{(u_t-1)(u_{t+p}-1), (u_s-1)(u_{s+q}-1)\right\},
\end{aligned}
$$

where the second step is by (E.35), the third step is because (E.41) and (E.53). Then because (E.38), (E.49), (E.44), and (E.54), we have that

$$\begin{aligned}
q^*_{2pr} = \beta_0^4 q_{pr} + \beta_0^4\sigma_u^2\Big[&E\{(X_t-\mu)^2(X_{t+p}-\mu)(X_{t+r}-\mu)\} \\
&+ E\{(X_t-\mu)(X_{t+p}-\mu)^2(X_{t+p+r}-\mu)\} \\
&+ E\{(X_{t-r}-\mu)(X_t-\mu)^2(X_{t+p}-\mu)\} \\
&+ E\{(X_t-\mu)(X_{t+p-r}-\mu)(X_{t+p}-\mu)^2\}\Big] \\
&+ \mu\beta_0^4\sigma_u^2\Big[E\{(X_t-\mu)(X_{t+p}-\mu)(X_{t+r}-\mu)\} \\
&+ E\{(X_t-\mu)(X_{t+p}-\mu)(X_{t+p+r}-\mu)\} \\
&+ E\{(X_{t-r}-\mu)(X_t-\mu)(X_{t+p}-\mu)\} \\
&+ E\{(X_t-\mu)(X_{t+p-r}-\mu)(X_{t+p}-\mu)\}\Big] \\
&+ 2\mu^2\beta_0^4\sigma_u^2(\gamma_{|p-r|}+\gamma_{p+r}).
\end{aligned} \tag{E.57}$$

4°. Finally, similar to the derivation of $q^*_{2pq}$, now we derive the value of $q^*_{2pp}$. By (E.35),

$$\begin{aligned}
q^*_{2pp} &= \lim_{T\to\infty} T\mathrm{Cov}\left\{\frac{1}{T}\sum_{t=1}^{T}(X_t^*-\mu^*)(X_{t+p}^*-\mu^*), \frac{1}{T}\sum_{s=1}^{T}(X_s^*-\mu^*)(X_{s+p}^*-\mu^*)\right\} \\
&= \lim_{T\to\infty}\frac{\beta_0^4}{T}\sum_{t=1}^{T}\sum_{s=1}^{T}\mathrm{Cov}\Big\{u_tu_{t+p}(X_t-\mu)(X_{t+p}-\mu)+\mu u_t(u_{t+p}-1)(X_t-\mu) \\
&\quad + \mu u_{t+p}(u_t-1)(X_{t+p}-\mu)+\mu^2(u_t-1)(u_{t+p}-1), \\
&\quad u_su_{s+r}(X_s-\mu)(X_{s+p}-\mu)+\mu u_s(u_{s+p}-1)(X_s-\mu) \\
&\quad + \mu u_{s+p}(u_s-1)(X_{s+p}-\mu)+\mu^2(u_s-1)(u_{s+p}-1)\Big\}.
\end{aligned}$$

Then, because (E.41) and (E.53), we have that,

$$
\begin{aligned}
q^*_{2pp} = & \lim_{T\to\infty} \frac{\beta_0^4}{T} \sum_{t=1}^{T}\sum_{s=1}^{T} \mathrm{Cov}\left\{u_t u_{t+p}(X_t-\mu)(X_{t+p}-\mu), u_s u_{s+p}(X_s-\mu)(X_{s+p}-\mu)\right\} \\
& + \lim_{T\to\infty} \frac{2\mu^2\beta_0^4}{T} \sum_{t=1}^{T}\sum_{s=1}^{T} \mathrm{Cov}\left\{u_t u_{t+p}(X_t-\mu)(X_{t+p}-\mu), (u_s-1)(u_{s+p}-1)\right\} \\
& + \lim_{T\to\infty} \frac{\mu\beta_0^4}{T} \sum_{t=1}^{T}\sum_{s=1}^{T} \mathrm{Cov}\left\{u_s u_{s+p}(X_s-\mu)(X_{s+p}-\mu), u_t(u_{t+p}-1)(X_t-\mu)\right\} \\
& + \lim_{T\to\infty} \frac{\mu\beta_0^4}{T} \sum_{t=1}^{T}\sum_{s=1}^{T} \mathrm{Cov}\left\{u_s u_{s+p}(X_s-\mu)(X_{s+p}-\mu), u_{t+p}(u_t-1)(X_{t+p}-\mu)\right\} \\
& + \lim_{T\to\infty} \frac{\mu\beta_0^4}{T} \sum_{t=1}^{T}\sum_{s=1}^{T} \mathrm{Cov}\left\{u_t u_{t+p}(X_t-\mu)(X_{t+p}-\mu), u_s(u_{s+p}-1)(X_s-\mu)\right\} \\
& + \lim_{T\to\infty} \frac{\mu\beta_0^4}{T} \sum_{t=1}^{T}\sum_{s=1}^{T} \mathrm{Cov}\left\{u_t u_{t+p}(X_t-\mu)(X_{t+p}-\mu), u_{s+p}(u_s-1)(X_{s+p}-\mu)\right\} \\
& + \lim_{T\to\infty} \frac{\mu^2\beta_0^4}{T} \sum_{t=1}^{T}\sum_{s=1}^{T} \mathrm{Cov}\left\{u_t(u_{t+p}-1)(X_t-\mu), u_s(u_{s+p}-1)(X_s-\mu)\right\} \\
& + \lim_{T\to\infty} \frac{\mu^2\beta_0^4}{T} \sum_{t=1}^{T}\sum_{s=1}^{T} \mathrm{Cov}\left\{u_t(u_{t+p}-1)(X_t-\mu), u_{s+p}(u_s-1)(X_{s+p}-\mu)\right\} \\
& + \lim_{T\to\infty} \frac{\mu^2\beta_0^4}{T} \sum_{t=1}^{T}\sum_{s=1}^{T} \mathrm{Cov}\left\{u_{t+p}(u_t-1)(X_{t+p}-\mu), u_s(u_{s+p}-1)(X_s-\mu)\right\} \\
& + \lim_{T\to\infty} \frac{\mu^2\beta_0^4}{T} \sum_{t=1}^{T}\sum_{s=1}^{T} \mathrm{Cov}\left\{u_{t+p}(u_t-1)(X_{t+p}-\mu), u_{s+p}(u_s-1)(X_{s+p}-\mu)\right\} \\
& + \lim_{T\to\infty} \frac{\mu^4\beta_0^4}{T} \sum_{t=1}^{T}\sum_{s=1}^{T} \mathrm{Cov}\left\{(u_t-1)(u_{t+p}-1), (u_s-1)(u_{s+p}-1)\right\}.
\end{aligned}
$$

Then, because (E.40), (E.50), (E.51) and (E.55), we have that,

$$
\begin{aligned}
q^*_{2pp} = & \beta_0^4 q_{pp} + \beta_0^4(\sigma_u^4 + 2\sigma_u^2)\mathrm{Var}\{(X_t-\mu)(X_{t+p}-\mu)\} \\
& + 2\beta_0^4 E\{(X_t-\mu)(X_{t+p}-\mu)^2(X_{t+2p}-\mu)\} \\
& + \mu\beta_0^4\sigma_u^2\Big[E\{(X_t-\mu)(X_{t+p}-\mu)^2\} + 2E\{(X_t-\mu)(X_{t+p}-\mu)(X_{t+2p}-\mu)\} \\
& + E\{(X_t-\mu)^2(X_{t+p}-\mu)\}\Big] \\
& + 2\mu^2\beta_0^4\sigma_u^4\gamma_p + 2\mu^2\beta_0^4\sigma_u^2(\gamma_0+\gamma_{2p}) + \mu^4\beta_0^4\sigma_u^4. \qquad \text{(E.58)}
\end{aligned}
$$