Virginia Commonwealth University

VCU Scholars Compass

Theses and Dissertations

Graduate School

2014

LATENT VARIABLE MODELS GIVEN INCOMPLETELY OBSERVED SURROGATE OUTCOMES AND COVARIATES

Chunfeng Ren
Virginia Commonwealth University

LATENT VARIABLE MODELS GIVEN INCOMPLETELY OBSERVED SURROGATE

OUTCOMES AND COVARIATES

A Dissertation submitted in partial fulfillment of the requirements for the degree of

Doctor of Philosophy at Virginia Commonwealth University

by

Chunfeng Ren

M.P.H. in Biostatistics, Georgia Southern University, 2009

M.S. in Computational Mathematics, Xi'an Jiaotong University, 1997

Director: Yongyun Shin

Assistant Professor, Department of Biostatistics

Acknowledgements

The process and success of my thesis require commitment, diligence, perseverance, as well as the help of many individuals without which I could not be able to complete the long journey in obtaining my Ph.D. in Biostatistics.

First and foremost, I would like to thank Dr. Yongyun Shin, who has introduced me to the wonders and inspiration of scientific research and has provided me with help, support, and ideas during the development of all stages in this dissertation. In addition, Dr. Robert Johnson, who has given financial support during my training at VCU Center on Society and Health, deserves my grateful appreciation. Drs. Johnson and Shin, who have taught me all about critical thinking, as well as the structure and format of scientific statistical work, need to be commended for their assistance. Other committees, Drs. Levent Dumenci, Nitai Mukhopadhyay and Roy Sabo have provided me with their encouragement, insightful comments, and challenging questions. I would be glad to thank you all for making my Ph.D. experience productive, stimulating, and enjoyable.

My team members, Amber Haley, Allison Phillip, Drs. Steven Woolf, Steven Cohen, and Emily Zimmerman have always given welcome excursions in order to refresh my mind throughout weekly meetings. The faculty in the Department of Biostatistics at VCU have guided my coursework and have helped me build up my knowledge and skills in the field of biostatistics. Helen Wang, Brain Bush, Gayle Spivey, and Yvonne Hargrove have offered me continuous support and help. I would like to acknowledge you all for making my life easier and more enjoyable.

My sincere thanks go to my classmates at VCU for their comments, discussion, and questions about my course and thesis work.

Last but not least, I would like to thank my parents, my husband and my children for unconditionally supporting me and understanding me spiritually throughout my life.

The education at VCU not only augments my experience and skills in biostatistics, but also makes my life colorful and pleasant. I feel blessed, and I will always cherish my memories here.

Table of Contents

List of Figures

List of Tables

Abstract

LATENT VARIABLE MODELS GIVEN INCOMPLETELY OBSERVED SURROGATE OUT-
COMES AND COVARIATES

By Chunfeng Ren

A dissertation submitted in partial fulfillment of the requirements for the degree of Doctor
of Philosophy at Virginia Commonwealth University

Virginia Commonwealth University, 2014

Director: Yongyun Shin, Ph.D., Assistant Professor, Department of Biostatistics

Latent variable models (LVMs) are commonly used in the scenario where the outcome of the main
interest is an unobservable measure, associated with multiple observed surrogate outcomes, and
affected by potential risk factors. This dissertation develops an approach of efficient handling
missing surrogate outcomes and covariates in two- and three-level latent variable models. How-
ever, corresponding statistical methodologies and computational software are lacking efficiently
analyzing the LVMs given surrogate outcomes and covariates subject to missingness in the LVMs.

We analyze the two-level LVMs for longitudinal data from the National Growth of Health
Study where surrogate outcomes and covariates are subject to missingness at any of the levels. A
conventional method for efficient handling of missing data is to reexpress the desired model as
a joint distribution of variables, including the surrogate outcomes that are subject to missingness
conditional on all of the covariates that are completely observable, and estimate the joint model
by maximum likelihood, which is then transformed to the desired model. The joint model, how-
ever, identifies more parameters than desired, in general. The over-identified joint model produces
biased estimates of LVMs so that it is most necessary to describe how to impose constraints on

the joint model so that it has a one-to-one correspondence with the desired model for unbiased estimation. The constrained joint model handles missing data efficiently under the assumption of ignorable missing data and is estimated by a modified application of the expectation-maximization (EM) algorithm.

There is evidence that reduced class size causes higher academic achievement for both African-American and white students. African-American students benefit more than white students from reduced class size. Studying the expected class size is interesting and contributes to moderate differences in academic achievement between African-American and white students. To draw causal inferences, three-level LVMs with an instrumental variable (IV) for the cluster-randomized study from the Tennessee Class Size Study are developed where the class size as an IV and class size as an endogenous regressor interacts with African-American student indicator. The approach extends the three-level multivariate causal effect model (Shin and Raudenbush, 2011; Shin, 2012), and is more powerful to identify causal effects and random effects across schools. The results show that the reduced class size provides higher achievement scores for African-American students than for white students, and there is no evidence that the causal minority differences are significantly different across schools.

1   Introduction

The missingness mechanism, in general, is of concern if the missingness is related to the study variables. Little and Rubin (1987) categorized the mechanism into three classes: missing completely at random (MCAR), missing at random (MAR), and not missing at random (NMAR). Let complete data $Y = (Y_{obs}, Y_{mis})$ where Y_{obs} and Y_{mis} are observed and missing variables, respectively. Define a missing pattern matrix M with elements ones and zeros indicating the missing values. The MCAR, MAR, and NMAR are defined as if M is independent of Y, given Y_{obs} if M is conditionally independent of Y_{mis}, and if M associated with Y, respectively. This dissertation is about maximum likelihood (ML) estimation of latent variable models (LVMs) in longitudinal or cluster-randomized studies given multivariate observed continuous surrogate outcomes and covariates MAR. The dissertation consists of four topics: the theoretical and practical investigation of two-level LVMs given surrogate outcomes MAR, two-level LVMs given surrogate outcomes and covariates MAR, three-level LVMs with surrogate outcomes MAR, and multivariate instrumental variable (IV) estimators in three-level LVMs with IV and surrogate outcomes MAR.

Different observed surrogate outcomes are often used to characterize an overall effect of interest. Specifically, the endpoint of the interest is a construct and cannot be directly measured. Instead, various observed surrogate outcomes are measured with error from different perspectives to quantify the overall endpoint (Pocock, Geller, and Tsiatis, 1987; Roy and Lin, 2000; Sammel and Ryan, 1996). It is also interesting to examine the covariate effects on the endpoint. There are some challenges to analyze this situation such as the chief unobservable endpoint, the various observed surrogate outcomes subject to missingness, and some potential risk factors affecting the central endpoint. The analysis of LVMs becomes most challenging if the covariates are also subject to missingness in longitudinal or high-level cluster-randomized studies and, to our knowledge, little work has studied the effects in this scenario.

Researchers commonly fit linear mixed models (LMM) for a single surrogate outcome in longitudinal studies. However, limited work has been done for multivariate surrogate outcomes. Some analysts have combined the different surrogate outcomes into a single composite and others have separately fitted the LMM for each of the surrogate outcomes. While simple and informative, the former does not reflect the uncertainty of the overall variable, and the latter fails to capture the correlation among the manifold surrogate outcomes. These univariate analyses are inefficient, not powerful to identify coefficient effects as well as they may result in biased inferences. Some authors have extended the LMM to multifarious surrogate outcomes by allowing that each surrogate outcome had different covariate effects (Shah, Laird, Schoenfeld, 1997; Shin and Raudenbush, 2011; Shin, 2012). Although this approach is desirable in some applications, it does not account for the feature that the different observed surrogate outcomes quantify the endpoint and is not robust enough to test coefficient effects or random effects unless a sample size is large enough. Roy and Lin (2000) proposed the LVMs approach for multiple continuous surrogate outcomes repeatedly measured over time. This approach provided a straightforward way to test the global covariate effects. It is useful for completely observed covariates or covariates MCAR. With the assumption of data MCAR, this process is subject to loss of information: loss of precision and bias, unless the missing data are MCAR or the completely observed observations are a random sample of all observations. In addition, this approach ignores some possible systematic difference between the total cases and incomplete cases. The biased conclusion becomes severe, in particular, with a small number of completely observed cases.

Some researchers developed methods of estimation for incomplete data according to the likelihood function under a mild assumption MAR. Under this assumption, model-based methods for missing data (Orchard and Woodbury, 1972), especially the expectation-maximization (EM) algorithm (Dempster et. al., 1977; Wu, 1993), provide the efficient estimation of parameters of the complete data through analysis of the observed data. The analysis of the LVMs, given surrogate outcomes MAR, has been well-established in some existing software such as Amos (Arbuckle,

2003), EQS (Bentler, 2007), and Mplus (Muthén and Muthén, 2010). However, to my knowledge, no work has been done given surrogate outcomes and covariates MAR for the LVMs in longitudinal studies. The idea is to reexpress hierarchical models as a joint model with the joint distribution of all variables and the surrogate outcomes subject to missingness, conditional on all completely observed variables, and to estimate the joint model in the normality framework. The unconstrained joint model over-identifies the LVMs so that it results in biased estimates. One-to-one transformations between the constrained joint model and LVMs are derived to correct the bias.

The examples that motivate the dissertation are National Growth of Health Study (NGHS) and Tennessee's Student/Teacher Achievement Ratio study (STAR). The NGHS was initiated in 1987 by the National Heart, Lung, and Blood Institute to study racial disparities in child obesity and obesity-related diseases. Though physicians usually screen obesity or obesity-related diseases via body mass index (BMI), a number of authors have detailed the disadvantages of BMI as a measure of child obesity (Garn et. al, 1986; Livingstone, 2001; Wang, 2004). The disadvantages include: (1) Unlike adults, children and adolescents have age- and gender- specific BMI. Consequently, nutritional status is identified based on percentiles; (2) BMI reflects both fat and fat-free components of body weight and it measures excess weight rather than excess fat; (3) No consensus cut-point is used to define obesity in children and adolescents since BMI does not measure fat directly. The definition of obesity is excess body fat and the other measures of body fat distribution, such as percent body fat, waist circumference, and skinfolds can also account for child obesity. The STAR randomly assigned teachers and students to small class size (13-17 classmates) and regular class size (22-25 classmates) whose objectives included studying the causal effects of racial differences in academic achievement. In these two studies, the two endpoints of our interest are child obesity and academic achievement which are associated with multiple observed surrogate outcomes and affected by some potential risk factors.

We consider the situation where surrogate outcomes measure two latent variables, child obesity and academic achievement, with error from different perspectives in the two studies, respectively.

3

It is beneficial to use scoring to classify obesity or achievement via LVMs to identify subjects with high obesity scores or low academic achievement. The approach provides optimal means of combining information and allows empirical assessment of the validity of the measures. The predictions for a unit not only depend on the measurement for that unit, but also rely on the likely distribution of the latent variables for the population of units. Prediction for a given component borrows strength from the measurements of other units because the latent variable distribution is calculated using information from all components (Rubin, 1983; Morris, 1983).

We organize the remainder of the dissertation as follows. Chapter 2 investigates the risk factors of child obesity and identifies its surrogate outcomes via a modified parameter expansion EM (PX-EM) algorithm based on the LVMs approach developed by Roy and Lin (2000) and the PX-EM algorithm proposed by Liu, Rubin and Wu (1998). Chapter 3 extends the LVMs in Chapter 2 to two-level LVMs given surrogate outcomes and covariates subject to missingness at any of the levels. An approach is developed efficiently to handle missing surrogate outcomes and covariates and to obtain unbiased estimates in the LVMs. A simulation study illustrates that an unconstrained joint model produces biased inference. The method is applied to NGHS to identify risk factors of latent child obesity given surrogate outcomes and covariates MAR. Chapter 4 expands the two-level LVMs to three-level LVMs for the STAR study where students nest within classes and classes nest within schools. It analyzes that a treatment (small class size) effect decreases the race differences of potential academic achievement. Chapter 5 continues three-level multivariate causal effect models (Shin and Raudenbush, 2011; Shin, 2012) to three-level latent variable causal effect models where class size as an IV and class size as an endogenous regressor interacts with African-American student indicator. The model is more robust to identify causal effects and significant differences randomly across schools than three-level multivariate causal effect models. Finally, chapter 6 concludes the dissertation with a short discussion.

2 Identifying Covariate Effects on Child Obesity via a Latent Variable Approach Given Incompletely Observed Biomarkers

2.1 Introduction

In the last three decades, obesity has increased rapidly among school-age children (Ogden et al., 2002). Because child obesity is associated with seminal diseases, for example, hypertension (Sabo et al., 2010), metabolic syndrome (Sun et al., 2008), cardiovascular diseases (Siervogel et al., 2000), and type 2 diabetes (Dean and Flett, 2002), the increased rate of child obesity demotes public health. It is well-known that an excess body fat defines obesity. Hu (2008) described some easy, inexpensive, but inaccurate measures of body fat and some reliable, but expensive means. The former includes body mass index (BMI), waist circumference, waist-to-hip ratio, skinfold thicknesses, and bioelectric impedance. The latter consists of underwater weighing, air-displacement plethysmography, dilution method, dual energy x-ray absorptiometry, computerized tomography, and magnetic resonance imaging. Many researchers have tried to identify the risk factors for child obesity which is measured by at least one of these surrogate outcomes (Biro et al., 2003; Huenemann,1969; Tybor et al., 2010; Mahoney, 2011; Patterson et al., 1997; Sue et al., 2005; Kiess, Marcus, and Wabitsch, 2008; Kriemler et al., 2010; Vani, 2007). Although useful and simple, some of these surrogate outcomes do not differentiate the fat mass from body mass, and some of them can be just measured with error. For example, BMI, the ratio of body weight in kilograms to height in meters squared, is widely used to define obesity (BMI$\geq$30) for men and women (WHO, 2000). Consequently, it is a broadly examined outcome variable as a surrogate body fat, but it is not an accurate assessment of body fat, in particular, for children and adolescents (Krebs et al., 2007; Maynard et al., 2001; Prentice and Jebb, 2001). The BMI of a muscular athlete, for example, will categorize the person as obese due to his/her heavy weight. Many studies have reported that body fat distribution is a more powerful predictor of diseases than

BMI (Laurie, 2002; Bjorntorp, 1988; Maynard et al., 2001; Tybor et al., 2010; Zamboni et al., 1992; Zeng et al., 2012). Other investigators analyzed the impact of covariates on multivariate surrogate outcomes and viewed the observed surrogate outcomes as measures of the latent variable with error (Sammel and Ryan, 1996; Pocock et al. 1987; Roy and Lin, 2000). Motivated by these findings and considering no surrogate outcomes of obesity accurately measuring child obesity, we use multiple observed surrogate outcomes to quantify child obesity and study its risk factors simultaneously.

In this chapter, we implement simultaneous two-level LVMs: a measurement model where multivariate surrogate outcomes measure the latent child obesity with error and a structural model where the latent obesity is related to time-varying as well as time-invariant covariates (Laird and Ware, 1982; Roy and Lin, 2000). The data for analysis include girls of age 9 to 19 years from NGHS. National Heart, Lung, and Blood Institute initiated the NGHS to investigate ethnic disparities in dietary, family, psychosocial and physical activity factors of obesity about 2,379 girls in 1985. It collected data on development of obesity and factors associated with the development from 1,213 African-American and 1,166 white girls. NGHS followed the subjects from 1987-1988 when they were 9 to 10 years old until 1996-1997 when they were 18 to 19 years old. The subjects were assessed on development of obesity and related factors annually (Morrison, 1992). The surrogate outcomes to describe the development of obesity are BMI, sum of skinfolds (SUMKIN), maximum below waist circumference (MAXBLOAV), percent fat by skinfolds (PCTFATSF), percent fat by bioelectrical impedance analysis (PFBIA), upper thigh circumference (UPTHIGAV), waist circumference (WAISTMIN). Covariates found to influence the development of obesity are age, race, number of parents in family, maturation stages, maximum parental education, household income, TV watching, and overall physical activity pattern score. Girls were ages 9 to 10 years at the first visit (1987-1988) and 18 to 19 years (1996-1997) at the tenth annual visit. Some girls missed visits or at least one of the covariates. Consequently, in this longitudinal study, occasions are nested within 2231 girls, and the number of times within each girl varies from one to seven

visits. We estimate these models simultaneously to produce efficient inferences by ML via the EM algorithm (Roy and Lin 2000). This process extends mixed linear models (Laird and Ware, 1982; Shah, Laird and Schoenfiled, 1997) to LVMs that efficiently calculate parameters.

EM algorithm (Dempster, Laird and Rubin, 1977) and EM-type algorithms (Fessler and Hero, 1994; He and Liu, 2009; Meng and Rubin 1993; Meng and Van Dyk, 1998) are easy to program and converge stably, but there are criticisms for their slow convergence. Many researchers have developed algorithms to hasten the convergence of EM and EM-type algorithms, for example, PX-EM algorithm (Liu, Rubin, and Wu, 1998), Aitken's acceleration method (Laird, Lange and Stram, 1987), conjugate gradient acceleration (Jamshidian and Jennrich, 1993), and Quasi-Newtonian acceleration (Lang, 1995a, 1995b). Among these algorithms, the PX-EM algorithm makes convergence dramatically faster than EM and EM-type algorithms and keeps their stability with simple modifications (Lewandowski, Liu, and Wiel, 2010; Liu, Rubin, and Wu, 1998). Seminal studies have implemented the algorithm to different scenarios (Gelman et al., 2008; Ghosh, Reid, and Frasser, 2010; Lavielle and Meza, 2007; Liu and Wu, 1999; Martin, Hwang, and Liu, 2010; Martin, Zhang, and Liu, 2010; Qi and Jaakkola, 2007; Yu and Meng, 2010; Zhang and Liu, 2011), but no studies extended it to LVMs for identifying risk factors of child obesity.

The objectives of this chapter are (1) to identify surrogate outcomes associated with child obesity, (2) to identify risk factors of child obesity, and (3) to define unit-specific scores of child obesity. It is challenging to achieve these goals due to the unobservable obesity and the various surrogate outcomes measured repeatedly over time with error. Section 2.2 introduces the LVMs. Section 2.3 describes the EM and PX-EM algorithms. With the assumption of surrogate outcomes MAR (Little and Rubin, 1987) or MCAR (Heitjan and Basu, 1996; Little and Rubin, 1987), Section 2.4 analyzes the NGHS data via the two algorithms. Section 2.5 concludes the chapter with a brief discussion. Finally, Section 2.6 describes detailed mathematical derivations.

2.2 Models

This section introduces the LVMs (Roy and Lin, 2000). The structural model for the latent child obesity is

$$U_{ik} = X_{ik}\alpha + Z_{ik}a_i + \epsilon_{ik}, \tag{2.1}$$

where U_{ik} is a univariate latent variable of child obesity; X_{ik} is a vector of covariates having fixed effects α; and Z_{ik} is a vector of covariates having level-2 unit-specific random effects $a_i \overset{iid}{\sim} N(0, D)$ independent of a level-1 unit-specific random error $\epsilon_{ik} \overset{iid}{\sim} N(0, 1)$ for level-1 unit or occasion $k = 1, \cdots, k_i$ nested within level-2 unit or girl $i = 1, \cdots, n$. If the latent variable U_{ik} were observable, we would be able to estimate the model by standard multilevel software. With the response variable unobservable, there are seven observable surrogate outcomes that are highly correlated and predict the latent score with accuracy. That is, the latent score is related to the surrogate outcomes by

$$Y_{ijk} = \beta_{0j} + \beta_{1j}U_{ik} + b_{ij} + e_{ijk}, \tag{2.2}$$

where Y_{ijk} $(j = 1, \cdots, 7)$ are seven observable surrogate outcomes; $\beta_j = [\beta_{0j}\ \beta_{1j}]^T$ is a vector of regression coefficients for the jth surrogate outcome; $b_{ij} \sim N(0, \xi_j)$ is a level-2 unit-specific random effect independent of level-1 unit-specific random error $e_{ijk} \sim N(0, \tau_j)$ for $j = 1, \cdots, 7$. To make parameters identifiable, we assume ϵ_{ik} is distributed as $N(0, 1)$ and X_{ik} does not contain an intercept. Figure 2.1 illustrates the feature of the models. At each time point, the vertical arrows indicate the covariates X_{ik} affect the latent variable U_{ik}, which then affects the seven surrogate outcomes $(Y_{i1k}\ Y_{i2k} \cdots Y_{i7k})$. The coefficients α and β_{1j} characterize the correlations of latent variable U_{ik} with covariates X_{ik} and the jth observed surrogate outcomes Y_{ijk}, respectively. The horizontal arrows show that how we model the random effects on the longitudinal multiple surrogate outcomes and the latent variable. Specifically, the jth random intercept b_{ij} is associated with the jth surrogate outcome with variance parameter ξ_j and the random effect a_i is associated with

the latent variable U_{ik} with covariance parameter D.

It is essential to aggregate the models at the individual level for deriving estimators and their standard errors. Define $U_i = [U_{i1}\ U_{i2} \cdots U_{ik_i}]^T$, $Y_{ij} = [Y_{ij1}\ Y_{ij2} \cdots Y_{ijk_i}]^T$, and $Y_i = [Y_{i1}^T\ Y_{i2}^T \cdots Y_{i7}^T]^T$, with ϵ_i, X_i, e_i, and Z_i defined similarly. Let $\beta_0 = [\beta_{01}\ \beta_{02} \cdots \beta_{07}]^T$ with β_1 similarly defined. Then we can write models (2.1) and (2.2) in matrix notation as

$$
\begin{aligned}
Y_i &= \beta_0 \otimes 1_{k_i} + \beta_1 \otimes U_i + b_i \otimes 1_{k_i} + e_i, \\
U_i &= X_i\alpha + Z_i a_i + \epsilon_i,
\end{aligned}
\tag{2.3}
$$

where $\otimes$ represents Kronecker product (Walter and Samuel, 2007), $b_i = [b_{i1}\ b_{i2} \cdots b_{i7}]^T$ follows $N(0, R)$ with $R(\xi) = \text{diag}(\xi_1, \xi_2 \cdots, \xi_7) \overset{\Delta}{=} \oplus_{j=1}^{7}\xi_j$. For the unit i, suppose we have $k_{ij} \leq k_i$ repeated measures on the jth surrogate outcome. Let O_{ij} be an index matrix to indicate the time points when the jth ($j = 1, 2, \cdots, 7$) surrogate outcome is observed. Specifically, O_{ij} is a $k_{ij} \times k_i$ matrix constructed by deleting rows of I_{k_i} which are corresponding to the missing observations on the jth surrogate outcome. Hence, $Y_{ij}^{\circ} = O_{ij}Y_{ij}$. Given $O_i = \oplus_{j=1}^{7}O_{ij}$, then the observed data $Y_i^{\circ} = O_iY_i$. The observed aggregate model (2.3) can be expressed as

$$
\begin{aligned}
Y_i^{\circ} &= O_i(\beta_0 \otimes 1_{k_i} + \beta_1 \otimes U_i + b_i \otimes 1_{k_i} + e_i), \\
U_i &= X_i\alpha + Z_i a_i + \epsilon_i,
\end{aligned}
\tag{2.4}
$$

2.3   EM and PX-EM Algorithms

It is difficult to estimate the model (2.3) directly via its actual log likelihood since β_1 enters both the marginal mean and variance of Y_i. Roy and Lin (2000) proposed the EM algorithm to estimate the LVMs. One advantage of using EM algorithm is that the multiple observed surrogate outcomes are conditionally independent given the latent variable. In the EM algorithm, we treat

the latent variables U_i, the random effects a_i and b_i as missing data. Therefore, the complete data are (Y_i, U_i, a_i, b_i) and the observed data are Y_i°. Given the initial values of the parameters, the EM algorithm iterates between its E- and M-steps until convergence. The E-step takes expectations of the sufficient statistics of the complete-data log likelihood, given the observed data. The M-step maximizes the expected complete-data log likelihood given parameters from the previous iteration. The method (see details in Sections 2.6.1 and 2.6.2) includes

E step: Calculate the conditional expectations related to U_i, a_i, b_i, e_{ij}, ϵ_i and $U_i^T e_{ij}$;

M step: Maximize the model parameters from the complete-data log likelihood.

The variances of the parameter estimators are computed by the expected Fisher information matrix (Section 2.6.3) based on the marginal log likelihood of the observed data Y_i° at convergence. A criticism of the EM algorithm is its slow convergence to maximum likelihood estimators (MLE). Liu, Rubin, and Wu (1998) developed a PX-EM algorithm by extending EM algorithm. We implement the PX-EM algorithm and extend it to the LVMs. The PX-EM algorithm is applied to the models (2.1) and (2.2) where the only change is an extension of the parameter $\epsilon_{ik} \overset{iid}{\sim} N(0, \sigma^2)$. The PX-EM algorithm (see details in Section 2.6.4) is

PX-E step: This is unchanged from EM;

PX-M step: Model parameters are estimated in the expanded space as

$$\gamma_\star^t = \left(\beta_0^t, \beta_1^t, \alpha_t, \tau_t, D_t, \xi_t, \sigma_t^2 \right)$$

and then $\gamma_\star^t$ is transformed to the desired model parameter space as

$$\gamma^t = \left(\beta_0^t, \beta_1^t \sigma_t, \frac{\alpha_t}{\sigma_t}, \tau_t, \frac{D_t}{\sigma_t^2}, \xi_t, \sigma^2 = 1 \right).$$

2.4 Data Analysis

In this section, we analyze the NGHS data described in Section 2.1 through the EM and PX-EM algorithms. First, we delineate how to choose the initial values of the two algorithms. Secondly, we summarize the data for analysis. Finally, we interpret the results.

The algorithms iterate the E- and M-steps given the initial values of parameters. It is crucial to choose their starting values carefully. In the LVMs, if we knew the latent variable, we could use standard statistics approaches to evaluate the models. We first perform a principal component analysis (PCA) to estimate latent obesity scores, called common factors or factor scores. A weighted least squares method is used to summarize the factor scores (Johnson and Winchern, 2007) based on the first factor, which explains 90% of the sample variance. Defining the factor scores as the latent scores and using PROC MIXED in SAS, we fit the models to estimate the initial values of the algorithms. The carefully estimated initial values help accelerate the convergence to MLE. An IML SAS program is written to implement the EM and PX-EM algorithms to estimate model parameters and their standard errors. The convergence criterion is the difference in log likelihoods of observed data between two-consecutive iterations, which is set as less than 10^{-6}.

The seven surrogate outcomes are highly correlated with correlations ranging from 0.53 to 0.98, and are useful to assess child obesity. Covariates associated with obesity include age, TV viewing and video game playing (hours per week), physical activity, maturation stages (prepuberty, puberty, post menarche, ≥ 2 years post menarche), maximum parental education (high school or less, some college or more), household yearly income ($\leq \$19,999$, $\$20,000 - \$39,999$, $\geq \$40,000$), race (white/black), and the number of parents in a family (two/one). We create dummy variables for the maturation stages, maximum parental education, household income and race by using prepuberty, high school or less, $\leq \$19,999$, Black, and two-parent families as the references, respectively. Table 2.1 displays the summary statistics of the surrogate outcomes for analysis. It indicates the number of missing values up to 30.12% for upper thigh circumference. Therefore, it encourages

to analyze the data assuming surrogate outcomes MAR.

The PX-EM algorithm converges 10 times faster than the EM-algorithm. We only present the results generated by the PX-EM algorithm in Tables 2.2 and 2.3 because the EM algorithm produces practically identical results. Table 2.2 shows the estimates of the random-intercept and -coefficient model (2.1) with $Z_{ik} = 1$ and $Z_{ik} = [1 \; \text{Age}_{ik}]$, respectively. Under the assumption of surrogate outcomes MCAR, the parameters are overestimated, and their standard errors are larger than these under the assumption of surrogate outcomes MAR regardless of the random-intercept or -coefficient model (2.1). Likelihood ratio tests, which has test statistics 184.38 and 2666.5 $\sim \chi_2^2$ with p-value< 0.0001, indicate that the random-coefficient model fits more adequately than the random-intercept model for both assumptions of surrogate outcomes MCAR and MAR, respectively. Compared with the counterparts in the random-coefficient model (2.1) in both assumptions of surrogate outcomes MCAR and MAR, the parameters in the random-intercept model (2.1) are underestimated. Tables 2.3 and 2.4 list the estimates and their standard errors of the measurement model (2.2) assuming surrogate outcomes MAR or MCAR under random-intercept and -coefficient model (2.1). Our analysis shows all seven outcomes are positively associated with the latent child obesity for both cases, but under the assumption of surrogate outcomes MCAR, it appears to underestimate all β_{1j} $(j = 1, 2, \cdots, 7)$-the slopes of latent obesity for the seven surrogate outcomes. In addition, the parameters in the model (2.2) are overestimated under the random-intercept model (2.1) than the counterparts under the random-coefficient model (2.1).

In the following, we explain the results in Table 2.2 assuming surrogate outcomes MAR for the random-coefficient model (2.1). The results show that controlling the other measures constant in the model (2.1), on average, one unit addition to physical activity scores decreases child obesity by 0.003 (p-value< 0.0001); girls' obesity significantly changes from prepuberty stage to the other three stages and units raise by 0.277 (p-value< 0.0001), 1.149 (p-value< 0.0001), and 1.191 (p-value< 0.0001), respectively; girls from families with household income greater than \$40,000 have 0.423 units (p-value=0.022) higher obesity scores than girls from families with lower house-

hold income; girls from single-parent families have 0.393 units (p-value=0.005) higher obesity scores than girls from two-parent families. However, unlike the overwhelming research, parental education is not associated with child obesity. The effects of age and race are displayed in Figure 2.2 due to the significant effects of age squared and the interaction between age and race. Figure 2.2 indicates that controlling the other measures constant in the model (2.1), the latent obesity scores for African-American students are higher than these for white students through age 9 to 19 and the difference of the latent obesity scores increases in age. It is of substantial interest to identify subjects whose latent obesity scores are higher than some typical points. The feature of the analysis provides the estimates of the latent obesity scores via posterior mean in equation (2.1) that can identify the subjects with high obesity scores at each age. Figure 2.3 shows the estimated unit-specific obesity score against age for African-American and white students. Subjects with the highest obesity score are on the top of the figure. The 2.5th and 97.5th percentiles for age 9 to 19 are included in the graph so that subjects with the highest obesity score can be easily identified.

2.5 Discussion

In this chapter, we implemented the measurement model where seven surrogate outcomes measure latent child obesity with error and the structural model where the child obesity is related to covariates (Roy and Lin, 2000). We analyzed girls with 9 to 19 years of age from NGHS and simultaneously estimated the models to yield efficient inferences by ML via the PX-EM algorithm. The convergence to ML by the PX-EM algorithm was shown to be 10 times faster than that by the conventional EM algorithm. Complete-case analysis stems from the loss of information in discarding incomplete cases, which has the loss of precision and bias if missing surrogate outcomes are not MCAR, and the complete cases are not a random sample for all cases. Assuming surrogate outcomes MAR, we calculated the parameters by conditional on observed data to reduce bias due to missingness and improve precision.

Child obesity was positively associated with the seven surrogate outcomes. Age, race, TV watching, physical activity, household income, maturation stages significantly were risk factors of child obesity. Our findings indicated that household income was inversely associated with child obesity, and some authors have reported this since 1969 (Huenemann, 1969; Sobal and Stunkard, 1989). These findings imply that a greater number of exercises is beneficial, increasing physical activity and decreasing the daily hours to watch TV and play video games are recommended as strategies for preventing obesity or obesity-related diseases in youth. Public funding of quality physical education and sports facilities are also helpful to decrease the prevalence of obesity in youth.

Though physicians screen overweight children through the 95th percentile of a BMI-for-age chart, other variables of body fat distribution have been studied about their association with diseases related to obesity. Roy and Lin (2000) concluded that it was challenging to perform global testing for continuous outcomes because the outcomes were often measured at different scales and units. The structural latent variable model provides a framework to address this issue and enables a global examination of covariate effects on child obesity. The PX-EM algorithm has all advantages of the EM algorithm, and, in addition, greatly speeds up the slow convergence of the EM algorithm.

We should note some limitations in this section. This chapter is limited to analyze thoroughly observed covariates. It will be interesting to investigate if ignoring the missing covariates leads to biased inferences. However, it is not our intention here to discuss the problems in depth. The next chapter analyzes the simultaneous equations and the latent obesity by handling missing surrogate outcomes and covariates efficiently under the assumption of data MAR. Girls in NGHS were recruited from three sites- San Francisco in California, Cincinnati in Ohio, and Washington, D.C. Therefore, the inferences drawn are not for the general adolescent girls, nor are for boys.

2.6 Miscellanea

2.6.1 *Conditional Expectations in E-step*

The conditional expectations in E-step are

$$\tilde{U}_i = E(U_i|Y_i^\circ) = X_i\alpha + \Lambda_i^T O_i^T (V_i^\circ)^{-1}(Y_i^\circ - \mu_i^\circ),$$

$$E(U_i^T U_i|Y_i^\circ) = \tilde{U}_i^T \tilde{U}_i + \text{trace}(\text{cov}(U_i|Y_i^\circ)),$$

$$\tilde{a}_i = E(a_i|Y_i^\circ) = (\beta_1 \otimes Z_i D)^T O_i^T (V_i^\circ)^{-1}(Y_i^\circ - \mu_i^\circ),$$

$$E(a_i a_i^T|Y_i^\circ) = \tilde{a}_i \tilde{a}_i^T + D - (\beta_1 \otimes Z_i D)^T O_i^T (V_i^\circ)^{-1} O_i (\beta_1 \otimes Z_i D), \qquad (2.5)$$

$$\tilde{b}_i = E(b_i|Y_i^\circ) = (R \otimes 1_{k_i})^T O_i^T (V_i^\circ)^{-1}(Y_i^\circ - \mu_i^\circ),$$

$$E(b_i b_i^T|Y_i^\circ) = \tilde{b}_i \tilde{b}_i^T + R - (R \otimes 1_{k_i})^T O_i^T (V_i^\circ)^{-1} O_i (R \otimes 1_{k_i}),$$

$$\text{cov}(U_i, e_{ij}|Y_i^\circ) = -\Lambda_i^T O_i^T (V_i^\circ)^{-1} O_i \nu^T,$$

$$\tilde{\epsilon}_i = E(\epsilon_i|Y_i^\circ) = \beta_1^T \otimes I_{k_i} O_i^T (V_i^\circ)^{-1}(Y_i^\circ - \mu_i^\circ),$$

$$\tilde{e}_{ij} = E(e_{ij}|Y_i^\circ) = \nu O_i^T (V_i^\circ)^{-1}(Y_i^\circ - \mu_i^\circ),$$

$$E(e_{ij}^T e_{ij}|Y_i^\circ) = \tilde{e}_{ij}^T \tilde{e}_{ij} + \text{trace}(\tau_j I_{k_i} - \nu O_i^T (V_i^\circ)^{-1} O_i \nu^T),$$

where

$$\Lambda_i = \beta_1 \otimes (I_{k_i} + Z_i D Z_i^T),$$

$$\text{cov}(U_i|Y_i^\circ) = I_{k_i} + Z_i D Z_i^T - \Lambda_i^T O_i^T (V_i^\circ)^{-1} O_i \Lambda_i,$$

$$\nu = [0_{k_i \times (j-1)k_i} \quad \tau_j I_{k_i} \quad 0_{k_i \times (7-j)k_i}],$$

$$E(Y_i) = \mu_i = \beta_0 \otimes 1_{k_i} + \beta_1 \otimes X_i\alpha,$$

$$\text{cov}(Y_i) = V_i = (\beta_1 \beta_1^T) \otimes (I_{k_i} + Z_i D Z_i^T) + R \otimes (1_{k_i} 1_{k_i}^T) + \oplus_{j=1}^7 \tau_j I_{k_i},$$

$$\mu_i^\circ = O_i \mu_i,$$

$$V_i^\circ = O_i V_i O_i^T.$$

2.6.2   *Parameter Estimates in the M-step*

The complete-data log likelihood for (Y_i, U_i, b_i, a_i) is, apart from a constant,

$$l(\beta_0, \beta_1, \alpha, \xi, \tau, D) = \sum_{i=1}^{n} \left(l(Y_i|U_i, b_i) + l(U_i|a_i) + l(a_i) + l(b_i) \right), \tag{2.6}$$

where $\xi = [\xi_1 \ \xi_2 \cdots \xi_7]$, $\tau = [\tau_1 \ \tau_2 \cdots \tau_7]$, and

$$l(Y_i|U_i, b_i) = \sum_{j=1}^{J} \left(-\frac{k_i}{2} \log \tau_j - \frac{1}{2\tau_j} \vartheta^T \vartheta \right),$$

$$l(U_i|a_i) = -\frac{1}{2}(U_i - X_i\alpha - Z_i a_i)^T (U_i - X_i\alpha - Z_i a_i),$$

$$l(a_i) = -\frac{1}{2}(\log |D| + a_i^T D^{-1} a_i),$$

$$l(b_i) = -\frac{1}{2}(\log |R| + b_i^T R^{-1} b_i),$$

where $\vartheta = Y_{ij} - \beta_{0j} 1_{k_i} - U_i \beta_{1j} - b_{ij} 1_{k_i}$.

Differentiating (2.6) with respect to the parameters $\beta_0, \beta_1, \alpha, \xi, \tau$ and D, respectively, taking expectations of the resulting forms conditional to the observed data Y_i°, setting them equal to zero, and solving these equations, we know

$$\hat{\beta}_j^{(k)} = \hat{\beta}_j^{(k-1)} + \left(\sum_{i=1}^{n} E(U_{i*}^T U_{i*}|Y_i^\circ) \right)^{-1} \sum_{i=1}^{n} E(U_{i*}^T e_{ij}|Y_i^\circ),$$

$$\hat{\tau}_j = \frac{1}{\sum_{i=1}^{n} k_i} \times \sum_{i=1}^{n} E(e_{ij}^T e_{ij}|Y_i^\circ),$$

$$\hat{\xi}_j = \frac{1}{n} \sum_{i=1}^{n} E(b_{ij}^2|Y_i^\circ), \tag{2.7}$$

$$\hat{\alpha}^{(k)} = \hat{\alpha}^{(k-1)} + \left(\sum_{i=1}^{n} X_i^T X_i \right)^{-1} \sum_{i=1}^{n} X_i^T \tilde{\epsilon}_i,$$

$$\hat{D} = \frac{1}{n} \sum_{i=1}^{n} E(a_i a_i^T|Y_i^\circ),$$

where $j = 1, \cdots, 7$, $\beta_j = [\beta_{0j} \ \beta_{1j}]^T$, $U_{i*} = [1_{k_i} \ U_i]$, $E(b_{ij}^2|Y_i^\circ)$ is the jth diagonal element in $E(b_i b_i^T|Y_i^\circ)$ and

$$E(U_{i*}^T U_{i*}|Y_i^\circ) = \begin{bmatrix} k_i & 1_{k_i}^T \tilde{U}_i \\ 1_{k_i}^T \tilde{U}_i & E(U_i^T U_i|Y_i^\circ) \end{bmatrix},$$

$$E(U_{i*}^T e_{ij}|Y_i^\circ) = \begin{bmatrix} 1_{k_i}^T \tilde{e}_{ij} \\ \tilde{U}_i^T \tilde{e}_{ij} + \text{tr}(\text{cov}(U_i, e_{ij}|Y_i^\circ)) \end{bmatrix}.$$

2.6.3 *Calculations of the Information Matrix*

The information matrix is obtained by differentiating twice the log likelihood for the observed data Y_i° with mean and variance given in (2.5) and taking the expectation of the resulting form. Let $G_i = O_i(I_7 \otimes 1_{k_i})$, $H_i = O_i(\beta_1 \otimes X_i)$, and $M_i = O_i(I_7 \otimes X_i \alpha)$. The expected information matrix for the MLE of $\theta_1 = (\beta_0, \beta_1, \alpha)$ is

$$I_{\theta_1 \theta_1} = \sum_{i=1}^n \begin{bmatrix} G_i^T(V_i^\circ)^{-1} G_i & G_i^T(V_i^\circ)^{-1} M_i & G_i^T(V_i^\circ)^{-1} H_i \\ M_i^T(V_i^\circ)^{-1} G_i & A + M_i^T(V_i^\circ)^{-1} M_i & M_i^T(V_i^\circ)^{-1} H_i \\ H_i^T(V_i^\circ)^{-1} G_i & H_i^T(V_i^\circ)^{-1} M_i & H_i^T(V_i^\circ)^{-1} H_i \end{bmatrix}, \quad (2.8)$$

where A has its (i, k)th element $\frac{1}{2}\text{tr}\left((V_i^\circ)^{-1}(\partial V_i^\circ/\partial \beta_{1i}) \times (V_i^\circ)^{-1}(\partial V_i^\circ/\partial \beta_{1k})\right)$.

Let $\theta_2 = (\tau, D, \xi)$. Then we know

$$I_{\theta_{2i}\theta_{2k}} = \frac{1}{2} \sum_{i=1}^n \text{tr}\left((V_i^\circ)^{-1}\frac{\partial V_i^\circ}{\partial \theta_{2i}}(V_i^\circ)^{-1}\frac{\partial V_i^\circ}{\partial \theta_{2k}}\right),$$

$$I_{\theta_{2i}\beta_{1k}} = \frac{1}{2} \sum_{i=1}^n \text{tr}\left((V_i^\circ)^{-1}\frac{\partial V_i^\circ}{\partial \theta_{2i}}(V_i^\circ)^{-1}\frac{\partial V_i^\circ}{\partial \beta_{1k}}\right),$$

$$(2.9)$$

and $I_{\theta_2\beta_0} = I_{\theta_2\alpha} = 0$, where

$$
\begin{aligned}
\frac{\partial V_i^\circ}{\partial D} &= O_i \left((\beta_1 \beta_1^T) \otimes (Z_i Z_i^T) \right) O_i^T, \\
\frac{\partial V_i^\circ}{\partial \xi_j} &= O_i \left((\Delta_j \Delta_j^T) \otimes (1_{k_i} 1_{k_i}^T) \right) O_i^T, \\
\frac{\partial V_i^\circ}{\partial \beta_{1j}} &= O_i \left((\Delta_j \beta_1^T + \beta_1 \Delta_j^T) \otimes (I_{k_i} + Z_i D Z_i^T) \right) O_i^T, \\
\frac{\partial V_i^\circ}{\partial \tau_j} &= O_i \left((\Delta_j \Delta_j^T) \otimes I_{k_i} \right) O_i^T.
\end{aligned}
$$

where Δ_j is a 7×1 vector with the jth element equal to one and zero otherwise.

2.6.4 *Parameter Estimates in the PX-EM Algorithm*

For the E-step in the PX-EM algorithm, besides all the conditional expectation in Section 2.6.1, we also estimate the conditional expectations related to ϵ_i as

$$
\begin{aligned}
\tilde{\epsilon}_i &= E(\epsilon_i | Y_i^\circ) = \beta_1^T \otimes (I_{k_i} \sigma^2) O_i^T (V_i^\circ)^{-1} (Y_i^\circ - \mu_i^\circ), \\
E(\epsilon_i^T \epsilon_i | Y_i^\circ) &= \tilde{\epsilon}_i^T \tilde{\epsilon}_i + \mathrm{tr} \left(\sigma^2 I_{k_i} - (\beta_1^T \otimes I_{k_i}) O_i^T (V_i^\circ)^{-1} O_i (\beta_1 \otimes I_{k_i}) \right).
\end{aligned}
\tag{2.10}
$$

For the M-step in the PX-EM algorithm, the estimated parameters are $\hat{\beta}_j^t = \hat{\beta}_j$ with $\hat{\beta}_j^t = [\hat{\beta}_0^t \ \hat{\beta}_1^t]^T$, $\hat{\alpha}_t = \hat{\alpha}$, $\hat{\tau}_t = \hat{\tau}$, $\hat{D}_t = \hat{D}$, $\hat{\xi}_t = \hat{\xi}$ and $\hat{\sigma}_t^2 = \frac{1}{\sum_{i=1}^n k_i} \sum_{i=1}^n E(\epsilon_i^T \epsilon_i | Y_i^\circ)$. The estimated variances of the parameters in the PX-EM algorithm are same as these in the EM algorithm.

Figure 2.1: *Illustration of the structure of the latent variable models*
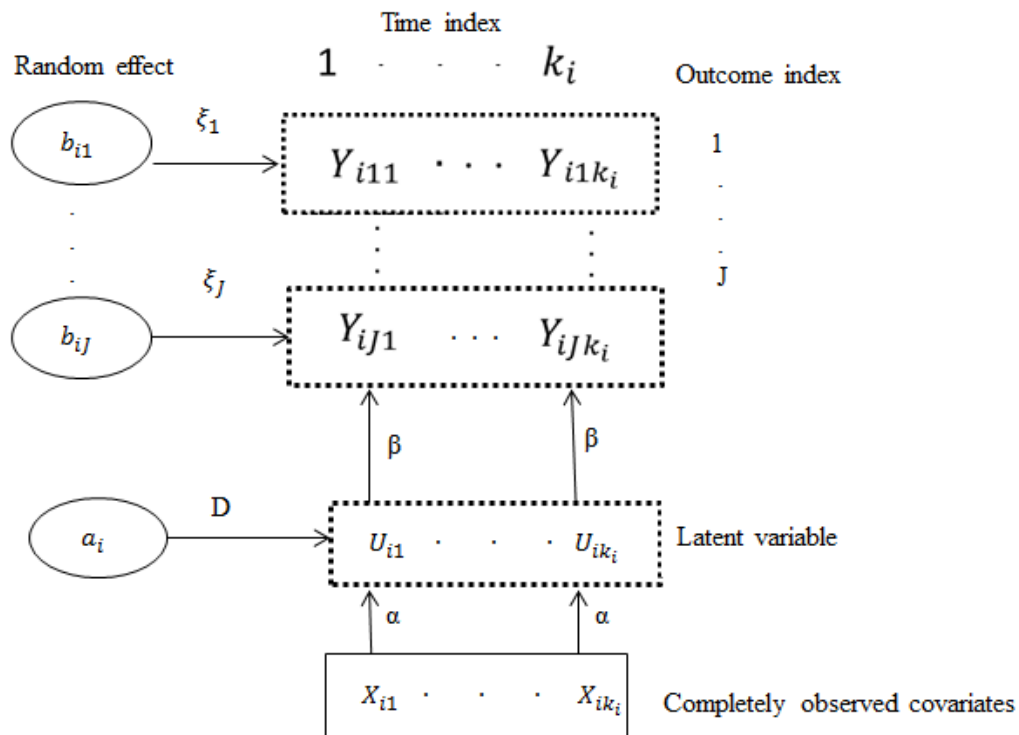
Table 2.1: Summary statistics of the seven surrogate outcomes

surrogate outcomes	N	Nmiss(%)	Mean	S.E.[a]
BMI	20580	320(1.53)	22.42	5.81
SUMSKIN	20104	796(3.81)	45.11	24.88
MAXBLOAV	18078	2822(13.50)	93.95	12.87
PCTFATSF	20322	578(2.77)	26.03	10.29
PFBIA	19419	1481(7.09)	24.59	17.86
UPTHIGAV	14604	6296(30.12)	53.61	8.88
WAISTMIN	18134	2766(13.23)	71.60	11.60

[a]standard error

Table 2.2: *Parameter estimates and their estimated standard errors in model (2.1)*

Variables	Random-intercept model		Random-coefficient model	
	MCAR	MAR	MCAR	MAR
age	0.475 (0.025)	0.365 (0.009)	0.656 (0.036)	0.531** (0.014)
age^2	-0.048 (0.007)	-0.010 (0.002)	-0.056 (0.008)	-0.013** (0.002)
age×white	-0.043 (0.020)	-0.076 (0.006)	-0.080 (0.031)	-0.125** (0.016)
TV viewing	0.006 (0.001)	0.005 (0.001)	0.007 (0.002)	0.005** (0.001)
physical activity	-0.006 (0.002)	-0.004 (0.001)	-0.005 (0.002)	-0.003* (0.001)
pubertal	0.418 (0.183)	0.027 (0.049)	0.629 (0.228)	0.277** (0.061)
postmenarchal	1.161 (0.200)	0.540 (0.070)	1.655 (0.252)	1.149** (0.087)
≥ 2 years post-menarchal	1.293 (0.222)	1.850 (0.086)	1.850 (0.279)	1.191** (0.106)
some college or more	-0.054(0.161)	-0.151 (0.122)	-0.011 (0.207)	0.108 (0.144)
$\$20,000 - \$39,999$	0.096(0.150)	0.110 (0.116)	0.129 (0.193)	0.154 (0.135)
$\geq \$40,000$	0.611 (0.206)	0.564 (0.157)	0.629 (0.267)	0.423* (0.185)
white	-0.728 (0.142)	-0.606 (0.108)	-1.027 (0.191)	-1.052** (0.161)
single-parent family	0.432 (0.157)	0.312 (0.120)	0.588 (0.202)	0.393* (0.141)

*p-value< 0.05

**p-value< 0.0001

Table 2.3: *Parameter estimates and their estimated standard errors in model (2.2) Given Random-intercept Model (2.1)*

| Assumption	Biomaker	Coefficient		Variance	
		$\hat{\beta}_{0j}$ (S.E.[a])	$\hat{\beta}_{1j}$ (S.E.[a])	$\hat{\tau}_j$ (S.E.[a])	$\hat{\xi}_j$ (S.E.[a])
MAR	BMI	22.75 (0.32)	1.97 (0.01)	0.28 (0.01)	0.88 (0.04)
	SUMSKIN	46.63 (1.20)	7.42 (0.07)	61.95 (0.90)	50.18 (2.02)
	MAXBLOAV	94.46 (0.75)	4.68 (0.04)	7.07 (0.12)	11.53 (0.47)
	PCTFATSF	26.71 (0.52)	3.22 (0.03)	9.74 (0.14)	7.02 (0.30)
	PFBIA	26.13 (0.55)	3.39 (0.03)	15.73 (0.23)	20.39 (0.74)
	UPTHIGAV	56.10 (0.58)	3.60 (0.03)	4.18 (0.09)	3.41 (0.19)
	WAISTMIN	71.40 (0.60)	3.71 (0.03)	2.54 (0.05)	9.63 (0.35)
MACR	BMI	21.29 (0.40)	1.44 (0.02)	0.30 (0.01)	0.97 (0.05)
	SUMSKIN	41.33 (1.82)	6.58 (0.10)	57.75 (1.60)	36.25 (2.24)
	MAXBLOAV	89.93 (0.95)	3.44 (0.05)	7.17 (0.21)	6.00 (0.36)
	PCTFATSF	24.10 (0.75)	2.71 (0.04)	8.65 (0.24)	6.52 (0.37)
	PFBIA	22.62 (0.80)	2.90 (0.05)	14.96 (0.41)	20.12 (0.89)
	UPTHIGAV	54.18 (0.69)	2.49 (0.04)	2.28 (0.07)	3.25(0.17)
	WAISTMIN	68.62 (0.76)	2.76 (0.04)	3.27 (0.10)	8.81 (0.36)

[a]standard error

Table 2.4: *Parameter estimates and their estimated standard errors in model (2.2)*
Given Random-intercept Model (2.1)

Assumption	Biomaker	Coefficient		Variance	
		$\hat{\beta}_{0j}$ (S.E.[a])	$\hat{\beta}_{1j}$ (S.E.[a])	$\hat{\tau}_j$ (S.E.[a])	$\hat{\xi}_j$ (S.E.[a])
MAR	BMI	22.14 (0.27)	1.31 (0.01)	0.22 (0.01)	0.86 (0.04)
	SUMSKIN	44.38 (1.01)	4.94 (0.05)	62.71 (0.90)	50.59 (2.04)
	MAXBLOAV	93.04 (0.64)	3.11 (0.03)	7.22 (0.12)	11.88 (0.48)
	PCTFATSF	25.73 (0.44)	2.14 (0.02)	10.00 (0.14)	7.07 (0.30)
	PFBIA	25.10 (0.47)	2.24 (0.02)	16.17 (0.23)	20.49 (0.74)
	UPTHIGAV	55.00 (0.49)	2.41 (0.02)	4.29 (0.09)	3.27 (0.18)
	WAISTMIN	70.27 (0.51)	2.48 (0.02)	2.58 (0.05)	9.51 (0.35)
MACR	BMI	21.14 (0.38)	1.08 (0.02)	0.27 (0.01)	0.97 (0.05)
	SUMSKIN	40.68 (1.73)	4.93 (0.12)	57.82 (1.59)	36.04 (2.23)
	MAXBLOAV	89.60 (0.90)	2.57 (0.06)	7.34 (0.21)	5.93 (0.36)
	PCTFATSF	23.83 (0.71)	2.03 (0.05)	8.72 (0.24)	6.47 (0.37)
	PFBIA	22.34 (0.76)	2.17 (0.05)	15.11 (0.41)	20.07 (0.89)
	UPTHIGAV	53.93 (0.65)	1.87 (0.04)	2.36 (0.07)	3.23 (0.17)
	WAISTMIN	68.34(0.72)	2.07 (0.05)	3.21 (0.10)	8.78 (0.36)

[a]standard error

Figure 2.2: *Age and race effects in model (2.1)*
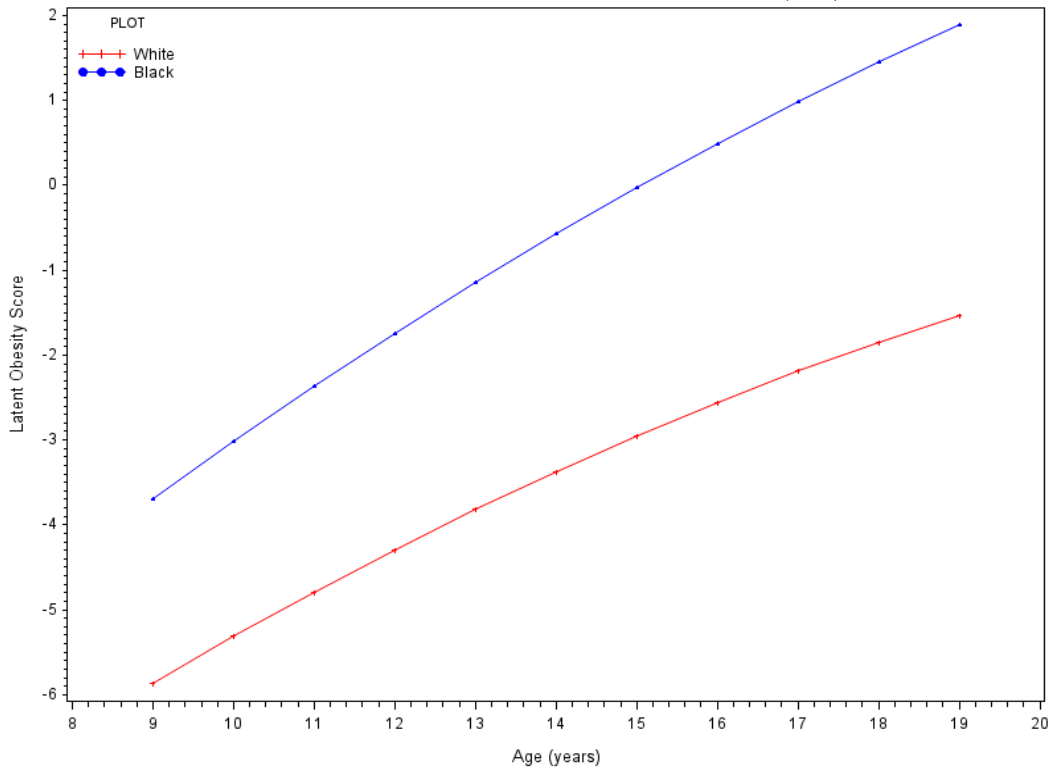
Figure 2.3: *Estimated latent scores at each age for each race*
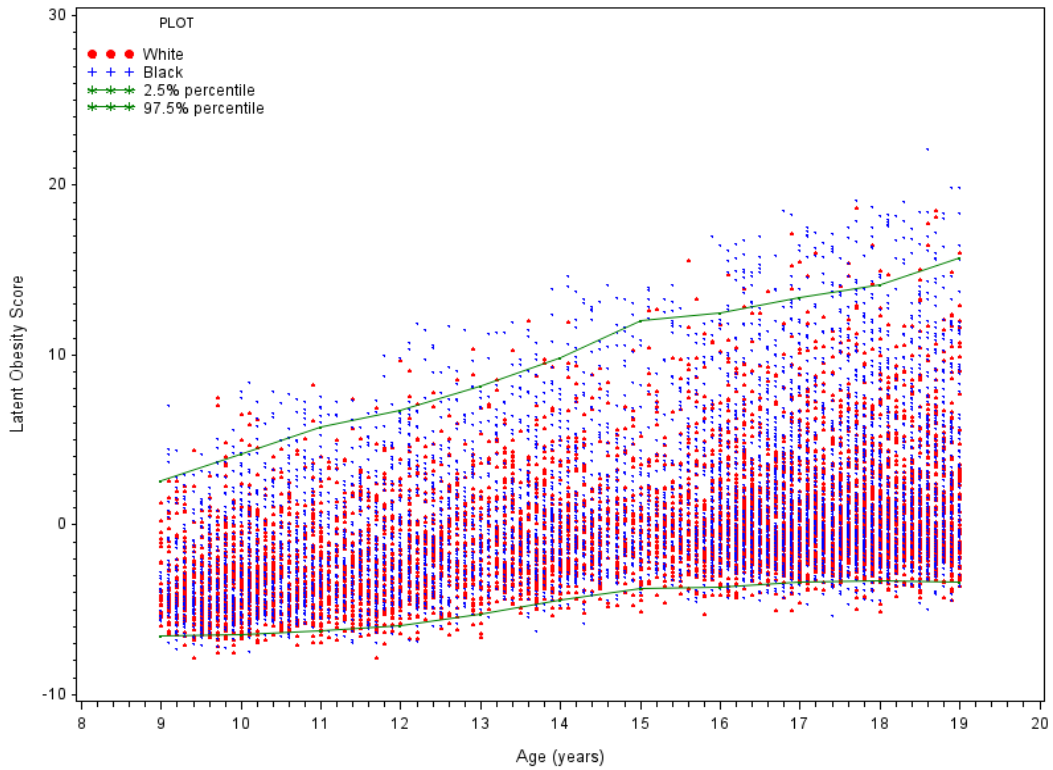
3   Longitudinal Latent Variable Models Given Incompletely Observed Biomarkers and Covariates

3.1   Introduction

In Chapter 2, we investigated risk factors of latent child obesity given missing surrogate outcomes and completely observed covariates. Our analysis in this chapter aims to identify the risk factors of child obesity, given the surrogate outcomes and covariates MAR. Specifically, we want to control for ethnic and social disparities in the growth of child obesity, and ask how environmental factors such as TV watching and mother's BMI influence the development of child obesity. Because child obesity is not directly observable, in Chapter 2 we used the seven surrogate outcomes to quantify it. Considering the highly correlated surrogate outcomes make the rate of convergence slow, we characterize the child obesity by the four surrogate outcomes: BMI, skinfold thickness, percent body fat, and waist circumference. We formulate LVMs where surrogate outcomes, given the latent obesity, are independent in a measurement model, and the obesity is regressed on covariates in a structural model (Catalano and Ryan, 1992; Cox and Wermuth, 1992; Fitzmaurice and Laird, 1995; Roy and Lin, 2000; Sammel, Ryan, and Legler, 1997; Sammel, Lin, and Ryan, 1999; Moustaki, 2003; Moustaki and Steele, 2005; Zhu, Eickhoff, and Yan, 2005; Song, Xia, and Lee, 2009).

Given completely observed covariates and surrogate outcomes having ignorable missing data (Little and Rubin 2002), LVMs may be estimated by ML via standard LVMs software such as Amos (Arbuckle, 2003), EQS (Bentler, 2007), and Mplus (Muthén and Muthén, 2010). However, little work has been done given surrogate outcomes and covariates MAR in LVMs for longitudinal studies. This chapter focuses on a longitudinal multilevel model where occasions at level 1 nest within individuals at level 2 and where missing data are present at both levels under the assumption of ignorable missing data (Rubin, 1976; Little and Rubin, 2002). Recent advances handle ignorable missing data in a hierarchical linear model (Raudenbush and Bryk 2002; Goldstein 2003)

24

efficiently by ML (Schafer and Yucel, 2002; Shin and Raudenbush 2007, 2010, 2011, 2013) or Bayesian approaches (Goldstein and Browne, 2002; Schafer and Yucel, 2002; Yucel, 2008; Goldstein et al., 2009; Goldstein and Kounali, 2009). Shin and Raudenbush (2007) reexpressed a univariate hierarchical linear model as a joint normal distribution of the variables, including the response, subject to missingness at both levels, conditional on the completely observed covariates, efficiently estimated the joint model by ML via the EM algorithm (Dempster, Laird, and Rubin, 1977), and then transformed the estimated joint model to the hierarchical model. They showed that the unconstrained joint model, in general, over-identifies the hierarchical model and that the over-identified hierarchical model may lead to biased inferences. Shin and Raudenbush estimated a constrained joint model to identify the hierarchical model for unbiased estimation. In this section, we extend this approach to efficient analysis of a longitudinal LVMs given multilevel incomplete data.

We analyze the LVMs given surrogate outcomes and covariates that are subject to missingness with a general missing pattern at any of the levels. A conventional method for efficient handling of the missing data is to reexpress the LVMs as a joint distribution of the variables, including the surrogate outcomes, which are subject to missingness conditional on all of the covariates that are completely observed, and estimate the joint model which is then transformed to the LVMs. We show that the unconstrained joint model overidentifies the LVMs leading to biased estimation of the LVMs, and explain how to characterize the joint model so that it is a one-to-one transformation of the LVMs for unbiased estimation. We efficiently estimate both the joint model and the LVMs via the EM algorithm, constraining the joint model according to the LVMs within each iteration of the EM algorithm, and demonstrate that the constrained joint model produces unbiased estimation of the LVMs.

The next section introduces a latent variable model of our interest given incomplete data. Section 3.3 explains a joint model for efficient handling of missing data in the LVMs and shows how to impose proper constraints on the joint model for unbiased estimation of the LVMs. Section 3.4

describes the EM algorithm for efficient handling of the constrained joint model. Section 3.5 simulates simple LVMs to show that the conventional method produces biased estimation of the LVMs and that our approach corrects the bias. Section 3.6 illustrates unbiased and efficient analysis of the desired LVMs given the NGHS data. Section 3.7 discusses the limitations and future extensions of our method. Section 3.8 describes detailed mathematical derivations.

3.2   Latent Variable Models

The LVMs are the same as these in Chapter 2 except for some notations. The structural model is

$$U_{ik} = X_{uik}^T \alpha + Z_{uik}^T a_i + \epsilon_{ik}, \tag{3.1}$$

where U_{ik} is a univariate latent obesity score, X_{uik} is a vector of covariates having fixed effects α, Z_{uik} is a vector of known covariates having level-2 unit-specific random effects $a_i \overset{iid}{\sim} N(0, D)$ independent of a level-1 unit-specific random error $\epsilon_{ik} \overset{iid}{\sim} N(0, 1)$, and level-1 unit or occasion k is nested within level-2 unit or subject i for $k = 1, \cdots, k_i$ and $i = 1, \cdots, n$. This model cannot be directly estimated due to unobservable U_{ik}. However, U_{ik} is related to surrogate outcomes by

$$R_{ik} = \beta_{r0} + \beta_{r1} U_{ik} + a_{ri} + e_{rik}, \tag{3.2}$$

where R_{ik} is a vector of J surrogate outcomes, β_{r0} is a vector of J intercepts, β_{r1} is a vector of the J effects or factor loadings of U_{ik}, and subject-specific random effects $a_{ri} \overset{iid}{\sim} N(0, \oplus_{j=1}^J \xi_j)$ are independent of level-1 random errors $e_{rik} \overset{iid}{\sim} N(0, \oplus_{j=1}^J \tau_j)$ for a diagonal matrix $\oplus_{\ell=1}^J A_\ell = \text{diag}(A_1, A_2, \cdots, A_J)$ with diagonal elements or submatrices $(A_1, A_2 \cdots, A_J)$ and all other elements equal to zero. To make parameters identifiable, we assume that var(ϵ_{ik})=1 and that X_{uik} does not contain an intercept. The feature of the LVMs with missing covariates is demonstrated in Figure 3.1 which can be described similarly as Figure 2.1 in Section 2.2 except for the four com-

ponents of covariates (Section 3.3). They include level-1 and -2 covariates subject to missingness and completely observed ones whose effects on the latent variable are separately displayed due to the effects from missing covariates can not be directly estimated.

3.3 Efficient Handling of Missing Data

Our analysis involves X_{uik} and R_{ik} subject to missingness. To handle the missing data efficiently, we decompose $X_{uik}^T = [S_{ik}^T \ Y_{2i}^T \ W_{1ik}^T \ W_{2i}^T]$ having fixed effects $\alpha = [\alpha_1^T \ \alpha_2^T \ \alpha_3^T \ \alpha_4^T]^T$ for vectors of p_1 level-1 covariates S_{ik} and p_2 level-2 covariates Y_{2i} subject to missingness, and vectors of p_3 level-1 covariates W_{1ik} and p_4 level-2 covariates W_{2i} completely observed. For a positive integer m, let I_m and 1_m denote a m-by-m identity matrix and a vector of m unities. If U_{ik} were observed, the missing data in model (3.1) would be efficiently handled by

$$
\begin{bmatrix} U_{ik} \\ S_{ik} \\ Y_{2i} \end{bmatrix} = \begin{bmatrix} \beta_{u1}^T & \beta_{u2}^T \\ \beta_{s1} & \beta_{s2} \\ 0 & \beta_{22} \end{bmatrix} \begin{bmatrix} W_{1ik} \\ W_{2i} \end{bmatrix} + \begin{bmatrix} Z_{uik}^T & 0 & 0 \\ 0 & I_{p_1} & 0 \\ 0 & 0 & I_{p_2} \end{bmatrix} \begin{bmatrix} b_{ui} \\ b_{si} \\ b_{2i} \end{bmatrix} + \begin{bmatrix} \epsilon_{uik} \\ \epsilon_{sik} \\ 0 \end{bmatrix}, \qquad (3.3)
$$

where β_{u1}^T and β_{s1} are 1-by-p_3 and p_1-by-p_3 matrices of the fixed effects of W_{1ik} on U_{ik} and S_{ik}, respectively, β_{u2}^T, β_{s2} and β_{22} are 1-by-p_4, p_1-by-p_4 and p_2-by-p_4 matrices of the fixed effects of W_{2i} on U_{ik}, S_{ik} and Y_{2i}, respectively, and $\begin{bmatrix} b_{ui} \\ b_{si} \\ b_{2i} \end{bmatrix} \overset{iid}{\sim} N \left(0, \begin{bmatrix} T_{uu} & T_{us} & T_{u2} \\ T_{su} & T_{ss} & T_{s2} \\ T_{2u} & T_{2s} & T_{22} \end{bmatrix} \right)$ are independent of

$\begin{bmatrix} \epsilon_{uik} \\ \epsilon_{sik} \end{bmatrix} \overset{iid}{\sim} N \left(0, \begin{bmatrix} \Sigma_{uu} & \Sigma_{us} \\ \Sigma_{su} & \Sigma_{ss} \end{bmatrix} \right)$. We center level-1 S_{ik} and W_{1ik} around respective sample means and level-2 Y_{2i} and W_{2i} around respective weighted sample means $\frac{\sum_i k_i Y_{2i}}{\sum_i k_i}$ and $\frac{\sum_i k_i W_{2i}}{\sum_i k_i}$ in equation (3.3), except for Z_{uik} that is centered around its group mean for precise estimation of the variance matrix (Raudenbush and Bryk, 2002). The centering ensures that we identify the model (3.1) with

no intercept and model (3.2). Shin and Raudenbush (2007) expressed $[\beta_{u1}^T\ \beta_{u2}^T]\begin{bmatrix} W_{1ik} \\ W_{2i} \end{bmatrix} = \beta_u^T W_{uik}$,

$\beta_{s1}W_{1ik} + \beta_{s2}W_{2i} = (I_{p_1} \otimes W_{uik}^T)\beta_s$ and $\beta_{22}W_{2i} = (I_{p_2} \otimes W_{2i}^T)\beta_2$, and efficiently estimated the model (3.3) given U_{ik} observed by ML via the EM algorithm.

Because U_{ik} is unobservable, the model (3.3) cannot be directly estimated. Instead, we formulate the joint distribution of (R_i, S_i, Y_{2i}) subject to missingness given completely observed covariates for $R_i = [R_{i1}^T\ R_{i2}^T \cdots R_{ik_i}^T]^T$ and $S_i = [S_{i1}^T\ S_{i2}^T \cdots S_{ik_i}^T]^T$ based on the aggregate models (3.2) and (3.3)

$$\begin{bmatrix} R_i \\ S_i \\ Y_{2i} \end{bmatrix} = \begin{bmatrix} 1_{k_i} \otimes \beta_{r0} + (W_{ui}\beta_u + Z_{ui}b_{ui} + \epsilon_{ui}) \otimes \beta_{r1} \\ W_{si}\beta_s + (1_{k_i} \otimes I_{p_1})b_{s_i} + \epsilon_{si} \\ X_{2i}\beta_2 + b_{2i} \end{bmatrix} + \begin{bmatrix} 1_{k_i} \otimes a_{ri} \\ 0 \\ 0 \end{bmatrix} + \begin{bmatrix} e_{ri} \\ 0 \\ 0 \end{bmatrix}, \quad (3.4)$$

for $W_{ui} = [W_{ui1}\ W_{ui2} \cdots W_{uik_i}]^T$, $Z_{ui} = [Z_{ui1}\ Z_{ui2} \cdots Z_{uik_i}]^T$ $\epsilon_{ui} = [\epsilon_{ui1}\ \epsilon_{ui2} \cdots \epsilon_{uik_i}]^T$, $e_{ri} = [e_{ri1}^T\ e_{ri2}^T \cdots e_{rik_i}^T]^T$, $W_{si} = [I_{p_1} \otimes W_{ui1}\ I_{p_1} \otimes W_{ui2}\ \cdots I_{p_1} \otimes W_{uik_i}]^T$, $\epsilon_{si} = [\epsilon_{si1}^T\ \epsilon_{si2}^T \cdots \epsilon_{sik_i}^T]^T$, and $X_{2i} = I_{p_2} \otimes W_{2i}^T$. To derive estimators, we reexpress model (3.4) parsimoniously as

$$\begin{bmatrix} Y_{1i} \\ Y_{2i} \end{bmatrix} = \begin{bmatrix} X_{1i} & 0 \\ 0 & X_{2i} \end{bmatrix} \begin{bmatrix} \beta_1 \\ \beta_2 \end{bmatrix} + \begin{bmatrix} Z_{1i} & 0 \\ 0 & I_{p_2} \end{bmatrix} \begin{bmatrix} b_{1i} \\ b_{2i} \end{bmatrix} + \begin{bmatrix} \epsilon_{1i} \\ 0 \end{bmatrix} + \begin{bmatrix} a_{1i} + e_{1i} \\ 0 \end{bmatrix}, \quad (3.5)$$

for $Y_{1i} = \begin{bmatrix} R_i \\ S_i \end{bmatrix}$, $X_{1i} = \begin{bmatrix} I_{J \times k_i} & W_{ui} \otimes I_J & 0 \\ 0 & 0 & W_{si} \end{bmatrix}$, $\beta_1 = \begin{bmatrix} 1_{k_i} \otimes \beta_{r0} \\ \beta_u \otimes \beta_{r1} \\ \beta_s \end{bmatrix}$, $Z_{1i} = \begin{bmatrix} Z_{ui} \otimes I_J & 0 \\ 0 & 1_{k_i} \otimes I_{p_1} \end{bmatrix}$,

$b_{1i} = \begin{bmatrix} b_{ui} \otimes \beta_{r1} \\ b_{si} \end{bmatrix}$, $\epsilon_{1i} = \begin{bmatrix} \epsilon_{ui} \otimes \beta_{r1} \\ \epsilon_{si} \end{bmatrix}$, $a_{1i} = \begin{bmatrix} 1_{k_i} \otimes a_{ri} \\ 0 \end{bmatrix}$, and $e_{1i} = \begin{bmatrix} e_{ri} \\ 0 \end{bmatrix}$, where $\text{var}(b_{1i}, b_{2i}) =$

$$\begin{bmatrix} \tau_{11} & \tau_{12} \\ \tau_{12}^T & \tau_{22} \end{bmatrix}, \text{var}(\epsilon_{1i}) = \begin{bmatrix} I_{k_i} \otimes (\Sigma_{uu}\beta_{r1}\beta_{r1}^T) & I_{k_i} \otimes (\beta_{r1}\Sigma_{us}) \\ I_{k_i} \otimes (\Sigma_{su} \otimes \beta_{r1}^T) & I_{k_i} \otimes \Sigma_{ss} \end{bmatrix}, \text{var}(a_{1i}) = \begin{bmatrix} (1_{k_i}1_{k_i}^T) \otimes (\oplus_{j=1}^{J}\xi_j) & 0 \\ 0 & 0 \end{bmatrix},$$

and $\text{var}(e_{1i}) = \begin{bmatrix} I_{k_i} \otimes (\oplus_{j=1}^{J}\tau_j) & 0 \\ 0 & 0 \end{bmatrix}$ for $\tau_{11} = \begin{bmatrix} T_{uu} \otimes (\beta_{r1}\beta_{r1}^T) & T_{us} \otimes \beta_{r1} \\ T_{su} \otimes \beta_{r1}^T & T_{ss} \end{bmatrix}, \tau_{12} = \begin{bmatrix} T_{u2} \otimes \beta_{r1} \\ T_{s2} \end{bmatrix},$

and $\tau_{22} = T_{22}$.

Although the conditional model (3.1) expresses a single effect of each covariate in S_{ik} on U_{ik}, the joint model (3.3) expresses a distinct covariance at each level between the covariate and U_{ik} to identify more parameters than desired in the model (3.1). The consequence is biased estimation of the LVMs as will be illustrated by a simulation study in this chapter. To correct the bias, we impose constraints on the joint model so that it is a one-to-one transformation of the LVMs. For clarity, we describe the constraints for a random-intercept model (3.1) having $Z_{uik} = 1$. Section 3.8.1 explains the constraints for a random-coefficient model (3.1). To simplify the notation, let $\text{cov}(b_{ui}, b_{s_i}|b_{2i}) = \begin{bmatrix} T_{uu|2} & T_{us|2} \\ T_{su|2} & T_{ss|2} \end{bmatrix}$. Given Y_{2i}, we constrain the covariances between U_{ik} and each covariate in S_{ik} to equal, i.e.

$$\alpha_1^T = T_{us|2}T_{ss|2}^{-1} = \Sigma_{us}\Sigma_{ss}^{-1}, \tag{3.6}$$

which says that, given Y_{2i}, the association between U_{ik} and the missing level-1 covariate is the same at each level. The constraints imply $\text{cov}(U_{ik}, S_{ik}|Y_{2i})[\text{var}(S_{ik}|Y_{2i})]^{-1} = (T_{us|2} + \Sigma_{us})(T_{ss|2} + \Sigma_{ss})^{-1} = \alpha_1^T$ for $T_{us|2} = \alpha_1^T T_{ss|2}$ and $\Sigma_{us} = \alpha_1^T \Sigma_{ss}$, and the one-to-one transformations between the LVMs and the joint model (3.5) as

$$\alpha_1 = \Sigma_{ss}^{-1}\Sigma_{su},$$

$$\alpha_2 = T_{22}^{-1}(T_{2u} - T_{2s}\alpha_1),$$

$$\alpha_3 = \beta_{u1} - \beta_{s_1}^T\alpha_1,$$

$$\alpha_4 = \beta_{u2} - \beta_{s_2}^T\alpha_1 - \beta_{22}^T\alpha_2, \tag{3.7}$$

$$1 = \Sigma_{uu} - \alpha_1^T \Sigma_{ss} \alpha_1,$$

$$D = T_{uu} - \alpha_2^T T_{22} \alpha_2 - 2\alpha_1^T T_{s2} \alpha_2 - \alpha_1^T T_{ss} \alpha_1.$$

To efficiently handle missing data, let O_{1i} and O_{2i} be matrices of zeros and ones indicating the observed values in Y_{1i} and Y_{2i} such that the observed values are $Y_{1i}^\circ = O_{1i}Y_{1i}$ and $Y_{2i}^\circ = O_{2i}Y_{2i}$, respectively (Shin and Raudenbush, 2007). The model (3.5) for the observed data is

$$
\begin{bmatrix} Y_{1i}^\circ \\ Y_{2i}^\circ \end{bmatrix} = \begin{bmatrix} X_{1i}^\circ & 0 \\ 0 & X_{2i}^\circ \end{bmatrix} \begin{bmatrix} \beta_1 \\ \beta_2 \end{bmatrix} + \begin{bmatrix} Z_{1i}^\circ & 0 \\ 0 & O_{2i} \end{bmatrix} \begin{bmatrix} b_{1i} \\ b_{2i} \end{bmatrix} + \begin{bmatrix} a_{1i}^\circ + \epsilon_{1i}^\circ + e_{1i}^\circ \\ 0 \end{bmatrix}, \qquad (3.8)
$$

for $X_{1i}^\circ = O_{1i}X_{1i}$, $X_{2i}^\circ = O_{2i}X_{2i}$, $Z_{1i}^\circ = O_{1i}Z_{1i}$, $a_{1i}^\circ = O_{1i}a_{1i}$, $\epsilon_{1i}^\circ = O_{1i}\epsilon_{1i}$, and $e_{1i}^\circ = O_{1i}e_{1i}$. Then $Y_i^\circ \sim N(\mu_i^\circ, V_i^\circ)$ for $Y_i^\circ = [Y_{1i}^{\circ T} \ Y_{2i}^{\circ T}]^T$,

$$
\mu_i^\circ = \begin{bmatrix} X_{1i}^\circ \beta_1 \\ X_{2i}^\circ \beta_2 \end{bmatrix}, \ V_i^\circ = \begin{bmatrix} Z_{1i}^\circ \tau_{11} Z_{1i}^{\circ T} + O_{1i}(\mathrm{var}(\epsilon_{1i}) + \mathrm{var}(a_{1i}) + \mathrm{var}(e_{1i}))O_{1i}^T & Z_{1i}^\circ \tau_{12} O_{2i}^T \\ O_{2i}\tau_{21}Z_{1i}^{\circ T} & O_{2i}\tau_{22}O_{2i}^T \end{bmatrix}. \ (3.9)
$$

3.4 Estimation via the EM Algorithm

This section sketches efficient estimation of the joint model (3.5) by a modified application of the EM algorithm (Dempster et al., 1977). The modification is due to the fact we efficiently determine the LVMs to find the constraints (3.6) that will be imposed on the estimated joint model (3.5) within each iteration of the EM algorithm. See Sections 3.8.2, 3.8.3 and 3.8.4 for details. Let $(Y_1, Y_2, U, b_u, b_s, a_r)$ and Y° aggregate $(Y_{1i}, Y_{2i}, U_i, b_{ui}, b_{si}, a_{ri})$ and Y_i°, respectively, for $U_i = [U_{i1} \ U_{i2} \cdots U_{ik_i}]$ in the entire sample. We view $(Y_1, Y_2, U, b_u, b_s, a_r)$ as the complete data and Y° observed. The constraints (3.6) require to evaluate the parameters α of the LVMs. Within each iteration of the EM algorithm, we estimate the parameters α and translate them

into the parameters of the joint model (3.5) according to the transformations (3.7). To estimate α, let $X_{ui} = [X_{ui1} \quad X_{ui2} \cdots X_{uik_i}]^T$, $\epsilon_i = [\epsilon_{i1} \quad \epsilon_{i2} \cdots \epsilon_{ik_i}]^T$, $\beta_j = [\beta_{r0j} \; \beta_{r1j}]^T$, $U_{ik}^* = [1 \; U_{ik}]^T$, $\epsilon_{1ik} = [\epsilon_{uik} \; \epsilon_{sik}]^T$, $\epsilon_{1i}^* = [\epsilon_{ui} \; \epsilon_{si}]$ for the LVMs; $b_{1i}^* = [b_{ui} \; b_{si}]^T$, $b_i^* = [b_{1i}^* \; b_{2i}]^T$,

$$\beta_1^* = [\beta_u \; \beta_s]^T, \; T_{11} = \begin{bmatrix} T_{uu} & T_{us} \\ T_{su} & T_{ss} \end{bmatrix}, \; T_{12} = \begin{bmatrix} T_{u2} \\ T_{s2} \end{bmatrix}, \; T = \begin{bmatrix} T_{11} & T_{12} \\ T_{12}^T & T_{22} \end{bmatrix}, \; \Sigma = \begin{bmatrix} \Sigma_{uu} & \Sigma_{us} \\ \Sigma_{su} & \Sigma_{ss} \end{bmatrix},$$

$W_{usi} = \begin{bmatrix} W_{ui} & 0 \\ 0 & W_{si} \end{bmatrix}$, and $T_{2|1} = T_{22} - T_{21} T_{11}^{-1} T_{12}$ for the joint model. The complete data ML

estimators are $\hat{\alpha}^{(k)} = \hat{\alpha}^{(k-1)} + \left(\sum_{i=1}^{n} \sum_{k=1}^{k_i} E(X_{uik} X_{uik}^T | Y_i^\circ) \right)^{-1} \sum_{i=1}^{n} \sum_{k=1}^{k_i} E(X_{uik} \epsilon_{ik} | Y_i^\circ)$ and

$\hat{D} = \sum_i E(a_i a_i^T | Y_i^\circ)/n$ for the structural model (3.1) and

$$\hat{\beta}_j^{(k)} = \hat{\beta}_j^{(k-1)} + \left(\sum_{i=1}^{n} \sum_{k=1}^{k_i} E(U_{ik}^* U_{ik}^{*T} | Y_i^\circ) \right)^{-1} \sum_{i=1}^{n} \sum_{k=1}^{k_i} E(U_{ik}^* e_{rikj} | Y_i^\circ),$$

$$\hat{\xi}_j = \frac{1}{n} \sum_{i=1}^{n} E(a_{rij}^2 | Y_i^\circ),$$

$$\hat{\tau}_j = \frac{1}{\sum_{i=1}^{n} k_i} \sum_{i=1}^{n} \sum_{k=1}^{k_i} E(e_{rikj}^2 | Y_i^\circ),$$

$$\hat{\Sigma} = \frac{1}{\sum_{i=1}^{n} k_i} \sum_{i=1}^{n} \sum_{k=1}^{k_i} E(\epsilon_{1ik} \epsilon_{1ik}^T | Y_i^\circ), \tag{3.10}$$

$$\hat{T} = \frac{1}{n} \sum_{i=1}^{n} E(b_i^* b_i^{*T} | Y_i^\circ),$$

$$\hat{\beta}_1^{*(k)} = \hat{\beta}_1^{*(k-1)} + \left(\sum_{i=1}^{n} \Sigma^{-1} \otimes (W_{usi}^T W_{usi}) \right)^{-1} \sum_{i=1}^{n} \Sigma^{-1} \otimes (W_{usi}^T \tilde{\epsilon}_{1i}^*),$$

$$\hat{\beta}_2^{(k)} = \hat{\beta}_2^{(k-1)} + \left(\sum_{i=1}^{n} T_{2|1}^{-1} \otimes (W_{2i} W_{2i}^T) \right)^{-1} \sum_{i=1}^{n} T_{2|1}^{-1} \otimes W_{2i} \left(\tilde{b}_{2i} - T_{21} T_{11}^{-1} \tilde{b}_{1i}^* \right)$$

for the joint model (3.5). At E step, we obtain conditional expectations, $E(X_{uik} X_{uik}^T | Y_i^\circ)$, $E(X_{uik} \epsilon_{ik} | Y_i^\circ)$, $E(a_i a_i^T | Y_i^\circ)$, $E(U_{ik} | Y_i^\circ)$, $E(U_{ik}^2 | Y_i^\circ)$, $E(U_{ik} e_{rikj} | Y_i^\circ)$, $E(e_{rikj} | Y_i^\circ)$, $E(e_{rikj}^2 | Y_i^\circ)$, $E(a_{rij}^2 | Y_i^\circ)$, $E(\epsilon_{1ik} \epsilon_{1ik}^T | Y_i^\circ)$, $E(b_i^* | Y_i^\circ)$, $E(b_i^* b_i^{*T} | Y_i^\circ)$, and $E(\epsilon_{1i}^* | Y_i^\circ)$ from the distribution of $Y_{1i}, Y_{2i}, U_i, e_{ri}, \epsilon_{1i}^*, b_i^*, a_{ri} | Y_i^\circ$. Let

31

$V(A)$ denote a vector of distinct elements in a variance covariance matrix A. At convergence, the expected Fisher information matrix is obtained from the observed log-likelihood of parameters $(\beta_{r0}, \beta_{r1}, \beta_1^*, \beta_2, \tau, T_{uu}, V(T_{ss}), V(T_{2s}), V(T_{22}), \xi, V(\Sigma_{ss}), \alpha_1, \alpha_2)$. The variance matrix associated with the parameter estimates in the constrained joint model (3.5) is produced by inverting the expected Fisher information matrix. We obtain the standard errors associated with the parameter estimates of the LVMs by the delta method.

The next two sections illustrate the approach by analysis of simulated and NGHS data. The convergence is taken to be the difference in the observed log-likelihoods between two consecutive iterations, which is set as less than 10^{-6}.

3.5 Simulation

In this section, we simulate the simple LVMs which involve two surrogate outcomes ($J = 2$), a single level-1 covariate S_{ik}, and a single level-2 covariate W_{2i}. The purpose of the simulation is to show that the over-identified joint model (3.5) of (R_{ik}, S_{ik}) given W_{2i} leads to biased estimation of the LVMs and that the constrained joint model (3.5), according to equations (3.6), corrects the bias. Five occasions ($k_i = 5$) are nested within each of 1000 subjects ($n = 1000$) in the simulated LVMs

$$
\begin{aligned}
U_{ik} &= S_{ik} + W_{2i} + a_i + \epsilon_{ik}, a_i \overset{iid}{\sim} N(0,1), \ \epsilon_i \overset{iid}{\sim} N(0,1), \\
R_{ik} &= 1_2 + 1_2 U_{ik} + a_{ri} + e_{rik}, \ a_{ri} \overset{iid}{\sim} N(0, 0.25 I_2), \ e_{rik} \overset{iid}{\sim} N(0, 0.25 I_2),
\end{aligned}
\tag{3.11}
$$

where $\alpha_2 = \alpha_3 = 0$, $\alpha_1 = \alpha_4 = D = \beta_{r01} = \beta_{r02} = \beta_{r11} = \beta_{r12} = 1$, $\tau_1 = \tau_2 = \xi_1 = \xi_2 = 0.25$, $S_{ik} \sim N(0,1)$, and $W_{2i} \sim \text{Bernoulli}(0.5)$. Given the simulated data, we estimate the LVMs by three different ML methods via the EM algorithm: the direct estimation of the LVMs; the evaluation of the corresponding unconstrained joint model (3.5); and the estimation of the constrained joint model (3.5) according to equations (3.6). We call the three approaches benchmark, over-

32

identified, and just-identified estimation methods. An estimation method works well if it produces all point estimates close to the benchmark counterparts. To illustrate that the over-identification problem causes biased inferences, we simulate no missing data.

Table 3.1 displays the results. The benchmark estimates are shown under column heading "Benchmark". All point estimates are close to their true values. The standard errors are relatively small. The next column under "Over-identified" shows the over-identified LVMs estimates. It is apparent that all point estimates of the model (3.1) and their standard errors are comparatively underestimated while the effects of U_{ik} and their standard errors in the model (3.2) appear over-estimated relative to the benchmark counterparts. On the other hand, the just-identified LVMs estimates and their standard errors in the next column under heading "Just-identified" are identical to the benchmark counterparts.

3.6 Analysis of NGHS Data

Now, we estimate just-identified LVMs to analyze the NGHS data. Each subject in the study was scheduled to visit a clinic for measurement once a year, but a number of subjects missed their visits to produce unit-nonresponse or had item-nonresponse. We consider multiple surrogate outcomes of obesity: BMI, sum of skinfolds at triceps, subscapular, and suprailiac sites (Skinfold), maximum below-waist circumference (Waist), and percent fat by bioelectrical impedance analysis (PercentFat). Many investigators have tried to identify the risk factors of childhood obesity where the obesity outcome variable is one of these surrogate outcomes (Patterson et al., 1997; Biro et al., 2003; Kimm et al., 2005; Vani, 2007; Kriemler et al., 2010; Mahoney, 2011). Although each surrogate outcome is a broadly examined obesity outcome variable, it is not an accurate measurement of body fat or obesity, in particular, for children and adolescents (Maynard et al., 2001; Prentice and Jebb, 2001; Krebs et al., 2007). These surrogate outcomes, however, have high correlations ranging from 0.81 to 0.92. We reason that the high positive correlations result because they are

the surrogate outcomes of obesity. The previous studies identified influential covariates of the obesity as age, race, the number of parents in a family (NumParents), maturation stages (Maturation categorizing prepuberty, puberty, post menarche, and ≥ 2 years after post menarche), maximum parental education (ParentEd classifying high school or less and some college or more), household yearly income (Income categorizing $\leq \$19,999$, $\$20,000 - \$39,999$, and $\geq \$40,000$), the number of weekly hours of TV watching (TV), overall physical activity pattern score (PhysicalAct), and mother's BMI (MotherBMI). Household income and maturation stages are coded as 0, 1, 2 and 0, 1, 2, 3, respectively, based on the preliminary analysis. We analyzed dummy indicator variables for white students (White), single-parent family (OneParent), and some college or more (ParentEd). The surrogate outcomes and covariates for analysis are summarized in Table 3.2. Nine variables are subject to missingness with each one missing up to 32% of the values. We use all available data efficiently to analyze a random-intercept model (3.1) and a random-coefficient model (3.1) in this section.

We use all available data efficiently to analyze the random-intercept LVMs and the random-coefficient LVMs (3.1). The random-intercept LVMs have R_{ik}=[BMI Skinfold PercentFat Waist]^T, S_{ik}=[Maturation TV PhysicalAct]^T, Y_{2i}=[MotherBMI Income]^T, W_{1ik}=[AGE AGE^2 AGE $\times$ White]^T, W_{2i}=[ParentEd White OneParent]^T, and $Z_{uik} = 1$, while the random-coefficient LVMs have every component the same as the random-intercept counterparts except for Z_{uik}=[1 AGE$_{ik}$]^T and $D = \begin{bmatrix} D_{00} & D_{01} \\ D_{10} & D_{11} \end{bmatrix}$. We compare the fitted models by the likelihood ratio test.

The estimated structural and measurement models of the random-intercept LVMs appear in the third column of Table 3.3 and the sixth-ninth rows of Table 3.4, respectively. From the fitted structural model, TV, maturation stage, mother's BMI, age and single-parent family are positively associated while the physical activity score, quadratic age and age by a white girl indicator interaction and the white girl indicator are negatively associated with obesity, ceteris paribus. Controlling for other covariates, the effects of household income and maximum parent education are not

statistically significant. The estimated measurement model in Table 3.4 shows that all surrogate outcomes are highly significant and, thus, predictive of the latent child obesity.

The estimated random-coefficient LVMs are displayed in the fourth column of Table 3.3 and the second-fifth rows of Table 3.4. The statistical inferences on all fixed effects stay the same as they are in the random-intercept LVMs. However, the effects of linear and quadratic ages, age by white interaction and white girl indicator strengthen, compared to the random-intercept counterparts. In particular, the negative gap of while girls' obesity relative to the African-American girls' triples. Besides, the variance of the random intercept in the random-coefficient LVMs doubles from that of the random-intercept LVMs. The measurement model in Table 3.4 shows that the surrogate outcomes have attenuating effects on child obesity, comparatively with the random-intercept counterparts. Under the null hypothesis $H_0 : D_{01} = D_{11} = 0$, the likelihood ratio test yields the p-value< 0.01 to show that the age effects vary randomly across individuals.

Figure 3.2 displays the effects of age for African-American and white girls based on the random-coefficient LVMs. Controlling for other measures in the model constant, age has a positive association with obesity overall (Obarzanek et al., 1994; Patterson et al., 1997; Chambers and Swanson, 2010). However, we find that the positive relationship weakens more rapidly for white girls than for African-American girls toward the later stage of adolescence, thereby widening the racial gap in obesity between the two subpopulations of girls. The gap starts extending rapidly from about age 14 where the 95% confidence interval -0.32 ± 0.27 $(0.05, 0.59)$.

Obesity scores are evaluated by the posterior distribution of the latent variable U_i. Apart from the need of the measure and the help of the model interpretation, another motivation of obtaining the scores is classification of units. Figure 3.3 demonstrates the plot of obesity scores against age in years for African-American and white students. The two green lines are the 2.5th and 97.5th percentiles which are the function of age because child obesity is age- and gender-specific. The subjects above the top green line and below the bottom green line have higher and lower obesity scores than the others, respectively. More African Americans than whites above the top green

line indicates that African Americans are more likely obese. Identifying these subjects and taking early prevention promote the public health and decrease the health care cost of obesity and obesity-related diseases.

3.7 Discussion

This chapter presented methods for efficient and unbiased analysis of LVMs given incomplete data with a general missing pattern at any of the levels under the assumption of ignorable missing data. Our process produces efficient estimation of the LVMs given surrogate outcomes and level-1 and level-2 covariates subject to missingness. To handle missing data efficiently, we reexpressed the LVMs as a joint distribution of the variables, including the surrogate outcomes, subject to missingness conditional on completely observed covariates. The joint model, however, over-identifies the desired LVMs when level-1 covariates are subject to missingness. The consequence is that the over-identified LVMs may provide considerably biased inferences as was illustrated in this chapter. To overcome the problem of over-identification, we constrained the joint model to be a one-to-one transformation of the LVMs and efficiently estimated the constrained joint model to produce unbiased and efficient estimation of the LVMs. We simulated LVMs to show that the just-identified LVMs estimates are unbiased while the over-identified LVMs counterparts are biased. We used a program written in SAS PROC IML in order to estimate both constrained and unconstrained joint models, which were then transformed to the desired LVMs via the multivariate Delta method. The convergence criterion was the difference in observed log likelihoods between two-consecutive iterations, which was set as 10^{-6}.

The EM algorithm (Dempster et al., 1977) and its extensions (Meng and Rubin, 1993; Fessler and Hero, 1994; Meng and Van Dyk, 1998; He and Liu, 2009) converge stably to ML, but slowly. Researchers have improved the slow convergence of the EM algorithm (Laird, Lang, and Stram, 1987; Jamshidian and Jennrich, 1993; Lang, 1995a, 1995b; Liu, Rubin, and Wu, 1998). In par-

ticular, the PX-EM algorithm speeds up the convergence with comparatively simple modifications of the EM algorithm (Liu et al., 1998; Liu and Wu, 1999; Lavielle, 2007; Lewandowski, Liu, and Wiel, 2010). We calculated the benchmark LVMs in Table 3.1 by the PX-EM algorithm, using the same convergence criterion. The computation time was increased by 10%, compared to that of the EM algorithm.

An alternative approach to our current method for the efficient estimation of LVMs, given incomplete data, is via multiple imputations (Rubin, 1987). Given the ML estimated joint distribution of variables subject to missingness conditional on covariates completely observed, we may randomly generate multiple imputations of completed data for subsequent analysis of the LVMs (Shin and Raudenbush, 2007, 2013). The multiple imputations may include the latent obesity. We would like to take on this research in the near future.

A limitation of the current approach is our assumption that the covariate having a random effect is completely observed. When such a covariate has missing values, it should be modeled on the left-hand side of the joint model in order to handle missing data efficiently. At the same time, the covariate should appear on the right-hand side of the joint model for estimation of the variance of the random effect. Such a joint model is non-normal so that the normal factorization of the joint model that leads to the desired LVMs as a conditional distribution of surrogate outcomes given covariates does not apply. Relaxing this assumption is beyond the current research.

Another limitation is that our approach bases on the multivariate normal joint model to handle missing data efficiently. We analyzed discrete covariates, household income and maturation stages, subject to missingness. Although it is not appropriate to handle such discrete missing values under the joint normality, the identified model is the desired LVMs we want to investigate, and previous studies dealt with the similar scenarios (Sammel et al., 1997; Moustaki, 2003; Song et al., 2009). The advantage is that we analyze the covariates subject to missingness by the efficient missing data method (Cox and Wermuth, 1992; Schafer, 1997; Shin and Raudenbush, 2007, 2011). Robust handling of a mixture of discrete and continuous missing data is in our future agenda of research.

3.8 Miscellanea

3.8.1 *Derivation of one-to-one transformations between models (3.1) and (3.3)*

It is easy to derive that the responses in models (3.1) and (3.3) are distributed as

$$U_{ik}|S_{ik}, Y_{2i} \sim N(\mu_{1ik}, V_{1ik}), \tag{3.12}$$

$$[U_{ik}\ S_{ik}^T\ Y_{2i}^T]^T \sim N(\mu_{2ik}, V_{2ik}), \tag{3.13}$$

respectively, where

$$\mu_{1ik} = S_{ik}^T \alpha_1 + Y_{2i}^T \alpha_2 + W_{1ik}^T \alpha_3 + W_{2i}^T \alpha_4,$$

$$V_{1ik} = Z_{uik}^T D Z_{uik} + 1,$$

$$\mu_{2ik} = \begin{bmatrix} \beta_{u1}^T W_{1ik} + \beta_{u2}^T W_{2i} \\ \beta_{s1} W_{1ik} + \beta_{s2} W_{2i} \\ \beta_{22} W_{2i} \end{bmatrix},$$

$$V_{2ik} = \begin{bmatrix} Z_{uik}^T T_{uu} Z_{uik} + \Sigma_{uu} & Z_{uik}^T T_{us} + \Sigma_{us} & Z_{uik}^T T_{u2} \\ T_{su} Z_{uik} + \Sigma_{su} & T_{ss} + \Sigma_{ss} & T_{s2} \\ T_{2u} Z_{uik} & T_{2s} & T_{22} \end{bmatrix}.$$

Let us express model (3.3) such that it recognizes the latent random effect b_{s_i} of S_{ik} as

$$[U_{ik}\ (S_{ik} - b_{s_i})^T\ b_{s_i}^T\ Y_{2i}^T]^T \sim N(\mu_{3ik}, V_{3ik}) \tag{3.14}$$

with

$$
\mu_{3ik} = \begin{bmatrix} \beta_{u1}^T W_{1ik} + \beta_{u2}^T W_{2i} \\ \beta_{s1} W_{1ik} + \beta_{s2} W_{2i} \\ 0 \\ \beta_{22} W_{2i} \end{bmatrix}, \quad
V_{3ik} = \begin{bmatrix} Z_{uik}^T T_{uu} Z_{uik} + \Sigma_{uu} & \Sigma_{us} & Z_{uik}^T T_{us} & Z_{uik}^T T_{u2} \\ \Sigma_{su} & \Sigma_{ss} & 0 & 0 \\ T_{su} Z_{uik} & 0 & T_{ss} & T_{s2} \\ T_{2u} Z_{uik} & 0 & T_{2s} & T_{22} \end{bmatrix}.
$$

Then, a regression of U_{ik} on the other variables leads to

$$
U_{ik} | S_{ik} - b_{s_i}, b_{s_i}, Y_{2i} \sim N(\mu_{4ik}, V_{4ik}) \tag{3.15}
$$

where

$$
\mu_{4ik} = (Z_{uik}^T T_{us|2} T_{ss|2}^{-1} - \Sigma_{us} \Sigma_{ss}^{-1}) b_{si} + S_{ik}^T \Sigma_{ss}^{-1} \Sigma_{su} + Y_{2i}^T T_{22}^{-1} \left(T_{u2} - T_{2s} T_{ss|2}^{-1} T_{su|2} \right) Z_{uik}
$$
$$
+ W_{1ik}^T (\beta_{u1} - \beta_{s1}^T \Sigma_{ss}^{-1} \Sigma_{su}) + W_{2i}^T \left(\beta_{u2} - \beta_{22}^T T_{22}^{-1} (T_{2u} - T_{2s} T_{ss|2}^{-1} T_{su|2}) Z_{uik} - \beta_{s2}^T \Sigma_{ss}^{-1} \Sigma_{su} \right),
$$
$$
V_{4ik} = \Sigma_{uu} - \Sigma_{us} \Sigma_{ss}^{-1} \Sigma_{su} + Z_{uik}^T (T_{uu|2} - T_{us|2} T_{ss|2}^{-1} T_{su|2}) Z_{uik}.
$$

Model (3.15) implies model (3.12) if $b_{si} = 0$. Model (3.15) with $b_{si} = 0$, however, has too strong assumption that S_{ik} does not vary across level-2 unit. The violation of the assumption leads to substantially biased inferences. Alternatively, model (3.15) implies model (3.12) if

$$
\alpha_1^T = Z_{uik}^T T_{us|2} T_{ss|2}^{-1} = \Sigma_{us} \Sigma_{ss}^{-1},
$$
$$
\Sigma_{uu} - \alpha_1^T \Sigma_{ss} \alpha_1 = 1. \tag{3.16}
$$

In the following, we discuss constraints and transformation formulas for two cases: $Z_{uik} = 1$ and $Z_{uik} = [1 \ X_{dik}^T]^T$ with p_5 covariates X_{dik} having random slopes in the structural model (3.1). If

$Z_{uik} = 1$, then the one-to-one transformations between models (3.12) and (3.15) are

$$\alpha_1 = \Sigma_{ss}^{-1}\Sigma_{su},$$

$$\alpha_2 = T_{22}^{-1}(T_{2u} - T_{2s}\alpha_1),$$

$$\alpha_3 = \beta_{u1} - \beta_{s_1}^T\alpha_1,$$

$$\alpha_4 = \beta_{u2} - \beta_{s_2}^T\alpha_1 - \beta_{22}^T\alpha_2, \tag{3.17}$$

$$D = T_{uu} - \alpha_2^T T_{22}\alpha_2 - 2\alpha_1^T T_{s2}\alpha_2 - \alpha_1^T T_{ss}\alpha_1,$$

$$1 = \Sigma_{uu} - \alpha_1^T\Sigma_{ss}\alpha_1,$$

$$T_{us} = \alpha_1^T T_{ss} + \alpha_2^T T_{2s}.$$

If $Z_{uik}^T = [1 \ X_{dik}^T]$, then let $b_{ui} = [b_{u0i} \ b_{u1i}^T]^T$, $T_{uu} = \begin{bmatrix} T_{u_0u_0} & T_{u_0u_1} \\ T_{u_1u_0} & T_{u_1u_1} \end{bmatrix}$, $T_{us} = \begin{bmatrix} T_{u_0s} \\ 0 \end{bmatrix}$, $T_{su} = T_{us}^T$,

and $T_{u2} = \begin{bmatrix} T_{u_02} \\ 0 \end{bmatrix}$. Note that we assume $\mathrm{cov}(b_{u_1i}, b_{si}) = \mathrm{cov}(b_{u_1i}, b_{2i}) = 0$. Non-zero covariances

can be estimated, but they introduce extraneous terms and make interpretable difficulty. Let $\tilde{T} = \begin{bmatrix} \alpha_2^T T_{22}\alpha_2 + 2\alpha_1^T T_{s2}\alpha_2 + \alpha_1^T T_{ss}\alpha_1 & 0 \\ 0 & 0 \end{bmatrix}$. The one-to-one transformations for α_2, D, and T_{u_0s} are

$$\alpha_2 = T_{22}^{-1}(T_{2u_0} - T_{2s}\alpha_1),$$

$$D = T_{uu} - \tilde{T}, \tag{3.18}$$

$$T_{u_0s} = \alpha_1^T T_{ss} + \alpha_2^T T_{2s},$$

and the others keep same as these in (3.17).

3.8.2 *Estimation*

The complete data $(Y_1, Y_2, U, b_u, b_s, a_r)$ discussed in Section 3.4 can be also viewed as $(Y_1, U, b_u, b_s, b_2, a_r)$ for b_2 aggregating b_{2i} in the entire sample and $Y_1 = (R, S)$. The log likelihood of the complete data is, apart from a constant,

$$\ell(\theta|R, U, S, b_u, b_s, b_2, a_r) = \sum_{i=1}^{n} \{\ell(R_i|U, a_r) + \ell(U_i, S_i|b_{ui}, b_{si}) + \ell(a_{ri}) + \ell(b_{ui}, b_{si}, b_{2i})\},$$

where

$$\ell(R_i|U, a_r) = \sum_{j=1}^{J} \left(-\frac{k_i}{2} \log \tau_j - \frac{1}{2\tau_j} \vartheta^T \vartheta \right),$$

$$\ell(U_i, S_i|b_{ui}, b_{si}) = -\frac{1}{2} \left(\ln|\Sigma \otimes I_{k_i}| + [\epsilon_{ui}^T \ \epsilon_{si}^T](\Sigma^{-1} \otimes I_{k_i}) \begin{bmatrix} \epsilon_{u_i} \\ \epsilon_{s_i} \end{bmatrix} \right),$$

$$\ell(a_{ri}) = -\frac{1}{2}(\log |R| + a_{ri}^T R^{-1} a_{ri}),$$

$$\ell(b_{ui}, b_{si}, b_{2i}) = -\frac{1}{2}(\log |T| + b_i^{*T} T^{-1} b_i^*).$$

where $\vartheta = R_{ij} - \beta_{0j} 1_{k_i} - U_i \beta_{1j} - a_{rij} 1_{k_i}$.

Differentiating the log likelihood with respect to the parameters, taking the expectation condition to the observed data, setting them equal to zero, and solving the equations, we know the MLEs of the complete data are

$$\hat{\alpha}^{(k)} = \hat{\alpha}^{(k-1)} + \left(\sum_{i=1}^{n} \sum_{k=1}^{k_i} E(X_{uik} X_{uik}^T | Y_i^\circ) \right)^{-1} \sum_{i=1}^{n} \sum_{k=1}^{k_i} E(X_{uik} \epsilon_{ik} | Y_i^\circ),$$

$$\hat{D} = \frac{1}{n} \sum_{i=1}^{n} E(a_i a_i^T | Y_i^\circ),$$

$$\hat{\beta}_j^{(k)} = \hat{\beta}_j^{(k-1)} + \left(\sum_{i=1}^{n} \sum_{k=1}^{k_i} E(U_{ik}^* U_{ik}^{*T} | Y_i^\circ) \right)^{-1} \sum_{i=1}^{n} \sum_{k=1}^{k_i} E(U_{ik}^* e_{rijk} | Y_i^\circ),$$

$$\hat{\xi}_j = \frac{1}{n} \sum_{i=1}^{n} E(a_{rij}^2 | Y_i^\circ),$$

$$\hat{\tau}_j = \frac{1}{\sum_{i=1}^{n} k_i} \sum_{i=1}^{n} \sum_{k=1}^{k_i} E(e_{rijk}^2 | Y_i^\circ),$$

$$\hat{\Sigma} = \frac{1}{\sum_{i=1}^{n} k_i} \sum_{i=1}^{n} \sum_{k=1}^{k_i} E(\epsilon_{1ik} \epsilon_{1ik}^T | Y_i^\circ),$$

$$\hat{T} = \frac{1}{n} \sum_{i=1}^{n} E(b_i^* b_i^{*T} | Y_i^\circ),$$

$$\hat{\beta}_1^{*(k)} = \hat{\beta}_1^{*(k-1)} + \left(\sum_{i=1}^{n} \Sigma^{-1} \otimes (W_{usi}^T W_{usi}) \right)^{-1} \sum_{i=1}^{n} \Sigma^{-1} \otimes (W_{usi}^T \tilde{\epsilon}_{1i}^*),$$

$$\hat{\beta}_2^{(k)} = \hat{\beta}_2^{(k-1)} + \left(\sum_{i=1}^{n} T_{2|1}^{-1} \otimes (W_{2i} W_{2i}^T) \right)^{-1} \sum_{i=1}^{n} (T_{2|1}^{-1} \otimes W_{2i}) \left(\tilde{b}_{2i} - T_{21} T_{11}^{-1} \tilde{b}_{1i} \right).$$

(3.19)

Note that the β_2 was estimated based on the distribution of $Y_2 | b_u, b_s$.

Given $\hat{\alpha}$ and $\hat{D}$, for the random-intercept model (3.1) we update the estimators, $\hat{\Sigma}_{us}$, $\hat{\Sigma}_{uu}$, $\hat{T}_{uu}$, $\hat{\beta}_{u1}$, $\hat{\beta}_{u2}$, $\hat{T}_{u2}$, and $\hat{T}_{us}$ in the joint model (3.5) according to the transformation formulas (3.17). Given $\hat{\alpha}$ and $\hat{D}$, for the random-coefficient model (3.1) we update the estimators, $\hat{\Sigma}_{us}$, $\hat{\Sigma}_{uu}$, $\hat{T}_{uu}$, $\hat{\beta}_{u1}$, $\hat{\beta}_{u2}$, $\hat{T}_{u_0 2}$, and $\hat{T}_{u_0 s}$ in the joint model (3.5) according to the transformation formulas (3.18) and set $T_{u_1 2} = T_{u_1 s} = 0$.

At E-step, we estimate the following conditional expectations.

(I) Calculate the conditional expectations for the latent variable U_i

$$\tilde{U}_{ik} = E(U_{ik} | Y_i^\circ) = \beta_{u1}^T W_{1ik} + \beta_{u2}^T W_{2i} + \Delta_u (V_i^\circ)^{-1} (Y_i^\circ - \mu_i^\circ),$$

$$E(U_{ik}^2 | Y_i^\circ) = \tilde{U}_{ik}^2 + \text{var}(U_{ik}) - \Delta_u (V_i^\circ)^{-1} \Delta_u^T,$$

(3.20)

where $\Delta_u = \begin{bmatrix} \Delta_{u1} & \Delta_{u2} & Z_{uik}^T T_{u2} \end{bmatrix} O_i^T$ and $\text{var}(U_{ik}) = Z_{uik}^T T_{uu} Z_{uik} + \Sigma_{uu}$ for $\Delta_{u1} = (Z_{uik}^T T_{uu} Z_{ui}^T + [0_{1 \times (k-1)} \ \Sigma_{uu} \ 0_{1 \times (k_i - k)}]) \otimes \beta_{r1}^T$ and $\Delta_{u2} = 1_{k_i}^T \otimes (Z_{uik}^T T_{us}) + [0_{1 \times (k-1)p_1} \ \Sigma_{us} \ 0_{1 \times (k_i - k)p_1}]$.

(II) Calculate the conditional expectation for e_{rikj}, e_{rikj}^2, and $U_{ik}e_{rikj}$ as

$$
\begin{aligned}
\tilde{e}_{rikj} &= E(e_{rikj}|Y_i^\circ) = \Delta_{er}(V_i^\circ)^{-1}(Y_i^\circ - \mu_i^\circ), \\
E(e_{rikj}^2|Y_i^\circ) &= \tilde{e}_{rikj}^2 + \text{var}(e_{rikj}) - \Delta_{er}(V_i^\circ)^{-1}\Delta_{er}^T, \\
E(U_{ik}e_{rikj}|Y_i^\circ) &= \tilde{U}_{ik}\tilde{e}_{rikj} - \Delta_u(V_i^\circ)^{-1}\Delta_{er}^T,
\end{aligned}
\tag{3.21}
$$

where $\text{var}(e_{rikj}) = \tau_j$ and $\Delta_{er} = [0_{1\times((k-1)J+j-1)} \ \tau_j \ 0_{1\times(J-j)} \ 0_{1\times((k_i-k)J+p_1k_i+p_2)}]O_i^T$.

(III) Calculate the conditional expectation for a_{rij} as

$$
\begin{aligned}
\tilde{a}_{rij} &= E(a_{rij}|Y_i^\circ) = \Delta_a(V_i^\circ)^{-1}(Y_i^\circ - \mu_i^\circ) \\
E(a_{rij}^2|Y_i^\circ) &= \tilde{a}_{rij}^2 + \xi_j - \Delta_a(V_i^\circ)^{-1}\Delta_a^T,
\end{aligned}
\tag{3.22}
$$

where $\Delta_a = [0_{1\times(j-1)k_i} \ 1_{k_i}^T\xi_j \ 0_{1\times((J-j)k_i+p_1k_i+p_2)}]O_i^T$.

(IV) Calculate the conditional expectations of $\epsilon_{1i}^* = [\epsilon_{u_i}^T \ \epsilon_{s_i}^T]^T$ and $\epsilon_{1ik} = [\epsilon_{u_{ik}} \ \epsilon_{s_{ik}}]^T$

$$
\begin{aligned}
E(\epsilon_{1i}^*|Y_i^\circ) &= \Delta_{es}(V_i^\circ)^{-1}(Y_i^\circ - \mu_i^\circ), \\
\tilde{\epsilon}_{1ik} &= E(\epsilon_{1ik}|Y_i^\circ) = \Delta_e(V_i^\circ)^{-1}(Y_i^\circ - \mu_i^\circ), \\
E(\epsilon_{1ik}\epsilon_{1ik}^T|Y_i^\circ) &= \tilde{\epsilon}_{1ik}\tilde{\epsilon}_{1ik}^T + cov(\epsilon_{1ik}) - \Delta_e(V_i^\circ)^{-1}\Delta_e^T,
\end{aligned}
\tag{3.23}
$$

where given Δ_k a vector with the kth element equal to 1 and zero otherwise, $cov(\epsilon_{1ik}) = \Sigma$,

$$
\Delta_{es} = \begin{bmatrix} I_{k_i} \otimes \Sigma_{uu} \otimes \beta_{r1}^T & I_{k_i} \otimes \Sigma_{us} & 0 \\ I_{k_i} \otimes \Sigma_{su} \otimes \beta_{r1}^T & I_{k_i} \otimes \Sigma_{ss} & 0 \end{bmatrix} O_i^T, \text{ and } \Delta_e = \begin{bmatrix} \Sigma_{uu} \otimes \Delta_k^T \otimes \beta_{r1}^T & \Delta_k^T \otimes \Sigma_{us} & 0 \\ \Sigma_{su} \otimes \Delta_k^T \otimes \beta_{r1}^T & \Delta_k^T \otimes \Sigma_{ss} & 0 \end{bmatrix} O_i^T.
$$

(V) Calculate the conditional expectations of the random effects b_i^* as

$$
\begin{aligned}
\tilde{b}_i^* &= E(b_i^*|Y_i^\circ) = \Delta_b(V_i^\circ)^{-1}(Y_i^\circ - \mu_i^\circ) \\
E(b_i^*b_i^{*T}|Y_i^\circ) &= \tilde{b}_i^*\tilde{b}_i^{*T} + T - \Delta_b(V_i^\circ)^{-1}\Delta_b^T,
\end{aligned}
\tag{3.24}
$$

where $\Delta_b = \begin{bmatrix} (T_{uu}Z_{ui}^T) \otimes \beta_{r1}^T & 1_{k_i}^T \otimes T_{us} & T_{u2} \\ (T_{su}Z_{ui}^T) \otimes \beta_{r1}^T & 1_{k_i}^T \otimes T_{ss} & T_{s2} \\ (T_{2u}Z_{ui}^T) \otimes \beta_{r1}^T & 1_{k_i}^T \otimes T_{2s} & T_{22} \end{bmatrix} O_i^T.$

In addition, we calculate $E(X_{uik}X_{uik}^T|Y_i^\circ)$, $E(X_{uik}\epsilon_{ik}|Y_i^\circ)$, and $E(a_i a_i^T|Y_i^\circ)$ in the LVMs. Let $\Delta_s = [\Delta_{s1}\ \Delta_{s2}\ 1_{k_i}^T \otimes T_{s2}]O_i^T$ for $\Delta_{s1} = \big((T_{su}Z_{ui}^T) + [0_{p_1\times(k-1)}\ T_{su}\ 0_{p_1\times(k_i-k)}]\big) \otimes \beta_{r1}^T$ and $\Delta_{s2} = 1_{k_i}^T \otimes T_{ss} + [0_{p_1\times(k-1)p_1}\ \Sigma_{ss}\ 0_{p_1\times(k_i-k)p_1}]$, $\Delta_y = [(T_{2u}Z_{ui}^T) \otimes \beta_{r1}^T\ 1_{k_i}^T \otimes T_{2u}\ T_{22}]O_i^T$, $\Delta_{ec} = [0_{1\times(k-1)J}\ \beta_{r1}^T\ 0_{1\times(k_iJ-kJ+p_1k_i+p_2)}]O_i^T$, and $\Delta_{ac} = [(DZ_{ui}^T) \otimes \beta_{r1}^T\ 0\ 0]O_i^T$.

$$E(X_{uik}X_{uik}^T|Y_i^\circ) = \begin{bmatrix} E(S_{ik}S_{ik}^T|Y_i^\circ) & E(S_{ik}Y_{2i}^T|Y_i^\circ) & \tilde{S}_{ik}W_{1ik}^T & \tilde{S}_{ik}W_{2i}^T \\ E(Y_{2i}S_{ik}^T|Y_i^\circ) & E(Y_{2i}Y_{2i}^T|Y_i^\circ) & \tilde{Y}_{2i}W_{1ik}^T & \tilde{Y}_{2i}W_{2i}^T \\ W_{1ik}\tilde{S}_{ik}^T & W_{1ik}Y_{2i}^T & W_{1ik}W_{1ik}^T & W_{1ik}W_{2i}^T \\ W_{2i}\tilde{S}_{ik}^T & W_{2i}\tilde{Y}_{2i}^T & W_{2i}W_{1ik}^T & W_{2i}W_{2i}^T \end{bmatrix}, \tag{3.25}$$

$$E(X_{uik}\epsilon_{ik}|Y_i^\circ) = \begin{bmatrix} \tilde{S}_{ik}^T\tilde{\epsilon}_{ik} - \Delta_s(V_i^\circ)^{-1}\Delta_{ec}^T \\ \tilde{Y}_{2i}^T\tilde{\epsilon}_{ik} - \Delta_y(V_i^\circ)^{-1}\Delta_{ec}^T \\ W_{1ik}\tilde{\epsilon}_{ik} \\ W_{2i}\tilde{\epsilon}_{ik} \end{bmatrix}, \tag{3.26}$$

where

$$\tilde{S}_{ik} = E(S_{ik}|Y_i^\circ) = \beta_{s1}W_{1ik} + \beta_{s2}W_{2i} + \Delta_s(V_i^\circ)^{-1}(Y_i^\circ - \mu_i^\circ)$$

$$E(S_{ik}S_{ik}^T|Y_i^\circ) = \tilde{S}_{ik}\tilde{S}_{ik}^T + T_{ss} + \Sigma_{ss} - \Delta_s(V_i^\circ)^{-1}\Delta_s^T,$$

$$\tilde{Y}_{2i} = E(Y_{2i}|Y_i^\circ) = \beta_{22}W_{2i} + \Delta_y(V_i^\circ)^{-1}(Y_i^\circ - \mu_i^\circ)$$

$$E(Y_{2i}Y_{2i}^T|Y_i^\circ) = \tilde{Y}_{2i}\tilde{Y}_{2i}^T + T_{22} - \Delta_y(V_i^\circ)^{-1}\Delta_y^T,$$

$$E(S_{ik}Y_{2i}^T|Y_i^\circ) = \tilde{S}_{ik}\tilde{Y}_{2i}^T + T_{s2} - \Delta_s(V_i^\circ)^{-1}\Delta_y^T,$$

$$\tilde{\epsilon}_{ik} = E(\epsilon_{ik}|Y_i^\circ) = \Delta_{ec}(V_i^\circ)^{-1}(Y_i^\circ - \mu_i^\circ),$$

$$\tilde{a}_i = \Delta_{ac}(V_i^\circ)^{-1}(Y_i^\circ - \mu_i^\circ),$$

44

$$E(a_i a_i^T | Y_i^\circ) = \tilde{a}_i \tilde{a}_i^T + D - \Delta_{ac}(V_i^\circ)^{-1}\Delta_{ac}^T.$$

3.8.3 *Calculation of the Information Matrix*

The expected Fisher information matrix is obtained by differentiating twice the observed marginal multivariate normal log-likelihood with mean and covariance given in (3.9), but we introduce new parameters α_1 and α_2, which are defined in (3.7). Consequently, parameters Σ_{us}, T_{u2}, T_{us}, and Σ_{uu} are the function of α_1, α_2 and the other elements in Σ and T as

$$\begin{aligned}
\Sigma_{us} &= \alpha_1^T \Sigma_{ss}, \\
T_{u2} &= \alpha_2^T T_{22} + \alpha_1^T T_{s2}, \\
T_{us} &= \alpha_1^T T_{ss} + \alpha_2^T T_{2s}, \\
\Sigma_{uu} &= 1 + \alpha_1^T \Sigma_{ss} \alpha_1.
\end{aligned} \tag{3.27}$$

Let $W(A)$ denote a vector by horizontally arranging the elements in the matrix A and $\gamma = (\beta_{r0}, \beta_{r1}, \beta^{**})$ in which $\beta^{**} = [\beta_u^T \ W(\beta_{s1})^T \ W(\beta_{s2})^T \ W(\beta_{22})^T]^T$. The arrangement makes us easily extract the covariances between $W(\beta_{s1})$, $W(\beta_{s2})$, $W(\beta_{22})$ and α_1, α_2 to estimate the variances of α_3, α_4 and D by Delta method. let $H_i = O_i \oplus_{j=1}^3 H_{ij}$ with $H_{i1} = [1_{k_i} \otimes I_J \ (W_{ui}\beta_u) \otimes I_J]$, $H_{i2} = [1_{k_i} \otimes I_{p_1} \ W_{1i} \otimes I_{p_1} \ W_{2i} \otimes 1_{k_i} \otimes I_{p_1}]$, and $H_{i3} = [I_{p_2} \ I_{p_2} \otimes W_{2i}]$, $F_i = O_i \oplus_{j=1}^3 F_{ij}$ with $F_{i1} = [1_{k_i} \otimes I_J \ W_{ui} \otimes \beta_{r1}]$, $F_{i2} = H_{i2}$, and $F_{i3} = H_{i3}$, $G_i = H_i \begin{bmatrix} I_J \\ 0_{(J+p_3 p_1 + p_4 p_2) \times J} \end{bmatrix}$,

$M_i = H_i \begin{bmatrix} 0_{J \times J} \\ I_J \\ 0_{(p_3 p_1 + p_4 p_2) \times J} \end{bmatrix}$, and $Q_i = F_i \begin{bmatrix} 0_{J \times (p_3 + p_3 p_1 + p_4 p_2)} \\ I_{p_3 + p_3 p_1 + p_4 p_2} \end{bmatrix}$. The expected Fisher informa-

tion matrix for the MLE of $\gamma = (\beta_{r0}, \beta_{r1}, \beta^{**})$ is

$$I_{\gamma\gamma} = \sum_{i=1}^{n} \begin{bmatrix} G_i^T(V_i^\circ)^{-1}G_i & G_i^T(V_i^\circ)^{-1}M_i & G_i^T(V_i^\circ)^{-1}Q_i \\ M_i^T(V_i^\circ)^{-1}G_i & A + M_i^T(V_i^\circ)^{-1}M_i & M_i^T(V_i^\circ)^{-1}Q_i \\ Q_i^T(V_i^\circ)^{-1}G_i & Q_i^T(V_i^\circ)^{-1}M_i & Q_i^T(V_i^\circ)^{-1}Q_{i,} \end{bmatrix} \tag{3.28}$$

where A has its $(j,k)th$ component $\frac{1}{2}\mathrm{tr}\left((V_i^\circ)^{-1}\frac{\partial V_i^\circ}{\partial \beta_{rj}}(V_i^\circ)^{-1}\frac{\partial V_i^\circ}{\partial \beta_{rk}}\right)$.

Define $V(A)$ is a vector by vertically arranging the distinct elements in a matrix A. Let $\delta = (\xi, \tau, T_{uu}, V(T_{ss}), V(T_{2s}), V(T_{22}), V(\Sigma_{ss}), \alpha_1, \alpha_2)$, then

$$I_{\delta_j \beta_{rk}} = \frac{1}{2}\sum_{i=1}^{n}\mathrm{tr}\left((V_i^\circ)^{-1}\frac{\partial V_i^\circ}{\partial \delta_j}(V_i^\circ)^{-1}\frac{\partial V_i^\circ}{\partial \beta_{rk}}\right), \tag{3.29}$$

$$I_{\delta_j \delta_k} = \frac{1}{2}\sum_{i=1}^{n}\mathrm{tr}\left((V_i^\circ)^{-1}\frac{\partial V_i^\circ}{\partial \delta_j}(V_i^\circ)^{-1}\frac{\partial V_i^\circ}{\partial \delta_k}\right), \tag{3.30}$$

and $I_{\delta\beta_{r0}} = 0$, $I_{\delta\beta^{**}} = 0$, where

$$\frac{\partial V_i^\circ}{\partial \beta_{rj}} = O_i \begin{bmatrix} ((Z_{u_i}T_{uu}Z_{u_i}^T + \Sigma_{uu}I_{k_i}) \otimes (\beta_{r1}\Delta_j^T + \Delta_j\beta_{r1}^T)) & M_1 & (Z_{ui}T_{u2}) \otimes \Delta_j \\ M_1^T & 0 & 0 \\ (T_{2u}Z_{ui}^T) \otimes \Delta_j^T & 0 & 0 \end{bmatrix} O_i^T,$$

$$\frac{\partial V_i^\circ}{\partial \xi_j} = O_i \begin{bmatrix} (1_{k_i}1_{k_i}^T) \otimes \Delta_j & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} O_i^T, \quad \frac{\partial V_i^\circ}{\partial \tau_j} = O_i \begin{bmatrix} I_{k_i} \otimes \Delta_j & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} O_i^T,$$

$$\frac{\partial V_i^\circ}{\partial V(T_{uu})_j} = O_i \begin{bmatrix} \left(Z_{u_i}\frac{\partial T_{uu}}{\partial V(T_{uu})_j}Z_{u_i}^T\right) \otimes (\beta_{r1}\beta_{r1}^T) & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} O_i^T,$$

46

$$\frac{\partial V_i^\circ}{\partial V(\Sigma_{ss})_j} = O_i \begin{bmatrix} \left(\alpha_1^T \frac{\partial \Sigma_{ss}}{\partial V(\Sigma_{ss})_j} \alpha_1 I_{k_i}\right) \otimes (\beta_{r1}\beta_{r1}^T) & I_{k_i} \otimes (\beta_{r1}\alpha_1^T \frac{\partial \Sigma_{ss}}{\partial V(\Sigma_{ss})_j}) & 0 \\ I_{k_i} \otimes (\frac{\partial \Sigma_{ss}}{\partial V(\Sigma_{ss})_j} \alpha_1 \beta_{r1}^T) & I_{k_i} \otimes \frac{\partial \Sigma_{ss}}{\partial V(\Sigma_{ss})_j} & 0 \\ 0 & 0 & 0 \end{bmatrix} O_i^T,$$

$$\frac{\partial V_i^\circ}{\partial V(T_{ss})_j} = O_i \begin{bmatrix} 0 & M_2 & 0 \\ M_2^T & (1_{k_i}1_{k_i}^T) \otimes \frac{\partial T_{ss}}{\partial V(T_{ss})_j} & 0 \\ 0 & 0 & 0 \end{bmatrix} O_i^T,$$

$$\frac{\partial V_i^\circ}{\partial V(T_{s2})_j} = O_i \begin{bmatrix} 0 & M_3 & \left(Z_{ui}\alpha_1^T \frac{\partial T_{s2}}{\partial V(T_{s2})_j}\right) \otimes \beta_{r1} \\ M_3^T & 0 & 1_{k_i} \otimes \frac{\partial T_{s2}}{\partial V(T_{s2})_j} \\ \left(\frac{\partial T_{2s}}{\partial V(T_{s2})_j} \alpha_1 Z_{ui}^T\right) \otimes \beta_{r1}^T & 1_{k_i}^T \otimes \frac{\partial T_{2s}}{\partial V(T_{s2})_j} & 0 \end{bmatrix} O_i^T,$$

$$\frac{\partial V_i^\circ}{\partial V(T_{22})_j} = O_i \begin{bmatrix} 0 & 0 & \left(Z_{ui}\alpha_2^T \frac{\partial T_{22}}{\partial V(T_{22})_j}\right) \otimes \beta_{r1} \\ 0 & 0 & 0 \\ \left(\frac{\partial T_{22}}{\partial V(T_{22})_j} \alpha_2 Z_{u_i}^T\right) \otimes \beta_{r1}^T & 0 & \frac{\partial T_{22}}{\partial V(T_{22})_j} \end{bmatrix} O_i^T,$$

$$\frac{\partial V_i^\circ}{\partial \alpha_1} = O_i \begin{bmatrix} \left(2\delta_{1j}^T \Sigma_{ss}\alpha_1 I_{k_i}\right) \otimes (\beta_{r1}\beta_{r1}^T) & M_4 & (Z_{ui}\delta_{1j}^T T_{s2}) \otimes \beta_{r1} \\ M_4^T & 0 & 0 \\ (T_{2s}\delta_{1j}Z_{ui}^T) \otimes \beta_{r1}^T & 0 & 0 \end{bmatrix} O_i^T,$$

$$\frac{\partial V_i^\circ}{\partial \alpha_{2j}} = O_i \begin{bmatrix} 0 & M_5 & (Z_{ui}\delta_{2j}^T T_{22}) \otimes \beta_{r1} \\ M_5^T & 0 & 0 \\ (T_{22}\delta_{2j}Z_{ui}^T) \otimes \beta_{r1}^T & 0 & 0 \end{bmatrix} O_i^T,$$

where Δ_j, δ_{1j}, δ_{2j} are J-by-1, p_1-by-1, and p_2-by-1 vectors with jth element equal to one and zero otherwise, $M_1 = ((Z_{ui}T_{us}) \otimes \Delta_j)(1_{k_i}^T \otimes I_{p_1}) + I_{k_i} \otimes (\Delta_j \Sigma_{us})$, $M_2 = \left((Z_{ui}\alpha_1^T \frac{\partial T_{ss}}{\partial V(T_{ss})_j}) \otimes \beta_{r1}\right)$ $(1_{k_i}^T \otimes I_{p_1})$, $M_3 = \left((Z_{ui}\alpha_2^T \frac{\partial T_{2s}}{\partial V(T_{s2})_j}) \otimes \beta_{r1}\right)(1_{k_i}^T \otimes I_{p_1})$, $M_4 = I_{k_i} \otimes (\beta_{r1}\delta_{1j}^T \Sigma_{ss}) + \left((Z_{ui}\delta_{1j}^T T_{ss}) \otimes \beta_{r1}\right)$ $(1_{k_i}^T \otimes I_{p_1})$, and $M_5 = \left((Z_{ui}\delta_{2j}^T T_{2s}) \otimes \beta_{r1}\right)(1_{k_i}^T \otimes I_{p_1})$. Note that the above formulas have unknown terms $\frac{\partial \Sigma_{ss}}{\partial V(\Sigma_{ss})_j}$, $\frac{\partial T_{ss}}{\partial V(T_{ss})_j}$, $\frac{\partial T_{s2}}{\partial V(T_{s2})_j}$, and $\frac{\partial T_{22}}{\partial V(T_{22})_j}$. We know for any p-by-p matrix ϖ_1 the first deriva-

tive of the $(l, k)th$ $(k > l)$ element is

$$\frac{\partial \varpi_1}{\partial \varpi_{1kl}} = \begin{cases} \delta_k \delta_l^T + \delta_l \delta_k^T & k > l \\ \\ \delta_k \delta_l^T & k = l, \end{cases} \tag{3.31}$$

and for any p-by-q $(p \neq q)$ matrix ϖ_2 the first derivative of the $(l, k)th$ element is

$$\frac{\partial \varpi_2}{\partial \varpi_{2kl}} = \delta_k \eta_l^T, k = 1, \cdots, p, l = 1, \cdots, q \tag{3.32}$$

where δ_h and η_h are p-by-1 and q-by-1 vectors with the hth element equal to one and zero otherwise, respectively. After we vertically arrange the distinct elements in ϖ_1 and ϖ_2, the first derivative of the jth element for $j = 1, \cdots, p(p+1)/2$ or $j = 1, \cdots, pq$ has a one-to-one transformation with equations (3.31) and (3.32), respectively.

3.8.4 *The Variance Calculation of the Parameters in the LVMs*

The variances of the estimators α_1, α_2, β_0, β_1, ξ and τ in the LVMs can be estimated from Section 3.8.3. The variances of the other estimates can be calculated by Delta method. Let $\theta_1 = [\beta_{u1}^T \ W(\beta_{s1})^T \ \alpha_1^T]^T$, $\theta_2 = [\beta_{u2}^T \ W(\beta_{s2})^T \ W(\beta_{22})^T \ \alpha_1^T \ \alpha_2^T]^T$, and $\theta_3 = [T_{uu} \ V(T_{ss})^T \ V(T_{s2})^T \ V(T_{22})^T \ \alpha_1^T \ \alpha_2^T]^T$. From the transformation formulas (3.7) and Delta method, the covariances of $\hat{\alpha}_3$, $\hat{\alpha}_4$, and $\hat{D}$ with $Z_{uik} = 1$ are calculated as

$$\begin{aligned} \text{cov}\hat{\alpha}_3 &= \widehat{\nabla f}_1 \text{cov}\hat{\theta}_1 \widehat{\nabla f}_1^T, \\ \text{cov}\hat{\alpha}_4 &= \widehat{\nabla f}_2 \text{cov}\hat{\theta}_2 \widehat{\nabla f}_2^T, \\ \text{cov}\hat{D} &= \widehat{\nabla f}_3 \text{cov}\hat{\theta}_3 \widehat{\nabla f}_3^T, \end{aligned} \tag{3.33}$$

where $\text{cov}\hat{\theta}_i$ $(i = 1, \cdots, 3)$ can be extracted from the inverse of the expected Fisher information matrix in Section 3.8.3 and

$$\nabla f_1 = \left[I_{p_3} \quad -\alpha_1^T \otimes I_{p_3} \quad -\beta_{s_1}^T \right],$$

$$\nabla f_2 = [I_{p_4} \quad -\left(\alpha_1^T \otimes I_{p_4}\right) \quad -\left(\alpha_2^T \otimes I_{p_4}\right) \quad -\beta_{s_2}^T \quad -\beta_{22}^T],$$

$$\nabla f_3 = \left[1 \quad \left(\frac{\partial D}{\partial V(T_{ss})}\right)^T \quad \left(\frac{\partial D}{\partial V(T_{s2})}\right)^T \quad \left(\frac{\partial D}{\partial V(T_{22})}\right)^T \quad \left(\frac{\partial D}{\partial \alpha_1}\right)^T \quad \left(\frac{\partial D}{\partial \alpha_2}\right)^T \right]$$

with

$$\frac{\partial D}{\partial V(T_{ss})_j} = -\alpha_1^T \frac{\partial T_{ss}}{\partial V(T_{ss})_j} \alpha_1,$$

$$\frac{\partial D}{\partial V(T_{s2})_j} = -2\alpha_1^T \frac{\partial T_{s2}}{\partial V(T_{s2})_j} \alpha_2,$$

$$\frac{\partial D}{\partial V(T_{22})_j} = -\alpha_2^T \frac{\partial T_{22}}{\partial V(T_{22})_j} \alpha_2,$$

$$\left(\frac{\partial D}{\partial \alpha_1}\right)^T = -2\alpha_2^T T_{2s} - 2\alpha_1^T T_{ss},$$

$$\left(\frac{\partial D}{\partial \alpha_2}\right)^T = -2\alpha_2^T T_{22} - 2\alpha_1^T T_{s2}.$$

The terms $\frac{\partial T_{ss}}{\partial V(T_{ss})_j}$, $\frac{\partial T_{s2}}{\partial V(T_{s2})_j}$, and $\frac{\partial T_{22}}{\partial V(T_{22})_j}$ are described in Section 3.8.3. Similarly, using multivariate Delta method, we could derive the variances of distinct elements $V(D)$ of D if we fit a random-coefficient (3.1).

Figure 3.1: *Illustration of the structure of the latent variable models*
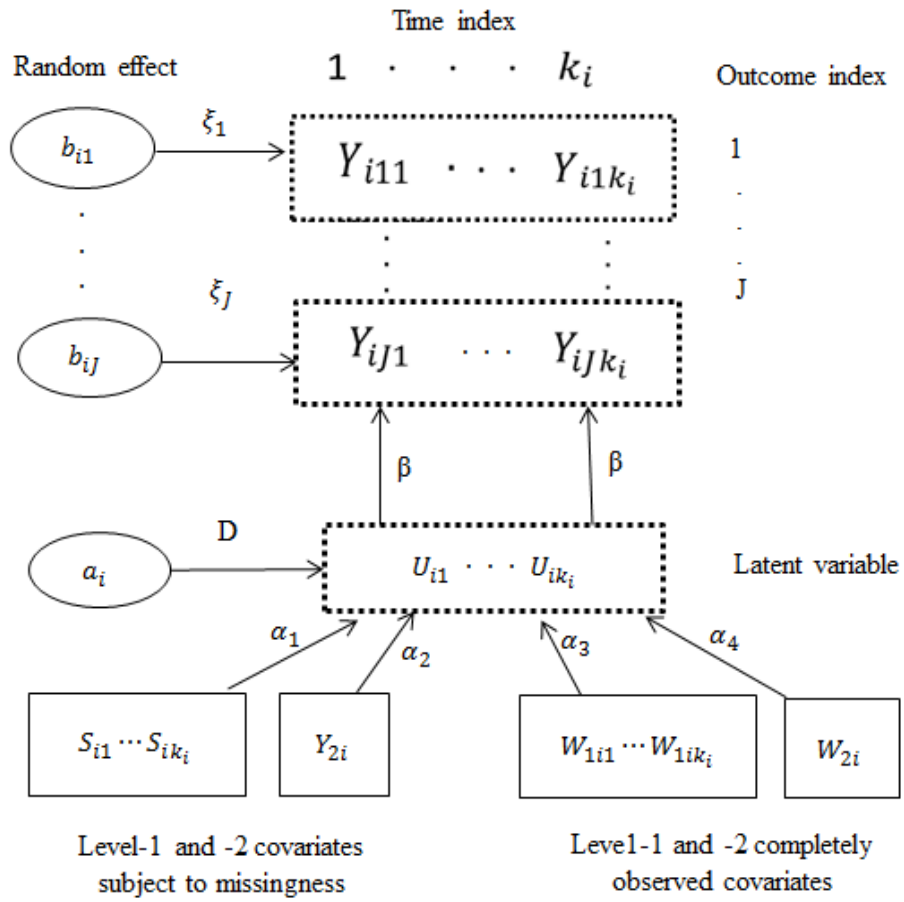
Table 3.1: *Estimation of the simulated LVMs (11) by three different estimation methods*

Model	Para.	True value	Benchmark	Just-identified	Over-identified
			Estimate (S.E.[a])		
(1)	α	1	1.031 (0.075)	1.032 (0.075)	**0.901** (0.065)
		1	1.007 (0.024)	1.007 (0.024)	**0.882** (0.021)
	D	1	0.999 (0.069)	0.999 (0.069)	**0.751** (0.052)
(2)	β_{r0}	1	0.993 (0.054)	0.993 (0.054)	0.993 (0.054)
		1	1.026 (0.055)	1.026 (0.055)	1.026 (0.054)
	β_{r1}	1	0.987 (0.014)	0.987 (0.014)	**1.129** (0.016)
		1	0.987 (0.014)	0.987 (0.014)	**1.129** (0.016)
	ξ	0.25	0.268 (0.035)	0.268 (0.035)	0.267 (0.035)
		0.25	0.291 (0.035)	0.291 (0.036)	0.291 (0.036)
	τ	0.25	0.240 (0.018)	0.240 (0.018)	0.240 (0.018)
		0.25	0.258 (0.019)	0.258 (0.019)	0.258 (0.018)

[a] standard error

Table 3.2: *NGHS data for analysis*

level	variable	description	mean (S.E.)	missing (%)
	BMI	BMI(kg/m^2)	22.42 (5.81)	308 (1.5)
	Skinfold	sum of skinfolds (mm)	45.11 (24.88)	783 (3.8)
	Waist	max. below-waist circumference(cm)	93.95 (12.87)	2807 (13.5)
level 1	PercentFat	percent fat by BIA	25.29 (11.49)	1694 (8.1)
	AGE	age in years at time of visit	14.36 (2.99)	0 (0.0)
	TV	TV watching (hours/week)	31.35 (21.32)	4834 (23.2)
	PhysicalAct	physical activity pattern score	17.35 (17.75)	6573 (31.5)
	Maturation	maturation stages	2.10 (1.03)	1063 (5.1)
	MotherBMI	mother's BMI	27.35 (6.91)	6772 (32.4)
	ParentEd	maximum parental education	0.75 (0.43)	0 (0.0)
level 2	Income	household income	1.06 (0.83)	1156 (5.5)
	RACE	race (white/black)	0.48 (0.50)	0 (0.0)
	NumParents	the number of parents	0.31 (0.46)	0 (0.0)

Table 3.3: *Parameter estimates and their estimated standard errors in model (3.1)*

Parameter	Covariate	Estimate(S.E.)	
		Random intercept	Random slope
α_1	TV	0.004**(0.001)	0.004**(0.001)
	PhysicalAct	-0.003**(0.001)	-0.002*(0.001)
	Maturation	0.347**(0.021)	0.387**(0.024)
α_2	MotherBMI	0.150**(0.011)	0.133**(0.013)
	Income	-0.183 (0.096)	0.078 (0.114)
α_3	AGE	0.502 (0.020)	0.713 (0.024)
	AGE^2	-0.025** (0.005)	-0.031** (0.005)
	AGE×White	-0.057* (0.026)	-0.124** (0.033)
α_4	ParentEd	0.012 (0.155)	0.144 (0.179)
	White	-0.309 (0.137)	-0.938 (0.186)
	OneParent	0.380*(0.159)	0.568**(0.185)
D_{00}^a		8.040 (0.386)	16.482(0.560)
D_{01}			0.942 (0.043)
D_{11}			0.155 (0.006)

[a] $D_{00} = D$ in a random-intercept model (3.1)

[*] p-value< 0.05, [**] p-value< 0.01

Table 3.4: *Parameter estimates and their estimated standard errors in model (3.2)*

Model (3.1) with	Biomarker	$\hat{\beta}_{r0j}$	$\hat{\beta}_{r1j}$	$\hat{\tau}_j$	$\hat{\xi}_j$
	BMI	22.74 (0.09)	1.46 (0.01)	1.06 (0.02)	1.07 (0.09)
random	Skinfold	47.30 (0.37)	5.24 (0.04)	75.05 (0.83)	73.38 (2.73)
intercept	Waist	93.46 (0.29)	4.37 (0.03)	2.02 (0.10)	29.72 (1.16)
	PercentFat	25.88 (0.19)	2.69 (0.02)	15.29 (0.18)	21.00 (0.75)
	BMI	22.74 (0.09)	1.08 (0.01)	0.54 (0.01)	0.86 (0.08)
random	Skinfold	47.37 (0.38)	3.95 (0.03)	65.06 (0.73)	69.98 (2.56)
coefficient	Waist	93.49 (0.28)	3.04 (0.02)	6.04 (0.11)	24.79 (0.97)
	PercentFat	25.85 (0.19)	1.95 (0.02)	15.40 (0.18)	21.88 (0.76)

Figure 3.2: *Obesity growth curves for blacks and whites*
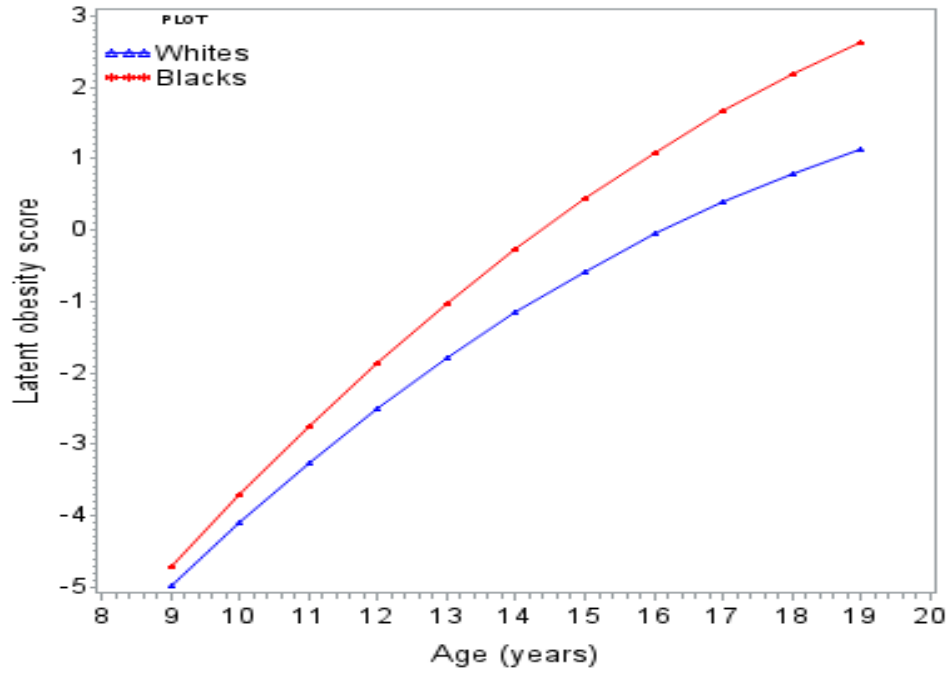
Figure 3.3: *Estimated latent scores at each age for each race*

53
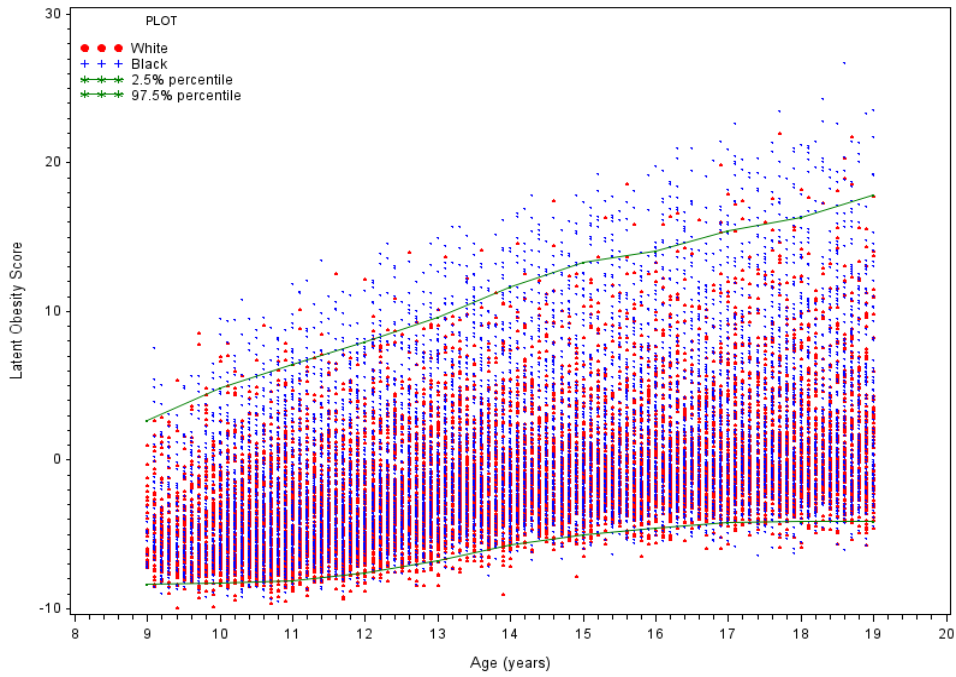
4 Three-Level Latent Variable Analysis Given Incompletely Observed Multivariate Markers in a Cluster-Randomized Study

4.1 Introduction

Multilevel data arise when units nest within clusters. It is of interest to study what risk factors and surrogate outcomes are associated with a latent variable like academic achievement or treatment effectiveness, in the setting of multilevel data. It is necessary to develop multilevel methods for this scenario, in particular, we cannot expect to include all cluster-specific influences as covariates in the analysis. This chapter is to extend the two-level LVMs to three-level LVMs that are implemented to study racial disparities in the academic achievement. Disparities in achievement scores between African-American and white students have been published for several decades. They declined steadily for most of the 20th century, but this progress has been halted or even reversed in recent years (Neal and Johnson, 1996). Understanding the reasons and utilizing effective strategies are crucial for designing policies to reduce racial inequality in achievement score and, therefore, potentially improves the well-being of African-American students later. A meta-analysis of several hundred studies by Glass and Smith (1978) and review by Robinson (1990) summarized that the small class size had a positive effect on student achievement. Many of the studies have poor quality, however, and none of them was a randomized experiment.

Tennessee's Student/Teacher Achievement Ratio Study (STAR) has been widely regarded as one of the most important experiments on education research (Mosteller, 1995). The data are publicly available and have been intensively investigated the relationship between small class size and academic achievement. Education researchers reported that there was a significant relationship between small class size and high academic performance, and small class size was beneficial to minority (Finn and Achilles, 1990; Goldstein and Blatchford, 1998; Goldstein et al., 2000; Krueger, 1999; Krueger and Whitmore, 2001; Mosteller, 1995; Nye et al., 1999, 2000, 2004; Shin,

2012; Shin and Raudenbush, 2011), though some other researchers disagreed with this proposition (Hanushek, 1999; Milesi and Gamoran, 2006). Some of these studies generated results from the analysis of a univariate outcome by completely observed cases or an ad hoc imputation like sample mean substitution. Such estimation requires a strong assumption of data MCAR (Heitjan and Basu, 1996; Little and Rubin, 2002; Rubin, 1976; Schafer, 1997). The univariate analysis under the assumption of MCAR is, in general, inefficient and might result in biased inferences. The other studies proposed multivariate simultaneous equation model to investigate the relationship between class size and achievement scores (Shin, 2012; Shin and Raudenbush, 2011) with a comparatively weak assumption of data MAR (Little and Rubin, 2002; Rubin, 1976). This method has three limitations. First, it neglects to address a covariate effect on the overall achievement score. Secondly, it does not account for the feature that the surrogate outcomes measure an overall interest, and, therefore, can not provide an estimate of a unit-specific achievement score. Finally, it needs more degrees to test coefficient effects and cannot identify if the treatment effect is random across schools. Although one can perform an analysis assuming a common effect on all outcomes, this assumption is inappropriate and misleading especially when the outcomes are measured on different scales and units.

In this chapter, we formulate simultaneous three-level LVMs: a measurement model where multivariate surrogate outcomes measure the latent achievement score with error and a structure model where the latent achievement score is associated with some potential risk factors (Laird and Ware, 1982; Pocock, Geller, and Tsiatis, 1987; Roy and Lin, 2000; Sammel and Ryan, 1996). We use these models to analyze third graders attending 75 elementary schools in Tennessee. In the fall before the school year started, the third graders were randomly assigned to classes of the treatment group (small class size) or control group (regular class size) within each school in STAR. Four surrogate outcomes are reading (R_1), math (R_2), listening (R_3), and word recognition (R_4) skill scores, which are highly correlated and thought to measure academic achievement accurately. Race (black/white) and treatment (yes/no) are possible risk factors for the academic achievement. Some

subjects have missed all scores. Consequently, in this cluster-randomized study, the data consist of 1270 third graders attending 310 classes in 75 schools. We estimate these models simultaneously to produce efficient estimates by ML via the PX-EM algorithm. The method extends the mixed linear models from Laird and Ware (1982) and Shah, Laird and Schoenfiled (1997), and EM-algorithm for LVMs in a longitudinal study proposed by Roy and Lin (2000).

The objectives of this chapter are (1) to identify surrogate outcomes associated with academic achievement, (2) to identify risk factors of academic achievement, and (3) to provide a unit-specific achievement score. Chapter 2 has discussed that it is challenging to achieve these goals due to the latent variable and the various surrogate outcomes measured with errors. The approach becomes challenging for three-level LVMs. In this section, we apply the EM and PX-EM algorithms to the three-level LVMs by assuming the multiple surrogate outcomes MCAR or MAR. In both cases, competing models are compared to identify if the magnitude of minority advantages vary significantly across schools. Section 4.2 introduces the model. Section 4.3 describes the EM and PX-EM algorithms. Section 4.4 analyzes the STAR data by the two algorithms. Section 4.5 concludes the chapter with a short discussion. Finally, Section 4.6 describes some detailed mathematical derivations.

4.2   Three-level Latent Variable Models

This section extends the two-level simultaneous equation models developed by Roy and Lin (2000) to three-level data. The structural model for the academic achievement is

$$U_{ikl} = X_{ikl}\alpha + E_{ikl}\lambda_i + Z_{ikl}a_{ik} + \epsilon_{ikl}, \tag{4.1}$$

where U_{ikl} is a univariate latent score, X_{ikl} is a vector of covariates having fixed effects α, E_{ikl} is a vector of covariates having level-3 unit-specific random effects $\lambda_i \overset{iid}{\sim} N(0, \Gamma)$, Z_{ikl} is a vector of covariates having level-2 unit-specific random effects $a_{ik} \overset{iid}{\sim} N(0, D)$. Both λ_i and a_{ik} are

independent of a level-1 unit-specific random error $\epsilon_{ikl} \overset{iid}{\sim} N(0,1)$ for level-1 unit $l = 1, 2, \cdots, n_{ik}$ nested within level-2 unit $k = 1, 2, \cdots, n_i$ nested within level-1 unit $i = 1, 2, \cdots, n$. If the latent score U_{ikl} were observed, we would be able to estimate the model by standard multilevel software. With the response variable unobservable, there are observable surrogate outcomes that are highly correlated and supposed to predict the latent score with accuracy. That is that the latent score is related to the surrogate outcomes by

$$R_{ijkl} = \beta_{0j} + \beta_{1j}U_{ikl} + c_{ij} + b_{ijk} + e_{ijkl}, \tag{4.2}$$

where R_{ijkl} are observable surrogate outcomes, $\beta_r = [\beta_{0j} \ \beta_{1j}]^T$ is a vector of regression coefficients for the jth surrogate outcome, $c_{ij} \overset{iid}{\sim} N(0, T_j)$, $b_{ijk} \overset{iid}{\sim} N(0, \xi_j)$, and $e_{ijk} \overset{iid}{\sim} N(0, \tau_j)$ are level-3, level-2, and level-1 unit-specific random effects, respectively. Given the latent variable U_{ikl}, the surrogate outcomes R_{ijkl} are mutually independent. We further assume the b_{ijk} are independent. To make parameters identifiable, we assume ϵ_{ikl} is distributed as $N(0,1)$ and X_{ikl} does not contain an intercept.

It is essential to aggregate models (4.1) and (4.2) at individual level for deriving estimates and their variances. For $k = 1, 2, \cdots, n_i$ and $J = 1, 2, \cdots, J$, let $U_{ik} = [U_{ik1} \ U_{ik2} \cdots U_{ikn_{ik}}]^T$, $U_i = [U_{i1}^T \ U_{i2}^T \cdots U_{in_i}^T]^T$, $R_{ijk} = [R_{ijk1} \ R_{ijk2} \cdots R_{ijkn_{ik}}]^T$, $R_{ij} = [R_{ij1}^T \ R_{ij2}^T \cdots R_{ijn_i}^T]^T$, $R_i = [R_{i1}^T \ R_{i2}^T \cdots R_{iJ}^T]^T$ and $\epsilon_i, X_i, e_i, b_i, c_i, E_i$ defined similarly. Let $\beta_0 = [\beta_{01} \ \beta_{02} \cdots \beta_{0J}]^T$ with β_1 similarly defined. Then we can write models (4.1) and (4.2) in matrix notation as

$$\begin{aligned}
R_i &= \beta_0 \otimes 1_{m_i} + \beta_1 \otimes U_i + c_i \otimes 1_{m_i} + (I_J \otimes W_i)b_i + e_i, \\
U_i &= X_i\alpha + E_i\lambda_i + Z_ia_i + \epsilon_i,
\end{aligned} \tag{4.3}$$

where $\otimes$ represents Kronecker product, $b_i = [b_{i1} \ b_{i2} \cdots b_{iJ}]^T \sim N(0, R(\xi) \otimes I_{n_i})$, $m_i = \sum_{k=1}^{n_i} n_{ik}$, $a_i = [a_{i1} \ a_{i2} \cdots a_{in_i}]^T$, $Z_i = \oplus_{k=1}^{n_i} Z_{ik}$, and $W_i = \oplus_{k=1}^{n_i} 1_{n_{ik} \times 1}$ for $b_{ij} = [b_{ij1} \ b_{ij2} \cdots b_{ijn_i}]^T$,

$R(\xi) = \oplus_{j=1}^{J} \xi_j$, $Z_{ik} = [Z_{ik1} \ Z_{ik2} \cdots Z_{ikn_{ik}}]^T$, and $\oplus_{j=1}^{J} A_j = \text{diag}(A_1, A_2, \cdots, A_J)$ representing a diagonal matrix with diagonal elements or submatrices $(A_1, A_2, \cdots, A_J)$ and all the other elements equal to zero. It follows $R_i \sim N(\mu_i, V_i)$ with

$$\mu_i = \beta_0 \otimes 1_{m_i} + \beta_1 \otimes (X_i \alpha),$$
$$V_i = (\beta_1 \beta_1^T) \otimes \text{cov}(U_i) + R(T) \otimes (1_{m_i} 1_{m_i}^T) + R(\tau) \otimes I_{m_i} + R(\xi) \otimes (W_i W_i^T) \tag{4.4}$$

where $\text{cov}(U_i) = Z_i D Z_i^T + E_i \Gamma E_i^T + I_{m_i}$, $R(T) = \oplus_{j=1}^{J} T_j$, and $R(\tau) = \oplus_{j=1}^{J} \tau_j$.

Following Shin and Raudenbush (2007), for unit i suppose we have $m_{ij} \leq m_i$ students on the jth surrogate outcome. Let O_{ij} be an index matrix to indicate the students when the jth ($j = 1, 2, \cdots, J$) surrogate outcome is observed. Specifically, O_{ij} is a $m_{ij} \times m_i$ matrix constructed by deleting rows of I_{m_i} which are corresponding to the missing observations on the jth surrogate outcome. Hence, $R_{ij}^{\circ} = O_{ij} R_{ij}$. Given $O_i = \oplus_{j=1}^{4} O_{ij}$, then $R_i^{\circ} = O_i R_i$. It follows that the marginal distribution of R_i° is multivariate normal with mean and covariance

$$\mu_i^{\circ} = E(R_i^{\circ}) = O_i E(R_i),$$
$$V_i^{\circ} = \text{cov}(R_i^{\circ}) = O_i \text{cov}(R_i) O_i^T. \tag{4.5}$$

4.3 EM and PX-EM Algorithms

Like the two-level LVMs, it is challenging to estimate the model (4.3) by directly using the actual log likelihood since β_1 enters both the marginal mean and variance of R_i. EM algorithm is implemented to estimate the model because R_i is conditionally independent given U_i. In the EM algorithm we treat the latent variables U_i, the random effects a_i, b_i, c_i and λ_i as missing data. Therefore, the complete data are $(R_i, X_i, Z_i, U_i, a_i, b_i, c_i, \lambda_i)$ and the observed data are R_i°. Given the initial values of the parameters, the EM algorithm iterates between its E- and M-steps until

convergence. The E-step takes expectations of the sufficient statistics of the complete-data log likelihood, given the observed data. The M-step maximizes the expected complete-data log likelihood given parameters from the previous iteration. The method (detailed in Subsections 4.6.1 and 4.6.2) is

E step : Calculate the conditional expectations related to the latent variable U_i, and

the random terms $a_i, b_i, e_{ij}, \epsilon_i, \lambda_i, c_{ij}$ and $U_i^T e_{ij}$;

M step : Maximize the model parameters from the complete-data log likelihood.

The variances of the parameter estimators are computed by the expected Fisher information matrix (Section 4.6.3) based on the marginal log likelihood of R_i° at convergence.

A criticism of the EM algorithm is its slow convergence to MLEs, which has already demonstrated for NGHS longitudinal study in Chapter 2. We implement the PX-EM algorithm and extend it to the three-level LVMs for the cluster-randomized STAR data to check how faster the PX-EM algorithm is than the EM-algorithm. The PX-EM algorithm is applied to the models (4.1) and (4.2) where the only change is an extension of the parameter $\epsilon_{ik} \overset{iid}{\sim} N(0, \sigma^2)$. The PX-EM algorithm is

PX-E step: This is unchanged from EM;

PX-M step: Estimate model parameters in the expanded space

$$\gamma_\star^t = \left(\beta_0^t, \beta_1^t, \alpha_t, \tau_t, D_t, \xi_t, \sigma_t^2, \Gamma_t, T_t \right)$$

and reduce $\gamma_\star^t$ to the desired model parameters by the following modification

$$\gamma^t = \left(\beta_0^t, \beta_1^t \sigma_t, \frac{\alpha_t}{\sigma_t}, \tau_t, \frac{D_t}{\sigma_t^2}, \xi_t, \sigma^2 = 1, \frac{\Gamma_t}{\sigma_t^2}, T_t \right).$$

Note that the general idea of two-level and three-level LVMs are similar except for estimating more components in the three-level LVMs.

4.4 Data Analysis

In this section, we implement the EM and PX-EM algorithms to analyze the STAR data described in the Introduction. Their initial values are calculated by the same approach as these in Chapter 2 for the two-level LVMs. Because the class type is randomly assigned to a class, it may have school-level confounders. In order to assess if such confounding seriously biases inferences, two-level LVMs are assessed and compared, which control for all school-level covariates by fixing school effects.

Surrogate outcomes R_1, R_2, R_3, R_4 are highly correlated with correlations ranging 0.49 to 0.87, and are useful to assess academic achievement. We are interested in if African-American and white students have a differential treatment effect on the latent score and if this differential effect is random across classes. Therefore, we fit the models (4.1) and (4.2) simultaneously by ML via the EM and PX-EM algorithm for two cases: (1) $Z_{ik\ell} = 1$, and (2) $Z_{ik\ell} = [1 \ B_{ik\ell} \times T_{ik\ell}]$ and $D = \begin{bmatrix} D_{00} & D_{01} \\ D_{10} & D_{11} \end{bmatrix}$ for $X_{ik\ell} = [B_{ik\ell} \ T_{ik\ell} \ B_{ik\ell} \times T_{ik\ell}]$. We call the LVMs with these two cases model A and model B, respectively. We analyze dummy indicator variables for treatment and African-American students. The surrogate outcomes and race-specific scores are summarized in Table 4.1. All race-specific scores are subject to missingness ranging from 11% to 25%. We analyze them under the assumption of MAR or MCAR and compare the results under these two assumptions.

We only shows the results by the PX-EM algorithm in Table 4.2 because the EM algorithm produces practically identical results. Table 4.2 lists the parameter estimates and their standard errors of the LVMs given surrogate outcomes MAR or MCAR. Under $H_0 : D_{01} = D_{11} = 0$, likelihood ratio tests give test statistics $0.1 \sim \chi_2^2$ and $0.01 \sim \chi_2^2$ with p-values> 0.05 under the assumptions of MCAR and MAR, respectively. The results show that the race-treatment effect is not randomly across classes. Our analysis shows four surrogate outcomes are positively associated with the latent score for both assumptions, but with the assumption of MCAR, all slopes of the

latent score appear to be underestimated, and the standard errors of the slopes are exaggerated. There is a significant interaction effect between race and treatment. The point estimate and its standard deviation under the assumption of MCAR are overestimated. Table 4.3 gives the treatment versus control effect on the latent score. We interpret the results under the comparatively weak assumption of MAR. African-American students in the treatment group have 0.32 units higher achievement score than these in the control group. After treatment, African-American students have 0.33 units higher achievement score than white students. Treatment seems not to have a significant effect on the achievement score for white students.

The estimation of unit-specific achievement scores is a by-product of the algorithms. Figure 4.1 shows the scatter plot of the scores against schools and Figure 4.2 is their QQ-plot. In Figure 4.1, the two horizontal yellow lines indicate the 2.5th and 97.5th percentiles of the academic achievement. While 2.5% of the subjects above the top line have higher achievement score than the others, 2.5% of the subjects below the bottom line have lower achievement score than the others. Identifying these subjects and studying their characteristics may be beneficial to reduce the gap of achievement between African-American and white students. The QQ plot of the academic achievement (Figure 4.2) allows researchers to examine the assumptions graphically and to identify cases for which the models provide a particularly poor fit. It appears to follow a 45-degree straight line fairly through the origin. Therefore, the assumption of normality for the latent variable might be tenable, and there is no evidence of some potential outliers.

To rule out confounders between T_{ik} and school-level covariates, two-level LVMs are fitted as

$$R_{kjl} = \beta_{0j} + \beta_{1j}U_{kl} + c_{kj} + e_{kjl}, \ c_{kj} \sim N(0, \xi_j), \ e_{kjl} \sim N(0, \tau_j), \tag{4.6}$$

$$U_{kl} = X_{kl}\alpha + Z_{kl}a_k + \epsilon_{kl}, \ a_k \sim N(0, D), \ \epsilon_{kl} \sim N(0, 1) \tag{4.7}$$

where $k = 1, 2, \cdots, n$ and $l = 1, 2, \cdots, n_k$, $X_{kl} = [B_{kl} \ T_k \ B_{kl} \times T_k \ S_k]$ with T_k is a vector of school indicators. If confounding is severe, then the results are differential between the two-

and three-level LVMs. Like three-level LVMs, two competing models are fitted under the assumptions of MAR and MCAR. The results are summarized in Table 4.4. Likelihood ratio test shows that the race-treatment effect is not randomly across classes for either the assumption of MAR or MCAR with p-value=0.62 and p-value=0.76, respectively. In Model (1), the results under the assumption of MCAR seem underestimate these under the assumption of MAR. In Model (2), the results appear under the assumption of MCAR overestimate these under the assumption of MAR. The estimated effects are in close range to their counterparts under three-level LVMs. The standard errors are relatively inflated under two-level LVMs, which indicates that three-level models are more efficient than two-level models. In addition, three-level models are more powerful to test coefficient effect than two-level model. Because the estimates are close to each other across the two- and three-level LVMs, the treatment effect due to the confounder of school-level variable seems implausible.

4.5   Discussion

In this chapter, we formulated a measurement model where four highly correlated observed surrogate outcomes measure the academic achievement with error and a structural model where the academic achievement is related to covariates (Roy and Lin, 2000). We analyzed third grade students from STAR and simultaneously estimated the models to yield valuable inferences by ML via the PX-EM algorithm. The convergence to ML by the PX-EM algorithm was shown to be 10% faster than that by the conventional EM algorithm. Complete-case analysis suffers from a loss of information due to discarding incomplete cases, which leads to a loss of precision and bias if missing surrogate outcomes are not MCAR, and the complete cases are not a random sample of all the cases. Under the assumption of surrogate outcomes MAR, we estimated the parameters conditionally on observed data to reduce bias due to missingness and to improve precision.

It is challenging to perform global testing for the outcomes since they are often measured at

different scales and units. In this chapter, we considered the LVMs as providing a framework to address this issue and enable a global examination of covariate effects on academic achievement. The results showed that the achievement score was positively associated with the four surrogate outcomes. Race and small class size were significantly associated with academic achievement. These findings imply policy makers that the reduced class size may be an important factor at improving African American education which is eventually useful to narrow or close the education gap between African American and white students. Small class may be introduced to be more inclusive of African American. Narrowing the gap not only moves the United States toward racial equality, but also has a significant positive economic and social impact.

Our estimates are based on ML assuming multivariate normality for the random effect at each level and constant variances for each surrogate outcome. The conclusions based on the LVMs depend for their validity on the tenability of assumptions about the structural and random parts. We generated QQ plots and scatter plots of empirical Bayesian residuals graphically to identify cases for which the models fit poorly. Overall the plots exhibited no clear pattern against normality, constant variance, and linearity.

Some limitations of this study should be noted. First, although the study is cluster-randomized, the substantial imbalance in baseline variables might occur by chance. The conclusions on the treatment effect might be confounded by such imbalance if not properly adjusted (Wei and Zhang, 2001). Therefore, we might account for baseline covariates and demographic characteristics in Model (4.1) to examine the treatment effect. Second, our study is limited to completely observed covariate analysis. The race-treatment effect on academic achievement in this article might not be the only or most interesting thing we want to evaluate. There is suggestive evidence that gender and social, economic status are reported to be significant for academic achievement (Goldstein and Blatchford, 1998; Guryan, Hurst, and Kearney, 2008; Krueger, 1999; Pamela, 2005). It will be interesting to develop an approach of handling missing surrogate outcomes and covariates. However, it is not our intention here to discuss these problems in depth. In our future agenda of research,

63

we will examine the simultaneous equations and the latent outcome by handling missing covariates efficiently under the assumption of data MAR. It is also of interest to extend the simultaneous equation models with discrete/ordinal surrogate outcomes (Zhang, Chen, and Albert, 2012). In this chapter, we found that the small class size caused higher academic achievement scores. In the next Chapter, we will continue our work to find if reduced class size has a causal effect on the academic achievement (Shin, 2012; Shin and Raudenbush, 2011).

4.6 Miscellanea

4.6.1 *Conditional Expectations*

The conditional expectations in E-step are

$$\tilde{U}_i = E(U_i|R_i^\circ) = X_i\alpha + \Lambda_i O_i^T(V_i^\circ)^{-1}(R_i^\circ - \mu_i^\circ),$$

$$E(U_i^T U_i|R_i^\circ) = \tilde{U}_i^T\tilde{U}_i + \text{tr}(\text{cov}(U_i|R_i^\circ)),$$

$$\tilde{a}_i = E(a_i|R_i^\circ) = (\beta_1^T \otimes ((I_{n_i} \otimes D)Z_i^T))O_i^T(V_i^\circ)^{-1}(R_i^\circ - \mu_i^\circ),$$

$$E(a_i a_i^T|R_i^\circ) = \tilde{a}_i\tilde{a}_i^T + I_{n_i} \otimes D - \Delta_{a_i}(V_i^\circ)^{-1}\Delta_{a_i}^T,$$

$$\tilde{b}_{ij} = E(b_{ij}|R_i^\circ) = \nu_1 O_i^T(V_i^\circ)^{-1}(R_i^\circ - \mu_i^\circ),$$

$$E(b_{ij}b_{ij}^T|R_i^\circ) = \tilde{b}_{ij}\tilde{b}_{ij}^T + \text{tr}\left(\xi_j I_{n_i} - \nu_1 O_i^T(V_i^\circ)^{-1}O_i\nu_1^T\right), \tag{4.8}$$

$$E(U_i^T e_{ij}|R_i^\circ) = \tilde{U}_i^T\tilde{e}_{ij} - \text{tr}(\Lambda_i O_i^T(V_i^\circ)^{-1}O_i\nu_2^T),$$

$$\tilde{\epsilon}_i = E(\epsilon_i|R_i^\circ) = \beta_1^T \otimes I_{m_i} O_i^T(V_i^\circ)^{-1}(R_i^\circ - \mu_i^\circ),$$

$$\tilde{e}_{ij} = E(e_{ij}|R_i^\circ) = \nu_2 O_i^T(V_i)^{-1}(R_i^\circ - \mu_i^\circ),$$

$$E(e_{ij}^T e_{ij}|R_i^\circ) = \tilde{e}_{ij}^T\tilde{e}_{ij} + \text{tr}(\tau_j I_{m_i} - \nu_2 O_i^T(V_i^\circ)^{-1}O_i\nu_2^T),$$

$$\tilde{\lambda}_i = E(\lambda_i|R_i^\circ) = \beta_1^T \otimes (\Gamma E_i^T)O_i(V_i^\circ)^{-1}(R_i^\circ - \mu_i^\circ),$$

$$E(\lambda_i\lambda_i^T|R_i^\circ) = \tilde{\lambda}_i\tilde{\lambda}_i^T + \beta_1^T \otimes (\Gamma E_i^T)O_i^T(V_i^\circ)^{-1}O_i\beta_1 \otimes (E_i\Gamma),$$

$$\tilde{c}_i = E(c_i|R_i^\circ) = R(T) \otimes 1_{m_i}^T O_i^T (V_i^\circ)^{-1} (R_i^\circ - \mu_i^\circ),$$

$$E(c_i^T c_i|R_i^\circ) = \tilde{c}_i \tilde{c}_i^T + R(T) - (R(T) \otimes 1_{m_i}^T) O_i^T (V_i^\circ)^{-1} O_i (R(T) \otimes 1_{m_i}),$$

where

$$\Delta_{a_i} = (\beta_1^T \otimes ((I_{n_i} \otimes D) Z_i^T)) O_i^T,$$

$$\Lambda_i = \beta_1^T \otimes (I_{m_i} + Z_i D Z_i^T + E_i \Gamma E_i^T),$$

$$\text{cov}(U_i|R_i) = I_{m_i} + Z_i D Z_i^T + E_i \Gamma E_i^T - \Lambda_i V_i^{-1} \Lambda_i^T,$$

$$\nu_1 = \left[(0_{n_i \times (j-1)n_i} \ \ \xi_j I_{n_i} \ \ 0_{n_i \times (J-j)n_i}) (I_J \otimes W_i^T) \right],$$

$$\nu_2 = [0_{m_i \times (j-1)m_i} \ \ \tau_j I_{m_i} \ \ 0_{m_i \times (J-j)m_i}].$$

4.6.2 *CD ML Estimates*

The complete-data log likelihood for $(R_i, U_i, b_i, a_i, c_i, \lambda_i)$ is, apart from a constant,

$$l = \sum_{i=1}^n (l(R_i|U_i, b_i, c_i) + l(U_i|a_i, \lambda_i) + l(a_i) + l(b_i) + l(c_i) + l(\lambda_i)), \qquad (4.9)$$

where $\xi = (\xi_1, \cdots, \xi_J)$, $\tau = (\tau_1, \cdots, \tau_J)$, $T = (T_1, \cdots, T_J)$, and

$$l(R_i|U_i, b_i, c_i) = \sum_{j=1}^J \left(-\frac{m_i}{2} \log \tau_j - \frac{1}{2\tau_j} e_{ij}^T e_{ij} \right),$$

$$l(U_i|a_i, \lambda_i) = -\frac{1}{2} \epsilon_i^T \epsilon_i,$$

$$l(a_i) = -\frac{1}{2} (n_i \log |D| + a_i^T D^{-1} a_i),$$

$$l(b_i) = -\frac{1}{2} (m_i \log |R(\xi)| + b_i^T R(\xi)^{-1} b_i),$$

$$l(c_i) = -\frac{1}{2} (n \log |R(T)| + c_i^T R(T)^{-1} c_i),$$

$$l(\Gamma) = -\frac{1}{2}(n \log |\Gamma| + \lambda_i^T \Gamma^{-1} \Lambda_i).$$

Differentiating (4.9) with the parameters β_0, β_1, α, ξ, τ, Γ, T and D, respectively, taking expectations of the resulting forms conditional to the observed data R_i°, setting them equal to zeros, and solving these equations, we know

$$\hat{\beta}_j^{(k)} = \hat{\beta}_j^{(k-1)} + \left(\sum_{i=1}^n E(U_{i*}^T U_{i*}|R_i^\circ) \right)^{-1} \sum_{i=1}^n E((U_{i*}^T e_{ij})|R_i^\circ),$$

$$\hat{\tau}_j = \frac{1}{\sum_{i=1}^n m_i} \times \sum_{i=1}^n E(e_{ij}^T e_{ij}|R_i^\circ),$$

$$\hat{\xi}_j = \frac{1}{\sum_{i=1}^n n_i} \sum_{i=1}^n E(b_{ij}^2|R_i^\circ), \tag{4.10}$$

$$\hat{\alpha}^{(k)} = \hat{\alpha}^{(k-1)} + \left(\sum_{i=1}^n X_i^T X_i \right)^{-1} \sum_{i=1}^n X_i^T \tilde{\epsilon}_i,$$

$$\hat{D} = \frac{1}{\sum_{i=1}^n n_i} \sum_{i=1}^n E(a_i a_i^T|R_i^\circ),$$

$$\hat{\Gamma} = \frac{1}{n} \sum_{i=1}^n E(\lambda_i \lambda_i^T|R_i^\circ),$$

$$\hat{T}_j = \frac{1}{n} \sum_{i=1}^n E(c_{ij} c_{ij}^T|R_i^\circ),$$

where $j = 1, \cdots, J$, $\beta_j = [\beta_{0j} \ \beta_{1j}]^T$, $U_{i*} = [1_{k_i} \ U_i]$, $E(b_{ij}^2|R_i^\circ)$ is the jth diagonal element in $E(b_i b_i^T|R_i^\circ)$ and $E(U_{i*}^T U_{i*}|R_i^\circ) = \begin{bmatrix} m_i & 1_{m_i}^T \tilde{U}_i \\ 1_{m_i}^T \tilde{U}_i & E(U_i^T U_i|R_i^\circ) \end{bmatrix}$, $E(U_{i*}^T e_{ij}|R_i^\circ) = \begin{bmatrix} 1_{k_i}^T \tilde{e}_{ij} \\ E(U_i^T e_{ij}|R_i^\circ)) \end{bmatrix}$.

4.6.3 *Calculations of the Information Matrix*

The information matrix is obtained by differentiating twice the log likelihood for the observed data R_i° with mean and variance given in (4.5) and taking an expectation of the resulting form. Let $G_i = O_i(I_J \otimes 1_{m_i})$, $H_i = O_i(\beta_1 \otimes X_i)$, and $M_i = O_i(I_J \otimes (X_i\alpha))$. The expected information

matrix for the MLE of $\theta_1 = (\beta_0, \beta_1, \alpha)$ is

$$
I_{\theta_1\theta_1} = \sum_{i=1}^{n} \begin{bmatrix} G_i^T(V_i^\circ)^{-1}G_i & G_i^T(V_i^\circ)^{-1}M_i & G_i^T(V_i^\circ)^{-1}H_i \\ M_i^T(V_i^\circ)^{-1}G_i & A + M_i^T(V_i^\circ)^{-1}M_i & M_i^T(V_i^\circ)^{-1}H_i \\ H_i^T(V_i^\circ)^{-1}G_i & H_i^T(V_i^\circ)^{-1}M_i & H_i^T(V_i^\circ)^{-1}H_i \end{bmatrix}, \tag{4.11}
$$

where A has its (i,k)th element $\frac{1}{2}tr\left((V_i^\circ)^{-1}(\partial V_i^\circ/\partial\beta_{1i}) \times (V_i^\circ)^{-1}(\partial V_i^\circ/\partial\beta_{1k})\right)$.

Let $\theta_2 = (D, \Gamma, \xi, T, \tau^2)$, then

$$
I_{\theta_{2i}\theta_{2k}} = \frac{1}{2}\sum_{i=1}^{n} \text{tr}\left((V_i^\circ)^{-1}\frac{\partial V_i^\circ}{\partial\theta_{2i}}(V_i^\circ)^{-1}\frac{\partial V_i^\circ}{\partial\theta_{2k}}\right), \tag{4.12}
$$

$$
I_{\theta_{2i}\beta_{1k}} = \frac{1}{2}\sum_{i=1}^{n} \text{tr}\left((V_i^\circ)^{-1}\frac{\partial V_i^\circ}{\partial\theta_{2i}}(V_i^\circ)^{-1}\frac{\partial V_i^\circ}{\partial\beta_{1k}}\right), \tag{4.13}
$$

and $I_{\delta_2\beta_0} = I_{\delta_2\alpha} = 0$, where

$$
\frac{\partial V_i^\circ}{\partial V(D)_k} = (\beta_1\beta_1^T) \otimes \left(Z_i\frac{\partial D}{\partial V(D)_k}Z_i^T\right),
$$

$$
\frac{\partial V_i^\circ}{\partial \xi_j} = (\Delta_j\Delta_j^T) \otimes (1_{m_i}1_{m_i}^T),
$$

$$
\frac{\partial V_i^\circ}{\partial \beta_{1j}} = (\Delta_j\beta_1^T + \beta_1\Delta_j^T) \otimes (I_{k_i} + Z_iDZ_i^T + E_i\Gamma E_i^T + I_{m_i}),
$$

$$
\frac{\partial V_i^\circ}{\partial \tau_j} = (\Delta_j\Delta_j^T) \otimes I_{m_i},
$$

$$
\frac{\partial V_i^\circ}{\partial \Gamma} = (\beta_1\beta_1^T) \otimes (E_iE_i^T),
$$

$$
\frac{\partial V_i^\circ}{\partial T_j} = (\Delta_j\Delta_j^T) \otimes I_{m_i},
$$

where $j = 1, \cdots, J$, $k = 1, \cdots, \frac{n_1(n_1+1)}{2}$) ($n_1$ is the dimension of D), Δ_j is a $J \times 1$ vector with the jth element equal to one and zero otherwise.

4.6.4 *Parameter Estimates in the PX-EM Algorithm*

For the E-step in the PX-EM algorithm, besides all the conditional expectation in Section 4.6.1, we also estimate the conditional expectations about ϵ_i as

$$\tilde{\epsilon}_i = E(\epsilon_i | R_i^\circ) = \beta_1^T \otimes (I_{m_i} \sigma^2) O_i^T (V_i^\circ)^{-1} (R_i^\circ - \mu_i^\circ), \tag{4.14}$$

$$E(\epsilon_i^T \epsilon_i | R_i^\circ) = \tilde{\epsilon}_i^T \tilde{\epsilon}_i + \mathrm{tr}\left(\sigma^2 I_{m_i} - (\beta_1^T \otimes I_{m_i}) O_i^T (V_i^\circ)^{-1} O_i (\beta_1 \otimes I_{m_i}) \right). \tag{4.15}$$

For the M-step in the PX-EM algorithm, the estimated parameters are $\hat{\beta}_j^t = \hat{\beta}_j$ with $\hat{\beta}_j^t = [\hat{\beta}_0^t \ \hat{\beta}_1^t]^T$, $\hat{\alpha}_t = \hat{\alpha}$, $\hat{\tau}_t = \hat{\tau}$, $\hat{D}_t = \hat{D}$, $\hat{\xi}_t = \hat{\xi}$ and $\hat{\sigma}_t^2 = \sum_{i=1}^n \frac{1}{m_i} E(\epsilon_i^T \epsilon_i | R_i^\circ)$. The estimated variances of the parameters in the PX-EM algorithm are same as these in the EM algorithm.

Table 4.1: *Data for analysis*

Level	Variable	Description	All Available Data	Observed Data
			Mean(Std.), missing %	Mean(Std.)
Child	$B_{ik\ell}$	1 if black	0.40 (0.49), 0%	0.42 (0.49)
	$R_{i1k\ell}$	score 1	605 (37), 21%	606 (37)
	$B_{ik\ell} = 0$		614 (37), 25%	614 (37)
	$B_{ik\ell} = 1$		593 (33), 16%	594 (33)
	$R_{i2k\ell}$	score 2	608 (37), 20%	608 (37)
	$B_{ik\ell} = 0$		615 (37), 22%	615 (37)
	$B_{ik\ell} = 1$		597 (35), 17%	598 (35)
	$R_{i3k\ell}$	score 3	617 (31), 21%	617 (32)
	$B_{ik\ell} = 0$		624 (30), 23%	624 (30)
	$B_{ik\ell} = 1$		607 (31), 18%	608 (31)
	$R_{i4k\ell}$	score 4	600 (43), 13%	600 (42)
	$B_{ik\ell} = 0$		609 (43), 15%	609 (43)
	$B_{ik\ell} = 1$		587 (38), 11%	587 (38)
Class	Z_{ik}	1 if treated	0.40 (0.5), 0%	0.38 (0.44)

Table 4.2: *Model coefficient estimates and their standard errors*

LVM	Coef.	$MCAR$		MAR	
		Model A	Model B	Model A	Model B
		MLE (S.E.)	MLE (S.E.)	MLE (S.E.)	MLE (S.E.)
	α_1	-0.63 (0.10)	-0.64 (0.10)	-0.67 (0.09)	-0.67 (0.21)
Model (4.1)	α_2	-0.05 (0.10)	-0.05 (0.10)	-0.01 (0.09)	-0.01 (0.17)
	α_3	**0.46** (0.17)	0.47 (0.17)	**0.33** (0.15)	0.33 (0.41)
	β_{11}	**32.62** (0.90)	32.59 (0.91)	**33.06** (0.87)	33.07 (1.89)
Model (4.2)	β_{12}	**24.49** (0.95)	24.47 (0.97)	**24.72** (0.94)	24.72 (2.01)
	β_{13}	**17.56** (0.84)	17.54 (0.85)	**17.63** (0.83)	17.63 (1.76)
	β_{14}	**33.60** (1.06)	33.57 (1.08)	**34.08** (1.03)	34.08 (2.20)
$-2logL$		-14699.26	-14699.16	-15527.26	-15527.25
# of Para.		25	27	25	27

Table 4.3: *The treatment versus control effect*

contrast[a]	effect	latent score[b]			
		MCAR		MAR	
		Est.	95% CI	Est.	95% CI
NC	0	0	0	0	0
BC	$\hat{\alpha}_1$	**-0.63**	(-0.83, -0.43)	**-0.67**	(-0.85, -0.49)
NT	$\hat{\alpha}_2$	-0.05	(-0.25, 0.15)	-0.01	(-0.19, 0.17)
BT-BC	$\hat{\alpha}_2 + \hat{\alpha}_3$	**0.41**	(0.09, 0.67)	**0.32**	(0.03, 0.55)
(BT-BC)-(NT-NC)	$\hat{\alpha}_3$	**0.46**	(0.13, 0.79)	**0.33**	(0.04, 0.62)

[a] B: black students, T: treatment, N: nonblack students, C: control

[b] $\hat{U} = \hat{\alpha}_1 B_{ij} + \hat{\alpha}_2 T_j + \hat{\alpha}_3 B_{ij} \times T_j$.

Table 4.4: *Two-level fixed model coefficient estimates and their standard errors*

LVM	Coef.	$MCAR$		MAR	
		Model A	Model B	Model A	Model B
		MLE (S.E.)	MLE (S.E.)	MLE (S.E.)	MLE (S.E.)
	β_{11}	32.53 (0.90)	31.86 (0.86)	**32.97** (0.87)	32.31 (0.84)
Model (4.6)	β_{12}	24.50 (0.95)	24.13 (0.93)	**24.71** (0.94)	24.37 (0.92)
	β_{13}	17.69 (0.84)	17.40 (0.83)	**17.77** (0.83)	17.51 (0.81)
	β_{14}	33.54 (1.06)	32.99 (1.03)	**34.04** (1.02)	33.51 (0.99)
	α_1	-0.67 (0.09)	-0.48 (0.13)	-0.70 (0.09)	-0.58 (0.13)
Model (4.7)	α_2	-0.07 (0.11)	0.006 (0.10)	-0.04 (0.09)	0.06 (0.09)
	α_3	0.45 (0.18)	0.44 (0.17)	**0.32** (0.16)	0.29 (0.15)
$-2logL$		-14901.94	-14837.61	-15746.05	-15677.24
# of Para.		20	93	20	93

Figure 4.1: *Unit-specific achievement score against school*
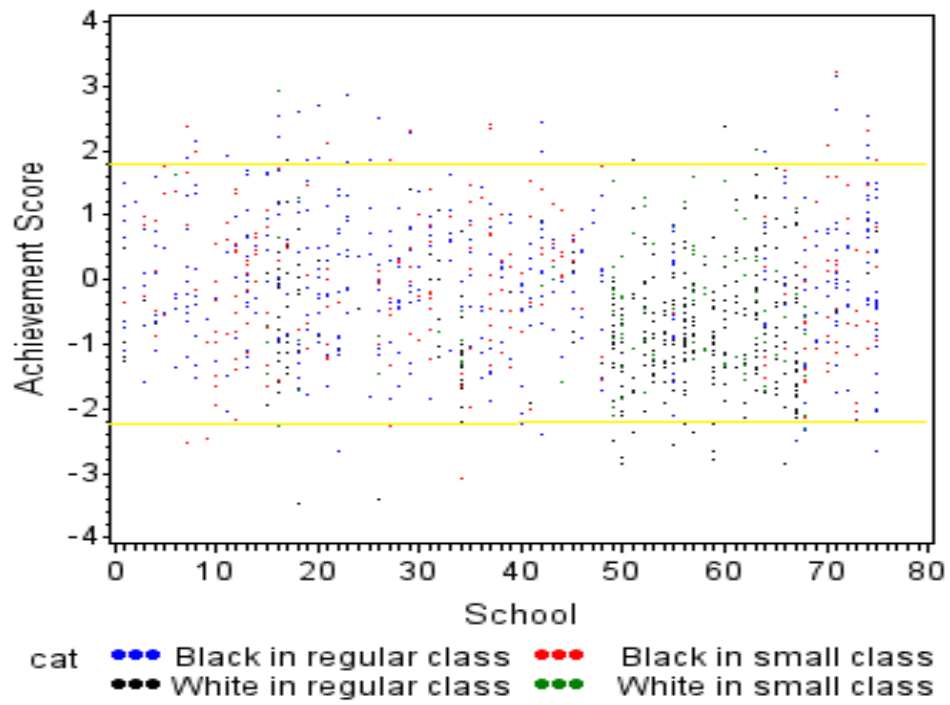
Figure 4.2: *QQ Plot for the latent achievement score*
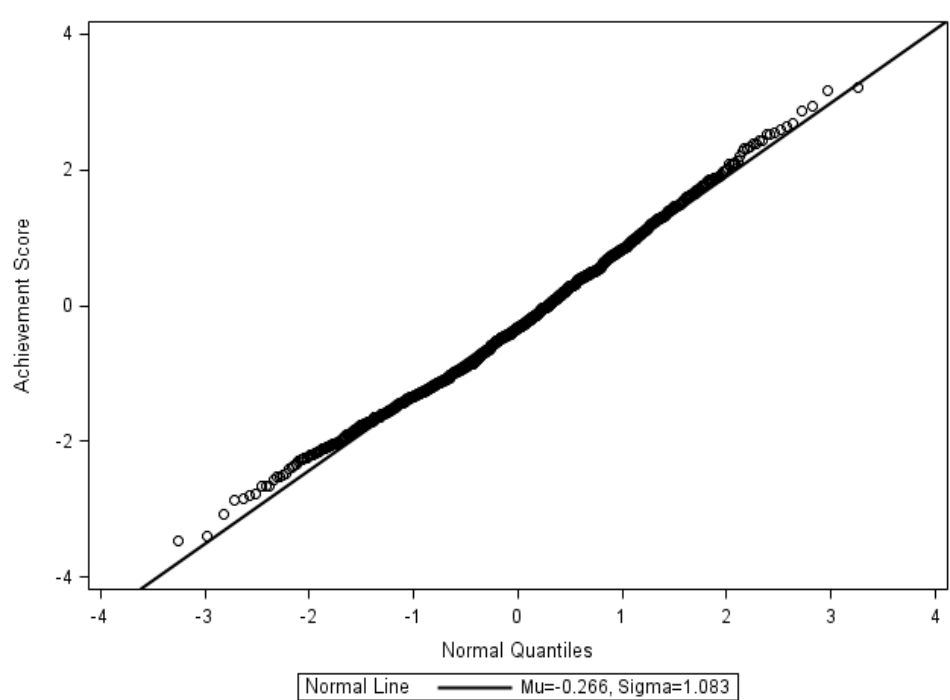
5   A Latent Variable Approach for Multivariate Instrumental Variable Estimators with Ignorable Missing Data

5.1   Introduction

Chapter 4 implied that the academic achievement score was associated with small class size especially for African-American students. The findings have found that reduced class size increases academic achievement scores, especially, for this minority. In this chapter, the causal effects of the academic achievement simultaneously studied by discussing two issues. Does reduced class size cause higher academic achievement score? If so, how large is the effect and does the magnitude of the effect vary significantly across schools? We discuss the two questions by an instrumental variable (IV) approach (Angrist and Imbens, 1995; Angrist, Imbens, and Rubin, 1996; Bollen, Kirby, Curran, Paxton, and Chen, 2007; Ecob and Goldstein, 1983; Greene, 2002; Krueger, 1999; Krueger and Whitmore, 2001; Nye, Konstantopoulos, and Hedges, 2004; Pearl, 2000; Rubin, 1978; Stolzenberg and Waite, 1977). The IV is the randomized assignment of students and teachers to a small or regular class. There is evidence that the reduced class size improves academic achievement (Finn and Achilles, 1990; Goldstein and Blatchford, 1998; Goldstein, Yang, Omar, Turner, and Thompson, 2000; Krueger,1999; Krueger and Whitmore, 2001; Mosteller, 1995; Nye, Hedges, and Konstantopoulos, 1999, 2000) or increases the likelihood of the students to take college entrance exams (Krueger and Whitmore, 2001). The results are based on univariate outcome analysis assuming data MCAR (Little and Rubin, 2002; Rubin, 1976; Schafer, 1997) or using ad-hoc imputation to deal with missing data. Such estimations may, in general, be inefficient and result in biased inferences. Shin and Raudenbush (2012, 2013) proposed a three-level multivariate simultaneous equation model (3LMSE) efficiently to handle missing data with a relatively mild assumption MAR. This 3LMSE evaluated the causal effects more efficiently than the two-stage least square method (2SLS; Bollen, 1996; Imbens and Rubin, 1997a, 1997b; Little and Yau, 1998).

However, there are two main disadvantages of this analysis: (1) it needs more degrees to test coefficient effects, and (2) it cannot provide a unit-specific achievement score.

In this chapter, three-level LVMs with an IV are developed to estimate the causal effects of reduced class size on the academic achievement given the observable reading, math, listening, and word recognition skills scores via PX-EM algorithm. The approach is built on model-based missing data at a single level (Dempster, Laird, and Rubin, 1977; Little and Rubin, 2002; Orchard and Woodbury, 1972; Rubin, 1976, 1987; Schafer, 1997), two levels (Dempster, Rubin, and Tsutakawa, 1981; Liu, Taylor, and Belin, 2000; Schafer and Yucel, 2002; Shin and Raudenbush, 2007), three levels (Shin and Raudenbush, 2011; Shin, 2012), and LVMs via EM algorithm (Roy and Lin, 2000). Under the ML framework, the causal effects, the heterogeneity of class size effect across schools, and the heterogeneity of treatment and class size across schools are more efficiently estimated and more powerfully tested. Section 5.2 introduces the models. Section 5.3 describes the algorithm. Section 5.4 analyzes the STAR data. Section 5.5 concludes the chapter with a short discussion. Finally, Section 5.6 describes detailed mathematical derivations.

Shin (2012) demonstrated that the conventional approach of causality analysis was laborious with complicated constraints in the estimation of a reduced-form structural model. He continued the single-population approach of Shin and Raudenbush (2011) to three-level causal inference involving multiple subpopulations (Angrist et al., 1996; Hong and Raudenbush, 2006; Imbens and Angrist, 1994; Shin and Raudenbush, 2011; Raudenbush, 2010) with the following assumptions:

1. Intact school: Given an observed school assignment $A = a$, $S_{ik}(T, B, A|A = a) = S_{ik}(T, B)$, $U_{ikl}(S, T, B, A|A = a) = U_{ikl}(S, T, B)$;

2. No interference between classes: $S_{ik}(T, B) = S_{ik}(T_{ik}, B_{ik})$ for all T and all B, $U_{ikl}(S, T, B) = U_{ikl}(S_{ik}, T_{ik}, B_{ik})$ for all S, for all T, and for all B;

3. Exclusion restriction: $U_{ikl}(S, T, B) = U_{ikl}(S_{ik}, T', B)$ for all S, all T and T', and for all B;

4: Random treatment assignment: The class type assignment T_{ik} is random;

5: Nonzero average causal effect of class type on class size: $E[S_{ik}(T_{ik}) - S_{ik}(T'_{ik})] \neq 0$ for all $T_{ik} \neq T'_{ik}$;

6: Linearity of academic achievement in class size: The achievement is linearly dependent on the class type size.

With the same assumptions, the new approach developed by Shin (2012) is extended to LVMs framework as

$$U_{ikl} = \gamma_{u1}B_{ikl} + \gamma_{u2}(\alpha_{s1}T_{ik}) + \gamma_{u3}(\alpha_{s1}T_{ik})B_{ikl} + \lambda_{ui}E_{uikl} + a_{uik}Z_{uikl} + \epsilon_{uikl},$$

$$S_{ik} = \alpha_{s0} + \alpha_{s1}T_{ik} + \lambda_{si}E_{sik} + a_{sik}, \tag{5.1}$$

where U_{ikl} is a univariate latent score, B_{ikl} is a black student indicator, class size S_{ik} is an endogenous regressor, class type T_{ik} is randomly assigned class type to students, $\alpha_{s1}T_{ik}$ explains the causal variability in class size induced by T_{ik}, $\begin{bmatrix} \lambda_{ui} \\ \lambda_{si} \end{bmatrix} \sim N\left(0, \begin{bmatrix} \Sigma_{uu} & \Sigma_{us} \\ \Sigma_{su} & \Sigma_{ss} \end{bmatrix}\right)$, $\begin{bmatrix} a_{uik} \\ a_{sik} \end{bmatrix} \sim N\left(0, \begin{bmatrix} \Lambda_{uu} & \Lambda_{us} \\ \Lambda_{su} & \Lambda_{ss} \end{bmatrix}\right)$, and $\epsilon_{uikl} \sim N(0,1)$ for students $l = 1, \cdots, n_{ik}$ attending classrooms $k = 1, \cdots, n_i$ in school $i = 1, \cdots, n$. Random effects are independent across different levels. The desired causal effects are γ_{u2} and γ_{u3} controlling for the pretreatment gaps γ_{u1} in academic achievement. Reduced class size causes higher academic achievement overall if both $\gamma_{u2} < 0$ and $\gamma_{u2} + \gamma_{u3} < 0$ and moderates a minority disparity of interest in academic achievement if $\gamma_{u3} < 0$. The structural models (SMs) (5.1) suggest the reduced-form models

$$U_{ikl} = \alpha_{u1}B_{ikl} + \alpha_{u2}T_{ik} + \alpha_{u3}B_{ikl}T_{ik} + \lambda_{ui}E_{uikl} + a_{uik}Z_{uikl} + \epsilon_{uikl},$$

$$S_{ik} = \alpha_{s0} + \alpha_{s1}T_{ik} + \lambda_{si}E_{sik} + a_{sik}, \tag{5.2}$$

for $\alpha_{u1} = \gamma_{u1}$, $\alpha_{u2} = \gamma_{u2}\alpha_{s1}$, and $\alpha_{u3} = \gamma_{u3}\alpha_{s1}$. Therefore, the desired causal effects are $\gamma_{u2} = \alpha_{u2}/\alpha_{s1}$ and $\gamma_{u3} = \alpha_{u3}/\alpha_{s1}$. If the academic achievement U_{ikl} were observed, we would be able to

estimate the model by standard multilevel software. With the response variable unobservable, there are observable surrogate outcomes that are highly correlated and believed to predict the academic achievement with accuracy. That is, the academic achievement is related to the surrogate outcomes by

$$R_{ijkl} = \beta_{0j} + \beta_{1j}U_{ikl} + c_{ij} + b_{ijk} + e_{ijkl}, \tag{5.3}$$

where R_{ijkl} are observable surrogate outcomes, $\beta_r = [\beta_{0j}\ \beta_{1j}]^T$ is a vector of regression coefficients for the jth surrogate outcome, $c_{ij} \overset{iid}{\sim} N(0, \Gamma_j)$, $b_{ijk} \overset{iid}{\sim} N(0, \xi_j)$, and $e_{ijk} \overset{iid}{\sim} N(0, \tau_j)$ are level-3, level-2, and level-1 unit-specific random effects, respectively. Given the unobserved variable U_{ikl}, the J surrogate outcome R_{ijkl} are mutually independent. We further assume the b_{ijk} are independent. To make parameters identifiable, we assume ϵ_{ikl} is distributed as $N(0, 1)$ and X_{ikl} does not contain an intercept. The illustration of the three-level LVMs is in Figure 5.1 which implies that the exogenous IV Z_{ik} has a nonzero causal effect α_{s1} on the endogenous regressor S_{ik} and affect U_{ikl} only through its effect on class size. The latent achievement U_{ikl} has nonzero effect β_{1j} on the jth observed surrogate outcome R_{ijkl} which has three-level random effect terms c_{ij}, b_{ijk}, and e_{ijkl}.

It is essential to aggregate models (5.2) and (5.3) at school level to derive estimates and their variances. Aggregating the models by student and then by class, we can write models (5.2) and (5.3) at school level in matrix notation as

$$\begin{bmatrix} U_i \\ S_i \end{bmatrix} = \begin{bmatrix} X_{ui} & \\ & X_{si} \end{bmatrix} \begin{bmatrix} \alpha_u \\ \alpha_s \end{bmatrix} + \begin{bmatrix} E_{ui} & \\ & E_{si} \end{bmatrix} \begin{bmatrix} \lambda_{ui} \\ \lambda_{si} \end{bmatrix} + \begin{bmatrix} \oplus_{k=1}^{n_i} Z_{uik} & \\ & 1 \end{bmatrix} \begin{bmatrix} a_{ui} \\ a_{si} \end{bmatrix} + \begin{bmatrix} \epsilon_{ui} \\ 0 \end{bmatrix}, \tag{5.4}$$

and

$$R_i = \beta_0 \otimes 1_{m_i} + \beta_1 \otimes U_i + c_i \otimes 1_{m_i} + (I_J \otimes W_i)b_i + e_i, \tag{5.5}$$

where $U_i = [U_{i1}^T\ U_{i2}^T \cdots U_{in_i}^T]^T$ with $U_{ik} = [U_{ik1}\ U_{ik2} \cdots U_{ikn_{ik}}]^T$, $X_{ui} = [X_{ui1}^T\ X_{ui2}^T \cdots X_{uin_i}^T]^T$ with $X_{uik} = [X_{uik1}^T\ X_{uik2}^T \cdots X_{uikn_{ik}}]$ and $X_{uikl} = [B_{ikl}\ T_{ik}\ B_{ikl}T_{ik}]$, $X_{si} = [X_{si1}^T\ X_{si2}^T \cdots X_{sin_i}^T]^T$,

$\alpha_u = [\alpha_{u1}\ \alpha_{u2}\ \alpha_{u3}]^T$, $\alpha_s = [\alpha_{s0}\ \alpha_{s1}]^T$, $E_{ui} = [E_{ui1}^T\ E_{ui2}^T \cdots E_{uin_i}^T]^T$ with $E_{uik} = [E_{uik1}\ E_{uik2} \cdots E_{uikn_{ik}}]^T$ and $E_{uikl} = [E_{uik1}\ E_{uik2} \cdots E_{uikn_i}]^T$, $E_{si} = [E_{si1}\ E_{si2} \cdots E_{sin_i}]^T$, $\oplus_{k=1}^{n_i} A_k$ represents a diagonal matrix with diagonal elements or submatrices $(A_1, \cdots, A_{n_i})$ and all the other elements equal to zero, $Z_{uik} = [Z_{uik1}\ Z_{uik2} \cdots Z_{uikn_{ik}}]^T$ with $Z_{uikl} = [Z_{uik1}\ Z_{uik2} \cdots Z_{uikn_i}]^T$, $\epsilon_{ui} = [\epsilon_{ui1}^T\ \epsilon_{ui2}^T \cdots \epsilon_{uin_i}^T]^T$ with $\epsilon_{uik} = [\epsilon_{uik1}\ \epsilon_{uik2} \cdots \epsilon_{uikn_{ik}}]^T$ and $\epsilon_{uikl} = [\epsilon_{uik1}\ \epsilon_{uik2} \cdots \epsilon_{uikn_i}]^T$, $R_i = [R_{i1}\ R_{i2} \cdots R_{iJ}]$ with $R_{ij} = [R_{ij1}^T\ R_{ij2}^T \cdots R_{ijn_i}^T]^T$ and $R_{ijk} = [R_{ijk1}\ R_{ijk2} \cdots R_{ijkn_{ik}}]^T$,, $\beta_0 = [\beta_{01}\ \beta_{02} \cdots \beta_{0J}]^T$, $\otimes$ represents Kronecker product, $m_i = \sum_{k=1}^{n_i} n_{ik}$, $\beta_1 = [\beta_{11}\ \beta_{12} \cdots \beta_{1J}]^T$, $c_i = [c_{i1}\ c_{i2} \cdots c_{iJ}]^T$, $W_i = \oplus_{k=1}^{n_i} 1_{n_{ik}}$, $b_i = [b_{i1}\ b_{i2} \cdots b_{iJ}]^T$ with $b_{ij} = [b_{ij1}\ b_{ij2} \cdots b_{ijn_i}]^T$, $e_i = [e_{i1}^T\ e_{i2}^T \cdots e_{iJ}^T]^T$ with $e_{ij} = [e_{ij1}^T\ e_{ij2}^T \cdots e_{ijn_i}^T]^T$ and $e_{ijk} = [e_{ijk1}\ e_{ijk2} \cdots e_{ijkn_{ik}}]^T$ for $i = 1, \cdots, n$, $k = 1, \cdots, n_i$, and $l = 1, \cdots, n_{ik}$. Since U_i is unobservable, the aggregate joint model for the multiple observed surrogate outcomes is

$$
\begin{bmatrix} R_i \\ S_i \end{bmatrix} = \begin{bmatrix} \beta_0 \times 1_{m_i} + \beta_1 \times (X_{u_i}\alpha_u) \\ X_{si}\alpha_s \end{bmatrix} + \begin{bmatrix} \beta_1 \times (E_{ui}\lambda_{u_i}) \\ E_{si}\lambda_{si} \end{bmatrix} + \begin{bmatrix} \beta_1 \otimes \oplus_{k=1}^{n_i}(Z_{uik}a_{u_{ik}}) \\ a_{si} \end{bmatrix}
$$
$$
+ \begin{bmatrix} \beta_1 \otimes \epsilon_{ui} \\ 0 \end{bmatrix} + \begin{bmatrix} c_i \otimes 1_{mi} + (I_J \otimes W_i)b_i + e_i \\ 0 \end{bmatrix}. \tag{5.6}
$$

Let $Y_i = \begin{bmatrix} R_i \\ S_i \end{bmatrix}$. It follows that $Y_i \sim N(\mu_i, V_i)$ with

$$
\mu_i = \begin{bmatrix} \beta_0 \otimes 1_{m_i} + \beta_1 \otimes (X_i\alpha) \\ X_{si}\alpha_s \end{bmatrix},
$$
$$
V_i = \begin{bmatrix} (\beta_1\beta_1^T) \otimes \text{cov}(U_i) & \beta_1 \otimes (E_{ui}\Sigma_{us}E_{si}^T + \oplus_{k=1}^{n_i}(Z_{uik}\Lambda_{us})) \\ \beta_1^T \otimes (E_{si}\Sigma_{su}E_{ui}^T + \oplus_{k=1}^{n_i}(\Lambda_{su}Z_{uik}^T)) & E_{si}\Sigma_{ss}E_{si}^T + I_{n_i} \otimes \Lambda_{ss} \end{bmatrix}
$$
$$
+ \begin{bmatrix} R(T) \otimes (1_{m_i}1_{m_i}^T) + R(\tau) \otimes I_{m_i} + R(\xi) \otimes (W_iW_i^T) & 0 \\ 0 & 0 \end{bmatrix},
$$

where $\text{cov}(U_i) = Z_{ui}\Sigma_{uu}Z_{ui}^T + E_{ui}\Gamma E_{ui}^T + I_{m_i}$ for $Z_{ui} = \oplus_{k=1}^{n_i} Z_{uik}$, $R(\Gamma) = \oplus_{j=1}^{J}\Gamma_j$, $R(\tau) = \oplus_{j=1}^{J}\tau_j$, and $R(\xi) = \oplus_{j=1}^{J}\xi_j$.

To efficiently handle missing data, define indicator matrices O_{1i} and O_{2i} of zeros and ones indicating the observed values in R_i and S_i such that $R_i^\circ = O_{1i}R_i$ and $S_i^\circ = O_{2i}S_i$, respectively. Let $O_i = \begin{bmatrix} O_{1i} & \\ & O_{2i} \end{bmatrix}$. It follows that the observed data $Y_i^\circ \sim N(\mu_i^\circ, V_i^\circ)$ with $\mu_i^\circ = O_i\mu_i$ and $V_i^\circ = O_iV_iO_i^T$.

5.2 PX-EM Algorithm

It is challenging to estimate the model (5.6) by directly using the actual log likelihood since β_1 enters both the marginal mean and variance of Y_i. Chapters 2 and 4 have implemented the EM algorithm (Roy and Lin, 2000) and the PX-EM algorithm (Liu, Rubin, and Wu, 1998) to estimate the two-level LVMs and the three-level LVMs. It has been showed that the PX-EM algorithm converged faster than the EM-algorithm. In this chapter, we just apply the PX-EM algorithm to estimate models (5.2) and (5.3) where the only change is an extension of the parameter $\epsilon_{ik} \overset{iid}{\sim} N(0, \sigma^2)$. The PX-EM algorithm is developed based on the complete data $(R_i, S_i, \lambda_{ui}, \lambda_{si}, a_{ui}, U_i, b_i, c_i)$ and the observed data Y_i°. Given the initial values of the parameters, the PX-EM algorithm iterates between its E-, M-steps in the expanded parameter space, and then reduces to the original parameter space until convergence. The E-step takes expectations of the sufficient statistics of the complete-data log likelihood, given the observed data. The M-step maximizes the expected complete-data log likelihood given parameters from the previous iteration. The method (detailed in Sections 5.6.1 and 5.6.2) is described as

PX-E step: Calculate the conditional expectations, $E(U_i^TU_i|Y_i^\circ)$, $E(U_i^Te_{ij}|Y_i^\circ)$, $E(e_{ij}^Te_{ij}|Y_i^\circ)$, $E(\epsilon_{ui}^T\epsilon_{ui}|Y_i^\circ)$, $E(\epsilon_{ui}|Y_i^\circ)$, $E(b_{ij}^Tb_{ij}|Y_i^\circ)$, $E(\lambda_i\lambda_i^T|Y_i^\circ)$, $E(a_{ik}a_{ik}^T|Y_i^\circ)$, $E(a_{2i}|Y_i^\circ)$, and $E(c_{ij}^2|Y_i^\circ)$ for $\lambda_i = \begin{bmatrix} \lambda_{ui} \\ \lambda_{si} \end{bmatrix}$ and $a_{ik} = \begin{bmatrix} a_{uik} \\ a_{sik} \end{bmatrix}$.

PX-M step: Estimate model parameters in the extended space

$$\gamma^t_\star = \left(\beta^t_0, \beta^t_1, \alpha^t_u, \alpha^t_s, \tau^t, \Sigma_t, \xi^t, \sigma^2_t, \Gamma^t, \Lambda^t\right)$$

and reduce $\gamma^t_\star$ to the desired model parameters by the following transformation

$$\gamma^t = \left(\beta^t_0, \beta^t_1\sigma_t, , \frac{\alpha^t_u}{\sigma_t}, \alpha^t_s, \frac{\Sigma_{uu}}{\sigma^2_t}, \frac{\Sigma_{us}}{\sigma_t}, \Sigma^t_{ss}, \frac{\Lambda^t_{uu}}{\sigma^2_t}, \frac{\Lambda^t_{us}}{\sigma_t}, \Lambda^t_{ss}, \tau^t, \xi^t, \sigma^2 = 1, \Gamma^t\right).$$

At convergence, the variance of the parameter estimates are computed by the expected Fisher information matrix (see details in Section 5.6.3) based on the marginal log likelihood of Y°_i.

The next section shows the approach to STAR data. The desired SMs are estimated by the PROC IML in SAS via ML. The convergence criterion is the difference in the observed log-likelihood between two consecutive iterations taken to be less than 10^{-6}. The statistical significance of an effect estimate is discussed at a significant level 0.05.

5.3   Data Analysis

This section explains the causal analysis to begin with the causal intent-to-treat (ITT) effect on the academic achievement in the LVMs. The model is the U_{ikl} equation of the reduced-form SMs (5.2) and is called "3L ITT LVMs (5.2)", three-level ITT LVMs to assess the causal impact of the ITT intervention to treat a student to reduced class size controlling for the pretreatment effect of race ethnicity. Next, the SMs (5.1) is estimated to study if reduced class size causes higher academic achievement overall and moderates a difference in academic achievement between African-American and white students. This model is referred to as a "3L Rand-Int LVMs (5.1)", three-level random-intercept LVMs. Two-level ITT LVMs are also analyzed by including school as an indicator variable. Finally, the analysis then extends to estimation of three-level random-coefficient LVMs, "Random Coef. LVMs (5.1)".

5.3.1 ITT Causal Effects

This analysis examines if the ITT intervention of assigning a student to reduced class size causes higher academic achievement overall and moderates a difference in academic achievement between African-American and white students. The ITT model is the first equation in the SM (5.2) where the desired causal effects are $E[U_{ikl}|B_{ikl} = 0, T_{ik} = 1] - E[U_{ikl}|B_{ikl} = 0, T_{ik} = 0] = \alpha_{u2}$ and $E[U_{ikl}|B_{ikl} = 1, T_{ik} = 1] - E[U_{ikl}|B_{ikl} = 1, T_{ik} = 0] = \alpha_{u2} + \alpha_{u3}$. Their difference α_{u3} is the causal minority disparity in academic achievement caused by the randomized ITT intervention.

The α_{u3} of the ITT LVMs displays significant pretreatment minority gaps in academic achievement scores (Finn and Achilles, 1990; Fryer and Levitt, 2004; Goldstein and Blatchford, 1998; Krueger, 1999; Word et al., 1990). The results are summarized in the second column in Table 5.1. For white students, the ITT treatment does not cause higher academic achievement while, for African-American students, it causes higher academic achievement, controlling for the effects of pretreatment race ethnicity fixed. The minority differences are pronounced in third-grades achievement. The ITT is subject-specific. An African-American third grader assigned to reduced class size, for example, improves his or her academic achievement score by 0.321 units on average. The improvement is similar to the corresponding expected pretreatment minority gap in academic achievement, 0.533 points lower than that of a white student.

To rule out a confounder between T_{ik} and a school-level covariate, two-level ITT LVMs are fitted as

$$R_{kjl} = \beta_{0j} + \beta_{1j}U_{kl} + c_{kj} + e_{kjl}, \ c_{kj} \sim N(0, \xi_j), \ e_{kjl} \sim N(0, \tau_j) \tag{5.7}$$

$$U_{kl} = \alpha_{u1}B_{kl} + \alpha_{u2}T_k + \alpha_{u3}B_{kl}T_k + \alpha_{u4}A_k + a_{uk} + \epsilon_{ukl}, \tag{5.8}$$

$$S_k = \alpha_{s0} + \alpha_{s1}T_k + a_{sk},$$

where A_k is a vector of school indicators having fixed effects α_{u4}, $\begin{bmatrix} a_{uk} \\ a_{sk} \end{bmatrix} \sim N \left(0, \begin{bmatrix} \Lambda_{uu} & \Lambda_{us} \\ \Lambda_{su} & \Lambda_{ss} \end{bmatrix} \right),$ $\epsilon_{ukl} \sim N(0,1)$, and all others are defined similarly as in the three-level counterparts in the LVMs (5.2) for student l attending the class k. If the confounder is severe, then the results are differential between the two- and three-level ITT LVMs. The results are summarized in the third column in Table 5.1. The estimated effects are in a close range to their counterparts under three-level ITT LVMs. The standard errors of some coefficients are relatively inflated under two-level LVMs, which indicates that the three-level model is more efficient than the two-level model. In addition, three-level model is more powerful to test coefficient effect than two-level model. Because the estimates are close to each other across the two- and three-level LVMs, the treatment effect due to the confounder of school-level variable seems implausible.

5.3.2 Causal Effects of Reduced Class Size

In this section, we examine if reduced class size causes higher achievement score overall and moderates a difference in academic achievement between African-American and white students. The LVMs (5.1) are the desired models where the causal effects are γ_{u2} and $\gamma_{u2} + \gamma_{u3}$ for African-American and white students controlling for the pretreatment minority gap γ_{u1} in academic achievement. Their difference γ_{u3} is the causal disparity induced by reduced class size. Models (5.1) with random-intercept effect, i.e. $E_{uikl} = E_{sik} = Z_{uikl} = 1$, are first fitted to identify the causal effects. To examine if the causal disparities randomly vary across schools, three more competing models are compared with the random-intercept model by defining (a) $E_{uikl} = \begin{bmatrix} 1 & B_{ikl} & T_{ikl} & B_{ikl} \times T_{ikl} \end{bmatrix}$ and $E_{sik} = Z_{uikl} = 1$, (b) $E_{uikl} = \begin{bmatrix} 1 & T_{ikl} & B_{ikl} \times T_{ikl} \end{bmatrix}$ and $E_{sik} = Z_{uikl} = 1$, (c) $E_{uikl} = \begin{bmatrix} 1 & T_{ikl} \end{bmatrix}$ and $E_{sik} = Z_{uikl} = 1$.

Likelihood ratio tests indicate that the random-intercept model fits adequately than the random-coefficient models (a), (b), and (c) with p-values 1, 0.97, and 0.96, respectively. The results for the random-intercept model (5.1) are presented under Rand-Int (5.1) in Table 5.2. The pretreatment

minority gap in academic achievement γ_{u1} is statistically significant. For white students, reduced class size does not cause higher academic achievement while, for African-American students, it causes higher academic achievement score, controlling for the pretreatment effects of race ethnicity. The causal disparity is the most noticeable because of no causal effect for white students.

To eliminate confounders between T_{ik} and school-level covariates, the three-level LVMs with a random-intercept effect are compared to an alternative LVMs controlling for the school effect

$$R_{kjl} = \beta_{0j} + \beta_{1j}U_{kl} + c_{kj} + e_{kjl}, \ c_{kj} \sim N(0, \xi_j), \ e_{kjl} \sim N(0, \tau_j) \tag{5.9}$$

$$U_{kl} = \gamma_{u1}B_{kl} + \gamma_{u2}(\alpha_{s1}T_k) + \gamma_{u3}(\alpha_{s1}T_k)B_{kl} + \gamma_{u4}A_k + a_{uk} + \epsilon_{ukl}, \tag{5.10}$$

$$S_k = \alpha_{s0} + \alpha_{s1}T_k + a_{sk},$$

where A_k is a vector of school indicators having fixed effects γ_{u4}, $\begin{bmatrix} a_{uk} \\ a_{sk} \end{bmatrix} \sim N \left(0, \begin{bmatrix} \Lambda_{uu} & \Lambda_{us} \\ \Lambda_{su} & \Lambda_{ss} \end{bmatrix} \right)$, $\epsilon_{ukl} \sim N(0, 1)$, and all others are defined similarly as in the three-level counterparts in the LVMs (5.1) for student l attending the class k. The models are called 2L LVMs, two-level models with school fixed effects. The estimates are displayed under 2L LVMs (5.10) in Table 5.2 where the estimated effects of 3L Rand-Int (5.2) are in close range to their counterparts under 2L LVMs (5.10). No significant differences are due to no school-level confounders. With similar inferences, the overestimates of statistical inferences in the two-level models are mainly due to the relative inefficiency of the two-level approach.

5.3.3 Surrogate Outcomes on Child Academic Achievement

In the measurement model (5.3), the most interesting parameters are β_{1j} $(j = 1, \cdots, 4)$- the slopes of academic achievement for the four various observed scores. The point estimates under 3L LVMs (5.3) in Table 5.3 indicate that the four subjects' scores are positively associated with academic achievement. Though we may draw the similar inferences from the 2L LVMs (5.7), the

point estimates of β_{1j} $(j = 1, \cdots, 4)$ are underestimated compared with the counterparts of the 3L LVMs (5.2).

5.3.4 Unit-Specific Child Academic Achievement Score

Like the previous three chapters, we impute the latent variable via estimating its posterior probability in the LVMs and demonstrate it in Figures 5.2 and 5.3. These scores are well-summarized and can serve as a vehicle for the further study of middle school graduation rates, middle school drop-out rates, and classification of subjects. The 2.5th and 97.5th percentiles are used reference curves to categorize students with low, normal, and high achievement scores. Efforts to improve performance could focus on the subjects whose scores are well below the typical curve, e.g. the 2.5th percentile. The QQ plot of the latent variable appears not to depart from a straight line with slope 1. Therefore, it is tenable for the normality of the latent variable and no potential outliers.

5.4 Discussion

The analysis in this chapter continued Shin and Rauderbush's three-level causal modeling framework to three-level LVMs having a continuous mediator whose value indicates the degree of compliance or the received treatment dosage and whose effects on the outcome variables may differ across multiple subpopulations of students. The extension enabled this study more powerfully to identify that reduced class size causes higher academic achievement for the African-American students in third grade at Tennessee. Hypothesis tests revealed that African-American students benefit more from reduced class size than white students in terms of academic achievement in third grade. The analysis was then extended to three-level random-coefficient LVMs where the minority differences in the causal effects of reduced class size on academic achievement were hypothesized to be heterogeneous across schools. This chapter did not find evidence that the minority differences varied randomly across schools for third graders.

The causal analysis in this chapter is based on six assumptions. Cases may be made to violate each assumption (Shin and Raudenbush, 2011). If students with prior exposure to a particular class type are more likely to study better in the class type, then they bias the causal effect. The assumption, however, appears reasonable within the context of the current application. The intact school assumption is realistic with existing school assignments. The assumption of no interference between class seems reasonable because students share academic experiences with classmates. The random treatment assignment assumption was violated due to the randomization within schools. This violation was shown to yield no bias in the causal inferences. The exclusion restriction assumption is reasonable because randomly labeling each student by class type cannot affect academic achievement unless it induces the dosage in class size. The assumption of nonzero average causal effect of class type on class size is very reasonable from the sample average dosage greater than 7. The no compliance-effect covariance assumption seems plausible from the fact that both students and teachers were randomly assigned to the class type so that their differences in ability to learn and teach are also randomized across class types. Consequently, the violating cases of this assumption above are unlikely. School differences due to randomization within schools have been shown to cause no serious bias in the causal inferences.

5.5   Miscellanea

5.5.1   *Conditional Expectations in E-step*

The conditional expectations in E-step are

$$\tilde{U}_i = E(U_i|Y_i^\circ) = X_{ui}\alpha_u + \Lambda_{ui}(V_i^\circ)^{-1}(Y_i^\circ - \mu_i^\circ),$$

$$E(U_i^T U_i|Y_i^\circ) = \tilde{U}_i^T \tilde{U}_i + \text{tr}(\Lambda_{1i} - \Lambda_{ui}(V_i^\circ)^{-1}\Lambda_{ui}^T),$$

$$\tilde{a}_i = E(a_i|Y_i^\circ) = \Lambda_{ai}(V_i^\circ)^{-1}(Y_i^\circ - \mu_i^\circ),$$

$$E(a_i a_i^T|Y_i^\circ) = \tilde{a}_i \tilde{a}_i^T + I_{n_i} \otimes \Lambda - \Delta_{a_i}(V_i^\circ)^{-1}\Delta_{a_i}^T,$$

83

$$\tilde{b}_{ij} = E(b_{ij}|Y_i^\circ) = \nu_1(V_i^\circ)^{-1}(Y_i^\circ - \mu_i^\circ),$$

$$E(b_{ij}b_{ij}^T|Y_i^\circ) = \tilde{b}_{ij}\tilde{b}_{ij}^T + \mathrm{tr}\left(\xi_j I_{n_i} - \nu_1(V_i^\circ)^{-1}\nu_1^T\right),$$

$$E(U_i^T e_{ij}|Y_i^\circ) = \tilde{U}_i^T \tilde{e}_{ij} - \mathrm{tr}(\Lambda_{ui}(V_i^\circ)^{-1}\nu_2^T),$$

$$\tilde{\epsilon}_{ui} = E(\epsilon_{ui}|Y_i^\circ) = \nu_3(V_i^\circ)^{-1}(Y_i^\circ - \mu_i^\circ), \tag{5.11}$$

$$E(\epsilon_{ui}^T \epsilon_{ui}|Y_i^\circ) = \tilde{\epsilon}_{ui}^T \tilde{\epsilon}_{ui} + \mathrm{tr}(\sigma^2 I_{m_i} - \nu_3(V_i^\circ)^{-1}\nu_3^T),$$

$$\tilde{e}_{ij} = E(e_{ij}|Y_i^\circ) = \nu_2(V_i)^{-1}(Y_i^\circ - \mu_i^\circ),$$

$$E(e_{ij}^T e_{ij}|Y_i^\circ) = \tilde{e}_{ij}^T \tilde{e}_{ij} + \mathrm{tr}(\tau_j I_{m_i} - \nu_2(V_i^\circ)^{-1}\nu_2^T),$$

$$\tilde{\lambda}_i = E(\lambda_i|Y_i^\circ) = \nu_4(V_i^\circ)^{-1}(Y_i^\circ - \mu_i^\circ),$$

$$E(\lambda_i\lambda_i^T|Y_i^\circ) = \tilde{\lambda}_i\tilde{\lambda}_i^T + \nu_4(V_i^\circ)^{-1}\nu_4^T,$$

$$\tilde{c}_i = E(c_i|Y_i^\circ) = \nu_5(V_i^\circ)^{-1}(Y_i^\circ - \mu_i^\circ),$$

$$E(c_i^T c_i|Y_i^\circ) = \tilde{c}_i\tilde{c}_i^T + R(\Gamma) - \nu_5(V_i^\circ)^{-1}\nu_5^T,$$

where

$$\Lambda_{ui} = [\beta_1^T \otimes \Lambda_{1i} \ E_{ui}\Sigma_{us}E_{si}^T + \oplus_{k=1}^{k_i}(Z_{uik}\Sigma_{us})]O_i^T,$$

$$\Delta_{a_i} = \begin{bmatrix} \beta_1^T \otimes [(I_{n_i} \otimes \Lambda_{uu})Z_{ui}^T] & I_{n_i} \otimes \Lambda_{us} \\ \beta_1^T \otimes [(I_{n_i} \otimes \Lambda_{su})Z_{ui}^T)] & I_{n_i} \otimes \Lambda_{ss} \end{bmatrix} O_i^T,$$

$$\nu_1 = \left[(0_{n_i \times (j-1)n_i} \ \xi_j I_{n_i} \ 0_{n_i \times (J-j)n_i})(I_J \otimes W_i^T) \ 0_{n_i \times 1}\right] O_i^T,$$

$$\nu_2 = [0_{m_i \times (j-1)m_i} \ \tau_j I_{m_i} \ 0_{m_i \times (J-j)m_i}]O_i^T,$$

$$\nu_3 = [\beta_1^T \otimes (I_{m_i}\sigma^2) \ 0_{m_i \times 1}]O_i^T,$$

$$\nu_4 = \begin{bmatrix} \beta_1^T \otimes (\Sigma_{uu} \otimes E_{ui}^T) & \Sigma_{us} \otimes E_{si}^T \\ \beta_1^T \otimes (\Sigma_{su} \otimes E_{ui}^T) & \Sigma_{ss} \otimes E_{si}^T \end{bmatrix} O_i^T,$$

$$\nu_5 = [R(\Gamma) \otimes 1_{m_i}^T \ 0]O_i^T,$$

for $a_i = [a_{ui}^T \ a_{si}^T]^T$ and $\Lambda_{1i} = I_{m_i}\sigma^2 + E_{ui}\Sigma_{uu}E_{ui}^T + \oplus_{k=1}^{k_i}(Z_{uik}\Sigma_{uu}Z_{uik}^T)$.

5.5.2 *Parameter Estimates in the M-step*

The complete-data log likelihood for $(Y_i, U_i, S_i, b_i, a_i, c_i, \lambda_i)$ is, apart from a constant,

$$l = \sum_{i=1}^{n} \left(l(Y_i|U_i, b_i, c_i) + l(U_i|a_{ui}, \lambda_{ui}) + l(a_i) + l(b_{ui}) + l(c_i) + l(\lambda_i) + l(S_i|\lambda_{si}, a_{ui}) \right), \quad (5.12)$$

where $\xi = [\xi_1 \cdots \xi_J]^T$, $\tau = [\tau_1 \cdots \tau_J]^T$, $\Gamma = [\Gamma_1 \cdots \Gamma_J]^T$, and

$$l(Y_i|U_i, b_i, c_i) = \sum_{j=1}^{J} \left(-\frac{m_i}{2}\log(\tau_j) - \frac{1}{2\tau_j}e_{ij}^T e_{ij} \right),$$

$$l(U_i|a_{ui}, \lambda_{ui}) = -\frac{1}{2}\epsilon_{ui}^T \epsilon_{ui},$$

$$l(a_i) = -\frac{1}{2}(n_i \log|\Lambda| + a_i^T \Lambda^{-1} a_i),$$

$$l(b_i) = -\frac{1}{2}(m_i \log|R(\xi)| + b_i^T R(\xi)^{-1} b_i),$$

$$l(c_i) = -\frac{1}{2}(n \log|R(\Gamma)| + c_i^T R(\Gamma)^{-1} c_i),$$

$$l(\lambda_i) = -\frac{1}{2}(n \log|\Sigma| + \lambda_i^T \Sigma^{-1} \Lambda_i),$$

$$l(S_i) = -\frac{1}{2}\left(n_i \log|V_s| + (S_i - \mu_s)^T V_s^{-1}(S_i - \mu_s) \right),$$

for $\mu_s = X_{si}\alpha_s + E_{2i}\lambda_{2i} + [I_{k_i} \otimes (\Lambda_{2u}\Lambda_{uu}^{-1})]a_{ui}$ and $V_s = I_{k_i} \otimes \Lambda_{ss} - I_{k_i} \otimes (\Lambda_{su}\Lambda_{uu}^{-1}\Lambda_{us})$.

Differentiating (5.12) with respect to the parameters β_0, β_1, α, ξ, τ, Λ, Γ and Σ, respectively, taking expectations of the resulting forms conditional to the observed data Y_i°, setting them equal to zeros, and solving these equations, we know

$$\hat{\beta}_j^{(k)} = \hat{\beta}_j^{(k-1)} + \left(\sum_{i=1}^{n} E(U_{i*}^T U_{i*}|Y_i^\circ) \right)^{-1} \sum_{i=1}^{n} E(U_{i*}^T e_{ij}|Y_i^\circ),$$

$$\hat{\tau}_j = \frac{1}{\sum_{i=1}^n m_i} \times \sum_{i=1}^n E(e_{ij}^T e_{ij}|Y_i^\circ),$$

$$\hat{\xi}_j = \frac{1}{\sum_{i=1}^n n_i} \sum_{i=1}^n E(b_{ij}^2|Y_i^\circ),$$

$$\hat{\alpha}_u^{(k)} = \hat{\alpha}_u^{(k-1)} + \left(\sum_{i=1}^n X_{ui}^T X_{ui}\right)^{-1} \sum_{i=1}^n X_{ui}^T \tilde{\epsilon}_{ui}, \tag{5.13}$$

$$\hat{\Lambda} = \frac{1}{\sum_{i=1}^n n_i} \sum_{i=1}^n \text{tr}\left(E(a_i a_i^T|Y_i^\circ)\right),$$

$$\hat{\Sigma} = \frac{1}{n} \sum_{i=1}^n E(\lambda_i \lambda_i^T|Y_i^\circ),$$

$$\hat{\Gamma}_j = \frac{1}{n} \sum_{i=1}^n E(c_{ij} c_{ij}^T|Y_i^\circ),$$

$$\hat{\alpha}_s^{(k)} = \hat{\alpha}_s^{(k-1)} + \left(\sum_{i=1}^n X_{si}^T V_s^{-1} X_{si}\right)^{-1} \sum_{i=1}^n X_{si}^T V_s^{-1} \left(a_{si} - [I_{k_i} \otimes (\Lambda_{2u}\Lambda_{uu}^{-1})]a_{ui}\right),$$

$$\hat{\sigma}_t^2 = \sum_{i=1}^n \frac{1}{m_i} E(\epsilon_i^T \epsilon_i|Y_i^\circ),$$

where $j = 1, \cdots, J$, $\beta_j = [\beta_{0j} \ \beta_{1j}]^T$, $U_{i*} = [1_{k_i} \ U_i]$, $E(b_{ij}^2|Y_i^\circ)$ is the jth diagonal element in $E(b_i b_i^T|Y_i^\circ)$, $E(U_{i*}^T U_{i*}|Y_i^\circ) = \begin{bmatrix} m_i & 1_{m_i}^T \tilde{U}_i \\ 1_{m_i}^T \tilde{U}_i & E(U_i^T U_i|Y_i^\circ) \end{bmatrix}$, $and E(U_{i*}^T e_{ij}|Y_i^\circ) = \begin{bmatrix} 1_{k_i}^T \tilde{e}_{ij} \\ E(U_i^T e_{ij}|Y_i^\circ)) \end{bmatrix}$.

5.5.3 Calculation of the Information Matrix

The information matrix is obtained by differentiating twice the log likelihood of the observed data Y_i° with mean and variance given in section 5.2 and taking the expectation of the resulting form. Let $G_i = O_i \begin{bmatrix} I_J \otimes 1_{m_i} \\ 0_{1 \times J} \end{bmatrix}$, $M_i = O_i \begin{bmatrix} \beta_1 \otimes X_{ui} \\ 0_{1 \times J} \end{bmatrix}$, $N_i = O_i \begin{bmatrix} I_J \otimes (X_{ui}\alpha_u) \\ 0_{1 \times J} \end{bmatrix}$, and $H_i =$

$O_i \begin{bmatrix} 0_{Jm_i \times 2} \\ X_{si} \end{bmatrix}$. The expected information matrix for the MLEs of $\theta_1 = (\beta_0, \beta_1, \alpha_u, \alpha_2)$ is

$$
I_{\theta_1 \theta_1} = \sum_{i=1}^{n} \begin{bmatrix} G_i^T(V_i^{\circ})^{-1}G_i & G_i^T(V_i^{\circ})^{-1}M_i & G_i^T(V_i^{\circ})^{-1}N_i & G_i^T(V_i^{\circ})^{-1}H_i \\ M_i^T(V_i^{\circ})^{-1}G_i & A + M_i^T(V_i^{\circ})^{-1}M_i & M_i^T(V_i^{\circ})^{-1}N_i & M_i^T(V_i^{\circ})^{-1}H_i \\ N_i^T(V_i^{\circ})^{-1}G_i & N_i^T(V_i^{\circ})^{-1}M_i & N_i^T(V_i^{\circ})^{-1}N_i & N_i^T(V_i^{\circ})^{-1}H_i \\ H_i^T(V_i^{\circ})^{-1}G_i & H_i^T(V_i^{\circ})^{-1}M_i & H_i^T(V_i^{\circ})^{-1}N_i & H_i^T(V_i^{\circ})^{-1}H_i \end{bmatrix}, \quad (5.14)
$$

where A has its (i,k)th element $\frac{1}{2}tr\left((V_i^{\circ})^{-1}\frac{\partial V_i^{\circ}}{\partial \beta_{1i}} \times (V_i^{\circ})^{-1}\frac{\partial V_i^{\circ}}{\partial \beta_{1k}}\right)$.

Let $\theta_2 = (\Sigma, \Lambda, \xi, \Gamma, \tau)$, then

$$
\begin{aligned}
I_{\theta_{2i}\theta_{2k}} &= \frac{1}{2}\sum_{i=1}^{n} \text{tr}\left((V_i^{\circ})^{-1}\frac{\partial V_i^{\circ}}{\partial \theta_{2i}}(V_i^{\circ})^{-1}\frac{\partial V_i^{\circ}}{\partial \theta_{2k}}\right), \\
I_{\theta_{2i}\beta_{1k}} &= \frac{1}{2}\sum_{i=1}^{n} \text{tr}\left((V_i^{\circ})^{-1}\frac{\partial V_i^{\circ}}{\partial \theta_{2i}}(V_i^{\circ})^{-1}\frac{\partial V_i^{\circ}}{\partial \beta_{1k}}\right),
\end{aligned} \quad (5.15)
$$

and $I_{\delta_2 \beta_0} = I_{\delta_2 \alpha_u} = I_{\delta_2 \alpha_s} = 0$, where

$$
\begin{aligned}
\frac{\partial V_i^{\circ}}{\partial V(D)_k} &= (\beta_1 \beta_1^T) \otimes \left(Z_i \frac{\partial D}{\partial V(D)_k} Z_i^T\right), \\
\frac{\partial V_i^{\circ}}{\partial \xi_j} &= (\Delta_j \Delta_j^T) \otimes (1_{m_i} 1_{m_i}^T), \\
\frac{\partial V_i^{\circ}}{\partial \beta_{1j}} &= (\Delta_j \beta_1^T + \beta_1 \Delta_j^T) \otimes (I_{k_i} + Z_i D Z_i^T + E_i \Gamma E_i^T + I_{m_i}), \\
\frac{\partial V_i^{\circ}}{\partial \tau_j} &= (\Delta_j \Delta_j^T) \otimes I_{m_i}, \\
\frac{\partial V_i^{\circ}}{\partial \Gamma} &= (\beta_1 \beta_1^T) \otimes (E_i E_i^T), \quad \frac{\partial V_i^{\circ}}{\partial T_j} = (\Delta_j \Delta_j^T) \otimes I_{m_i},
\end{aligned}
$$

for $j = 1, \cdots, J$, $k = 1, \cdots, \frac{n_1(n_1+1)}{2})$ (n_1 is the dimension of D), Δ_j is a $J \times 1$ vector with the jth element equal to one and zero otherwise.

5.5.4 Estimates of the Desired Causal Effects

From the invariance property of MLEs, the point estimates of the desired causal effects are $\hat{\gamma}_{u2} = \frac{\hat{\alpha}_{u2}}{\hat{\alpha}_{s1}}$ and $\hat{\gamma}_{u2} + \hat{\gamma}_{u3} = \frac{\hat{\alpha}_{u2} + \hat{\alpha}_{u3}}{\hat{\alpha}_{s1}}$. According to multivariate Delta method, their estimated variances are

$$
cov(\hat{\gamma}_{u2}) = [\frac{1}{\hat{\alpha}_{s1}} \quad -\frac{\hat{\alpha}_{u2}}{\hat{\alpha}_{s1}^2}] \begin{bmatrix} cov(\hat{\alpha}_{u2}) & cov(\hat{\alpha}_{u2}, \hat{\alpha}_{s1}) \\ cov(\hat{\alpha}_{s1}, \hat{\alpha}_{u2}) & cov(\hat{\alpha}_{s1}) \end{bmatrix} \begin{bmatrix} \frac{1}{\hat{\alpha}_{s1}} \\ -\frac{\hat{\alpha}_{u2}}{\hat{\alpha}_{s1}^2} \end{bmatrix}, \tag{5.16}
$$

$$
cov(\hat{\gamma}_{u3}) = [\frac{1}{\hat{\alpha}_{s1}} \quad -\frac{\hat{\alpha}_{u3}}{\hat{\alpha}_{s1}^2}] \begin{bmatrix} cov(\hat{\alpha}_{u3}) & cov(\hat{\alpha}_{u3}, \hat{\alpha}_{s1}) \\ cov(\hat{\alpha}_{s1}, \hat{\alpha}_{u3}) & cov(\hat{\alpha}_{s1}) \end{bmatrix} \begin{bmatrix} \frac{1}{\hat{\alpha}_{s1}} \\ -\frac{\hat{\alpha}_{u3}}{\hat{\alpha}_{s1}^2} \end{bmatrix}, \tag{5.17}
$$

$$
cov(\hat{\gamma}_{u2} + \hat{\gamma}_{u3}) = [\frac{1}{\hat{\alpha}_{s1}} \quad \frac{1}{\hat{\alpha}_{s1}} \quad -\frac{\hat{\alpha}_{u2} + \hat{\alpha}_{u3}}{\hat{\alpha}_{s1}^2}] \zeta \begin{bmatrix} \frac{1}{\hat{\alpha}_{s1}} \\ \frac{1}{\hat{\alpha}_{s1}} \\ -\frac{\hat{\alpha}_{u2} + \hat{\alpha}_{u3}}{\hat{\alpha}_{s1}^2} \end{bmatrix}, \tag{5.18}
$$

where $\zeta = \begin{bmatrix} cov(\hat{\alpha}_{u2}) & cov(\hat{\alpha}_{u2}, \hat{\alpha}_{u3}) & cov(\hat{\alpha}_{u2}, \hat{\alpha}_{s1}) \\ cov(\hat{\alpha}_{u3}, \hat{\alpha}_{u2}) & cov(\hat{\alpha}_{u3}) & cov(\hat{\alpha}_{u3}, \hat{\alpha}_{s1}) \\ cov(\hat{\alpha}_{s1}, \hat{\alpha}_{u2}) & cov(\hat{\alpha}_{s1}, \hat{\alpha}_{u3}) & cov(\hat{\alpha}_{s1}) \end{bmatrix}$ and the variances and covariances of

the estimates can be extracted from Fisher information matrix.

Figure 5.1: *Illustration of the structure of the LVMs with an IV Z_{ik} given B_{ikl}*
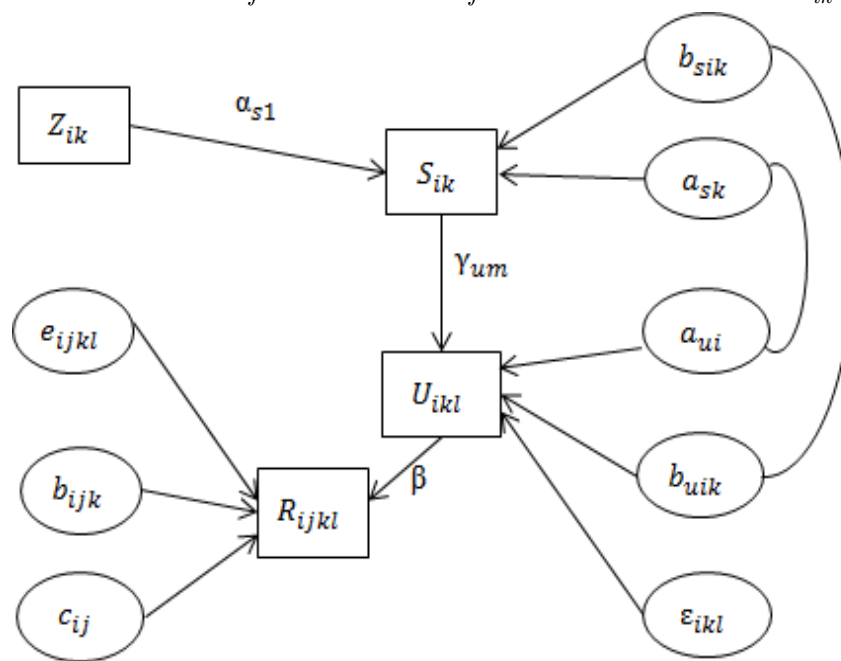
$$m=0,\ 1,\ \gamma_{u0}=\gamma_{u2}\ ,\ \gamma_{u1}=\gamma_{u2}+\gamma_{u3}$$

Table 5.1: *Fixed coefficient estimates and their standard errors for the ITT causal effect*

Effect	3L ITT (5.2)	2L ITT (5.8)
α_{u1}	-0.533 (0.085)	-0.576 (0.125)
α_{u2}	-0.013 (0.094)	0.066 (0.093)
α_{u3}	0.334 (0.155)	0.294 (0.153)
$\alpha_{u2} + \alpha_{u3}$	0.321 (0.126)	0.360 (0.124)

Table 5.2: *Fixed coefficient estimates and their standard errors for the causality of reduced class size*

Effect	3L Rand-Int (5.1)	2L LVMs (5.10)
γ_{u1}	-0.533 (0.085)	-0.576 (0.125)
γ_{u2}	0.002 (0.012)	-0.008 (0.012)
γ_{u3}	-0.041 (0.019)	-0.036 (0.019)
$\gamma_{u2} + \gamma_{u3}$	-0.040 (0.016)	-0.045 (0.015)

Table 5.3: *Fixed coefficient estimates and their standard errors for LVMs (5.3)*

Effect	3L LVMs (5.2)	2L LVMs (5.7)
β_{11}	33.05 (0.87)	32.30 (0.84)
β_{12}	24.71 (0.94)	24.36 (0.92)
β_{13}	17.63 (0.63)	17.51 (0.81)
β_{14}	34.07 (1.03)	33.50 (0.99)

Figure 5.2: *Unit-specific achievement score against school*
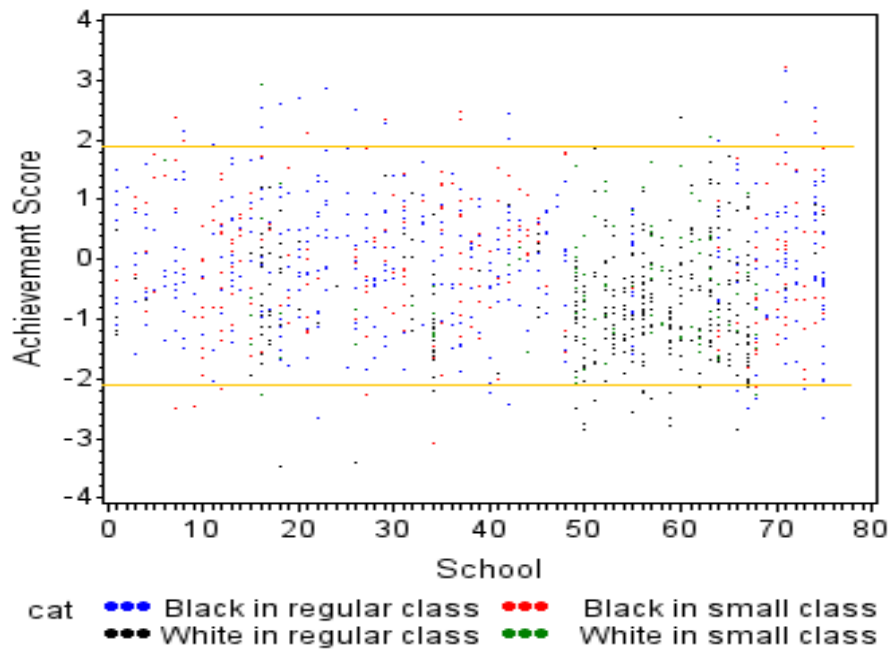
Figure 5.3: *QQ Plot for the latent achievement score*
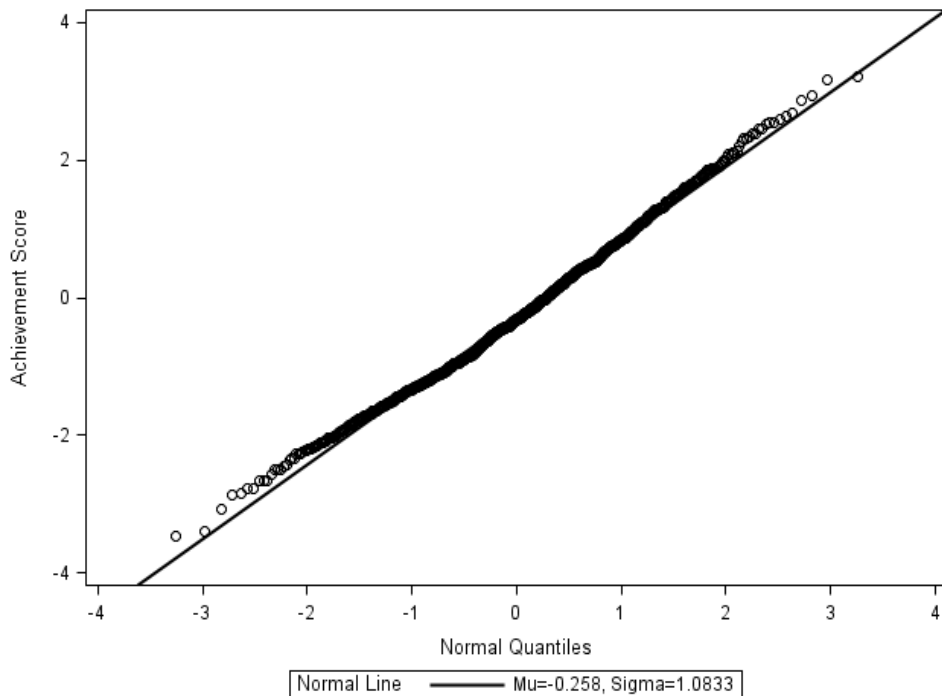
6   Discussion

This dissertation investigated the two-level and three-level LVMs in a longitudinal study and a cluster-randomized study, respectively. One advantage of the LVMs is that they reduce the dimensionality of data. The other one is that the latent variable in the LVMs are included to more correctly model the data. Therefore, they are more powerful to test and assess the coefficient effects. These coefficients represent the global effects of the covariates on the overall interest, child obesity or academic achievement. The LVMs borrow the information of the different observed surrogate outcomes by modeling their correlations and exploiting the nature of the data-that the surrogate outcomes measure the same quantity. The LVMs not only provide a straightforward way to address the most challenging issue of surrogate outcomes measured in different units or scales, but also produce efficiency by estimating many fewer parameters than a direct modeling of covariate effects on the various surrogate outcomes.

In Chapter 2, the two-level LVMs were analyzed to identify risk factors and surrogate outcomes of the child obesity. Given surrogate outcomes subject to missingness, a modified PX-EM algorithm was implemented to NGHS for identifying risk factors. An often mentioned advantage of the EM algorithm is ease of implementation as compared to another optimization method. Theoretical advantages include the fact that each iteration increases the likelihood. One disadvantage of the EM algorithm is that its convergence can be very slow whenever there is a large fraction of missing information. Therefore, the PX-EM algorithm was implemented to accelerate the convergence, and it converged 10 time faster than the EM-algorithm. In contrast to the case for gradient methods such as Newton-Raphson, the EM algorithm does not need to calculate information matrix at each step. That makes the algorithm easier, in particular, for the multivariate surrogate outcomes in longitudinal or high-level cluster-randomized studies with general variance-covariance structure.

Given surrogate outcomes and covariates subject to missingness, Chapter 3 proposed an ap-

proach efficiently to handle missing data in a longitudinal study of the LVMs in Chapter 3. This approach has four advantages: efficient analysis, unbiased estimation, the identification of covariate and random effects, and mild restrictive assumption. The method was generated by defining a normality model for the observed data and basing the inferences on the likelihood and posterior distribution under the model, with parameters estimated by ML. The advantages of model-based approaches are flexibility, efficiency, and unbiased estimation. With covariates subject to missingness, however, the joint model over-identifies the LVMs because there is a single fixed effect of each of level-1 covariates subject to missingness on the latent variable, but there are distinct covariances between the level-1 covariates subject to missingness and the latent variables. Technically, the joint model was imposed some constraints so that one-to-one transformation formula were derived to obtain the unbiased estimation. The simulation study in section 3.5 illustrated that the unconstrained joint model produced biased inferences for the LVMs and the constrained joint model just-identifies the LVMs.

In Chapter 4, three-level LVMs were proposed to investigate if small class had a significant effect on the racial disparities in academic achievement and if the interaction effect between race and small class was random across school with the assumption of subject scores MCAR or MAR. The findings showed that the small class caused high academic achievement for African-American students while for white students reduced class size does not cause higher academic achievement. In addition, small class did not have a random effect on school. These results imply policy-makers that assigning more African Americans to the small class may be a factor of decreasing the education gap of achievement between African-American and white students.

In Chapter 5, we developed an approach via three-level LVMs with IV to study the causality of racial differences in achievement. The approach extended the causal modeling framework of Shin and Raudenbush (2011) and Shin (2012). The expansion enabled the study to find that for African-American students reduced class size causes higher academic achievement and more robust to identify if minority differences randomly across schools through fitting random-coefficient

LVMs. There are some challenges to analyze the models. First, the four observed highly-correlated outcomes are subject to missingness in the three-level cluster-randomized study where schools nest classes and classes then nest students. Second, the academic achievement is influenced by race and class size, and class size is an endogenous variable. Finally, class size is a regressor given class type. With a less restrictive assumption, MAR, than MCAR, the three-level LVMs with IV were efficiently evaluated, and random-coefficient school effects were investigated. By overcoming these challenges, we showed reduced class size has a causal effect on achievement for African-American students, but the treatment effect did not randomly across school.

The modern approach of estimating LVMs is to calculate parameters without resorting to imputing the latent variable. The fact explains a remarkable paucity of research on scoring the latent variable. Unit-specific obesity score was calculated by the posterior probability in Chapters 2-5. Though the trend of the obesity score is similar for the two-level LVMs between the assumption of covariates MCAR and MAR, the scores appear significantly different (Figure 6.1). It seems that the difference is positively associated with age for both African-American and white students and the difference for white students is more positively associated with age than for African Americans.One possible reason is that the case-deletion analysis of missing covariates may result in biased inference. Figure 6.2 indicates that the trend of unit-specific achievement score generated by three-level LVMs (Chapter 4) is similar as that estimated by three-level LVMs with IV (Chapter 5).

Some weaknesses and extensions in this dissertation should be noted. First, the continuous surrogate outcomes were measured to quantify the overall interest. It is very interesting to develop methodologies efficiently to handle missing covariates and surrogate outcomes which are categorical, ordinal, or mixed-type variables. Two broad approaches are implemented for analysis of the scenario on the basis of how to factorize the joint model of these surrogate outcomes. The first one is to postulate a marginal model for binary or ordinal outcomes and then to formulates a conditional model on the categorical or ordinal one(s) for continuous outcomes. For the former, one can

fit logistic or loglinear regression, whereas for the latter conditional models are a straightforward choice. The other one starts from the reverse factorization by combining a marginal model for the continuous outcome and a conditional one for the categorical outcome. However, EM algorithm works best if the complete-data distribution belongs to an exponential family. Like the steps in the normal structure, the E-step consists of calculating the complete-data sufficient statistics by their posterior expectations. Given these estimates, the likelihood equations for the M-step then take the same form as for complete data.

Secondly, the random error terms in the two-level LVMs were independent no matter they are within the subject or between subjects assuming random coefficient effects of age or time. This method is useful for the traditional latent growth curve models in a longitudinal study. A possible extension is to provide a structural model where random error terms are dependent with AR(1) variance-covariance structure. The combination of these two models is called the Autoregressive Latent Trajectory (ALT) model (Bollen and Curran, 2004). However, the extension is challenging to estimate the parameter of the variance-covariance structure, in particular, for the LVMs. The traditional approach for estimating the nuisance parameter is Newton-Raphson method, which, however, is not useful under EM-algorithm with considering the latent variable as MAR.

Thirdly, we assume that the random effects in the structure model are completely observed, and there are no interaction effects among covariates subject to missingness. Otherwise, for the first case, the structure model (2.1) includes products between two random normal variables, which makes the estimation of parameters more difficult due to non-normal framework. One possible way is a Bayesian approach where the covariates are estimated from their prior distributions and the missing data are imputed from their posterior distributions. Another possible approach is based on iteration: imputing the missing covariates with random effects from a marginal model, estimating the LVMs conditional on the multiple imputations of the covariates, and combining the estimates (Rubin, 1987). Though the relaxing assumption will make the general approach more applicable, it is beyond the work in this dissertation. For the second case, the interaction terms should be fitted

in the left-hand side of the joint model, resulting in challenging estimation.

Finally, high-level longitudinal or cluster-randomized data are ubiquitous and subject to missingness, which encourages us to develop an approach efficiently to handle missing surrogate outcomes and covariates in Q -level LVMs ($Q > 3$). Facing high level data, conventionally we expect to include all cluster-specific influences as covariates in the analysis. However, it is always impossible because of limited knowledge regarding relevant covariates and furthermore dataset lacking information on these covariates. Consequently, it is necessary to fit random-intercept or random-coefficient models to account for unobserved heterogeneity leading to the correlation between responses for units in the same cluster after conditioning on covariates. It is of interest to develop an approach efficiently to handle missing surrogate outcomes and covariates in high level LVMs. As the analysis of two-level LVMs in Chapter 3, the joint model over-identifies the conditional model with constraints. The challenge is that how to represent the constraints within the framework of the Q-level model in a standard way regardless of the level Q and how to obtain unbiased and efficient analysis of the conditional model.

Figure 6.1: *The difference of unit-specific obesity score*
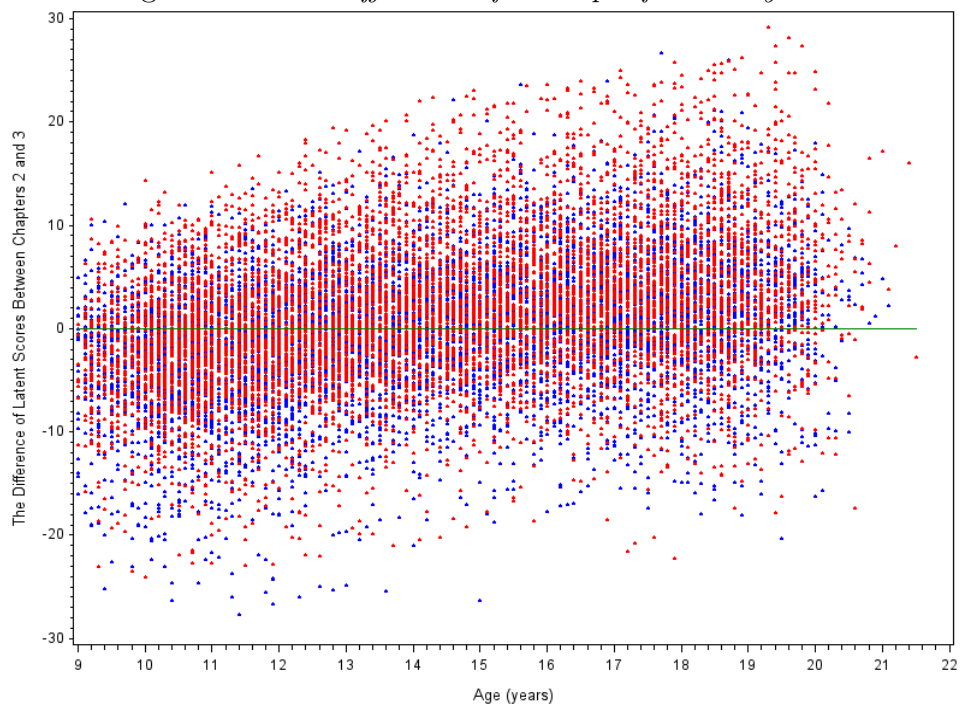
Figure 6.2: *The difference of unit-specific achievement score*
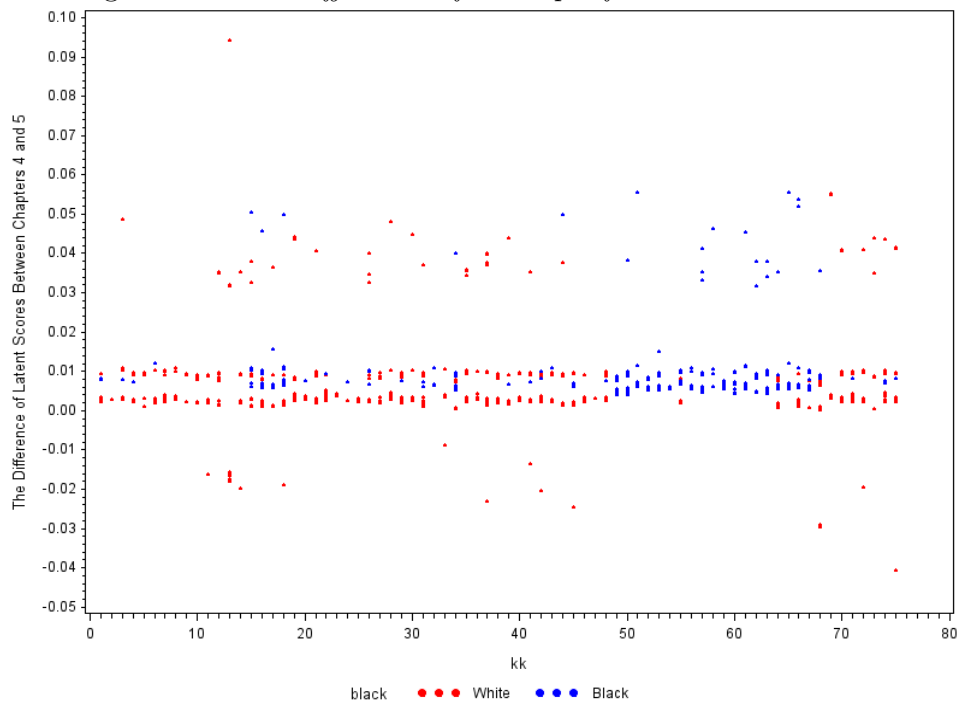
Bibliography

Angrist, J. D., Imbens, G. W., and Rubin, D. B. (1996). Identification of causal effects using instrumental variables. *Journal of the Acoustical Society of America* 91: 444-455.

Arbuckle, J. L. (2003). *Amos 5.0 Update to the Amos 5.0 Users Guide* [Computer Software and Manual], Chicago: Smallwaters.

Bentler, P. M. (2007). *EQS 6 Structural Equations Program Manual*. Encino, CA: Multivariate Software.

Biro, F. M., Lucky, A. W., Simbartl, L. A., Barton, B. A., Daniels, S. R. Striegel-Moore R. et al. (2003). Pubertal maturation in girls and the relationship to anthropometric changes: pathways through puberty. *Journal of Pediatrics* 142(6): 643-646.

Bjorntorp, P. (1988). Abdominal obesity and the development of noninsulin dependent diabetes mellitus. *Diabetes/Metabolism Research and Reviews* 4: 615-622.

Bollen, K. A. (1996). An alternative two stage least squares estimator for latent variable equations. *Psychometrika* 61: 109-121.

Bollen, K. A. and Curran, P. J. (2004). Autoregressive Latent Trajectory (ALT) models: A synthesis of two tradition. *Sociological methods and research* 32: 336-383.

Catalano, P. J. and Ryan, L. M. (1992). Bivariate latent variable models for clustered discrete and continuous outcomes. *Journal of American Statistical Association* 87: 651-658.

Chambers, J. A. and Swanson, V. (2010). A health assessment tool for multiple risk factors for obesity: age and sex differences in the prediction of body mass index. *British Journal of Nutrition* 104(2): 298-307.

Cox, D. R., and Wermuth, N. (1992). Response models for mixed binary and quantitative variables. *Biometrika* 79: 441-461.

Curran, P. J. and Bollen, K. A. (2001). The best of both worlds: Combining autoregressive and latent growth curve models. In Collins L. M. and Sayar, A. G. (Eds.),*New methods for the analysis of change* (PP 105-136). Washington, D.C.: American Psychological Association.

Dean, H. and Flett, B. (2002). Natural history of type 2 diabetes diagnosed in childhood: long term follow-up in young adult years. *Diabetes*, 51: A24 (Abstract) [cited in Ebbeling C. B., Pawlak D.B., Ludwig D.S.. Childhood obesity: public health crisis, common sense cure. Lancet 2002; 360: 473-482].

Dempster, A. P., Laird, N. M. and Rubin, D. B. (1977). Maximum likelihood estimation from incomplete data via the EM algorithm (with Discussion). *Journal of the Royal Statistical Society, Series B* 39: 1-38.

Fessler, J. A. and Hero, A. O. (1994). Space-alternating generalized expectation-maximization algorithm. *IEEE Transactions on Signal Processing* 42: 2664-2677.

Finn, J. D. and Achilles, C. M. (1990). Answers and questions about class size: A statewide experiment. *American Educational Research Journal* 27: 557-577.

Fitzmaurice, G. M., and Laird, N. M. (1995). Regression models for a bivariate discrete and continuous outcome with clustering. *Journal of American Statistical Association* 90: 845-852.

Garn, S. M., Leonard, W. R., Hawthorne, V. M. (1986). Three limitations of the body mass index. *American Journal of Clinical Nutrition* 44: 996-997.

Gelman, A., Van Dyk, A. A., Huang, Z. and Boscard, J. W. (2008). Using redundant parameters to fit hierarchical models. *Journal of Computational and Graphical Statistics* 17: 95-122.

Ghosh, M., Reid, N., and Frasher, D. A. S. (2010). Ancillary statistics: A review. *Statistical Sinica* 20: 1309-1332.

Glass, G. V. and Smith, M. L. (1978). *Meta-analysis of research of the relationship of class size and achievement*. San Francisco: Far West Laboratory for Educational Research and Development.

Goldstein, H. (2003). *Multilevel Statistical Models*. London: Edward Arnold.

Goldstein, H. and Blatchford, P. (1998). Class size and educational achievement: A review of methodology with particular reference to study design. *British Educational Research Journal* 24: 255-268.

Goldstein, H. and Browne, W. (2002). Multilevel factor analysis modeling using markov chain monte carlo estimation in G. Marcoulides and Moustaki I., eds., *Latent variable and latent structure models*, London: Lawrence Erlbaum.

Goldstein, H., Carpenter, J., Kenward, M. G., and Levin K. A. (2009). Multilevel models with multivariate mixed response types. *Statistical Modeling* 9: 173-197.

Goldstein, H. and Kounali, D. (2009). Multilevel multivariate modeling of childhood growth, numbers of growth measurements and adult characteristics. *Journal of the Royal Statistical Society, Series A* 172: 599-613.

Goldstein, H., Yang, M., Omar, R., Turner, R., and Thompson, S. (2000). Meta-analysis using multilevel models with an application to the study of class size effects. *Journal of Reliability and Statistical Studies* 49: 399-412.

Guryan, J., Hurst, E., and Kearney, M. (2008). Parental education and parental time with children. *Journal of Economic Perspectives* 22(3): 23-46.

Hanushek, E. A. (1999). Some findings from an independent investigation of the Tennessee STAR experiment and from other investigations of class size effects. *Educational Evaluation and Policy Analysis* 21: 143-163.

He, Y. and Liu, C. (2009).The dynamic ECME algorithm. Technical report, Department of Statistics, Purdue University.

Heitjan, D. F. and Basu, S. (1996). Distinguishing missing at random and missing completely at random. *The American Statistician* 50(3): 207-213.

Hong, G., and Raudenbush, S. W. (2006). Evaluating Kindergarten retention policy: A case study of causal inference for multilevel observational data. *Journal of the Acoustical Society of America* 101: 901-910.

Hu, F. (2008). Measurements of adiposity and body composition. In: Hu F, ed. *Obesity Epidemiology*. New York City: Oxford University Press: 53-83.

Huenemann, R. L. (1969). Factors associated with teenage obesity. In: Wilson NL, ed. *Obesity*. Philadelphia, Pa: FA Davis Co: 55-66.

Imbens, G. W., and Angrist, J. D. (1994). Identification and estimation of local average treatment effects. *Econometrica* 62: 467-475.

Imbens, G. W., and Rubin, D. B. (1997a). Bayesian inference for causal effects in randomized experiments with noncompliance. *The Annals of Statistics* 25: 305-327.

Imbens, G. W., and Rubin, D. B. (1997b). Estimating outcome distributions for compliers in instrumental variables models. *Review of Economic Studies* 64: 555-574.

Jamshidian, M. and Jennrich, R. I. (1993). Conjugate gradient acceleration of the EM algorithm. *Journal of the American Statistical Association* 88: 221-228.

Johnson, R. A. and Winchern, D. W. (2007). *Applied multivariate statistical analysis*. Upper Saddle River, NJ: Pearson Education, Inc.

Kimm, Y. S. S., Glynn, N. W., Obarzanek, E., Kriska, A. M., Daniels, S., Barton, B. A. et al. (2005), Relation between the changes in physical activity and body-mass index during adolescence: a multicentre longitudinal study. *Lancet* 366: 301-307.

Kiess, W., Marcus, C., Wabitsch, M. (2004). Obesity in Childhood and Adolescence. *Pediatric and Adolescent Medicine* 9: 1-19

Krebs, N. F., Himes, J. H., Jacobson, D., and et al. (2007). Assessment of child and adolescent overweight and obesity. *Pediatrics* 120(Suppl. 4): S193-S228.

Kriemler, S., Jardena, P., Lukas, Z., and et al. (2010). Estimation of percentage body fat in 6- to 13-year-old children by skinfold thickness, body mass index and waist circumference. *British Journal of Nutrition* 104: 1565-1572.

Krueger, A. B. (1999). Experimental estimates of education production functions. *Quarterly Journal of Economics* 114: 497-532.

Krueger, A. B. and Whitmore, D. M. (2001). The effect of attending a small class in the early grades on college-test taking and middle school test results: Evidence from project STAR. *The Economic Journal* 111: 1-28.

Laird, N., Lang, N. and Stram, D. (1987). Maximizing likelihood computations with repeated measures: application of the EM algorithm. *Journal of the American Statistical Association* 82: 97-105.

Laird, N. M. and Ware, J. H. (1982). Random-effects models for longitudinal data. *Biometrics* 38: 963-974.

Lang, K. (1995a). A gradient algorithm locally equivalent to the EM algorithm.*Journal of the Royal Statistical Society, Series B* 57: 425-38.

Lang, K. (1995b). A quasi-Newtonian acceleration of the EM algorithm. *Statistical Sinica* 5: 1-18.

Laurie, B. (2002). Waist girth predicts cardiovascular risk better than BMI. *American Journal Clinical Nutrition* 76: 743-749.

Lavielle, M. and Meza, C. (2007). A parameter expansion version of the SAEM algorithm. *Statistics and Computing* 17: 121-130.

Lewandowski, A., Liu, C., and Wiel, V. S. (2010). Parameter expansion and efficient inference. *Statistical Science* 25(4): 533-544.

Little, R. and Rubin, D. (2002). *Statistical Analysis with Missing Data*, New York: Wiley.

Liu, C., Rubin, B. D., and Wu, Y. (1998). Parameter expansion to accelerate EM: The PX-EM algorithm. *Biometrika* 85(4): 755-770.

Little, R. J. A. and Yau, L. H. Y. (1998). Statistical techniques for analyzing data from prevention trials: Treatment of no-shows using Rubins causal model. *Psychological Methods* 3: 147-159.

Liu, J. S. and Wu, Y. N. (1999). Parameter expansion for data augmentation. *Journal of the American Statistical Association* 94: 1264-1274.

Livingstone, M. B. (2001). Childhood obesity in Europe: a growing concern. *Public Health Nutrition* 4(1A): 109-116.

Mahoney, J. L. (2011). Adolescent summer care arrangements and risk for obesity the following school year. *Journal of Adolescence* 34(4): 737-749.

Martin, R., Hwang, J. S. and Liu, C. (2010). General theory of inferential models II. Marginal inference. Technical report, Department of Statistics, Purdue University.

Martin, R., Zhang, J. and Liu, C. (2010). Dempster-Shafer theory and statistical inference with weak beliefs. *Statistical Science* 25: 72-87.

Maynard, L. M., Wayne, W., and et al. (2001). Childhood body composition in relation to body mass index. *Pediatrics* 107(2): 344-350.

Milesi, C. and Gamoran, A. (2006). Effects of class size and instruction on kindergarten achievement. *Educational Evaluation and Policy Analysis* 28: 287-313.

Meng, X. L. and Rubin, D. B. (1993). Maximum likelihood estimation via the ECM algorithm: a general framework. *Biometrika* 80: 267-278.

Meng, X. L. and Van Dyk, D. A. (1998). Fast EM-type implementation for mixed effects models. *Journal of the Royal Statistical Society, Series B* 60: 559-578.

Morris, C. 1983. Parametric empirical Bayes inference, theory and applications. *Journal of the American Statistical Association* 78: 47-65.

Morrison, J. A. (1992). Obesity and cardiovascular disease risk factors in black and white girls: the NHLBI Growth and Health Study. *American Journal of Public Health* 82: 1613-1620.

Moustaki, I. (2003). A general class of latent variable models for ordinal manifest variables with covariate effects on the manifest and latent variables. *British Journal of Mathematical and Statistical Psychology* 56: 337-357.

Moustaki, I. and Steele, F. (2005). Latent variable models for mixed categorical and survival responses, with an application to fertility preferences and family planning in Bangladesh. *Statistical Modeling* 5: 327-342.

Mosteller, F. (1995). The Tennessee study of class size in the early school grades. *The Future of Children: Critical Issues for Children and Youths* 5: 113-127.

Muthén, L. and Muthén, B. (2010). *Mplus user's guide* (6th ed.), Los Angeles, CA: Muthén and Muthén, 6th edition.

Neal, D. A. and Johnson, W. R. (1996). The role of premarket factors in black-white wage differences. *Journal of Political Economy* 104(5): 869-895.

Nye, B., Hedges, L. V. and Konstantopoulos, S. (1999). The long-term effects of small classes: A five-year follow-up of the Tennessee class size experiment. *Educational Evaluation and Policy Analysis* 21: 127-142.

Nye, B., Hedges, L. V. and Konstantopoulos, S. (2000). The effects of small classes on academic achievement: The results of the Tennessee class size experiment. *American Educational Research Journal* 1: 123-151.

Nye, B., Konstantopoulos, S. and Hedges, L. V. (2004). How large are teacher effects? *Educational Evaluation and Policy Analysis* 26: 237-257.

Obarzanek, E., Schreiber, G. B., Crawford, P. B., Goldman, S. R., Barrier, P. M., Frederick, M. M. et al. (1994). Energy intake and physical activity in relation to indexes of body fat: the National Heart, Lung, and Blood Institute Growth and Health Study. *American Journal of Clinical Nutrition* 60(1): 15-22.

Ogden, C. L., Flegal, K. M., Carroll, M. D., and Johnson, C. L. (2002). Prevalence and trends in overweight among US children and adolescents, 1999-2000. *The Journal of American Association* 288: 1728-1732.

Orchard, T., and Woodbury, M. A. (1972). A missing information principal: theory and ap-

plications. *Proceedings of the Sixth Berkeley Symposium on Mathematical Statistics and Probability* 1: 697-715.

Pamela, E. D. (2005). The influence of parent education and family income on child achievement: The indirect role of parental expectations and the home environment. *Journal of Family Psychology* 19 (2): 294-304.

Patterson, M. L., Stern, S., and et al. (1997). Sociodemographic factors and obesity in preadolescent black and white girls: NHLBI's Growth and Health Study. *Journal of the National Medical Association* 89(9): 594-600.

Pocock, S. T., Geller, N. L., and Tsiatis, A. A. (1987). The analysis of multiple endpoints in clinical trials. *Biometrics* 43(3): 487-498.

Prentice, A. M. and Jebb, S. A. (2001). Beyond body mass index. *Obesity Review* 2: 141-147.

Qi, A. and Jaakola, T. S. (2007). Parameter expanded variational Bayesian methods. In Advances in Neural Information Processing Systems 19. MIT Press, Cambridge.

Roy, J. and Lin, X. (2000). Latent variable models for longitudinal data with multiple continuous outcomes. *Biometrics* 56(4): 1047-1054.

Raudenbush, S. W. (2010). Strategies for modeling interference between units in multisite trials. New Orleans, LA: Presentation at ENAR.

Raudenbush, S. W. and Bryk, A. S. (2002). *Hierarchical Linear Models*. Newbury Park, CA: Sage.

Robinson, G. E. (1990). Synthesis of research on effects of class size. *Educational Leadership* 47(7): 80-90.

Rubin, D. B. (1976). Inference and missing data. *Biometrika* 63: 581-592.

Rubin, D. B. 1983. Some applications of Bayesian statistics to educational data. *Journal of the Royal Statistical Society, Series D* 32: 155-167.

Rubin, D. B. (1987). *Multiple Imputation for Nonresponse in Surveys*. New York, Wiley.

Sabo, T. R., Zheng, L., Stephen, D., and Shumei, S. S. (2010). Relationships between Serial Childhood Adiposity Measures and Adult Blood Pressure: the Fels Longitudinal Study.*American Journal of Human Biology* 22(6): 830-835.

Sammel, M. D., Lin, X., and Ryan, L. (1999). Multivariate linear mixed models for multiple outcomes. *Statistics in Medicine* 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. 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: 667-678.

Schafer, J. L. (1997). *Analysis of Incomplete Multivariate Data*. London: Chapman & Hall.

Schafer, J. and Yucel, R. (2002). Computational strategies for multivariate linear mixed-effects models with missing values. *Journal of Computational and Graphical Statistics* 11: 437-457.

Shah, A., Lairdm N., and Schoenfeld, D. (1997). A random effects model for multiple characteristics with possibly missing data. *Journal of the American Statistical Association* 92: 775-779.

Shin, Y. (2012). Do black children benefit more from small classes? multivariate instrumental variable estimators with ignorable missing data. *Journal of Educational and Behavioral Statistics* 1(1): 1-36.

Shin, Y. and Raudenbush, S. W. (2007). Just-identified versus over-identified two-level hierarchical linear models with missing data. *Biometrics* 63(4): 1262-1268.

Shin, Y. and Raudenbush, S. W. (2010). A latent cluster mean approach to the contextual effects model with missing data, *Journal of Educational and Behavioral Statistics* 35: 26-53.

Shin, Y. and Raudenbush, S. W. (2011). The causal effect of class size on academic achievement: Multivariate instrumental variable estimators with data missing at random, *Journal of Educational and Behavioral Statistics* 36: 154-185.

Shin, Y. and Raudenbush, S. W. (2013). Efficient analysis of Q-level nested hierarchical general linear models given ignorable missing data. *International Journal of Biostatistics* 9(1): 109-133.

Siervogel, R. M., Wisemandle, W, Maynard, L. M., and et al. (2000). Lifetime overweight status in relation to serial changes in body composition and risk factors for cardiovascular disease: The Fels Longitudinal Study. *Obesity Research* 8: 422-430.

Sobal, J. and Stunkard, A. J. (1989). Socioeconomic status and obesity: a review of the literature. *Psychological Bulletin* 105: 260-275.

Song, X. Y., Xia, Y. M. and Lee, S. Y. (2009). Bayesian semi-parametric analysis of structural equation models with mixed continuous and unordered categorical variables. *Statistics in Medicine* 28: 2253-2276.

Sue, Y.S. K., Nancy, W. G., Eva, O., and et al. (2005), Relation between the changes in physical

activity and body-mass index during adolescence: a multicentre longitudinal study. *Lancet* 366: 301-307.

Sun, S. S., Ruohong, L., Terry, T-K H., and et al. (2008). Childhood obesity predicts adult metabolic syndrome: the Fels Longitudinal Study. *Journal of Pediatrics* 152(2): 191-200.

Tybor, J. D., Alice, H. L. and et al. (2010). Racial differences in central adiposity in a longitudinal cohort of black and white adolescent females. *BMC Pediatrics* 10, 2.

Vani, H. R. (2007). Longitudinal associations between television viewing and body mass index among white and black girls. *Journal of Adolescent Health* 41: 544-550.

Wang, Y. (2004). Epidemiology of childhood obesity - Methodological aspects and guidelines: What is new? *International Journal of Obesity* 28(SUPPL. 3): S21-S28.

Walter, E. S. and Samuel, S. W. (2007). *Matrix differential calculus with applications in statistics and econometrics*. John Wiley & Sons Ltd, Baffins Lane, Chichester, West Sussex PO19 1UD, England.

Wei, L. and Zhang, J. (2001). Analysis of Data with Imbalance in the Baseline Outcome Variable for Randomized Clinical Trials. *Drug Information Journal* 35: 1201-1214.

World Health Organization (WHO, 2000). p.9.

Wu, C. F. J. (1993). On the convergence properties of the EM algorithm. *The Analysis of Statistics* 11: 95-103.

Yu, Y. and Meng, X. L. (2011). To center or not to center, that is not the question: An ancillary-sufficiency interweaving strategy (ASIS) for boosting MCMC efficiency. *Journal of Computational and Graphical Statistics* 20(3): 531-570.

Yucel, R. M. (2008). Multiple imputation inference from multivariate multilevel continuous data with ignorable nonresponse. *Philosophical Transactions of the Royal Society, Series A* 366(1874): 2389-2403.

Zamboni, M., Armellini, F., and et al. (1992). Relation of body fat distribution in men and degree of coronary narrowing in coronary artery disease. *American Journal of Cardiology* 70: 1135-1138.

Zeng, Q., Dong, S., and et al. (2012). Percent body fat is a better predictor of cardiovascular risk factors than body mass index. *Brazilian Journal Of Medical and Biological Research* 45(7): 591-600.

Zhang, B., Chen, Z., Albert, P. S. (2012). Latent class models for joint analysis of disease prevalence and high-dimensional semi-continuous biomarker data. *Biostatistics* 13(1): 74-88.

Zhang, J. and Liu, C. (2011). Dempster-Shafer inference with weak beliefs. *Statistical Sinica* 19: 491-544.

Zhu, J., Eickhoff, J. C., Yan, P. (2005). Generalized linear latent variable models for repeated measures of spatially correlated multivariate data. *Biometrics* 61: 674-683.

Appendix

SAS Codes for the Random-coefficient model in Chapter 3

```
∗Final dataset based on NGHS data;
∗PX–EM algorithm for the laten variable analysis based on NGHS data;
∗This program is for just 4 biomarkers and for both
   missing biomarkers and covariates age and the others assuming
   there is 2 missing level−2 covariates;
∗ this program is for random age centered at sample mean and for
  just identified model;
options nofmterr;
libname V1 "~/";
data stack1;
  set v1.thesis;
  by id0;
  keep id0 BMI SUMSKIN MAXBLOAV PCTFATSF PFBIA UPTHIGAV WAISTMIN
          VIDTVWK F12score BMI2 race age parents CATINC MATSTAGE CATEDUC
          income1−income3 educ1−educ2 matur1−matur3 incom1−incom2 CATEDUC1
          CATINC1 maturation;
  if PFBIA<0 then PFBIA=.;
  if VIDTVWK<0 then VIDTVWK=.;
  if nmiss(of race age parents CATEDUC)>=1 then delete;
  ∗if all missing then it does not contribute to estimate;
  if nmiss(of BMI SUMSKIN MAXBLOAV PFBIA VIDTVWK F12score BMI2 CATINC1
          maturation)=9 then delete;
run;
∗center the level−1 covariates at grandmean, level−2 covariates at grandmean
 and weighted average mean, respectively;
proc means data=stack1; ∗getting grand mean;
 var age VIDTVWK F12score  maturation;
run;
∗try to get completely observed level−2 covariates' weighted mean;
ods output onewayfreqs=aa(keep=id0 frequency);
proc freq data=stack1;
 table id0;
run;
data stack0;
 merge stack1 aa;
 by id0;
run;
proc means data=stack0;
  var race parents cateduc1 BMI2 CATINC1;
  weight frequency;
```

```
run ;
proc summary data=stack0 noprint ;
 class id0 ;
 var age ;
 output out=age mean=ageM ;
run ;
data age ; set age ; if id0 =. then delete ; run ;
data stack0 ; merge stack0 age ( keep=ID0 ageM ); by ID0 ; run ;
data stack1 ;
 set stack0 ;
 ager=age −ageM ;
 age=age − 14.3633689;
 VIDTVWK=VIDTVWK − 31.3475711;
 F12score=F12score − 17.3481753;
 maturation=maturation − 2.0972236;
 race=race − 0.4761598;
 parents=parents − 0.3119893;
 cateduc1=cateduc1 − 0.7562707;
 BMI2=BMI2 − 27.3658793;
 CATINC1=CATINC1 − 1.0672173;
run ;

data stack1 ;
   set stack1 ;
   by id0 ;
   if first.id0  then time =1; else time +1;
run ;
proc sort data=stack1 out=p1 ; by time ; run ; *get the max(time );
*Deleting the missing value : for the two interested variable :BMI,SUMSKIN;
%macro level1 ( var1 , var2 , var3 , var4 );
 proc sort data=stack1 ; by id0 age race ;
 data &var1 ( keep=ID0 race &var1 level count status o1−o10 time );
   set stack1 ;
   by ID0 ;
   level=&var2 ;
   count=&var3 ;
   status ="&var4 ";
   array o{*} o1−o10; *the largest time occation is 10;
   do i =1 to 10;
    o[ i ]=( time=i );
   end ;
run ;
%mend ;
%level1 (BMI,1 ,1 ,b );
%level1 (SUMSKIN,1 ,2 ,b );
```

112

```
%level1 (MAXBLOAV, 1 , 3 , b );
%level1 (PFBIA  , 1 , 4 , b );
%level1 (VIDTVWK, 1 , 1 , c );
%level1 ( F12score , 1 , 2 , c );
%level1 ( maturation , 1 , 3 , c );
 data  level1 ;
        set  BMI( rename =(BMI=y ))  SUMSKIN( rename =(SUMSKIN=y ))
            MAXBLOAV( rename =(MAXBLOAV=y ))  PFBIA ( rename =(PFBIA=y ))
            VIDTVWK( rename =(VIDTVWK=y ))  F12score ( rename =( F12score=y ))
             maturation ( rename =( maturation=y ));
                label  y="Observed  response  variable ";
          if  y=.  then  do ;
       o1 = . ;  race = . ;
          end ;
          int =1;
run ;
proc  sort  data=level1  out=o1i ; by ID0    status  count  time ; run ;
%macro  level2 ( var1 , var2 , var3 );
proc  sort  data=stack1  nodupkey  out=rrr ; by  id0 ; run ;
data &var1 ( keep=id0 &var1  race  age  parents  CATEDUC1 &var1  level  count );
   set  rrr ;
   level=&var2 ;
   count=&var3 ;
run ;
%mend ;
%level2 (BMI2 , 2 , 1 );
%level2 (CATINC1 , 2 , 2 );
 data  level2 ( drop=i );
   set  BMI2( rename =(BMI2=y ))  CATINC1( rename =(CATINC1=y ));
         array  o{∗}  o1−o2 ;
         do  i =1  to  2 ;
          o[ i ]=( count=i );
         end ;
    label  y="y";
run ;
proc  sort  data=level2 ; by  id0  count ; run ;
proc  sort  data=level2  nodupkey
 out=idd ( keep=id0  y  level  race  age  parents  CATEDUC1 );
 by  id0 ;
data  level2 ;
   set  level2 ;
   if  y=.  then  do ;
     o1 = . ;              ∗ this  is  to  subtract  the  missing  values  in  IML;
     race = . ;
   end ;
```

```
run;
data final; set level1 level2; run;
*o1io2i is for subtracting o1ij, o2ij, and o2i, x1ij and x2j;
proc sort data=final out=o1io2i;
by id0 level status count time; run;
data o1io2ia; set o1io2i; keep o1-o10; run;

*this dataset is for observed y;
data final (keep=id0 y level status count); set o1io2i;
if y=. then delete; run;
*to get the frequence for each subject to subtract observed y;
proc summary data=final; class id0; var id0; output out=freq n=n; run;
data freq; set freq; if id0=. then delete; run;

*To fet the frequency for each subject after stack the data together;
data freq1; set o1io2i (keep=id0); by id0;
if first.id0 then n=1; else n+1; run;
data freq1; set freq1; by id0; if last.id0; run;

*to get the time visits for each subject;
proc sort data=o1i nodupkey out=freq2; by id0 time; run;
data freq2 (keep=id0 time); set freq2; by id0; if last.id0; run;

*to read Wi: only keep completely observed level-1 and -2 covariates;
data covL1;
  set stack1;
  int=age*race;
  ages=age*age;
  agec=ages*age;
  keep id0 age ages agec CATEDUC1 race parents int ager;
run;
*to read W2i: only keep completely observed level2 covariates;
proc sort data=stack1 out=covL2(keep=id0 cateduc1 race parents) nodupkey;
 by id0 CATEDUC1 race parents; run;

*====================================================================*;
*IML program;
proc iml;
  use final;
  read all var{y} into YY;
  read all var{ID0} into YYY;

  use covL1;
  read all var{age, ages, agec, int, cateduc1, race, parents} into x1;
  read all var{ager} into x3;
```

114

```
nrow=nrow(x1);
D1=j(nrow,1,1)||x3;

x11=x1;
use covL2; read all var{cateduc1,race, parents} into x2;
x21=x2;

use o1io2i;
read all var {y} into YM;

use o1io2ia;
read all var _all_ into o1i2i;

*control the matrix for each subject to get observed y;
use freq;
read all var{n} into cont;

*control the # of observations for each subject after statck data together;
use freq1;
read all var{n} into cont1;

*This is to control time observations for each subject;
use freq2;
read all var{time} into cont2;

vector=unique(yyy);
n=ncol(vector);   *how many subject in the dataset;
nb=4;              *# of biomarkers;
no1=3;              *# of level-1 covariates subject to missing values;
no2=2;             *# of level-2 covariates subject to missing values;
nu=ncol(x1);
ns=no1*(nu);
n2=no2*(ncol(x2));              *there are two columns in x2i;
nd1=ncol(D1);
nT=nd1+no1+no2;
nv=no1;
*# of parameters in the constraint model: this is for information matrix;
IFMD=4*nb+nu+ns+n2+nT*(nT+1)/2+nv*(nv+1)/2-(nd1-1)*(no1+no2);
nv=no1+1;
*# of parameters in the unconstraint model;
IFMD1=4*nb+nu+ns+n2+nT*(nT+1)/2+nv*(nv+1)/2;
print n nb no1 no2 IFMD1 IFMD;
m1=no1; m2=(ncol(x1)-ncol(x2))*no1;
m3=ncol(x2)*no1; m4=ncol(x1)-ncol(x2);m5=ncol(x2);
*initial values based on factor scores;
```

```
Tss=I(no1);
Sigmass=I(no1);
Tau={0.4470,29.2181,7.2996,109.01};
beta0={21.6261,43.9471 ,89.6789,22.7237};
betaR={4.5353,21.3535 ,12.7544,10.1767};
psi={1.1133,30.2868,11.3110,20.3500};
betau={ 0.1678,-0.02125,-0.00129,-0.00733,-0.09182,-0.2534,0.1219};
betaS={-1.8640,-0.08688,0.07507,-0.7811,-5.1856,-16.3731,2.8573,-3.4951,
    0.1990,0.05460,0.9081,2.7883,4.7327,-0.7338,0.3508,-0.03050,-0.00304,
    0.03827,-0.3271,0.03420,0.03053};
beta2={-0.7559,-3.5316, -0.2934 ,0.4845 ,0.3393 ,-0.7387};
sigmauu=0.09684;
sigmas={252.20,181.46,0.1369};
do i=1 to nrow(sigmas);
     sigmass[i,i]=sigmas[i];
end;
Tuu={1.5207  -0.1,-0.1  1.5};
Ts={102.10,  55.1771,11.3342};
T22={44.5275 0,0  0.4203};
do i=1 to nrow(Ts);
     Tss[i,i]=Ts[i];
end;
T=block(Tuu,Tss,T22);
sigma=block(sigmauu,sigmass);
*using the previous results as initial values in clean4.sas;
/*beta0={22.4809 ,46.4339, 92.8110, 25.3868 };
betaR={ 1.1159 ,4.0867, 3.1052, 2.0127};
Tau ={0.5108 ,64.3026, 6.4426 ,15.3091};
psi={ 0.8532, 69.9671, 23.3754, 21.9124 };
betau={0.5435, -0.02112,-0.00128,-0.0514, -0.2440, -1.0114, 0.41594};
betaS={-0.5765, -0.08141, 0.07496,-0.7538, -5.1660, -16.6121, 2.76510,
    -2.47419,0.1925,0.05464,0.83511, 2.64047, 4.85624, -0.75406,0.29237,
    -0.03078,-0.00303,0.045462, -0.020844, -0.19659, -.004158};
beta2 ={-0.7057, -3.4958, -0.4848, 0.4862, 0.34195, -0.72900};
sigma={1.117677 1.1117264 -1.313338 0.1568072,
        1.1117264 256.68172 5.6590696 -0.020782,
        -1.313338 5.6590696 189.21725 -0.916512,
        0.1568072 -0.020782 -0.916512 0.2232569};
T={15.702923  0.901303 1.8871239 -0.507164 0.0620783 5.6753258 -0.005689,
      0.901303 0.1481001           0          0          0          0          0,
      1.8871239          0 99.287296 -16.81754 0.1283915 11.234376 -1.111404,
     -0.507164          0 -16.81754 56.973693 -0.121676 -1.549731 0.6887753,
      0.0620783          0 0.1283915 -0.121676 0.0527143 0.1986027  0.004746,
      5.6753258          0 11.234376 -1.549731 0.1986027 44.764096 -0.290886,
     -0.005689          0 -1.111404 0.6887753  0.004746 -0.290886 0.4255986};
```

116

```
 D={15.10496   0.901303, 0.901303  0.1481001};
 alpha={0.0044694,-0.003745,0.6874041,0.1230068,0.0807719, 0.5305924,
 -0.130339, 0.1164613,-0.902599, 0.5033269};*/
 epi=0.5;
 iter=0;
 rpsi=diag(J(1,nb,1)@psi);
 rtau=diag(J(1,nb,1)@tau);
 sigmauu=sigma[1:1,1:1];
 sigmaus=sigma[1:1,2:(1+no1)];
 row=2;
 col=1+no1;
 sigmass=sigma[row:col,row:col];
 Tus=T[1:nd1,(nd1+1):(nd1+no1)];
 Tu2=T[1:nd1,(nd1+no1+1):(nd1+no1+no2)];
 row=nd1+1;
 col=nd1+no1;
 Tss=T[row:col,row:col];
 Ts2=T[row:col,(col+1):(col+no2)];
 row=nd1+no1+1;
 col=nd1+no1+no2;
 T22=T[row:col,row:col];
 Tuu=T[1:nd1,1:nd1];
*clculating initial value in the conditional model and these initial
 values are from completely observed data from above initial values;
 alpha1=(sigmaus*inv(sigmass))';
 Tus[1,]=alpha1'*(Tss-Ts2*inv(T22)*Ts2')+Tu2[1,]*inv(T22)*Ts2';
 alpha2=inv(T22)*(Tu2[1,]'-Ts2'*alpha1);
 m1=no1; m2=(ncol(x1)-ncol(x2))*no1;
 m3=ncol(x2)*no1; m4=ncol(x1)-ncol(x2); m5=ncol(x2);
 *betas1=j(m1,1,0); *intercept effect;
 *fixed effect on missing leve-1 covariates from completely observed
 level-1 covariates;
 betas2=j(m2,1,0);
*fixed effect on missing leve-1 covariates from completely observed
 level-2 covariates;
 betas3=j(m3,1,0);
*fixed effect on latent variable of completely observed level-1 covariates;
 betau1=j(m4,1,0);
*fixed effect on latent variable of completely observed level-2 covariates;
 betau2=j(m5,1,0);
 betas2[1:m4]=betas[1:m4]; betas2[(m4+1):(2*m4)]=betas[(1+m4+m5):(2*m4+m5)];
 betas2[(2*m4+1):(3*m4)]=betas[(1+2*m4+2*m5):(3*m4+2*m5)];
 betas3[1:m5]=betas[(m4+1):(m4+m5)];
 betas3[(m5+1):(2*m5)]=betas[(2*m4+m5+1):(2*m4+2*m5)];
 betas3[(2*m5+1):(3*m5)]=betas[(3*m4+2*m5+1):(3*m4+3*m5)];
```

```
beta22=j(ncol(x2)*no2,1,0);
beta22=beta2;
betau1=betaU[1:m4];
betau2=betaU[(m4+1):(m4+m5)];
alpha3=betau1-alpha1`@I(m4)*betas2;
alpha4=betau2-(alpha1`@I(m5))*betas3-(alpha2`@I(m5))*beta22;
Dhat=Tuu-alpha2`*T22*alpha2-2*alpha1`*Ts2*alpha2-alpha1`*Tss*alpha1;
alpha=alpha1//alpha2//alpha3//alpha4;
Tus[2,]=0; *random slope effect model has this constraints;
Tu2[2,]=0;
tau11=(betaR*betaR`)@Tuu;
tau12=betaR@Tus;
tau13=betaR@Tu2;
tau21=tau12`;
tau22=Tss;
tau23=Ts2;
tau31=tau13`;
tau32=tau23`;
tau33=T22;
tau7=(tau11||tau12||tau13)//(tau21||tau22||tau23)//(tau31||tau32||tau33);
a41a2=(Tuu||Tus)//(Tus`||Tss); *related to beta2;
a41a1=Tu2//Ts2;
a41a=T22-a41a1`*inv(a41a2)*a41a1;

*a function to delete rows with at least one missing value;
start delrow(x);
 c = cmiss(x);            /** matrix of zeros and ones **/
 count = c[,+];           /** add across columns **/
 mIdx = loc(count>0);     /** find rows with one or more missing values **/
 test=nrow(x)-ncol(mIdx);
 if  test>0 then do;
  NMIdx = setdif( 1:nrow(x),mIdx);   /** find nonmissing rows **/
  return(x[NMIdx,]);
 end;
  else if test=0 then return(test);
 finish;

LogLH1=-10**16; *setting the -2logL at initial values;
Betast=j(ncol(x11),no1,0);
beta2t=j(ncol(x21),no2,0);

do while (epi>.00001|iter<8); *this is for the log likelihood function;
   iter=iter+1;
   LogLH=LogLH1;
   *These are for iteration to get the summation;
```

```
a1 =0; a2 =0; a3 =0; a4 =0;  a5 =0;
a6=j(nb,1,0);  *related to beta0[j];
a7=j(nb,1,0);  *related to beta1[j];
a8 =0;
a9 =0;
a10=j(nv,nv,0);
a11=j(nt,nt,0);
a12=j((nu+ns),1,0);
a13=j(n2,1,0);
Es1iey=j(no1,1,0);
Ey2ey=j(no2,1,0);
ey2jiy=j(no2,1,0);
alphaE=0;
alphaD=0;
covy2jy=j(no2,1,0);
term2=j(no1,no2,0);
ca12=0;
cbeta0=0;
cbeta1=0;
cbeta11=0;
EXXY11a=j(no1,no1,0);
cdhat=0;
*these are related to E(xi) in the conditional model;
do j=1 to no1;
    Betast[,j]=betaS[((j-1)*ncol(x11)+1):(j*ncol(x11))];
end;
do j=1 to no2;
 Beta2t[,j]=beta2[((j-1)*ncol(x21)+1):(j*ncol(x21))];
end;
do i=1 to n;
 nn=i;
 a=0;
 do k=1 to i-1;a=cont[k]+a;end;a=a+1;
 b=0;
 do k=1 to i; b=cont[k]+b;end;   *a and b here are to restrict YY[i];

 aa=0;
 do k=1 to i-1;aa=cont2[k]+aa;end;aa=aa+1;
 bb=0;
 do k=1 to i; bb=cont2[k]+bb;end;   *aa and bb are to restrict XX[i];

 *aaa and bbb restrict olio2i for each subject;
 aaa=0;
 do k=1 to i-1;aaa=cont1[k]+aaa;end;aaa=aaa+1;
 bbb=0;
```

```
  do k=1 to i; bbb=cont1[k]+bbb;end;

*read the map matrix from completely data to observed data;
 ni=cont1[i];
 nni=cont2[i];   es1jiy=j(no1,nni,0);
 ob1=o1i2i[aaa:(aaa+nni−1), 1:nni];
 ob2=o1i2i[(aaa+nni):(aaa+2*nni−1), 1:nni];
 ob3=o1i2i[(aaa+2*nni):(aaa+3*nni−1), 1:nni];
 ob4=o1i2i[(aaa+3*nni):(aaa+4*nni−1), 1:nni];
 o1ij=block(ob1,ob2,ob3,ob4);
 o11i= delrow(o1ij);

 os1=o1i2i[(aaa+4*nni):(aaa+5*nni−1), 1:nni];
 os2=o1i2i[(aaa+5*nni):(aaa+6*nni−1), 1:nni];
 os3=o1i2i[(aaa+6*nni):(aaa+7*nni−1), 1:nni];
 o2ij=block(os1,os2,os3);
 o12i= delrow(o2ij);

 o2ai=o1i2i[(aaa+7*nni):(aaa+7*nni+no2−1),1:no2];
 c=cmiss(o2ai);
 count=c[,+];
 mIdx=loc(count>0);
 aaaa=no2−ncol(mIdx);
 if aaaa>0 then   do;
  NMIdx = setdif( 1:nrow(o2ai), mIdx );
  o22i = o2ai[NMIdx,];
 end;
  else if aaaa=0 then o22i=0;
*based on the three map indicator matrices to get the mean and variance
of the observed data fro each subject;
*totally, there are 8 cases. Basically, the  first case can be ignored
 because we select data by excluding it;
 ki1=j(nni,1,1);
 *use o11i=0 & o12i=0 &o22i=0 doesn't work;
 if o11i=0 then n11=1000; else n11=ncol(o11i);
 if o12i=0 then n12=1000; else n12=ncol(o12i);
 if o22i=0 then n22=1000; else n22=ncol(o22i);

 covui=D1[aa:bb,]*Tuu*D1[aa:bb,]'+sigmauu@I(nni); *related to EUY and EUUY;
 CUY1=betaR'@covui;
 CUY2=D1[aa:bb,]*Tus*(I(no1)@ki1')+sigmaus@I(nni);
 CUY3=D1[aa:bb,]*Tu2;

 covey2=j(nni,no1*nni,0); *related to covey;
 covey3=j(nni,no2,0);
```

```
covby2=j(1,no1*nni,0);      *related to covby;
covby3=j(1,no2,0);
*related to covb1iy and covb2iy;
covb1iy1=(betaR'@(Tuu*D1[aa:bb,]'))//(betaR'@(Tus'*D1[aa:bb,]'));
covb1iy2=(Tus@ki1')//(Tss@ki1');
covb1iy3=Tu2//Ts2;
covb2iy1=betaR'@(Tu2'*D1[aa:bb,]');
covb2iy2=Ts2'@ki1';
covb2iy3=T22;


a311a=x1[aa:bb,];
a311d=I(no1)@x11[aa:bb,];
a311b=j(nrow(a311a),ncol(a311d),0);
a311c=j(nrow(a311d),ncol(a311a),0);
a311=((a311a||a311b)//(a311c||a311d));
a312=inv(sigma)@I(nni);
a313=a311';
a31=a313*a312*a311;
a3=a3+a31;            *related to beta1*;


a41b=X21[i,]'*X21[i,];
a41=inv(a41a)@a41b;
a4=a4+a41;         *related to beta2;
*related to alpha in the reduced model;
ccovye=(betaR@I(nni))//j(nni*no1+no2,nni,0);


a61=j(nb,1,0);a71=j(NB,1,0);a81=j(nb,1,0);a91=j(nb,1,0);
Eustey1=j(nb,1,0);  Eustey2=j(nb,1,0);
yi=yy[a:b,];

if n11=1000 & n12=1000 & n22=1000 then oi="";
 else if n11^=1000 & n12=1000 & n22=1000 then
  oi=o11i||j(nrow(o11i),no1*nni+no2,0);
   else if n11=1000 & n12^=1000 & n22=1000 then
    oi=j(nrow(o12i),nb*nni,0)||o12i||j(nrow(o12i),no2,0);
     else if n11=1000 & n12=1000 & n22^=1000 then
      oi=j(nrow(o22i),nb*nni,0)||j(nrow(o22i),no1*nni,0)||o22i;
       else if n11^=1000 & n12^=1000 & n22=1000 then
        oi=(o11i||j(nrow(o11i),no1*nni+no2,0))
           //(j(nrow(o12i),nb*nni,0)||o12i||j(nrow(o12i),no2,0));
          else if n11^=1000 & n12=1000 & n22^=1000 then
           oi=(o11i||j(nrow(o11i),no1*nni+no2,0))
              //(j(nrow(o22i),nb*nni+no1*nni,0)||o22i);
             else if n11=1000 & n12^=1000 & n22^=1000 then
```

```
                  oi =( j ( nrow ( o12i ) , nb * nni , 0 ) || o12i || j ( nrow ( o12i ) , no2 , 0 ))
                      //( j ( nrow ( o22i ) , nb * nni+no1 * nni , 0 ) || o22i );
                    else  if  n11^=1000 & n12^=1000 & n22^=1000 then
                      oi=block ( o11i , o12i , o22i );

    mui1 =( beta0 @ ki1+betaR @ ( x1 [ aa : bb , ] * betaU ));
    mui2 =( I ( no1 ) @ x11 [ aa : bb , ]) * betaS ;
    mui3 =( I ( no2 ) @ x21 [ i , ]) * beta2 ;
    mui=oi *( mui1 // mui2 // mui3 );
    zi=block ( I ( nb ) @ D1 [ aa : bb , ] , I ( no1 ) @ ki1 , I ( no2 ));
    v1=zi * tau7 * zi ';
    v2=block ( rpsi @ j ( nni , nni , 1 ) , j ( nni * no1 , nni * no1 , 0 ) , j ( no2 , no2 , 0 ));
    v31 =((( betaR * betaR ') @ sigmauu @ I ( nni )) || ( betaR @ sigmaus @ I ( nni )))
         //(( betaR @ sigmaus @ I ( nni )) ' || ( sigmass @ I ( nni )));

  v3=block ( v31 , j ( no2 , no2 , 0 ));
  v4=block ( rtau @ I ( nni ) , j ( nni * no1 , nni * no1 , 0 ) , j ( no2 , no2 , 0 ));
  vit=v1+v2+v3+v4 ;  vit =1/2 *( vit+vit ');
  vi=oi * vit * oi ';  vi =1/2 *( vi+vi ');
  CUY=(CUY1 || CUY2 || CUY3) * Oi ';
   svidf=solve ( vi , yi−mui );
  EUY=x1 [ aa : bb , ] * betaU +CUY * svidf ;
  do  j =1 to nb ;
    if  j =1 then   do ;
     covey1 =( tau [ j ] * I ( nni )) || j ( nni , ( nb−j ) * nni , 0 );
     covby1 =( ki1 ' @ psi [ j ]) || j ( 1 , ( nb−j ) * nni , 0 );
     end ;
     else  if  j=nb then do ;
      covey1=j ( nni , ( j −1) * nni , 0 ) || ( tau [ j ] * I ( nni ));
      covby1=j ( 1 , ( j −1) * nni , 0 ) || ( ki1 ' @ psi [ j ]);
         end ;
      else  if  ( j^=1 & j^=nb ) then  do ;
       covey1=j ( nni , ( j −1) * nni , 0 ) || ( tau [ j ] * I ( nni )) || j ( nni , ( nb−j ) * nni , 0 );
       covby1=j ( 1 , ( j −1) * nni , 0 ) || ( ki1 ' @ psi [ j ]) || j ( 1 , ( nb−j ) * nni , 0 );
      end ;
   covey =( covey1 || covey2 || covey3 ) * oi ';
  EUEY=EUY ' * covey * svidf − trace (CUY * solve ( vi , covey '));
  EEY=covey * svidf ;
   cove=tau [ j ] * I ( nni );
  EEEY=EEY ' * EEY+ trace ( cove−covey * solve ( vi , covey '));
  a61 [ j ]=ki1 ' * EEY ;
  a71 [ j ]=EUEY ;
  a81 [ j ]=EEEY ;
  covby =( covby1 || covby2 || covby3 ) * oi ';
  EBY=covby * svidf ;
```

122

```
    EB2Y=EBY*EBY+psi[j]-covby*solve(vi,covby');
     a91[j]=EB2Y;
   end;

  do k=1 to nni;
    deltak=j(nni,1,0);
    deltak[k]=1;
    coviya1=((betaR'@deltak'@sigmauu)//(betaR'@deltak'@sigmaus'));
    coviya2=((sigmaus@deltak')//(sigmass@deltak'));
    coviya3=j(nrow(coviya1),no2,0);
    coviya=(coviya1||coviya2||coviya3)*oi';
    EiYa=coviya*svidf;            *the conditional expectation of epsilon*;
    cic=coviya*solve(vi,coviya');
    cic=(cic+cic')/2;
    EiiYa=EiYa*EiYa'+sigma-cic;
    a10=a10+EiiYa;
  end;

coviy1=((betaR'@sigmauu@I(nni))//(betaR'@sigmaus'@I(nni)));
coviy2=((sigmaus@I(nni))//(sigmass@I(nni)));
coviy3=j(nrow(coviy1),no2,0);
coviy=(coviy1||coviy2||coviy3)*oi';
covbsy1=(betaR'@(Tuu*D1[aa:bb,]'))//(betaR'@(Tus'*D1[aa:bb,]'))
        //(betaR'@(Tu2'*D1[aa:bb,]'));
covbsy2=(Tus@ki1')//(Tss@ki1')//(Ts2'@ki1');
covbsy3=Tu2//Ts2//T22;
covbsy=(covbsy1||covbsy2||covbsy3)*oi';

covb1iy=(covb1iy1||covb1iy2||covb1iy3)*oi';
covb2iy=(covb2iy1||covb2iy2||covb2iy3)*oi';
svicuy=solve(vi,CUY');
EUUY=EUY'*EUY+trace(covui-CUY*svicuy);
aa11=ki1'*ki1;
aa12=ki1'*EUY;
aa13=aa12';
aa14=EUUY;
aa1=(aa11||aa12)//(aa13||aa14);

EiY=coviy*svidf;              *the conditional expectation of epsiloni*;
a121=a313*a312*EiY;

EBSY=covbsy*svidf;
svibsy=solve(vi,covbsy');
EBSBSY=EBSY*EBSY'+T-covbsy*svibsy;
a111=EBSBSY;
```

```
EB1iy=covb1iy*svidf;
EB2iy=covb2iy*svidf;
a131=inv(a41a)@x21[i,]'*(EB2iy-a41a1'*inv(a41a2)*EB1iy);
a1=a1+aa1;  *related to betaj;
a2=a2+nni;    *related to tauj;


a6=a6+a61;    *related to beta0;
a7=a7+a71;    *related to beta1;
a8=a8+a81;    *related to tau;
a9=a9+a91; *related to xi;
*a10=a10+a101; *related to sigma;
a11=a11+a111; *related to T;
a12=a12+a121; *related to beta1;
a13=a13+a131; *related to beta2;


*following is about the estimates of alpha in the conditional model;
CEEY=ccovye'*oi'*svidf;
do j=1 to no1;
 cou1=(j-1)*(ncol(x1))+1;
 cou2=j*(ncol(x1));
 covs1jy=vit[((nb+(j-1))*nni+1):((nb+j)*nni),];
 es1jiy[j,]=(x11[aa:bb,]*betas[cou1:cou2]+covs1jy*oi'*svidf)';
 item2=es1jiy[j,]*CEEY;
 item1=-trace(covs1jy*oi'*solve(vi,oi*ccovye));
 Es1iey[j]=item1+item2;
end;


do j=1 to no2;
 covy2jy=(betaR'@(Tu2[,j]'*D1[aa:bb,]'))||(Ts2[,j]'@ki1')||T22[j,];
 ey2jiy[j]=x21[i,]*beta2[((j-1)*ncol(x21)+1):(j*ncol(x21))]+covy2jy*oi'*svidf;
 item3=ey2jiy[j]*ki1'*CEEY;
 item4=-trace(covy2jy@ki1*oi'*solve(vi,oi*ccovye));
 Ey2ey[j]=item3+item4;
end;
EXEY=Es1iey//Ey2ey//(x1[aa:bb,1:m4]'*CEEY)//(x2[i,]'@ki1'*CEEY);
alphaE=alphaE+EXEY;
do j=1 to no1;
 delta1=vit[((nb+(j-1))*nni+1):((nb+j)*nni),];
 do k=1 to no1;
  delta2=vit[((nb+(k-1))*nni+1):((nb+k)*nni),];
  EXXY11a[j,k]=trace(Tss[j,k]*j(nni,nni,1)+sigmass[j,k]*I(nni)
               -delta1*oi'*inv(vi)*oi*delta2');
 end;
end;
```

```
EXXY11=es1jiy*es1jiy '+EXXY11a;EXXY11=1/2*(EXXY11+EXXY11');
 do k=1 to no1;
  covs1jy=vit[((nb+(k-1))*nni+1):((nb+k)*nni),]*oi ';
  do j=1 to no2;
    covy2jy=((betaR'@(Tu2[,j]'*D1[aa:bb,]')@ki1)||(Ts2[,j]'@ki1'@ki1)
             ||(T22[j,]@ki1))*oi ';
    term2[k,j]=trace(Ts2[k,j]*j(nni,nni,1)-covs1jy*inv(vi)*covy2jy ');
  end;
 end;

 EXXY12=es1jiy*(ey2jiy'@ki1)+term2;
 EXXY13=es1jiy*x1[aa:bb,1:m4];
 EXXY14=es1jiy*(x2[i,]@ki1);
 pass=((betaR'@(Tu2'*D1[aa:bb,]'))||(Ts2'@ki1')||T22)*oi ';
 Ey2iTyi=(I(no2)@x21[i,])*beta2+pass*svidf;
 Covy2iTyi=T22-pass*inv(vi)*pass ';
 EXXY22=(Ey2iTyi*Ey2iTyi')@(ki1'*ki1)+Covy2iTyi@(ki1'*ki1);
 EXXY22=1/2*(EXXY22+EXXY22');
 EXXY23=Ey2iTyi@ki1'*x1[aa:bb,1:m4];
 EXXY24=Ey2iTyi@ki1'*(x2[i,]@ki1);
 EXXY34=x1[aa:bb,1:m4]'*(x2[i,]@ki1);
 EXXY33=x1[aa:bb,1:m4]'*x1[aa:bb,1:m4]; EXXY33=1/2*(EXXY33+EXXY33');
 EXXY44=(x2[i,]'*x2[i,])@(ki1'*ki1); EXXY44=1/2*(EXXY44+EXXY44');
 EXXY=(EXXY11||EXXY12||EXXY13||EXXY14)//(EXXY12'||EXXY22||EXXY23||EXXY24)
       //(EXXY13'||EXXY23'||EXXY33||EXXY34)
       //(EXXY14'||EXXY24'||EXXY34'||EXXY44);
 EXXY=1/2*(EXXY+EXXY');
 alphaD=alphaD+EXXY;
 *estimate the variance in the conditional model;
 covaiy1=(betaR'@(Dhat*D1[aa:bb,]')||j(nd1,nni*no1+no2,0))*oi ';
 covaiy=Dhat-covaiy1*inv(vi)*covaiy1 ';
 Eaiyi=covaiy1*svidf;
 Eaiaiyi=Eaiyi*Eaiyi'+covaiy;
 cdhat=cdhat+Eaiaiyi;
end;

do j=1 to nb;
 beta01=beta0[j]//betaR[j];
 beta01=beta01+inv(a1)*(a6[j]//a7[j]);
 beta0[j]=beta01[1];
 betaR[j]=beta01[2];
end;
tau=(1/a2)*a8;
psi=a9/n;
sigma=a10/a2;
```

```
T=a11/n;
beta1=betaU // betaS;
beta1=beta1+inv(a3)*a12;
beta2=beta2+inv(a4)*a13;
betaU=beta1[1:nu,];
betaS=beta1[(nu+1):(nu+ns),];
alpha=alpha+inv(alphaD)*alphaE;
Dhat=1/n*cdhat;
sigmass=sigma[2:(1+no1),2:(1+no1)];
Tuu=T[1:nd1,1:nd1];
Tus=T[1:nd1,(nd1+1):(nd1+no1)];
Tu2=T[1:nd1,(nd1+no1+1):(nd1+no1+no2)];
Tss=T[(nd1+1):(nd1+no1),(nd1+1):(nd1+no1)];
Ts2=T[(nd1+1):(nd1+no1),(nd1+no1+1):(nd1+no1+no2)];
T22=T[(nd1+no1+1):(nd1+no1+no2),(nd1+no1+1):(nd1+no1+no2)];

*use the constraint to update the parameters in the joint model;
alpha1=alpha[1:no1]; alpha2=alpha[(no1+1):(no1+no2)];
alpha3=alpha[(no1+no2+1):(no1+no2+ncol(x1)-ncol(x2))];
alpha4=alpha[(no1+no2+ncol(x1)-ncol(x2)+1):(no1+no2+ncol(x1))];
sigmaus=alpha1'*sigmass;
m1=no1; m2=(ncol(x1)-ncol(x2))*no1;
m3=ncol(x2)*no1; m4=ncol(x1)-ncol(x2); m5=ncol(x2);
betas2=j(m2,1,0);
betas3=j(m3,1,0);
betau1=j(m4,1,0);
betau2=j(m5,1,0);

*remember to change the following program if we have different
# of completely observed level-1 and level-2 covariates;
betas2[1:m4]=betas[1:m4]; betas2[(m4+1):(2*m4)]=betas[(1+m4+m5):(2*m4+m5)];
betas2[(2*m4+1):(3*m4)]=betas[(1+2*m4+2*m5):(3*m4+2*m5)];
betas3[1:m5]=betas[(m4+1):(m4+m5)];
betas3[(m5+1):(2*m5)]=betas[(2*m4+m5+1):(2*m4+2*m5)];
betas3[(2*m5+1):(3*m5)]=betas[(3*m4+2*m5+1):(3*m4+3*m5)];
beta22=beta2;
betau1=alpha3+alpha1'@I(m4)*betas2;
betau2=alpha4+alpha1'@I(m5)*betas3+alpha2'@I(m5)*beta22;
Tu2[1,]=alpha2'*T22+alpha1'*Ts2;
Tus[1,]=alpha1'*Tss+alpha2'*Ts2';
Tuu=Dhat+block(Tu2[1,]*inv(T22)*Tu2[1,]'
    +alpha1'*(Tss-Ts2*inv(T22)*Ts2')*alpha1,0);
sigmauu=1+alpha1'*sigmass*alpha1;
betau=betau1 // betau2;
*use constraint for delta at each M step to make the model identifiable;
```

126

```
/*          temp=sqrt(1+sigmaus*solve(sigmass,sigmaus'));
            sigmaus=(sigmaus*temp)/sqrt(sigmauu);
            betau=(betau*temp)/sqrt(sigmauu);
            betaR=betaR*sqrt(sigmauu)/temp;
            Tuu=Tuu*temp*temp/sigmauu;
            Tus=Tus*temp/sqrt(sigmauu);
            Tu2=Tu2*temp/sqrt(sigmauu);
            sigmauu=temp*temp;   */

 *updating these two values because when we fit a random slope effect model,
  we have these constraints;
 Tu2[2,]=0;
 Tus[2,]=0;
 sigma=(sigmauu||sigmaus)//(sigmaus'||sigmass);
 T=(Tuu||Tus||Tu2)//(Tus'||Tss||Ts2)//(Tu2'||Ts2'||T22);


  ***********************************************************************;
rpsi=diag(J(1,nb,1)@psi);
rtau=diag(J(1,nb,1)@tau);
tau11=(betaR*betaR')@T[1:nd1,1:nd1];
tau12=betaR@T[1:nd1,(nd1+1):(nd1+no1)];
tau13=betaR@T[1:nd1,(nd1+no1+1):(nd1+no1+no2)];
tau22=T[(nd1+1):(nd1+no1),(nd1+1):(nd1+no1)];
tau23=T[(nd1+1):(nd1+no1),(nd1+no1+1):(nd1+no1+no2)];
tau33=T[(nd1+no1+1):(nd1+no1+no2),(nd1+no1+1):(nd1+no1+no2)];
tau7=(tau11||tau12||tau13)//(tau12'||tau22||tau23)//(tau13'||tau23'||tau33);
sigmauu=sigma[1,1];
sigmaus=sigma[1,2:(1+no1)];
sigmass=sigma[2:(1+no1),2:(1+no1)];
Tuu=T[1:nd1,1:nd1];
Tus=T[1:nd1,(nd1+1):(nd1+no1)];
Tu2=T[1:nd1,(nd1+no1+1):(nd1+no1+no2)];
Tss=T[(nd1+1):(nd1+no1),(nd1+1):(nd1+no1)];
Ts2=T[(nd1+1):(nd1+no1),(nd1+no1+1):(nd1+no1+no2)];
T22=T[(nd1+no1+1):(nd1+no1+no2),(nd1+no1+1):(nd1+no1+no2)];

*calculate the observed log-likelihood function;
a14=0;
do i=1 to n;
 a=0;do k=1 to i-1;a=cont[k]+a;end;a=a+1;
 b=0;do k=1 to i; b=cont[k]+b;end;  *a and b here are to restrict YY[i];

 aa=0;do k=1 to i-1;aa=cont2[k]+aa;end;aa=aa+1;
 bb=0;do k=1 to i; bb=cont2[k]+bb;end;  *aa and bb here are to restrict XX[i];
```

127

```
aaa=0;do k=1 to i-1;aaa=cont1[k]+aaa;end;aaa=aaa+1;
bbb=0;do k=1 to i; bbb=cont1[k]+bbb;end;

*read the map matrix from completely data to observed data;
ni=cont1[i];
nni=cont2[i];
ob1=o1i2i[aaa:(aaa+nni-1), 1:nni];
ob2=o1i2i[(aaa+nni):(aaa+2*nni-1), 1:nni];
ob3=o1i2i[(aaa+2*nni):(aaa+3*nni-1), 1:nni];
ob4=o1i2i[(aaa+3*nni):(aaa+4*nni-1), 1:nni];
o1ij=block(ob1,ob2,ob3,ob4);
o11i= delrow(o1ij);

os1=o1i2i[(aaa+4*nni):(aaa+5*nni-1), 1:nni];
os2=o1i2i[(aaa+5*nni):(aaa+6*nni-1), 1:nni];
os3=o1i2i[(aaa+6*nni):(aaa+7*nni-1), 1:nni];
o2ij=block(os1,os2,os3);
o12i= delrow(o2ij);

o2ai=o1i2i[(aaa+7*nni):(aaa+7*nni+no2-1),1:no2];
c=cmiss(o2ai);
count=c[,+];
mIdx=loc(count>0);
aaaa=no2-ncol(mIdx);
if aaaa>0 then do;
 NMIdx = setdif( 1:nrow(o2ai), mIdx );
 o22i = o2ai[NMIdx,];
end;
   else if aaaa=0 then o22i=0;
ki=nni;
ki1=j(nni,1,1);
if o11i=0 then n11=1000; else n11=ncol(o11i);
if o12i=0 then n12=1000; else n12=ncol(o12i);
if o22i=0 then n22=1000; else n22=ncol(o22i);
   if n11=1000 & n12=1000 & n22=1000 then oi="";
      else if n11^=1000 & n12=1000 & n22=1000 then
        oi=o11i||j(nrow(o11i),no1*nni+no2,0);
         else if n11=1000 & n12^=1000 & n22=1000 then
           oi=j(nrow(o12i),nb*nni,0)||o12i||j(nrow(o12i),no2,0);
            else if n11=1000 & n12=1000 & n22^=1000 then
             oi=j(nrow(o22i),nb*nni,0)||j(nrow(o22i),no1*nni,0)||o22i;
              else if n11^=1000 & n12^=1000 & n22=1000 then
               oi=(o11i||j(nrow(o11i),no1*nni+no2,0))
                  //(j(nrow(o12i),nb*nni,0)||o12i||j(nrow(o12i),no2,0));
                 else if n11^=1000 & n12=1000 & n22^=1000 then
```

```
              oi=(o11i||j(nrow(o11i),no1*nni+no2,0))
                //(j(nrow(o22i),nb*nni+no1*nni,0)||o22i);
               else if n11=1000 & n12^=1000 & n22^=1000 then
                oi=(j(nrow(o12i),nb*nni,0)||o12i||j(nrow(o12i),no2,0))
                  //(j(nrow(o22i),nb*nni+no1*nni,0)||o22i);
                 else if n11^=1000 & n12^=1000 & n22^=1000 then
                    oi=block(o11i,o12i,o22i);

        mui1=(beta0@ki1+betaR@(x1[aa:bb,]*betaU));
        mui2=(I(no1)@x11[aa:bb,])*betaS;
        mui3=(I(no2)@x21[i,])*beta2;
        mui=oi*(mui1//mui2//mui3);
        zi=block(I(nb)@Dl[aa:bb,],I(no1)@ki1,I(no2));
        v1=zi*tau7*zi'; v1=1/2*(v1+v1');
        v2=block(rpsi@j(nni,nni,1),j(nni*no1,nni*no1,0),j(no2,no2,0));
        v31=(((betaR*betaR')@sigmauu@I(nni))||(betaR@sigmaus@I(nni)))
            //((betaR@sigmaus@I(nni))'||(sigmass@I(nni)));v31=1/2*(v31+v31');
        v3=block(v31,j(no2,no2,0));
        v4=block(rtau@I(nni),j(nni*no1,nni*no1,0),j(no2,no2,0));
    vit=v1+v2+v3+v4;
    vi=oi*vit*oi';
    df1=yy[a:b,]-mui;
    vi=1/2*(vi+vi');
    logvi=2*sum(log(vecdiag(root(vi)))); *avoid to overflow;
    a141=logvi+df1'*inv(vi)*df1;
    a14=a14+a141;
    end;
    LogLH1=-0.5*(n*log(2*3.14159265358979)+a14);
    epi=LogLH1-LogLH;
    rpsi=diag(J(1,nb,1)@psi);
    rtau=diag(J(1,nb,1)@tau);
    tau11=(betaR*betaR')@Tuu;
    tau12=betaR@Tus;
    tau13=betaR@Tu2;
    tau22=Tss;
    tau23=Ts2;
    tau33=T22;
    tau7=(tau11||tau12||tau13)//(tau12'||tau22||tau23)//(tau13'||tau23'||tau33);
    a41a2=(Tuu||Tus)//(Tus'||Tss);
    a41a1=Tu2//Ts2;
    a41a=T22-a41a1'*inv(a41a2)*a41a1; *print LogLH1 LogLH epi iter;
    if mod(iter,500)=0 then print LogLH1 LogLH epi iter
    beta0 betaR tau psi betaU betaS beta2 sigma T alpha Dhat;
    *print LogLH1 LogLH epi iter;
end;
```

```
print iter epi logLH LogLH1;
print beta0 betaR tau psi betaU betaS beta2;
print sigma;
print T;

Tuu=T[1:nd1,1:nd1];
Tus=T[1:nd1,(nd1+1):(nd1+no1)];
Tu2=T[1:nd1,(nd1+no1+1):(nd1+no1+no2)];
row=nd1+1;
col=nd1+no1;
Tss=T[row:col,row:col];
Ts2=T[row:col,(col+1):(col+no2)];
row=nd1+no1+1;
col=nd1+no1+no2;
T22=T[row:col,row:col];
row=2;
col=1+no1;
sigmass=sigma[row:col,row:col];
sigmauu=sigma[1,1];
sigmaus=sigma[1,2:(1+no1)];
alpha=sigmaus*inv(sigmass);
alpha1=alpha';
alpha2=inv(T22)*(Tu2[1,]'-Ts2'*alpha1);

*getting the intercept and save them in betas1, getting the slopes
of level-1 covariates and save them into betas2 and getting the
slopes of level-2 covariates and save them into betas3;
m1=no1; m2=(ncol(x1)-ncol(x2))*no1;
m3=ncol(x2)*no1; m4=ncol(x1)-ncol(x2); m5=ncol(x2);
betas1=j(m1,1,0); *intercept effect;
betas2=j(m2,1,0);
betas3=j(m3,1,0);
betau1=j(m4,1,0);
betau2=j(m5,1,0);
*remember to change the following program if we have different #
of completely observed level-1 and level-2 covariates;
betas2[1:m4]=betas[1:m4];  betas2[(m4+1):(2*m4)]=betas[(1+m4+m5):(2*m4+m5)];
betas2[(2*m4+1):(3*m4)]=betas[(1+2*m4+2*m5):(3*m4+2*m5)];
betas3[1:m5]=betas[(m4+1):(m4+m5)];
betas3[(m5+1):(2*m5)]=betas[(2*m4+m5+1):(2*m4+2*m5)];
betas3[(2*m5+1):(3*m5)]=betas[(3*m4+2*m5+1):(3*m4+3*m5)];
beta22=beta2;
betau1=betaU[1:m4];
betau2=betaU[(m4+1):(m4+m5)];
alpha3=betau1-alpha1'@I(m4)*betas2;
```

```
 alpha4=betau2 −(alpha1 '@I(m5))∗ betas3 −(alpha2 '@I(m5))∗ beta22 ;
 D=Tuu−block ( alpha2 '∗T22∗alpha2 −2∗alpha1 '∗Ts2∗alpha2 −alpha1 '∗Tss∗alpha1 ,0 ) ;
 alpha=alpha1 // alpha2 // alpha3 // alpha4 ;
 print beta0 betaR alpha1 alpha2 alpha3 alpha4 D tau psi alpha ;

∗ calculating the information matrix ;
  n=ncol ( vector ) ;   ∗how many subject in the dataset ;
  nb=4;              ∗# of biomarkers ;
  no1=3;             ∗# of level −1 covariates subject to missing values ;
  no2=2;             ∗# of level −2 covariates subject to missing values ;
  nu=ncol ( x1 ) ;
  ns=no1 ∗(nu ) ;
  n2=no2 ∗( ncol ( x2 ) ) ; ∗ there are two columns in x2i ;
  nd1=ncol (D1 ) ;
  nT=nd1+no1+no2 ;
  nv=no1 ;
  ∗# of parameters in the constraint model: this is for information matrix ;
  IFMD=4∗nb+nu+ns+n2+nT ∗(nT+1)/2+nv ∗(nv+1)/2 −(nd1 −1)∗( no1+no2 ) ;

∗ the function to create the derivative of betaR_j ;
 start DbetaR ( j , Di , Oi , betaR ,T, Sigma , nb , nd1 , no1 , no2 , ki , ss6 ) ;
   Tuu=T[ 1 : nd1 , 1 : nd1 ] ;
   Tus=T[ 1 : nd1 , ( nd1 +1):( nd1+no1 ) ] ;
   Tu2=T[ 1 : nd1 , ( nd1+no1 +1):( nd1+no1+no2 ) ] ;
   row=nd1 +1;
   col=nd1+no1 ;
   Tss=T[ row : col , row : col ] ;
   Ts2=T[ row : col , ( col +1):( col+no2 ) ] ;
   row=nd1+no1 +1;
   col=nd1+no1+no2 ;
   T22=T[ row : col , row : col ] ;
   row=2;
   col=1+no1 ;
   sigmass=sigma [ row : col , row : col ] ;
   sigmauu=sigma [ 1 , 1 ] ;
   sigmaus=sigma [ 1 ,2:(1+no1 ) ] ;
   deltaj=j ( nb , 1 ,0 ) ;
   deltaj [ j ]=1;
   ss1 =(( deltaj ∗betaR '+betaR ∗ deltaj ')@Tuu ) || ( deltaj@Tus ) || ( deltaj@Tu2 ) ;
   ss2 =( deltaj '@Tus ') || j ( no1 , no1 ,0 ) || j ( no1 , no2 ,0 ) ;
   ss3 =( deltaj '@Tu2 ') || j ( no2 , no1 ,0 ) || j ( no2 , no2 ,0 ) ;
   ss4 =(( deltaj ∗betaR '+betaR ∗ deltaj ')@( I ( ki )@sigmauu ) )
       || ( deltaj@sigmaus@I ( ki ) ) || j ( ki ∗nb , no2 ,0 ) ;
   ss5 =( deltaj '@sigmaus '@I ( ki ) ) || j ( ki ∗no1 , ki ∗no1 ,0 ) || j ( ki ∗no1 , no2 ,0 ) ;
   DVB=( oi ∗Di )∗( ss1 // ss2 // ss3 )∗( oi ∗Di ) '+oi ∗( ss4 // ss5 // ss6 )∗oi ';
```

131

```
    free ss1 ss2 ss3 ss4 ss5 deltaJ;
    return(DVB);
finish;

*the function to create the derivative of tau_j; ;
start Dtau(j,oi,ki,nb,no1,no2);
  deltaJ=j(nb,1,0);
  deltaJ[j]=1;
  ss1=block((deltaj*deltaj')@I(ki),j(ki*no1,ki*no1,0),j(no2,no2,0));
  DVT=oi*ss1*oi';
  return(DVT); free deltaJ ss1;
finish;

*the function to create the derivative of psi_j; ;
start Dpsi(j,oi,ki,nb,no1,no2,ki1);
  deltaj=j(nb,1,0);
  deltaj[j]=1;
  ss1=block((deltaj*deltaj')@(ki1*ki1'),j(ki*no1,ki*no1,0),j(no2,no2,0));
  DVP=oi*ss1*oi';
  return(DVP);   free ss1 deltaj;
finish;

*a subroutine for the derivative of a symmetric matrix about its distinct
 elements and aggregate the derivative matrix horizontally;
start MM(nn); *here the nn is the row dimension of the symmetric matrix;
  mat=j(nn,nn*nn*(nn+1)/2,0); k=0;
  do j=1 to nn;
     deltaj=j(nn,1,0);
     deltaj[j]=1;
     do i=j to nn;
            deltai=j(nn,1,0);
            deltai[i]=1;
            k=k+1;
            if i=j then mat1=deltaj*deltaj';
             else mat1=deltai*deltaj'+deltaj*deltai';
             cStart = (k-1)*nn + 1;
             cEnd = cStart + nn - 1;
             mat[,cStart:cEnd] = mat1;
          end;
  end;
  return(mat); free mat1;
finish;

*a function to create the derivative of the distinct elements in sigma_k
(columnwisly arranging the distinct elements);
```

132

```
start DSigma(k,no1,oi,ki,no2,DVS1,alpha1);
 mat2=MM(no1);
 cStart=(k−1)*no1 + 1;
 cEnd =cStart +no1 − 1;
 mat1=mat2[,cstart:cend];
 mat=((alpha1'*mat1*alpha1)||(alpha1'*mat1))//((mat1'*alpha1)||mat1);
 DVS=oi*DVS1*block(mat@I(ki),j(no2,no2,0))*DVS1'*oi';
 return(DVS);
finish;

*a subroutine for the derivative of a m*n nonsymmetric matrix and aggragate
 the derivative matriax column by column;
start NNN(m,n);
 mat=j(m,m*n*n,0);
 k=0;
 do i=1 to m;
    delta1=j(m,1,0);
    delta1[i]=1;
    do j=1 to n;
      k=k+1;
      delta2=j(n,1,0);
      delta2[j]=1;
      mat1=delta1*delta2';
      cstart=(k−1)*n+1;
      cend=cstart+n−1;
      mat[,cstart:cend]=mat1;
    end;
 end;
 return(mat);
finish;

*a subroutine to get the deverative about T;
*nn is the row dimension of the symmetric matrix;
start OO(kk,nb,nd1,no1,no2,alpha1,pi1,DVT1,ki);
  nn=nd1+no1+no2;
  mm=nn*(nn+1)/2−no1−no2;
  pp=nb*ki+no1*ki+no2;
  mat=j(pp,pp*mm,0);
  j=kk;
  if j<=nd1*(nd1+1)/2 then do;
     mat11=MM(nd1);
     cStart=(j−1)*nd1 + 1;
     cEnd =cStart +nd1 − 1;
     mat1=(mat11[,cstart:cend]||j(nd1,no1+no2,0))//j(no1+no2,nn,0);
   end;
```

133

```
        else if  j<=no1*(no1+1)/2+nd1*(nd1+1)/2 then  do;
          mat12=MM(no1);
          cStart=(j-nd1*(nd1+1)/2-1)*no1 + 1;
          cEnd =cStart +no1 - 1;
          mat2=mat12[,cstart:cend];
          dtusv=(alpha1'*mat2)//j(nd1-1,no1,0);
          mat1=(j(nd1,nd1,0)||dtusv||j(nd1,no2,0))//(dtusv'||mat2
                ||j(no1,no2,0))//j(no2,nn,0);
        end;
          else if  j<=no1*(no1+1)/2+nd1*(nd1+1)/2+no1*no2 then  do;
            mat11=NNN(no1,no2);
            cStart=(j-no1*(no1+1)/2-nd1*(nd1+1)/2-1)*no2 + 1;
            cEnd =cStart +no2 - 1;
            mat3=mat11[,cstart:cend];
            dtus=(pi1'*mat3')//j(nd1-1,no1,0);
            dtu2=(alpha1'*mat3)//j(nd1-1,no2,0);
            mat1=(j(nd1,nd1,0)||dtus||dtu2)//(dtus'||j(no1,no1,0)||mat3)
                  //(dtu2'||mat3'||j(no2,no2,0));
          end;
           else if  j<=no1*(no1+1)/2+nd1*(nd1+1)/2+no1*no2+no2*(no2+1)/2 then  do;
            mat41=MM(no2);
            cStart=(j-(no1*(no1+1)/2+nd1*(nd1+1)/2+no1*no2)-1)*no2 + 1;
            cEnd =cStart +no2 - 1;
            mat4=mat41[,cstart:cend];
            dtu22=(pi1'*mat4)//j(nd1-1,no2,0);
            mat1=(j(nd1,nd1+no1,0)||dtu22)//j(no1,nn,0)
              //(dtu22'||j(no2,no1,0)||mat4);
        end;
    term=DVT1*mat1*DVT1';
    cStart = (j-1)*pp + 1;
    cEnd = cStart + pp - 1;
    mat[,cStart:cEnd]=term;
    return(mat);
finish;

*a function to create the derivative of the distinct elements in T_k
 (columnwisly arranging the distinct elements);
start DT(k,nb,nd1,no1,no2,alpha1,pi1,DVT1,oi,ki);
  kk=k;
  matt=OO(kk,nb,nd1,no1,no2,alpha1,pi1,DVT1,ki);
  pp=nb*ki+no1*ki+no2;
  cStart=(k-1)*pp+1;
  cEnd =cStart +pp-1;
  mat=matt[,cStart:cEnd];
  DVT=oi*mat*oi';
```

134

```
    return (DVT);
finish;

*a function to create the derivative of alpha1;
start Dalpha(k,Ts2,Tss,sigmass,alpha1,nd1,no1,no2,DVT1,DVS1,oi,ki);
 deltak=j(no1,1,0);
 deltak[k]=1;
 a1=(deltak'*sigmass*alpha1+alpha1'*sigmass*deltak)@I(ki);
 a2=(deltak'*sigmass)@I(ki);
 a3=(deltak'*Tss)//j(nd1-1,no1,0);
 a4=(deltak'*Ts2)//j(nd1-1,no2,0);
 mat1=(a1||a2||j(ki,no2,0))//(a2'||j(no1*ki,no1*ki+no2,0))
       //j(no2,ki+no1*ki+no2,0);
 mat2=(j(nd1,nd1,0)||a3||a4)//(a3'||j(no1,no1+no2,0))
       //(a4'||j(no2,no1+no2,0));
 mat=DVS1*mat1*DVS1'+DVT1*mat2*DVT1';
 Dalph=oi*mat*oi';
 return (Dalph);
finish;

*a function to create the derivative of pi;
start Dpi(k,T22,Ts2,nd1,no1,no2,DVT1,oi);
  deltak=j(no2,1,0);
  deltak[k]=1;
  mat1=j(nd1+no1+no2,nd1+no1+no2,0);
  mat1[1:nd1,(nd1+1):(nd1+no1)]=(deltak'*Ts2')//j(nd1-1,no1,0);
  mat1[(nd1+1):(nd1+no1),1:nd1]=(Ts2*deltak)||j(no1,nd1-1,0);
  mat1[1:nd1,(nd1+no1+1):(nd1+no1+no2)]=(deltak'*T22)//j(nd1-1,no2,0);
  mat1[(nd1+no1+1):(nd1+no1+no2),1:nd1]=(T22*deltak)||j(no2,nd1-1,0);
  Dp=oi*DVT1*mat1*DVT1'*oi';
  return (Dp);
finish;

  rpsi=diag(J(1,nb,1)@psi);
  rtau=diag(J(1,nb,1)@tau);
  Tuu=T[1:nd1,1:nd1];
  Tus=T[1:nd1,(nd1+1):(nd1+no1)];
  Tu2=T[1:nd1,(nd1+no1+1):(nd1+no1+no2)];
  row=nd1+1;
  col=nd1+no1;
  Tss=T[row:col,row:col];
  Ts2=T[row:col,(col+1):(col+no2)];
  row=nd1+no1+1;
  col=nd1+no1+no2;
  T22=T[row:col,row:col];
```

135

```
sigmauu=sigma[1,1];
sigmaus=sigma[1,2:(1+no1)];
row=2;
col=1+no1;
sigmass=sigma[row:col,row:col];
alpha1=alpha[1:no1];
alpha2=alpha[(no1+1):(no1+no2)];
alpha3=alpha[(no1+no2+1):(no1+no2+ncol(x1)-ncol(x2))];
alpha4=alpha[(no1+no2+ncol(x1)-ncol(x2)+1):nrow(alpha)];
sigma=(sigmauu||sigmaus)//(sigmaus'||sigmass);
T=(Tuu||Tus||Tu2)//(Tus'||Tss||Ts2)//(Tu2'||Ts2'||T22);
tau11=(betaR*betaR')@Tuu;
tau12=betaR@Tus;
tau13=betaR@Tu2;
tau21=tau12';
tau22=Tss;
tau23=Ts2;
tau31=tau13';
tau32=tau23';
tau33=T22;
tau7=(tau11||tau12||tau13)//(tau21||tau22||tau23)//(tau31||tau32||tau33);
IFM=J(IFMD,IFMD,0);
A1=j(nb,nb,0);
IMD=2*nb+nT*(nT+1)/2+nv*(nv+1)/2-(nd1-1)*(no1+no2);
A12=j(nb,IMD+nb,0);
A22=j(IMD,IMD,0);    pi1=alpha2;

do i=1 to n;
 a=0; do k=1 to i-1;a=cont[k]+a;end;a=a+1;
 b=0; do k=1 to i; b=cont[k]+b;end;  *a and b are to restrict YY[i];
 aa=0; do k=1 to i-1;aa=cont2[k]+aa;end;aa=aa+1;
 bb=0; do k=1 to i; bb=cont2[k]+bb;end; *aa and bb are to restrict XX[i];
 aaa=0; do k=1 to i-1;aaa=cont1[k]+aaa;end;aaa=aaa+1;
 bbb=0; do k=1 to i; bbb=cont1[k]+bbb;end;
 *read the map matrix from completely data to observed data;
  ni=cont1[i];
  ki=cont2[i];
  ki1=j(ki,1,1);
  DVS1=block(betaR@I(ki),I(ki*no1),j(no2,no2,0));
  zi=block(I(nb)@D1[aa:bb,],I(no1)@ki1,I(no2));
  ss6=j(no2,ki*nb+ki*no1+no2,0);
  DVT1=block(betaR@D1[aa:bb,],I(no1)@ki1,I(no2));
  H1a=(I(no1)@x1[aa:bb,]);
  H2a=(I(no2)@X2[i,]);
  H1=block((I(nb)@ki1)||(I(nb)@(x1[aa:bb,]*betaU)),H1a,H2a);
```

```
    F1=block (( I (nb)@ki1)||( betaR@x1[aa:bb,]) ,H1a ,H2a );

    v1=zi*tau7*zi ';
    v2=block(rpsi@(ki1*ki1 '), j(ki*no1,ki*no1,0), j(no2,no2,0));
    v31 =(((betaR*betaR ')@sigmauu@I(ki ))||( betaR@sigmaus@I(ki )))
        //(( betaR@sigmaus@I(ki )) '||( sigmass@I(ki )));
    v3=block(v31 , j(no2,no2,0));
    v4=block(rtau@I(ki ), j(ki*no1,ki*no1,0), j(no2,no2,0));
    vit=v1+v2+v3+v4;
    free   v1 v2 v31 v3 v4;

ob1=o1i2i[aaa:(aaa+nni −1),  1:nni ];
 ob2=o1i2i[(aaa+nni ):(aaa+2*nni −1),  1:nni ];
 ob3=o1i2i[(aaa+2*nni ):(aaa+3*nni −1),  1:nni ];
 ob4=o1i2i[(aaa+3*nni ):(aaa+4*nni −1),  1:nni ];
 o1ij=block(ob1 ,ob2 ,ob3 ,ob4 );
 o11i= delrow(o1ij );

 os1=o1i2i[(aaa+4*nni ):(aaa+5*nni −1),  1:nni ];
 os2=o1i2i[(aaa+5*nni ):(aaa+6*nni −1),  1:nni ];
 os3=o1i2i[(aaa+6*nni ):(aaa+7*nni −1),  1:nni ];
 o2ij=block(os1 ,os2 ,os3 );
 o12i= delrow(o2ij );

 o2ai=o1i2i[(aaa+7*nni ):(aaa+7*nni+no2 −1),1:no2 ];
 c=cmiss(o2ai );
 count=c[,+];
 mIdx=loc(count >0);
 aaaa=no2−ncol(mIdx );
 if  aaaa>0 then   do;
  NMIdx = setdif( 1:nrow(o2ai), mIdx );
  o22i = o2ai[NMIdx,];
 end;
    else  if aaaa=0 then  o22i=0;
ki=nni;
ki1=j(nni ,1 ,1);
if o11i=0 then n11=1000; else n11=ncol(o11i );
if o12i=0 then n12=1000; else n12=ncol(o12i );
if o22i=0 then n22=1000; else n22=ncol(o22i );
    if  n11=1000 & n12=1000 & n22=1000 then  oi ="";
      else  if n11^=1000 & n12=1000 & n22=1000 then
       oi=o11i||j(nrow(o11i ),no1*nni+no2 ,0);
        else  if n11=1000 & n12^=1000 & n22=1000 then
         oi=j(nrow(o12i ),nb*nni ,0)||o12i||j(nrow(o12i ),no2 ,0);
          else  if n11=1000 & n12=1000 & n22^=1000 then
```

```
                    oi=j(nrow(o22i),nb*nni,0)||j(nrow(o22i),no1*nni,0)||o22i;
                     else if n11^=1000 & n12^=1000 & n22=1000 then
                      oi=(o11i||j(nrow(o11i),no1*nni+no2,0))
                         //(j(nrow(o12i),nb*nni,0)||o12i||j(nrow(o12i),no2,0));
                        else if n11^=1000 & n12=1000 & n22^=1000 then
                         oi=(o11i||j(nrow(o11i),no1*nni+no2,0))
                            //(j(nrow(o22i),nb*nni+no1*nni,0)||o22i);
                          else if n11=1000 & n12^=1000 & n22^=1000 then
                           oi=(j(nrow(o12i),nb*nni,0)||o12i||j(nrow(o12i),no2,0))
                              //(j(nrow(o22i),nb*nni+no1*nni,0)||o22i);
                            else if n11^=1000 & n12^=1000 & n22^=1000 then
                             oi=block(o11i,o12i,o22i);
vi=oi*vit*oi';
do j=1 to nb;
 PVJ=DbetaR(j,zi,oi,betaR,T,Sigma,nb,nd1,no1,no2,ki,ss6);
 do k=1 to IMD+nb;
  if k<=nb then PVK=DbetaR(k,zi,oi,betaR,T,Sigma,nb,nd1,no1,no2,ki,ss6);
   else if k<=2*nb then PVK=Dtau(k-nb,oi,ki,nb,no1,no2);
    else if k<=2*nb+nt*(nt+1)/2-nd1*(no1+no2) then
   PVK=DT(k-2*nb,nb,nd1,no1,no2,alpha1,pi1,DVT1,oi,ki);
    else if k<=3*nb+nt*(nt+1)/2-nd1*(no1+no2) then
    PVK=Dpsi(k-2*nb-nt*(nt+1)/2+nd1*(no1+no2),oi,ki,nb,no1,no2,ki1);
     else if k<=3*nb+nt*(nt+1)/2-nd1*(no1+no2)+no1*(no1+1)/2 then
      PVK=DSigma(k-(3*nb+nt*(nt+1)/2-nd1*(no1+no2)),no1,oi,ki,no2,DVS1,alpha1);
       else if k<=3*nb+nt*(nt+1)/2-nd1*(no1+no2)+no1*(no1+1)/2+no1 then
        PVK=Dalpha(k-(3*nb+nt*(nt+1)/2-nd1*(no2+no1)+no1*(no1+1)/2),Ts2,Tss,
             sigmass,alpha1,nd1,no1,no2,DVT1,DVS1,oi,ki);
         else if k<=3*nb+nt*(nt+1)/2-(nd1-1)*(no2+no1)+no1*(no1+1)/2 then
          PVK=Dpi(k-(3*nb+nt*(nt+1)/2-nd1*(no2+no1)+no1*(no1+1)/2+no1),T22,
               Ts2,nd1,no1,no2,DVT1,oi);
     A12[j,k]=0.5*trace(solve(vi,PVJ)*solve(vi,PVK));
    end;
  end;
do j=1 to IMD;
 if j<=nb then PVJ=Dtau(j,oi,ki,nb,no1,no2);
  else if j<=nb+nt*(nt+1)/2-nd1*(no1+no2) then
   PVJ=DT(j-nb,nb,nd1,no1,no2,alpha1,pi1,DVT1,oi,ki);
    else if j<=2*nb+nt*(nt+1)/2-nd1*(no1+no2) then
    PVJ=Dpsi(j-nb-nt*(nt+1)/2+nd1*(no1+no2),oi,ki,nb,no1,no2,ki1);
     else if j<=2*nb+nt*(nt+1)/2-nd1*(no1+no2)+no1*(no1+1)/2 then
    PVJ=DSigma(j-(2*nb+nt*(nt+1)/2-nd1*(no1+no2)),no1,oi,ki,no2,DVS1,alpha1);
      else if j<=2*nb+nt*(nt+1)/2-nd1*(no1+no2)+no1*(no1+1)/2+no1 then
       PVJ=Dalpha(j-(2*nb+nt*(nt+1)/2-nd1*(no1+no2)+no1*(no1+1)/2),Ts2,Tss,
            sigmass,alpha1,nd1,no1,no2,DVT1,DVS1,oi,ki);
        else if j<=2*nb+nt*(nt+1)/2-(nd1-1)*(no1+no2)+no1*(no1+1)/2 then
```

```
                PVJ=Dpi(j −(2∗nb+nt ∗(nt+1)/2−nd1∗(no1+no2)+no1∗(no1+1)/2+no1 ),T22,Ts2 ,
                      nd1 ,no1 ,no2 ,DVT1, oi );
        do  k=1  to  IMD;
          if  k<=nb  then  PVK=Dtau(k , oi , ki , nb , no1 , no2 );
            else  if  k<=nb+nt ∗(nt+1)/2−nd1∗(no1+no2)  then
              PVK=DT(k−nb , nb , nd1 , no1 , no2 , alpha1 , pi1 ,DVT1, oi , ki );
                else  if  k<=2∗nb+nt ∗(nt+1)/2−nd1∗(no1+no2)  then
                  PVK=Dpsi(k−nb−nt ∗(nt+1)/2+nd1∗(no1+no2), oi , ki , nb , no1 , no2 , ki1 );
                    else  if  k<=2∗nb+nt ∗(nt+1)/2−nd1∗(no1+no2)+no1∗(no1+1)/2  then
                      PVK=DSigma(k−(2∗nb+nt ∗(nt+1)/2−nd1∗(no1+no2)), no1 , oi , ki , no2 ,DVS1,
                          alpha1 );
                        else  if  k<=2∗nb+nt ∗(nt+1)/2−nd1∗(no1+no2)+no1∗(no1+1)/2+no1  then
                          PVK=Dalpha(k−(2∗nb+nt ∗(nt+1)/2−nd1∗(no1+no2)+no1∗(no1+1)/2),Ts2 ,
                          Tss , sigmass , alpha1 , nd1 , no1 , no2 ,DVT1,DVS1, oi , ki );
                            else  if  k<=2∗nb+nt ∗(nt+1)/2−(nd1−1)∗(no1+no2)+no1∗(no1+1)/2  then
                              PVK=Dpi(k−(2∗nb+nt ∗(nt+1)/2−nd1∗(no1+no2)+no1∗(no1+1)/2+no1 ),T22 ,
                              Ts2 , nd1 , no1 , no2 ,DVT1, oi );
            A22[j , k]=0.5∗ trace ( solve ( vi ,PVJ)∗ solve ( vi ,PVK));
          end ;
      end ;
    H=oi ∗H1;
    F=oi ∗F1;
    G=H∗(I (nb )// j (nb+ns+n2 , nb , 0 ));
    M=H∗( j (nb , nb , 0 )// I (nb )// j (ns+n2 , nb , 0 ));
    Q=F∗( j (nb , nu+ns+n2 , 0 )// I (nu+ns+n2 ));  free H F;
    IM1=(G‘∗ inv ( vi )∗G)||(G‘∗ inv ( vi )∗M)||(G‘∗ inv ( vi )∗Q);
    IM2=(M‘∗ inv ( vi )∗G)||(A12[ ,1:nb]+M‘∗ inv ( vi )∗M)||(M‘∗ inv ( vi )∗Q);
    IM3=(Q‘∗ inv ( vi )∗G)||(Q‘∗ inv ( vi )∗M)||(Q‘∗ inv ( vi )∗Q);
    IFM11=IM1// IM2// IM3;   ∗First  block ;
    IFM12=j (nb , 2∗nb+nt ∗(nt+1)/2+nv ∗(nv+1)/2−no1−no2 , 0 )// A12[ , (nb+1):(nb+IMD)]
            // j (nu+ns+n2 , 2∗nb+nt ∗(nt+1)/2+nv ∗(nv+1)/2−(no1+no2 ) , 0 );
    IFM22=A22;
    C=(IFM11 ||IFM12 )//( IFM12 ‘||IFM22 );
  IFM=IFM+C;
end ;

∗getting  the  variance  for  the  parameters ;
print  IFM;
InIFM=inv (IFM);  print  InIFM;
test=diag (InIFM);  test1=diag (InIFM)∗ j (nrow(InIFM) , 1 , 1 );  print  test1 ;
∗do  j=1  to  IFMD;
∗   if  test [ j , j ]<0 then  test [ j , j ]=0;
∗end ;
∗the  standard  deviation  for  beta0 , betaR ,  beta ∗, tau ,V(T), psi  and  V( sigma );
Std=sqrt ( test )∗ j (IFMD, 1 , 1 );
```

```
stdbeta0=std [1:nb ];  stdbetaR=std [( nb+1):(2*nb ) ];
stdbetaS=std [(2*nb+1):(2*nb+nu+ns+n2 ) ];
stdtau=std [(2*nb+nu+ns+n2+1):(2*nb+nu+ns+n2+nb ) ];
stdT=std [(3*nb+nu+ns+n2+1):(3*nb+nu+ns+n2+nt *( nt+1)/2 − nd1 *( no1+no2 ) ) ];
stdpsi=std [(3*nb+nu+ns+n2+nt *( nt+1)/2 − nd1 *( no1+no2 )+1)
          :(4*nb+nu+ns+n2+nt *( nt+1)/2 − nd1 *( no1+no2 ) ) ];
stdsig=std [(4*nb+nu+ns+n2+nt *( nt+1)/2 − nd1 *( no1+no2 )+1)
          :(4*nb+nu+ns+n2+nt *( nt+1)/2 − nd1 *( no1+no2 )+no1 *( no1+1)/2 ) ];
stdalpha1=std [(4*nb+nu+ns+n2+nt *( nt+1)/2 − nd1 *( no1+no2 )+no1 *( no1+1)/2+1)
            :(4*nb+nu+ns+n2+nt *( nt+1)/2 − nd1 *( no1+no2 )+no1 *( no1+1)/2+no1 ) ];
stdalpha2=std [(4*nb+nu+ns+n2+nt *( nt+1)/2 − nd1 *( no1+no2 )+no1 *( no1+1)/2+no1+1)
          :(4*nb+nu+ns+n2+nt *( nt+1)/2 − nd1 *( no1+no2 )+no1 *( no1+1)/2+no1+no2 ) ];
CIL1=betaR −1.96* stdbetaR ; CIU1=betaR +1.96* stdbetaR ;
stdU=stdbetaS [1: nu ]; stdS=stdbetaS [( nu+1):( nu+ns ) ];
std2=stdbetaS [( nu+ns+1):( nu+ns+n2 ) ];
print beta0 stdbeta0 betaR stdbetaR tau stdtau psi stdpsi
 betaU stdU betaS stdS beta2 std2 sigma stdsig T stdT ;

*getting the variance for the parameters in the conditional model ;
m1=no1 ; m2=( ncol ( x1)−ncol ( x2 ))* no1 ;
m3=ncol ( x2 )* no1 ; m4=ncol ( x1)−ncol ( x2 ); m5=ncol ( x2 );
betau1=betau [1: m4 ];        betau2=betau [( m4+1):( m4+m5 ) ];
betas1=j ( no1 ,1 ,0 );
betas2a=j ( m2,1 ,0 );
betas3a=j ( m3,1 ,0 );
betas2=j ( m4, no1 ,0 );
betas3=j ( m5, no1 ,0 );
beta21=j ( no2 ,1 ,0 );
beta22a=j ( m5* no2 ,1 ,0 );
beta22=j ( m5, no2 ,0 );        print no1 no2 nu ns n2 betas ;
do i =1 to no1 ;
  betas2a [(1+( i −1)* m4 ):( i *m4 )]= betas [(1+( i −1)* m4+( i −1)* m5)
        :(( i −1)* m4+( i −1)* m5+m4 ) ];
  betas3a [(1+( i −1)* m5 ):( i *m5 )]= betas [( i *m4+( i −1)* m5+1):( i *m4+i *m5 ) ];
  betas2 [ , i ]= betas [(1+( i −1)* m4+( i −1)* m5 ):(( i −1)* m4+( i −1)* m5+m4 ) ];
  betas3 [ , i ]= betas [(( i −1)* m4+( i −1)* m5+m4+1):( i *m4+i *m5 ) ];
end ;

do i =1 to no2 ;
  beta22 [ , i ]= beta2 [(( i −1)* m5+1):( i *m5 ) ];
end ;

beta22a=beta2 ;
print beta0 betaR alpha1 alpha2 alpha3 alpha4 D tau psi ;
ftheta=I (m4)||( − alpha1 '@I(m4))||( − betas2 );
```

140

```
cov11=InIFM[(2*nb+1):(2*nb+m4),(2*nb+1):(2*nb+m4)];
cov12=InIFM[(2*nb+1):(2*nb+m4),(2*nb+nu+1):(2*nb+nu+no1*m4)];
cov13=InIFM[(2*nb+1):(2*nb+m4),(IFMD−no1−no2+1):(IFMD−no2)];
cov22=InIFM[(2*nb+nu+1):(2*nb+nu+no1*m4),(2*nb+nu+1):(2*nb+nu+no1*m4)];
cov23=InIFM[(2*nb+nu+1):(2*nb+nu+no1*m4),(IFMD−no1−no2+1):(IFMD−no2)];
cov33=InIFM[(IFMD−no1−no2+1):(IFMD−no2),(IFMD−no1−no2+1):(IFMD−no2)];
cov1=(cov11||cov12||cov13)//(cov12'||cov22||cov23)//(cov13'||cov23'||cov33);
covalpha3=ftheta*cov1*ftheta';
stdalpha3=sqrt(diag(covalpha3))*j(m4,1,1);

*alpha2=pi1;
ftheta1=I(m5)||(−alpha1'@I(m5)*I(m3))||(−alpha2'@I(m5)*I(m5*no2))
         ||(−betas3)||(−beta22);
covv11=InIFM[(2*nb+m4+1):(2*nb+m4+m5),(2*nb+m4+1):(2*nb+m4+m5)];
covv12=InIFM[(2*nb+m4+1):(2*nb+m4+m5),(2*nb+nu+no1*m4+1)
   :(2*nb+nu+no1*m4+no1*m5)];
covv13=InIFM[(2*nb+m4+1):(2*nb+m4+m5),(2*nb+nu+ns+1):(2*nb+nu+ns+no2*m5)];
covv14=InIFM[(2*nb+m4+1):(2*nb+m4+m5),(IFMD−no1−no2+1):(IFMD−no2)];
covv15=InIFM[(2*nb+m4+1):(2*nb+m4+m5),(IFMD−no2+1):IFMD];

covv22=InIFM[(2*nb+nu+no1*m4+1):(2*nb+nu+no1*m4+no1*m5),
(2*nb+nu+no1*m4+1):(2*nb+nu+no1*m4+no1*m5)];
covv23=InIFM[(2*nb+nu+no1*m4+1):(2*nb+nu+no1*m4+no1*m5),
(2*nb+nu+ns+1):(2*nb+nu+ns+no2*m5)];
covv24=InIFM[(2*nb+nu+no1*m4+1):(2*nb+nu+no1*m4+no1*m5),
(IFMD−no1−no2+1):(IFMD−no2)];
covv25=InIFM[(2*nb+nu+no1*m4+1):(2*nb+nu+no1*m4+no1*m5),(IFMD−no2+1):IFMD];

covv33=InIFM[(2*nb+nu+ns+1):(2*nb+nu+ns+no2*m5),(2*nb+nu+ns+1)
  :(2*nb+nu+ns+no2*m5)];
covv34=InIFM[(2*nb+nu+ns+1):(2*nb+nu+ns+no2*m5),(IFMD−no1−no2+1):(IFMD−no2)];
covv35=InIFM[(2*nb+nu+ns+1):(2*nb+nu+ns+no2*m5),(IFMD−no2+1):IFMD];
covv44=InIFM[(IFMD−no1−no2+1):(IFMD−no2),(IFMD−no1−no2+1):(IFMD−no2)];
covv45=InIFM[(IFMD−no1−no2+1):(IFMD−no2),(IFMD−no2+1):IFMD];
covv55=InIFM[(IFMD−no2+1):IFMD,(IFMD−no2+1):IFMD];
covv=(covv11||covv12||covv13||covv14||covv15)//
     (covv12'||covv22||covv23||covv24||covv25)//
     (covv13'||covv23'||covv33||covv34||covv35)//
     (covv14'||covv24'||covv34'||covv44||covv45)//
     (covv15'||covv25'||covv35'||covv45'||covv55);
covalpha4=ftheta1*covv*ftheta1';
stdalpha4=sqrt(diag(covalpha4))*j(m5,1,1);

*alpha3=betau1−(alpha1'@I(m4))*betas2a;
*alpha4=betau2−(alpha1'@I(m5))*betas3a−(pi1'@I(m5))*beta22a;
```

```
*getting the variance for D;
D1=j(no1*(no1+1)/2,1,.);
D2=j(no1*no2,1,.);
D3=j(no2*(no2+1)/2,1,.);
D4=j(no1,1,.);
D5=j(no2,1,.);
mat2=MM(no1);
do k=1 to no1*(no1+1)/2;
   cStart=(k-1)*no1 + 1;
   cEnd =cStart +no1 − 1;
   mat1=mat2[,cstart:cend];
   D1[k]=−alpha1'*mat1*alpha1;
end;
mat4=NNN(no1,no2);
do k=1 to no1*no2;
         cStart=(k-1)*no2 + 1;
         cEnd =cStart +no2 − 1;
         mat3=mat4[,cstart:cend];
         D2[k]=−2*alpha1'*mat3*alpha2;
end;
mat6=MM(no2);
do k=1 to no2*(no2+1)/2;
   cStart=(k-1)*no2 + 1;
   cEnd =cStart +no2 − 1;
   mat5=mat6[,cstart:cend];
   D3[k]=−alpha2'*mat5*alpha2;
end;
D4=−2*alpha2'*Ts2'−2*alpha1'*Tss;
D5=−2*alpha2'*T22−2*alpha1'*Ts2;
dim=no1*(no1+1)/2+no1*no2+no2*(no2+1)/2+no1+no2;
fthet=(D1'||D2'||D3'||D4||D5)//j(nd1*(nd1+1)/2−1,dim,0);
ftheta=I(nd1*(nd1+1)/2)||fthet;
a=(3*nb+nu+no1*m4+no1*m5+no2*m5+1);
b=(3*nb+nu+no1*m4+no1*m5+no2*m5+nd1*(nd1+1)/2);
covv11=InIFM[a:b,a:b]; *Tuu;
covv12=InIFM[a:b,(b+1):(b+no1*(no1+1)/2)];
covv13=InIFM[a:b,(b+no1*(no1+1)/2+1):(b+no1*(no1+1)/2+no1*no2)];
covv14=InIFM[a:b,((b+no1*(no1+1)/2+no1*no2)+1):
  ((b+no1*(no1+1)/2+no1*no2)+no2*(no2+1)/2)];
covv15=InIFM[a:b,(IFMD−no1−no2+1):(IFMD−no2)];
covv16=InIFM[a:b,(IFMD−no2+1):IFMD];

covv22=InIFM[(b+1):(b+no1*(no1+1)/2),(b+1):(b+no1*(no1+1)/2)];
covv23=InIFM[(b+1):(b+no1*(no1+1)/2),(b+no1*(no1+1)/2+1)
```

```
       :(b+no1*(no1+1)/2+no1*no2)];
covv24=InIFM[(b+1):(b+no1*(no1+1)/2),((b+no1*(no1+1)/2+no1*no2)+1)
   :((b+no1*(no1+1)/2+no1*no2)+no2*(no2+1)/2)];
covv25=InIFM[(b+1):(b+no1*(no1+1)/2),(IFMD−no1−no2+1):(IFMD−no2)];
covv26=InIFM[(b+1):(b+no1*(no1+1)/2),(IFMD−no2+1):IFMD];

covv33=InIFM[(b+no1*(no1+1)/2+1):(b+no1*(no1+1)/2+no1*no2),
   (b+no1*(no1+1)/2+1):(b+no1*(no1+1)/2+no1*no2)];
covv34=InIFM[(b+no1*(no1+1)/2+1):(b+no1*(no1+1)/2+no1*no2),
   ((b+no1*(no1+1)/2+no1*no2)+1):((b+no1*(no1+1)/2+no1*no2)+no2*(no2+1)/2)];
covv35=InIFM[(b+no1*(no1+1)/2+1):(b+no1*(no1+1)/2+no1*no2),
   (IFMD−no1−no2+1):(IFMD−no2)];
covv36=InIFM[(b+no1*(no1+1)/2+1):(b+no1*(no1+1)/2+no1*no2),(IFMD−no2+1):IFMD];

covv44=InIFM[((b+no1*(no1+1)/2+no1*no2)+1):((b+no1*(no1+1)/2+no1*no2)
   +no2*(no2+1)/2),((b+no1*(no1+1)/2+no1*no2)+1):((b+no1*(no1+1)/2+no1*no2)
   +no2*(no2+1)/2)];
covv45=InIFM[((b+no1*(no1+1)/2+no1*no2)+1):((b+no1*(no1+1)/2+no1*no2)
   +no2*(no2+1)/2),(IFMD−no1−no2+1):(IFMD−no2)];
covv46=InIFM[((b+no1*(no1+1)/2+no1*no2)+1):((b+no1*(no1+1)/2+no1*no2)
   +no2*(no2+1)/2),(IFMD−no2+1):IFMD];

covv55=InIFM[(IFMD−no1−no2+1):(IFMD−no2),(IFMD−no1−no2+1):(IFMD−no2)];
covv56=InIFM[(IFMD−no1−no2+1):(IFMD−no2),(IFMD−no2+1):IFMD];

covv66=InIFM[(IFMD−no2+1):IFMD,(IFMD−no2+1):IFMD];

covv=(covv11||covv12||covv13||covv14||covv15||covv16)//
       (covv12`||covv22||covv23||covv24||covv25||covv26)//
       (covv13`||covv23`||covv33||covv34||covv35||covv36)//
       (covv14`||covv24`||covv34`||covv44||covv45||covv46)//
       (covv15`||covv25`||covv35`||covv45`||covv55||covv56)//
       (covv16`||covv26`||covv36`||covv46`||covv56`||covv66);
covD=ftheta*covv*ftheta`;
stdD=sqrt(diag(covD))*j(nd1*(nd1+1)/2,1,1);
print alpha1 stdalpha1 alpha2 stdalpha2 alpha3 stdalpha3 alpha4
stdalpha4 D stdD;
```

VITA

Chunfeng Ren was born on September 20, 1972 in Dengzhou, Henan Province, Peoples Republic of China. She received her Bachelor of Science in Mathematics from Zhengzhou University, Henan, China in 1994, a Master of Science in Computational Mathematics from Xi'an Jiaotong University, Shaanxi, China in 1997, and a Master of Public Health in Biostatistics from Georgia Southern University, Statesboro, Georgia in 2009. In August of 2009, she pursued her Ph.D. in Biostatistics at Virginia Commonwealth University in Richmond, Virginia.