Some Statistical Methods for Causal Mediation Pathway Analysis

by

Wei Hao

A dissertation submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
(Biostatistics)
in The University of Michigan
2021

Doctoral Committee:

Professor Peter X.K. Song, Chair
Professor Bhramar Mukherjee
Professor Karen E. Peterson
Associate Professor Xiang Zhou

Wei Hao

weihao@umich.edu

ORCID iD: 0000-0002-5367-1479

ACKNOWLEDGEMENTS

During my graduate studies, I would like to express my sincere thanks to faculty members, staff and my peers in Biostat Department; my friends and my family.

I especially want to thank my thesis advisor, Dr. Peter Song for his guidance, encouragement, patience, and support during these past six years. Peter introduced me to the exciting research area of mediation analysis. With his strong support and brilliant intuitions, I am able to make some contributions to this field.

I would also like to thank all my current collaborators, especially Dr Karen Peterson, PI of the ELEMENT study, for her encouragement and knowledge in the nutritional field. Peter and Karen introduced me to the ELEMENT team, and I have gained so much knowledge in the nutritional science field via different lines of research by Dr. Erica Jansen, Dr. Wei Perng, Dr. Deborah Watkins, and Dr. Jaclyn Goodrich from my data manager position. Thanks go to former ELEMENT member Dr Jennifer Labarre for the critical input for my data analysis in Chapter II. Thanks go to the study manager Laura Arboleda Merino for sharing her invaluable knowledge in the ELEMENT study.

I would like to thank my former collaboraters: Dr. Alan Baptist and Dr. Minal Patel for their expertise in Asthma, and co-authoring several manuscripts. I also would like to thank Dr. David Garabrant, Dr. Olivier Jolliet, and Dr. Alfred Franzblau in the environmental health science department, who encouraged me to pursue a PhD when I first started my research analyst position back in 2009. They demonstrated great examples as scientific researchers, and their support for present-

ing in the conferences and wring manuscripts haven benefited me substantially later during my graduate studies. My committee members Dr Bhramar Mukherjee, Dr Xiang Zhou and Dr Karen Peterson had provided me constructive input during my proposal, and I thank them for their insights in the mediation analyses and real data problems.

Next I would like to thank my friends/peers from Peter's lab, who are Yiwang Zhou, Margaret Banker, Mengtong Hu, Leyao Zhang, Lili Wang, Lan Luo and Lu Tang. They provided valuable feedback for my job presentation and defense rehearsals.

Lastly, great gratitude goes to my husband Jian, my father Jinglin Hao, and my two kids Violet and Vincent for their love and encouragement. Thanks to Jian for his companion during the many nights I had to work late, for his comfort in my many ups and downs during this journey of pursuing a PhD. The covid years of 2020-2021 have been very difficult where we all share the same space daily, I thank my kids for their understanding that they cannot be too loud during the weekdays at home. I thank my father Jinglin who always puts me in the first place during his retired life, and his strong support by travelling to US and helping me with two kids where they were little. I thank my late mother Fenglan Lu for her forever love and support!

TABLE OF CONTENTS

LIST OF FIGURES

Figure

LIST OF TABLES

Table

LIST OF APPENDICES

Appendix

ABSTRACT

Mediation analysis has been undertaken pervasively in practice. The primary goal of this analysis is to study whether the effect of an exposure on an outcome of interest is mediated by some intermediate factors such as epigenetic variants and metabolomic biomarkers. In this dissertation I develop new statistical methods to address some of statistical and scientific challenges arising from causal mediation pathway analyses.

Chapter II develops a simultaneous likelihood ratio (LR) test in the presence of multiple mediators. Statistical inference on the joint mediation effect is challenging due to the involvement of composite null hypotheses with a large number of parameter configurations. With an application of the Lagrange Multiplier approach, simultaneous LR test utilizes a block coordinate descent algorithm to solve the constrained likelihood under the irregular null parameter space. I establish the asymptotic null distribution and examine the finite-sample performance of the proposed joint test statistic via extensive simulations with comparisons to existing tests. The simulation results show that the joint testing method controls type I error properly and in general provides better power than existing tests. I apply this new method to investigate whether a group of glucose metabolites and acetylamino acids mediate the effect of nutrient intakes on insulin resistance.

Chapter III presents a unified framework of generalized structural equation models (GSEMs) for mediation analyses with data of mixed types to address practical needs in the analysis of biomedical data. This new class of models accommodates continuous, categorical, count variables. Using the Fréchet's construction of multivariate

distributions, I formulate GSEM as a hierarchical model consisting of (i) a Gaussian copula dependence model to characterize a directed acyclic graph (DAG) relationship among outcome variable, mediator and exposure variable, and (ii) generalized linear models (GLMs) to adjust confounding factors in marginal distributions. This new framework provides valid joint probability distributions and well-defined mediation effects for interpretation. I develop a pseudo-maximum likelihood estimation for various scenarios of mixed data types. I illustrate this new methodology via a dataset collected from a cohort study in environmental health sciences, where I study whether the tempo of reaching infancy BMI peak, an important early life growth milestone that may be measured as either a continuous variable or a binary variable (delay or not), may mediate the association between prenatal exposure to phthalates and pubertal health outcomes.

Chapter IV concerns a conceptual framework of generalized direct and indirect effects to relax the current definitions of causal mediation effects in the presence of categorical intervention or categorical exposure. I utilize the latent variable presentation to describe the role of a categorical "action" in a causal study. Specially, I focus on two important types of models, namely the effective dose model and the latent exposure model. I demonstrate that the proposed generalized direct and indirect effects are more desirable to quantify and interpret direct and indirect effects than the conventional approach. I develop maximum likelihood estimation for the model parameters, and examine numerically the performance of the estimation via simulation studies. Also, I illustrate this new causal mediation paradigm via a randomized trial from the ELEMENT study where I investigate whether the association of mother's calcium supplementation on offspring birth weight is mediated by mother's blood lead level measured during the third trimester.

CHAPTER I

Introduction

1.1 Mediation Analysis

In many disciplines such as Psychology, Sociology, Epidemiology, Environmental Health Sciences, Political Science, researchers aim to understand certain causal mediation mechanisms through which exposure variables affect outcome variables of interest. Mediators are typically referred to as some intermediate variables that lie in causal pathways between exposure and outcomes. For instance, obesity or overweight is deemed as the primary cause for insulin resistance, a health condition that human body cannot operate insulin secretion properly and consequently, fails to effectively move glucose from bloodstreams to cells. In a cohort study "Early Life Exposures in Mexico to ENvironmental Toxicants" (ELEMENT), investigators hypothesized that an individual's metabolomics may mediate the relationship between nutrient intakes and insulin resistance (*LaBarre et al.*, 2020). Figure 1.1 illustrates a hypothesized causal mediation mechanism involves exposure (i.e.,nutrient intakes), mediator (i.e., metabolome) and outcome (i.e., insulin resistance), in addition to some confounding factors that are not shown for the sake of simplicity. This figure is also known as directed acyclic graph (DAG), which presents a scientific hypothesis based on certain directed relationships via directed edges. In a causal mediation framework, the exposure variable (e.g., nutrient intake) is often referred to a variable that may vary

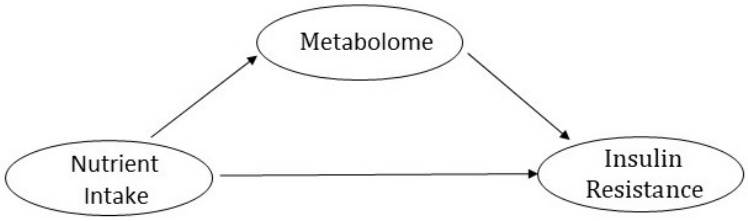

Figure 1.1: Hypothesized causal associations among food intakes, metabolome and insulin resistance.

under different conditions, such as nutrient intakes that measure the consumption of food (fats and carbohydrates), whose amount varies for each participant. Other examples of exposure include concentration levels of a toxic agent in environmental health sciences and a drug treatment in a randomized control trial.

The ELEMENT cohort study is a multi-institutional, internationally teamed project which was initiated in early 1990s. It recruited three birth cohorts of 1,643 child-mother pairs during pregnancy or delivery from maternity hospitals in Mexico city, between 1993 and 2004 (*Perng et al.*, 2019). The ELEMENT cohorts have since developed more than ten individual projects to answer some important questions related to the Developmental Origins of Health and Disease (DOHaD)' hypothesis in environmental health sciences. The main study aims are to understand how environmental toxicants, such as phthalates, chemicals and metals affect maternal and child health. A wide array of data were collected from mothers and children, including their dietary intakes, anthropometry, prenatal exposures, fasting metabolites, and DNA mythelation and so on.

In a DAG, causal effects pertain to three primary quantities, termed as total effect (TE), direct effect (DE) and indirect effect (IE). TE of an exposure on an outcome describes the total amount of changes of the outcome caused by the changes of the exposure. In the literature, TE usually is decomposed into two component (or effects), DE and IE. DE is the effect that is only explained by the exposure, while IE

is the effect that manifests through mediators. Sometimes, IE is also called mediation effect. In our motivating example, DE is the change in the level of insulin resistance when nutrient intakes change in which metabolome were held fixed. IE expresses the change in the level of insulin resistance resulted from the change of the metabolome under the condition of fixed exposure. A key objective in the mediation analysis is to assess the three types of effects, TE, DE and IE, especially the mediation effect IE. In such analysis, many methodological challenges arise, including (i) the presence of non-normal data such as discrete, count, survival variables; (ii) there are potentially many latent variables as a DAG is too simple to represent the underlying causal mediation pathways; (iii) the existence of unmeasured confounding factors; (iv) the presence of potential multiple or even high-dimensional correlated mediators, just name a few. Some of these methodological challenges motivated the development of statistical methods in this dissertation.

Specifically, this dissertation is motivated by three scientific research questions of interest:

- First, we are interested in understanding if the associations of dietary intakes on the insulin resistance is mediated through a cluster of glucose metabolites and acetylamino acids.

- Second, we are interested in finding how, and to what extent, a delay of reaching the BMI infancy peak on time or not may affect the associations of prenatal phthalate exposures on children's health outcomes during peripuberty.

- Lastly, we investigate if the associations of mother's calcium intakes in a calcium supplementation randomized control trial on birth weight is mediated by mother's blood lead level during third trimester.

1.2 Existing Work

The classical mediation analysis approach, first proposed by Baron and Kenny (*Baron and Kenny*, 1986), is implemented under the linear structural equation models (SEM, (*Ullman and Bentler*, 2003)) with multivariate normally distributed data. Recently, utilizing the counterfactual (potential) outcome framework in the causal inference literature (*Robins and Greenland*, 1992; *Pearl*, 2013; *VanderWeele and Vansteelandt*, 2014), the classical mediation approach has been extended to represent a causal mediation pathway via DAG under a certain scientific hypothesis. With a few extra assumptions on causality, the notions of DE and IE have been extended to natural direct effect (NDE), and natural indirect effect (NIE), which allows for a decomposition of the total effect under the settings with exposure-mediator interactions and/or settings with non-linearity (*VanderWeele and Vansteelandt*, 2009; *Vansteelandt and VanderWeele*, 2012; *Pearl*, 2013; *Valeri and VanderWeele*, 2013).

Since decades ago, most of the mediation analyses are only able to quantify and test one mediator in a DAG. MacKinnon et al. (*MacKinnon et al.*, 2002) summarize and compare fourteen different approaches to mediation analyses, including joint significance test(*MacKinnon et al.*, 2002), product test based on normality assumption (*Sobel*, 1982), and bootstrap test (*Bollen and Stine*, 1990). The challenge of the hypothesis testing for mediation effect lies on the composite nature of the null hypothesis, leading to conservative type I error. This remains an unsolved problem.

With the advent of advanced omics technologies, there has been an increasing need in analyzing the mediation effects in biomedical applications with multiple or even high-dimensional mediators, such as in epigenetic studies. In such settings, many works aim to decompose total effect into different path-specific effects involving multiple mediators (*Daniel et al.*, 2015; *Zhao et al.*, 2020). Other works aim to select the mediation pathways under high-dimensional settings (*Zhao and Luo*, 2016; *Song et al.*, 2020). Other procedures of causal mediator selection (*Barfield*

et al., 2017; *Huang et al.*, 2019; *Dai et al.*, 2020; *Liu et al.*, 2021) perform univariate analyses of mediators one by one, without acknowledging the fact that mediators under investigation are intrinsically correlated. These approaches assume that mediators are conditionally independent given both exposures and confounders, which is an unrealistic assumption. As noted in the literature, if correlations among mediators are present, analyzing their mediation effects one at a time would usually lead to biased effect estimates due to over-counted pathways and the violation of model assumptions (*VanderWeele and Vansteelandt*, 2014; *VanderWeele*, 2015). Thus, it is of critical importance to analyze a set of multiple mediators jointly, and test for their joint mediation effect simultaneously. Huang et al (*Huang and Pan*, 2016; *Huang et al.*, 2018) proposed two methods: Product Test based on Normal Product distribution (PT-NP), and Product Test based on Normality (PT-N) that test the mediation effect with multiple mediators. Although both methods work numerically well in simulations, there is a lack of rigorous theoretical understanding on the justification for several approximations taken in the derivations. This motivates the research of Chapter II, where I developed a likelihood ratio test for mediation effect with multiple mediators.

Non-normal data are frequently encountered in a mediation analysis, including one-dimensional dichotomous mediator (*Albert and Nelson*, 2011), one-dimensional dichotomous outcome (*VanderWeele and Vansteelandt*, 2010), and one-dimensional time-to-event outcome (*VanderWeele*, 2011). Due to the different types of non-normal data, different assumptions are required to fulfill the effect decomposition and to perform parameter estimation. There lacks a unified approach that conducts mediation analysis of exposure, mediator and outcome with different data types collectively. This motivates the research of Chapter III, where I proposed a unified framework via copula dependence models to analyze the data with mixed types in the mediation analysis.

Mediation analyses based on DAG may be too simple to represent the underlying causal mediation pathway. One challenge pertains to the influence of underlying latent variables. For example, Derkach et al. (*Derkach et al.*, 2019) considered a setting where a group of unmeasured variables are influenced by the exposure and would in turn impact mediators and outcome. However, little methodological work has been done to deal with potential latent exposure variables in mediation analyses. This motivates the research of Chapter IV, where I developed two models with latent exposure variables, and proposed generalized natural direct effect and generalized natural indirect effect.

1.3 A Summary of New Contributions

This dissertation develops new statistical methodologies to address the aforementioned limitations and challenges.

Chapter II develops a likelihood ratio (LR) test for multi-dimensional mediation effect that accounts for causally related mediators via the Lagrange Multiplier method. We decompose the parameter space under the composite null hypothesis into two disjoint spaces, and derive asymptotic null distributions of the test statistics in each sub-space. We develop a block coordinate descent algorithm to obtain the constrained maximum likelihood estimate, and perform extensive simulations to compare the LR method with two existing alternatives of PT-NP and PT-N. The simulation results demonstrate that our method can control type I error rate properly and provides higher or similar power with PT-N and PT-NP. A data from ELEMENT study is analyzed to examine whether a cluster consisting of seven glucose metabolites and acetylamino acids mediate the effect of fat or carbohydrate intakes on the scores of insulin resistance for Mexican children in the study.

Chapter III uses a copula dependence model to construct a unified mediation analysis approach to analyzing the data of mixed types, including continuous, categorical,

count variables. We establish joint parametric distributions of exposure, mediator and outcome featuring a lower-triangular dependence matrix reflecting DAG in the mediation pathway and a class of generalized linear models adjusting for confounding factors in the respective marginal distributions. We develop estimation procedures for the model parameters, as well as the causal effects of NDE and NIE. Simulation studies are performed to investigate the performance under three different settings with data of mixed types. A data from ELEMENT study with a binary mediator of whether a child has a delayed BMI infancy peak or not is examined to understand the association between mother's phthalate exposure during pregnancy and offspring's health outcomes during peripuberty.

Chapter IV introduces a conceptual framework of generalized total, direct and indirect effects (GTE/GNDE/GNIE) to relax the conventional definition of NDE and NIE, when the treatment or exposure is dichotomous. We utilize the latent exposure variable presentation to investigate the "actions" of the dichotomous treatment in a biomedical study. We propose two important types of models: effective dose model and latent exposure model, in which we demonstrate that the proposed GNDE and GNIE are more desirable to quantify and interpret compared with the conventional definitions. The simulation studies are carried out to examine the estimation procedure of the two models, and the mean bias, mean squared error and 95% coverage rate all suggest our methods provide accurate effect estimation and valid statistical inference. A randomized trial from the ELEMENT study is analyzed to illustrate our method, where we examine whether the relation between mother's calcium supplementation and children's birth weight is mediated by the mother's blood lead level during third trimester.

CHAPTER II

A Likelihood-based Test for Multi-dimensional Mediation Effects

2.1 Introduction

Mediation analysis is undertaken pervasively in practice to understand whether or not the effect of an exposure on an outcome has been mediated through some intermediate variables, which are, in short, called mediators. The mediation analysis approach, first proposed by *Baron and Kenny* (1986), has been extensively applied in many disciplines to perform pathway analyses. Utilizing the counterfactual outcome framework in the causal inference literature (*Rubin*, 1978; *Robins and Greenland*, 1992; *Pearl*, 2001), the mediation approach has been recently extended to study causal mediation pathways via directed acyclic graphs (DAG) formed under a certain scientific hypothesis as shown in Figure 2.1. With a few extra assumptions of causation, such extension allows to decompose the total causal effect into a sum of direct effect and indirect effect in the presence of interactions and non-linearities (*Pearl*, 2001; *VanderWeele and Vansteelandt*, 2009). This new causal framework has received much attention in the literature.

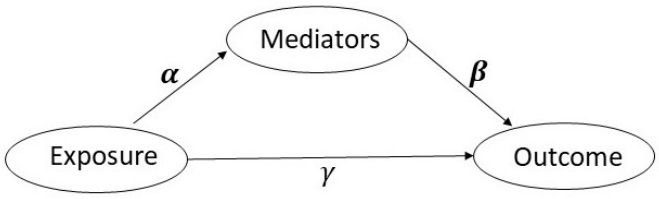

Figure 2.1: A DAG involving exposure, mediators and outcome.

There are many existing methods in the literature developed to test the existence of mediation effect (or the indirect effect) in the case of a single potential mediator, including Sobel's test (*Sobel*, 1982), bootstrap method (*Bollen and Stine*, 1990), joint significant test (*MacKinnon et al.*, 2002). Recently, with the advent of advanced omics technologies, there has been an increasing need in testing for mediation effects in applications with a group of multiple or even high-dimensional mediators, for which several methods have been developed. For examples, multiple testing approaches for genome-wide association analysis have been proposed based on simultaneous single mediator tests with multiple comparison correction (*Huang*, 2019; *Huang et al.*, 2019; *Djordjilović et al.*, 2019; *Dai et al.*, 2020). In such methods, test for a causal mediation effect has been focused on a single mediator via a univariate screening analysis of mediators one by one, ignoring the dependence among multiple mediators. Although multiple testing corrections have been adjusted to identify the potential mediators, the interpretation of the causal effect is still limited to each of the selected mediators, instead of a simultaneous inference for the group-level mediation effect. However, in many applications when there exist multiple correlated mediators, in particular a cohesive cluster of biologically relevant mediators, the group-level mediation effect does not simply equal to a summation of individual mediation effects, as pointed out by *VanderWeele* (2015). Therefore, the conclusion drawn from the univariate screening test with multiple comparison correction does not necessarily produce a valid statistical inference for the group-level mediation effect. While these univariate screening procedures are useful to discover individually potential mediators, it is of

critical importance to analyze a cluster of correlated multiple mediators jointly. This analytic objective calls for a test for their group-level mediation effect.

The mediation relationships of a DAG in Figure 2.1 are extensively analyzed by the linear normal structural equation model (SEM). When exposure-mediator interaction terms are absent in the SEM, the group-level mediation effect is expressed as the product, $\boldsymbol{\alpha}^\top\boldsymbol{\beta}$, where $\boldsymbol{\alpha}$ is the vector of coefficients for exposure-mediator association and $\boldsymbol{\beta}$ is the vector of coefficients for mediator-outcome association. In this chapter, we aim to develop a simultaneous test for the joint group-level mediation effect under the null hypothesis of no mediation effect $H_0 : \boldsymbol{\alpha}^\top\boldsymbol{\beta} = 0$. A key technical challenge of performing this hypothesis test pertains to the involvement of composite hypotheses; that is, $\boldsymbol{\alpha}^\top\boldsymbol{\beta} = 0$ may arise from a large number of combinations in α_q and β_q, $q = 1, \cdots, Q$, where Q is the number of mediators. One example of possible combination is $\boldsymbol{\alpha} = \boldsymbol{\beta} = \mathbf{0}$, which is of great interest in practice. More subtle cases may arise from cancellations among some individual products of $\alpha_q\beta_q, q = 1, \ldots, Q$ to satisfy $\boldsymbol{\alpha} = \boldsymbol{\beta} = \mathbf{0}$. Two existing approaches to testing this group-level mediation effect include: Product Test based on Normal Product distribution (PT-NP)(*Huang and Pan*, 2016; *Huang et al.*, 2018), and Product Test based on Normality (PT-N) (*Huang and Pan*, 2016; *Huang et al.*, 2018). Although these two methods have shown satisfactory performances numerically via simulation studies, the rigorous theoretical justification, such as the results of asymptotic distributions of such test statistics under the null remain little explored, especially under the case of $\boldsymbol{\alpha} = \boldsymbol{\beta} = \mathbf{0}$. To bridge this gap, in this chapter we investigate a simultaneous likelihood ratio (LR) test for the joint group-level mediation effect under the null hypothesis $\boldsymbol{\alpha}^\top\boldsymbol{\beta} = 0$ in that we establish asymptotic distributions of the proposed test statistics as well as confirm the theoretical results by numerical analyses.

This chapter makes two methodological contributions. First, we develop a constrained optimization to compute the likelihood ratio test statistic under an irregular

null parameter space using the Lagrange Multiplier. This computation is implemented by an efficient block coordinate decent algorithm. Second, we derive the asymptotic distributions of the proposed LR test statistic under the composite null hypothesis $H_0 : \boldsymbol{\alpha}^\top \boldsymbol{\beta} = 0$, and show theoretically that our LR test can properly control the type I error. Through numerical experiments, including both simulation studies and a data application example, we demonstrate that our LR test can not only have a proper type I error control but also improve the power in the cases considered in the simulation studies, in comparison to the two existing tests, PT-NP and PT-N.

The remainder of the chapter is organized as follows: Section 2.2 introduces the linear structural equation model. Section 2.3 concerns the development of likelihood ratio test, including the Lagrange Multiplier and the asymptotic null distributions for the LR test statistic. Section 2.4 presents an iterative procedure for implementing the LR test. Section 2.5 shows the numerical performance of the LR test in terms of type I error rate and power, and its comparison to the existing methods. Section 2.6 demonstrates an application of testing for a group-level mediation effect of a metabolite cluster on the association between dietary intakes and insulin resistance. Section 2.7 concludes the chapter with discussions on both advantages and limitations of the proposed LR method. Detailed technical derivations and proofs are included in the Appendix A.

2.2 Framework

2.2.1 Structure Equation Model

Consider a data set of n observations, $(X_i, M_{i,j}, Y_i)$, $i = 1, \ldots, n$, randomly sampled from n subjects. For the i-th subject, Y_i represents an outcome variable of interest, X_i represents an exposure variable, and $M_i = \{M_{i,j}\}_{j=1}^{Q}$ represents a Q-dimensional vector of mediators. In addition, $Z_i = \{Z_{i,l}\}_{l=1}^{L}$ represents an L-dimensional vector

of confounding variables with the first element $Z_{i,1} \equiv 1$ for the intercept. In this chapter, we consider the case of both Q and L being fixed and $Q + L + 1 < n$. A linear structural equation model (SEM) takes the following form:

$$Y_i = X_i\gamma + \mathbf{M}_i^\top \boldsymbol{\beta} + \mathbf{Z}_i^\top \boldsymbol{\eta} + \epsilon_{Y,i}, \qquad \mathbf{M}_i^\top = X_i\boldsymbol{\alpha}^\top + \mathbf{Z}_i^\top \boldsymbol{\zeta} + \epsilon_{M,i}^\top, \tag{2.1}$$

where $\mathbf{M}_i = (M_{i,1}, \ldots, M_{i,Q})^\top$, $\mathbf{Z}_i = (Z_{i,1}, \ldots, Z_{i,L})^\top$, γ is a scalar, $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_Q)^\top$, $\boldsymbol{\eta} = (\eta_1, \ldots, \eta_L)^\top$, $\boldsymbol{\alpha} = (\alpha_1, \ldots, \alpha_Q)^\top$, $\boldsymbol{\zeta} = (\zeta_{l,j})_{L\times Q}$, $\epsilon_{Y,i} \overset{i.i.d.}{\sim} N(0, \sigma_Y^2)$, $\epsilon_{M,i} \overset{i.i.d.}{\sim} MVN(0, \boldsymbol{\Sigma}_M)$, and $\boldsymbol{\Sigma}_M$ is a $Q \times Q$ positive definite covariance matrix, $i = 1, \ldots, n$.

Denote the collection of model parameters by $\boldsymbol{\theta} = \{\boldsymbol{\alpha}, \boldsymbol{\beta}, \gamma, \boldsymbol{\eta}, \boldsymbol{\zeta}, \boldsymbol{\Sigma}_M, \sigma_Y^2\}$ and $\boldsymbol{\Theta}$ is a generic notation for the parameter space. In the counterfactual outcome paradigm (*Robins and Greenland*, 1992; *Pearl*, 2001), under the fundamental assumptions of consistency and the absence of unmeasured confounders, VanderWeele (*VanderWeele and Vansteelandt*, 2014) shows that exposure variable X changes from a value x_0 to another value x_1, the Natural Direct Effect (NDE) and Natural Indirect Effect (NIE) in model (2.1) take the following forms: $\text{NDE}(x_0, x_1) = \gamma(x_1 - x_0)$, and $\text{NIE}(x_0, x_1) = \boldsymbol{\alpha}^\top\boldsymbol{\beta}(x_1 - x_0)$.

2.2.2 Unconstrained Parameter Estimation

To establish a likelihood ratio test for the null hypothesis of no group-level mediation effect, $H_0 : \boldsymbol{\alpha}^\top\boldsymbol{\beta} = 0$, we need to perform both unconstrianed and constrained maximum likelihood estimations (MLE) under the null and alternative hypotheses, respectively. SEM (2.1) may be rewritten as a matrix form:

$$\mathbf{Y} = \mathbf{W}\bar{\boldsymbol{\beta}} + \boldsymbol{\epsilon}, \qquad \mathbf{M} = \mathbf{B}\bar{\boldsymbol{\alpha}} + \mathbf{E}, \tag{2.2}$$

where $\bar{\boldsymbol{\beta}} = (\beta_1, \ldots, \beta_Q, \eta_1, \ldots, \eta_L, \gamma)^\top$, $\mathbf{Y}$ is an $n \times 1$ vector of the outcomes, $\mathbf{W}$ is an $n \times (Q + L + 1)$ matrix of mediators, confounders and exposure variable with

$\mathbf{W}_i = (M_{i,1}, \ldots, M_{i,Q}, Z_{i,1}, \ldots, Z_{i,L}, X_i)^\top$, $i = 1, \ldots, n$, and $\boldsymbol{\epsilon} \sim MVN(\mathbf{0}, \sigma_Y^2 \mathbf{I}_n)$. Similarly, $\mathbf{M}$ is an $n \times Q$ matrix of mediators, $\mathbf{B}$ is an $n \times (L+1)$ matrix of exposure and confounders with $\mathbf{B}_i = (X_i, Z_{i,1}, \ldots, Z_{i,L})$, and $\mathbf{E} = (\mathbf{E}_1^\top, \ldots, \mathbf{E}_n^\top)^\top$ with $\mathbf{E}_i \sim MVN(\mathbf{0}, \boldsymbol{\Sigma}_M)$. Here $\bar{\boldsymbol{\alpha}}$ is an $(L+1) \times Q$ matrix of parameters, with its first row vector being $\boldsymbol{\alpha}^\top$ in the model (2.1), and its remaining $L \times Q$ submatrix being the parameter matrix of $\boldsymbol{\zeta}$. It follows that the two times negative log likelihood function is given by

$$\begin{aligned} -2\ell(\boldsymbol{\theta}) =& n \log(\sigma_Y^2) + n \log(|\boldsymbol{\Sigma}_M|) + \sigma_Y^{-2} (\mathbf{Y} - \mathbf{W}\bar{\boldsymbol{\beta}})^\top (\mathbf{Y} - \mathbf{W}\bar{\boldsymbol{\beta}}) \\ &+ \mathrm{tr}\{(\mathbf{M} - \mathbf{B}\bar{\boldsymbol{\alpha}}) \boldsymbol{\Sigma}_M^{-1} (\mathbf{M} - \mathbf{B}\bar{\boldsymbol{\alpha}})^\top\}. \end{aligned}$$

The standard theory of the MLE leads to the following unconstrained maximum likelihood estimators of $\boldsymbol{\theta}$, denoted as $\hat{\boldsymbol{\theta}} = \{\hat{\bar{\boldsymbol{\alpha}}}, \hat{\bar{\boldsymbol{\beta}}}, \hat{\sigma}_y^2, \hat{\boldsymbol{\Sigma}}_M\}$, where

$$\begin{aligned} \hat{\bar{\boldsymbol{\alpha}}} &= (\mathbf{B}^\top \mathbf{B})^{-1} \mathbf{B}^\top \mathbf{M}, \text{ and } \hat{\bar{\boldsymbol{\beta}}} = (\mathbf{W}^\top \mathbf{W})^{-1} \mathbf{W}^\top \mathbf{Y}; \\ \hat{\sigma}_y^2 &= (\mathbf{Y} - \mathbf{W}\hat{\bar{\boldsymbol{\beta}}})^\top (\mathbf{Y} - \mathbf{W}\hat{\bar{\boldsymbol{\beta}}})/n, \text{ and } \hat{\boldsymbol{\Sigma}}_M = (\mathbf{M} - \mathbf{B}\hat{\bar{\boldsymbol{\alpha}}})^\top (\mathbf{M} - \mathbf{B}\hat{\bar{\boldsymbol{\alpha}}})/n. \end{aligned}$$

2.2.3 Constrained Parameter Estimation

Let $\tilde{\boldsymbol{\theta}}$ denote the constrained MLE under the null $H_0 : \boldsymbol{\alpha}^\top \boldsymbol{\beta} = 0$, which will be obtained by the method of Lagrange Multiplier. We consider a Lagrange objective function of the following form, with tuning parameter λ,

$$g(\bar{\boldsymbol{\alpha}}, \bar{\boldsymbol{\beta}}, \sigma_Y^2, \boldsymbol{\Sigma}_M, \lambda) = -2\ell(\boldsymbol{\theta}) - 2\lambda \boldsymbol{\alpha}^\top \boldsymbol{\beta}. \tag{2.3}$$

Differentiating the function $g(\cdot)$ with respect to the model parameters yields the following equations of the regression coefficients,

$$\bar{\boldsymbol{\alpha}} = (\mathbf{B}^\top \mathbf{B})^{-1}\mathbf{B}^\top \mathbf{M} + \lambda(\mathbf{B}^\top \mathbf{B})^{-1}\boldsymbol{\beta}^* \boldsymbol{\Sigma}_M = \hat{\bar{\boldsymbol{\alpha}}} + \lambda(\mathbf{B}^\top \mathbf{B})^{-1}\boldsymbol{\beta}^* \boldsymbol{\Sigma}_M, \tag{2.4}$$

$$\bar{\boldsymbol{\beta}} = (\mathbf{W}^\top \mathbf{W})^{-1}\mathbf{W}^\top \mathbf{Y} + \lambda\sigma_Y^2(\mathbf{W}^\top \mathbf{W})^{-1}\boldsymbol{\alpha}^* = \hat{\bar{\boldsymbol{\beta}}} + \lambda\sigma_Y^2(\mathbf{W}^\top \mathbf{W})^{-1}\boldsymbol{\alpha}^*, \tag{2.5}$$

and the equations of variance parameters,

$$\sigma_y^2 = (\mathbf{Y} - \mathbf{W}\bar{\boldsymbol{\beta}})^\top(\mathbf{Y} - \mathbf{W}\bar{\boldsymbol{\beta}})/n, \text{ and } \boldsymbol{\Sigma}_M = (\mathbf{M} - \mathbf{B}\bar{\boldsymbol{\alpha}})^\top(\mathbf{M} - \mathbf{B}\bar{\boldsymbol{\alpha}})/n, \tag{2.6}$$

where $\boldsymbol{\beta}^*$ is an $(L+1)\times Q$ matrix with the first row being $\boldsymbol{\beta}^\top$ and the rest of elements are zeros, and $\boldsymbol{\alpha}^*$ is a $(Q+L+1)\times 1$ vector with the first Q elements being $\boldsymbol{\alpha}$ and the rest of elements being zero. Given that $\boldsymbol{\alpha}^\top$ appears in the first row of $\bar{\boldsymbol{\alpha}}$, we denote the first row of $\hat{\bar{\boldsymbol{\alpha}}}$ by $\mathbf{a}_1^\top$, and the first row of $(\mathbf{B}^\top \mathbf{B})^{-1}\boldsymbol{\beta}^* \boldsymbol{\Sigma}_M$ by $\mathbf{b}_1^\top$. It follows that $\boldsymbol{\alpha}^\top = \mathbf{a}_1^\top - \lambda\mathbf{b}_1^\top$. Similarly, given $\boldsymbol{\beta}$ being in the first Q rows of vector $\bar{\boldsymbol{\beta}}$, denote the first Q rows of vector $\hat{\bar{\boldsymbol{\beta}}}$ by $\mathbf{a}_2$, and the first Q rows of $(\mathbf{W}^\top \mathbf{W})^{-1}\boldsymbol{\alpha}^*$ by $\mathbf{b}_2$. Under the constraint $\boldsymbol{\alpha}^\top\boldsymbol{\beta} = 0$, we obtain $(\mathbf{a}_1^\top + \lambda\mathbf{b}_1^\top)(\mathbf{a}_2 + \lambda\mathbf{b}_2) = 0$. This leads to two possible solutions of λ given in (2.7), and we shall choose the one that yields the higher log-likelihood,

$$\tilde{\lambda} = \frac{-(\mathbf{a}_1^\top\mathbf{b}_2 + \mathbf{b}_1^\top\mathbf{a}_2) \pm \sqrt{(\mathbf{a}_1^\top\mathbf{b}_2 + \mathbf{b}_1^\top\mathbf{a}_2)^2 - 4\mathbf{b}_1^\top\mathbf{b}_2\mathbf{a}_1^\top\mathbf{a}_2}}{2\mathbf{b}_1^\top\mathbf{b}_2}. \tag{2.7}$$

Remark II.1. After we obtain the constrained MLE solutions $(\tilde{\boldsymbol{\theta}}, \tilde{\lambda})$ by the method of the Lagrange Multiplier above, we then evaluate the Hessian matrix of the function $g(\cdot)$ in (2.3). It is easy to show that in the setting of the linear SEM the Hessian matrix is positive definite, guaranteeing the convexity of the penalized objective function $g(\cdot)$ and thus the unique minimum given by the solutions $(\tilde{\boldsymbol{\theta}}, \tilde{\lambda})$.

2.3 Likelihood Ratio Test for Joint Mediation Effect

2.3.1 Test Statistic

To simultaneously assess the joint mediation effect of multi-dimensional mediators, the first analytic task is to test the null hypothesis $H_0 : \boldsymbol{\alpha}^\top \boldsymbol{\beta} = 0$ versus $H_1 : \boldsymbol{\alpha}^\top \boldsymbol{\beta} \neq 0$, where the null hypothesis corresponds to the case of zero NIE under SEM (2.1). As pointed above, since the null hypothesis allows internal cancellation, it does not preclude the possibility of component-wise nonzero mediation effects in the sense that $\alpha_q \beta_q \neq 0, q = 1, \cdots Q$ but $\boldsymbol{\alpha}^\top \boldsymbol{\beta} = 0$. Following the classical Wilks' theory of likelihood ratio (LR) test, we construct a LR test statistic of the form:

$$T_n = -2\{\sup_{\boldsymbol{\theta} \in \boldsymbol{\Theta}: \boldsymbol{\alpha}^\top \boldsymbol{\beta} = 0} \ell(\boldsymbol{\theta}) - \sup_{\boldsymbol{\theta} \in \boldsymbol{\Theta}} \ell(\boldsymbol{\theta})\} = -2\{\ell(\tilde{\boldsymbol{\theta}}) - \ell(\hat{\boldsymbol{\theta}})\}, \tag{2.8}$$

where $\hat{\boldsymbol{\theta}}$ and $\tilde{\boldsymbol{\theta}}$ denote, respectively, the unconstrained MLE under H_1 and the constrained MLE under H_0 obtained in Sections 2.2.2 and 2.2.3.

2.3.2 Properties of the LR test

This section concerns the asymptotic distributions of the likelihood ratio statistic T_n in (2.8) under the null hypothesis $H_0 : \boldsymbol{\alpha}^\top \boldsymbol{\beta} = 0$. Using the large-sample properties, we propose a new test that can properly control the type I error with theoretical guarantees. For all lemmas and theorems presented in this section, their technical proofs are given in the Appendix. We begin with some notations. For the ease of exposition, we redefine $\boldsymbol{\theta} = (\boldsymbol{\alpha}^\top, \boldsymbol{\zeta}, \boldsymbol{\beta}^\top, \boldsymbol{\eta}^\top, \gamma)^\top$, where $\boldsymbol{\zeta}$ denotes the row vector of LQ elements vectorized from the matrix $\boldsymbol{\zeta}_{L \times Q}$. Define the constraint function by $h(\boldsymbol{\theta}) = \boldsymbol{\alpha}^\top \boldsymbol{\beta}$. It is easy to see that its gradient $\dot{h}(\boldsymbol{\theta}) = \nabla_{\boldsymbol{\theta}} h(\boldsymbol{\theta}) = (\boldsymbol{\beta}^\top, \mathbf{0}_{LQ}^\top, \boldsymbol{\alpha}^\top, \mathbf{0}_{L+1}^\top)^\top$. Let

$$\mathbf{H}(\boldsymbol{\theta}) = \nabla_{\boldsymbol{\theta}} \dot{h}(\boldsymbol{\theta}) = \begin{pmatrix} \mathbf{0}_{(L+1)Q \times (L+1)Q} & \tilde{\mathbf{H}}_{(L+1)Q \times (Q+L+1)} \\ \tilde{\mathbf{H}}^\top_{(L+1)Q \times (Q+L+1)} & \mathbf{0}_{(Q+L+1) \times (Q+L+1)} \end{pmatrix},$$

where

$$\tilde{\mathbf{H}}_{(L+1)Q\times(Q+L+1)} = \begin{pmatrix} \mathbf{I}_Q & \mathbf{0}_{Q\times(L+1)} \\ \mathbf{0}_{LQ\times q} & \mathbf{0}_{LQ\times(L+1)} \end{pmatrix}.$$

The information matrix $\mathbf{I}(\boldsymbol{\theta}) = -\mathrm{E}\left(\frac{1}{n}\frac{\partial^2 \ell(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}\boldsymbol{\theta}^\top}\right)$ has a closed-form, presented in Appendix A.1. Let $\mathbf{A}(\boldsymbol{\theta}) = \mathbf{I}(\boldsymbol{\theta})^{-\frac{1}{2}}\mathbf{H}(\boldsymbol{\theta})\mathbf{I}(\boldsymbol{\theta})^{-\frac{1}{2}}$. To derive the asymptotic properties, we first introduce a lemma that establishes the eigenvalue bounds of matrices $\mathbf{H}(\boldsymbol{\theta})$ and $\mathbf{A}(\boldsymbol{\theta})$.

Lemma II.2. *For any $\boldsymbol{\theta} \in \mathbb{R}^{2Q+LQ+L+1}$, we have the following results.*

(i) The matrix $\mathbf{H}(\boldsymbol{\theta}) = \nabla_{\boldsymbol{\theta}}\dot{h}(\boldsymbol{\theta})$ has $2Q$ nonzero eigenvalues equal to 1 or -1. If nonzero eigenvalues are arranged in a descending order as of the form $h_1 \geq h_2 \geq \cdots \geq h_{2Q}$, then $h_1 = \cdots = h_Q = 1$, $h_{Q+1} = \cdots = h_{2Q} = -1$.

(ii) The matrix $\mathbf{A}(\boldsymbol{\theta})$ has $2Q$ nonzero eigenvalues. If nonzero eigenvalues are arranged in a descending order as of the form $\upsilon_1 \geq \upsilon_2 \geq \cdots \geq \upsilon_Q > 0 > \upsilon_{Q+1} \geq \cdots \geq \upsilon_{2Q}$, then they satisfy $\sum_{i=1}^{2Q} \upsilon_i = 0$, and $\upsilon_1 = -\upsilon_{2Q}, \upsilon_2 = -\upsilon_{2Q-1}, \ldots, \upsilon_Q = -\upsilon_{Q+1}$.

The above properties for the eigenvalues of $\mathbf{A}(\boldsymbol{\theta})$ are used to establish asymptotic null distributions of the LR test statistic. The proof of Lemma II.2 is presented in Appendix A.2.

Lemma II.3. *In the case of $\boldsymbol{\alpha} = \boldsymbol{\beta} = \mathbf{0}$, let $\boldsymbol{\theta}_0$ be the true parameters that generate the data, and the asymptotic distributions of the constrained MLE $\tilde{\boldsymbol{\theta}}$ and $\tilde{\lambda}$ are given by, as $n \to \infty$,*

$$\tilde{\lambda} \xrightarrow{d} \Lambda_0, \text{ where } \Lambda_0 \overset{d}{\equiv} -\frac{\sum_{q=1}^{Q} \upsilon_q(\xi_q - \xi_{q+Q})}{2\sum_{q=1}^{Q} \upsilon_q^2(\xi_q + \xi_{q+Q})},$$

with $\xi_q \overset{i.i.d.}{\sim} \chi_1^2, \quad q = 1, \cdots, 2Q$.

For any $\lambda^* \in \mathbb{R}$*, conditional on a value* $\tilde{\lambda} = \lambda^*$*,*

$$\sqrt{n}(\tilde{\boldsymbol{\theta}} - \boldsymbol{\theta}_0) \mid \tilde{\lambda} = \lambda^* \xrightarrow{d} \mathrm{N}\left(\mathbf{0}, \{\mathbf{I}(\boldsymbol{\theta}_0) - \lambda^* \mathbf{H}(\boldsymbol{\theta}_0)\}^{-1} \mathbf{I}(\boldsymbol{\theta}_0) \{\mathbf{I}(\boldsymbol{\theta}_0) - \lambda^* \mathbf{H}(\boldsymbol{\theta}_0)\}^{-1}\right),$$

where $\upsilon_1, \ldots, \upsilon_Q$ *are* Q *positive eigenvalues of* $\mathbf{A}(\boldsymbol{\theta}_0)$*.*

Lemma II.3 leads to an asymptotic joint distribution of $\tilde{\boldsymbol{\theta}}$ and $\tilde{\lambda}$ due to the fact $[\tilde{\boldsymbol{\theta}}, \tilde{\lambda}] = [\tilde{\boldsymbol{\theta}}|\tilde{\lambda}][\tilde{\lambda}]$. Thus, we obtain the asymptotic distribution of the LR test statistic in the scenario of $\boldsymbol{\alpha} = \boldsymbol{\beta} = \mathbf{0}$. The proof of Lemma II.3 is presented in Appendix A.3.

Theorem II.4. *Under* $H_0 : \boldsymbol{\alpha}^\top \boldsymbol{\beta} = 0$*, the asymptotic distributions of the likelihood ratio test statistic* T_n *are given by,*

(i) when $(\boldsymbol{\alpha}^\top, \boldsymbol{\beta}^\top)^\top \neq \mathbf{0}$*, as* $n \to \infty$*,* $T_n \xrightarrow{d} \chi^2_1$*,*

(ii) when $\boldsymbol{\alpha} = \boldsymbol{\beta} = \mathbf{0}$*, as* $n \to \infty$*,* $T_n \xrightarrow{d} \Lambda_1$ *with* $\Lambda_1 \overset{d}{\equiv} \frac{\left\{\sum_{q=1}^{Q} \upsilon_q (\xi_q - \xi_{q+Q})\right\}^2}{4 \sum_{q=1}^{Q} \upsilon_q^2 (\xi_q + \xi_{q+Q})}$*, where* $\xi_q \overset{i.i.d.}{\sim} \chi^2_1, q = 1, \cdots, 2Q$*.*

In this chapter, we write $\Lambda_1 \sim \kappa_Q$ distribution. The proof of Theorem II.4 involves deriving the asymptotic distributions of the constrained MLE. Although the classical large-sample work for the LR test, e.g. (*Aitchison et al.*, 1958; *Wolak*, 1989), may be directly applied to prove part (i) of Theorem II.4, the proof of part (ii) is non-trivial and needs specific technical arguments and treatments on manipulating asymptotic distribution of $\tilde{\lambda}$, similar to those given in the proof of Lemma II.3. The proof of Theorem II.4 is presented in Appendix A.4. To implement the κ_Q distribution after both matrix $\mathbf{A}(\boldsymbol{\theta})$ and its Q eigenvalues are estimated, we invoke the Monte Carlo simulation with a large number of draws (say 10,000) independently from $2Q$ χ^2_1 distributed variables $\xi_q, q = 1, \cdots, 2Q$.

It follows from Theorem II.4 that we propose a test for $H_0 : \boldsymbol{\alpha}^\top \boldsymbol{\beta} = 0$, termed as

LR test, given by the decision function:

$$\phi_n = I[T_n > (\chi^2_{1,(1-\alpha)} \vee \kappa_{Q,(1-\alpha)})], \tag{2.9}$$

where $a \vee b = \max(a, b)$, $\kappa_{Q,(1-\alpha)}$ is the $(1-\alpha)$ quantile of the null distribution given in part (ii) of Theorem II.4, and $\chi^2_{1,(1-\alpha)}$ is the $(1-\alpha)$ quantile of the χ^2_1 distribution. When $\phi_n = 1$, we reject the null H_0; otherwise, accept the null H_0.

Theorem II.5. *The LR test in* (2.9) *controls the type I error; that is*

$$\sup_{\boldsymbol{\theta} \in \boldsymbol{\Theta}: \boldsymbol{\alpha}^\top \boldsymbol{\beta} = 0} P_{\boldsymbol{\theta}}(\phi_n = 1) \leq \alpha,$$

where $0 < \alpha < 1$ *is a prefixed type I error rate.*

Proof. Divide the parameter space under the H_0, $\boldsymbol{\Theta} = \{(\boldsymbol{\alpha}, \boldsymbol{\beta}) : \boldsymbol{\alpha}^\top \boldsymbol{\beta} = 0\}$ into two disjoint sub-spaces: $\boldsymbol{\Theta}_1 = \{(\mathbf{0}, \mathbf{0})\}$ and $\boldsymbol{\Theta}_2 = \boldsymbol{\Theta} \setminus \boldsymbol{\Theta}_1$. Then,

$$\begin{aligned}
&\sup_{\boldsymbol{\theta} \in \boldsymbol{\Theta}: \boldsymbol{\alpha}^\top \boldsymbol{\beta} = 0} P_{\boldsymbol{\theta}}(\phi_n = 1) \\
&= \sup_{\boldsymbol{\theta} \in \boldsymbol{\Theta}_1 \cup \boldsymbol{\Theta}_2} P_{\boldsymbol{\theta}}(T_n > \chi^2_{1,(1-\alpha)} \vee \kappa_{Q,(1-\alpha)}) \\
&= \max\{ \sup_{\boldsymbol{\theta} \in \boldsymbol{\Theta}_1} P_{\boldsymbol{\theta}}(T_n > \chi^2_{1,(1-\alpha)} \vee \kappa_{Q,(1-\alpha)}), \sup_{\boldsymbol{\theta} \in \boldsymbol{\Theta}_2} P_{\boldsymbol{\theta}}(T_n > \chi^2_{1,(1-\alpha)} \vee \kappa_{Q,(1-\alpha)}) \} \\
&\leq \max\{ \sup_{\boldsymbol{\theta} \in \boldsymbol{\Theta}_1} P_{\boldsymbol{\theta}}(T_n > \kappa_{Q,(1-\alpha)}), \sup_{\boldsymbol{\theta} \in \boldsymbol{\Theta}_2} P_{\boldsymbol{\theta}}(T_n > \chi^2_{1,(1-\alpha)}) \} \\
&\leq \alpha
\end{aligned}$$

□

2.4 Implementation

In practice, to perform the LR test ϕ_n, we first compute two p-values of $p_1 = 1 - F_{\chi^2_1}(T_n)$ and $p_2 = 1 - F_{\kappa_Q}(T_n)$, where $F_{\chi^2_1}$ is the CDF of the χ^2_1 distribution, and F_{κ_Q} is

the CDF of the κ_Q distribution. Then we reject the null hypothesis if the $\max(p_1, p_2)$ is smaller than significance level α.

To obtain the constrained MLE, we develop a block coordinate descent algorithm given as follows. We partition $\boldsymbol{\theta}$ into two sets: $\boldsymbol{\theta}_1 = \{\tilde{\tilde{\boldsymbol{\alpha}}}, \tilde{\tilde{\boldsymbol{\beta}}}\}$, and $\boldsymbol{\theta}_2 = \{\tilde{\sigma}_Y^2, \tilde{\boldsymbol{\Sigma}}_M\}$, as well as λ. The unconstrained MLE $\hat{\boldsymbol{\theta}} = \{\hat{\tilde{\boldsymbol{\alpha}}}, \hat{\tilde{\boldsymbol{\beta}}}, \hat{\sigma}_Y^2, \hat{\boldsymbol{\Sigma}}_M\}$ are used as the initial values to start the algorithm. This updating scheme consists of three steps: given $\boldsymbol{\theta}_1$ and $\boldsymbol{\theta}_2$, maximize the likelihood with respect to λ; given $\boldsymbol{\theta}_2$ and λ, update $\boldsymbol{\theta}_1$ until convergence; given $\boldsymbol{\theta}_1$, update $\boldsymbol{\theta}_2$. The algorithm is detailed below in Algorithm 1, where the default number of Monte Carlo simulations is set at 10,000.

Algorithm 1

- Compute the unconstrained MLE $\hat{\boldsymbol{\theta}} = \{\hat{\tilde{\boldsymbol{\alpha}}}, \hat{\tilde{\boldsymbol{\beta}}}, \hat{\sigma}_Y^2, \hat{\boldsymbol{\Sigma}}_M\}$, and evaluate the log-likelihood $\ell(\hat{\boldsymbol{\theta}})$. At the jth-iteration, let $\boldsymbol{\theta}_1^{(j)} = \{\tilde{\tilde{\boldsymbol{\alpha}}}^{(j)}, \tilde{\tilde{\boldsymbol{\beta}}}^{(j)}\}$, and let $\boldsymbol{\theta}_2^{(j)} = \{\tilde{\sigma}_Y^{2(j)}, \tilde{\boldsymbol{\Sigma}}_M^{(j)}\}$. Set $\boldsymbol{\theta}_1^{(0)} = \{\hat{\tilde{\boldsymbol{\alpha}}}, \hat{\tilde{\boldsymbol{\beta}}}\}$ and $\boldsymbol{\theta}_2^{(0)} = \{\hat{\sigma}_Y^2, \hat{\boldsymbol{\Sigma}}_M\}$ as the initial values.

- For $j = 0, 1, \ldots, J$

  - calculate $\lambda^{(j)} = \underset{\lambda}{\operatorname{argmax}}\{\ell(\boldsymbol{\theta}_1^{(j)}, \boldsymbol{\theta}_2^{(j)}, \lambda)\}$ from (2.7);
  - calculate $\boldsymbol{\theta}_1^{(j+1)} = \underset{\boldsymbol{\theta}_1}{\operatorname{argmax}}\{\ell(\boldsymbol{\theta}_1, \boldsymbol{\theta}_2^{(j)}, \lambda^{(j)})\}$ from (2.4) and (2.5);
  - calculate $\boldsymbol{\theta}_2^{(j+1)}$ from $\boldsymbol{\theta}_1^{(j+1)}$ based on (2.6);
  - calculate $\delta = \| \boldsymbol{\theta}_1^{(j+1)} - \boldsymbol{\theta}_1^{(j)} \|$;
  - **If** $|\delta| < tol$

    break

    end if

  end for

- **Output**: $\tilde{\boldsymbol{\theta}} = \left\{\tilde{\tilde{\boldsymbol{\alpha}}}^{(j+1)}, \tilde{\tilde{\boldsymbol{\beta}}}^{(j+1)}, \tilde{\sigma}_Y^{2(j+1)}, \tilde{\boldsymbol{\Sigma}}_M^{(j+1)}\right\}$, and calculate the log-likelihood.

- Calculate the test statistic $T = -2\left\{\ell(\tilde{\boldsymbol{\theta}}) - \ell(\hat{\boldsymbol{\theta}})\right\}$, and compute the p-value p_1 under the null distribution of χ_1^2.

- Estimate $\mathbf{A}(\boldsymbol{\theta}_0)$ based on $\hat{\sigma}_Y^2$ and $\hat{\boldsymbol{\Sigma}}_M$, and calculate its Q positive eigenvalues that are then used to simulate the κ_Q distribution, and compute its p-value p_2.

- Report $\max(p_1, p_2)$ as the final p-value.

2.5 Simulation Studies

2.5.1 Setup

We conduct extensive simulation studies to evaluate the performance of the proposed LR test. In particular, we compare the type I error control and power of our method with two existing methods: PT-N and PT-NP tests proposed by Huang et al (*Huang and Pan*, 2016). In addition, we consider a comparison to a recent method of High-Dimensional Multiple Testing (HDMT) proposed by (*Dai et al.*, 2020). HDMT was originally developed for a univariate screening of mediators with controlled false discovery rate in genome studies, representing a typical kind of testing approach widely adopted in practice to avoid simultaneous inference. Because HDMT is not a method established in the paradigm of the Neyman-Pearson hypothesis testing, we present the comparison results in the Appendix A.5.

The SEM is set up as follows. The exposure variable X is simulated from $N(0, 1)$, and two confounding variables Z_1 and Z_2 are generated from $BVN(0, I_2)$. Given X and (Z_1, Z_2), throughout the entire simulation experiments in this section, Q mediators M and outcome Y are generated according to the SEM (2.1), with $Q = 30$, $\gamma = -2$, $\boldsymbol{\eta} = (2, -3, 2)^\top$, $\sigma_Y^2 = 1$, and $\text{vec}(\boldsymbol{\zeta})$ consists of 18 repeated sequences of $(-2, 3, -3, 1, 1)$. The sample size n varies over 200, 500, and 1000. For each sample size, we run 10,000 replicates.

2.5.2 Type I Error

We consider the following four scenarios of the null hypotheses: (i) sparse pathways with no cancellation; (ii) sparse pathways with cancellation; (iii) non-sparse pathways with cancellation; and (iv) fully sparse pathways $\boldsymbol{\alpha} = \boldsymbol{\beta} = \mathbf{0}$. Here sparsity refers to the number of zero parameters in $\boldsymbol{\alpha}$ and/or $\boldsymbol{\beta}$. The detailed specifications of $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ can be found in Table 2.1. We report in Table 2.2 the estimated empirical type I error rate as the proportion of rejections from the 10,000 replicates. For the four null cases (i)-(iv), our LR test as well as two existing PT-N test and PT-NP test showed a proper control of the type I error. In the cases (i)-(iii), these three methods show their empirical type I error rates close to the nominal level 0.05, as desired. In the case (iv), they are all conservative, but our LR test appears to be the least conservative among the three.

2.5.3 Power Comparison

We evaluate and compare power under the same basic model specifications above, in which $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ are specified in four sets of alternative scenarios different from the null hypothesis; see the detail in Table 2.1. The design for the four alternative hypotheses corresponds the following scenarios of pathways: (v) both $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ are sparse; (vi) $\boldsymbol{\alpha}$ is sparse and $\boldsymbol{\beta}$ is not sparse; (vii) $\boldsymbol{\alpha}$ is not sparse and $\boldsymbol{\beta}$ is sparse; and (viii) both $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ are not sparse. Regardless of specific settings, the overall absolute group-level effect is fixed at 0.16, i.e. $|\boldsymbol{\alpha}^\top \boldsymbol{\beta}| = 0.16$. Table 2.2 reports the estimated empirical power by the proportion of rejections to the null from 10,000 replicates.

We calculate the percent of power increase of LR over a competing method by $\frac{\text{power of LR}}{\text{power of competitor}} - 1$. For all cases, our LR method demonstrates clearly higher power than existing PT-N and PT-NP tests, especially when the sample sizes are small or moderate, say 500 or less. It is also noteworthy that even though the mediation effect size is fixed constantly at 0.16 across four cases, the power varies according to

Table 2.1: Designed specifications for $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ for null and alternative hypotheses.

Mediator	Null Hypothesis ($\boldsymbol{\alpha}^\top\boldsymbol{\beta}=0$)								Alternative Hypothesis ($\lvert\boldsymbol{\alpha}^\top\boldsymbol{\beta}\rvert=0.16$)							
	i		ii		iii		iv		v		vi		vii		viii	
	α	β	α	β	α	β	α	β	α	β	α	β	α	β	α	β
1	0.2	0	0.2	0	0.2	-0.2	0	0	0.4	0.4	0	0.3	0.3	0	0.4	0.4
2	0.5	0	0.2	0.5	0.3	0.1	0	0	0	-0.8	0	0.3	0.3	0	0.2	-0.2
3	0	0.2	0.5	-0.2	0.1	0.1	0	0	0	0	0	0.3	0.3	0	0.3	0.1
4	0	0.5	0.2	0.5	0.2	-0.2	0	0	0	0	0	0.3	0.3	0	0.1	0.1
5	0	0	-0.2	0.5	0.3	0.1	0	0	0	0	0	0.3	0.3	0	0.2	-0.2
6	0	0	0	0	0.1	0.1	0	0	0	0	0.2	-0.8	-0.8	0.2	0.3	0.1
7	0	0	0	0	0.2	-0.2	0	0	0	0	0	0	0	0	0.1	0.1
8	0	0	0	0	0.3	0.1	0	0	0	0	0	0	0	0	0.2	-0.2
9	0	0	0	0	0.1	0.1	0	0	0	0	0	0	0	0	0.3	0.1
10	0	0	0	0	0.2	-0.2	0	0	0	0	0	0	0	0	0.1	0.1
11	0	0	0	0	0.3	0.1	0	0	0	0	0	0	0	0	0.2	-0.2
12	0	0	0	0	0.1	0.1	0	0	0	0	0	0	0	0	0.3	0.1
13	0	0	0	0	0.2	-0.2	0	0	0	0	0	0	0	0	0.1	0.1
14	0	0	0	0	0.3	0.1	0	0	0	0	0	0	0	0	0.2	-0.2
15	0	0	0	0	0.1	0.1	0	0	0	0	0	0	0	0	0.3	0.1
16	0	0	0	0	0.2	-0.2	0	0	0	0	0	0	0	0	0.1	0.1
17	0	0	0	0	0.3	0.1	0	0	0	0	0	0	0	0	0.2	-0.2
18	0	0	0	0	0.1	0.1	0	0	0	0	0	0	0	0	0.3	0.1
19	0	0	0	0	0.2	-0.2	0	0	0	0	0	0	0	0	0.1	0.1
20	0	0	0	0	0.3	0.1	0	0	0	0	0	0	0	0	0.2	-0.2
21	0	0	0	0	0.1	0.1	0	0	0	0	0	0	0	0	0.3	0.1
22	0	0	0	0	0.2	-0.2	0	0	0	0	0	0	0	0	0.1	0.1
23	0	0	0	0	0.3	0.1	0	0	0	0	0	0	0	0	0.2	-0.2
24	0	0	0	0	0.1	0.1	0	0	0	0	0	0	0	0	0.3	0.1
25	0	0	0	0	0.2	-0.2	0	0	0	0	0	0	0	0	0.1	0.1
26	0	0	0	0	0.3	0.1	0	0	0	0	0	0	0	0	0.2	-0.2
27	0	0	0	0	0.1	0.1	0	0	0	0	0	0	0	0	0.3	0.1
28	0	0	0	0	0.2	-0.2	0	0	0	0	0	0	0	0	0.1	0.1
29	0	0	0	0	0.3	0.1	0	0	0	0	0	0	0	0	0.2	-0.3
30	0	0	0	0	0.1	0.1	0	0	0	0	0	0	0	0	0.3	0.2

Table 2.2: Empirical type I error under four null hypotheses, and power under four alternative hypotheses with 10,000 replicates. The sample size varies from 200, 500, and 1,000. The exchangeable correlation of mediators is set with correlation 0.5. Power increase (%) $=\dfrac{\text{power of LR test}}{\text{power of competing test}}-1$.

n	Method	Null Hypothesis				Alternative Hypothesis				Percent of power increase			
		i	ii	iii	iv	v	vi	vii	viii	v	vi	vii	viii
200	LR	0.050	0.052	0.051	0.009	0.591	0.562	0.312	0.507	-	-	-	-
	PT-N	0.038	0.042	0.037	0.005	0.550	0.536	0.255	0.458	7.46%	4.78%	22.51%	10.72%
	PT-NP	0.032	0.038	0.028	0.001	0.507	0.497	0.242	0.426	16.55%	13.07%	28.72%	19.20%
500	LR	0.046	0.049	0.045	0.007	0.970	0.954	0.648	0.928	-	-	-	-
	PT-N	0.041	0.045	0.039	0.005	0.967	0.953	0.624	0.922	0.30%	0.12%	3.78%	0.66%
	PT-NP	0.038	0.044	0.035	0.001	0.962	0.947	0.620	0.916	0.78%	0.75%	4.5%	1.36%
1000	LR	0.051	0.049	0.046	0.006	1.000	1.000	0.917	0.999	-	-	-	-
	PT-N	0.048	0.048	0.043	0.004	1.000	1.000	0.911	0.999	0.00%	0.00%	0.67%	0.01%
	PT-NP	0.046	0.049	0.041	0.000	1.000	1.000	0.910	0.998	0.00%	0.00%	0.71%	0.05%

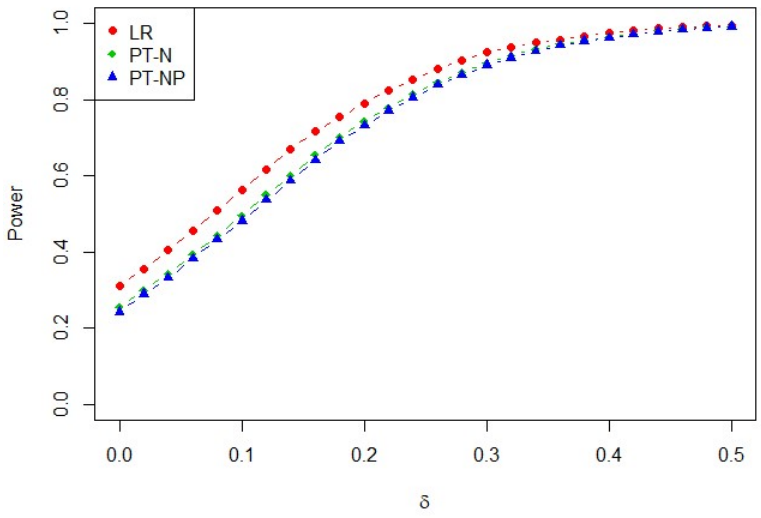

Figure 2.2: Power curves of three tests (LR, PT-N, and PT-NP) under the simulation case (vii) with sample size $n = 200$ and δ varing from 0 to 0.50 by an increment unit of 0.02.

the underlying parameter configurations and sparsity. Among these four cases, case (vii) appears to be the most challenging scenario, where $\boldsymbol{\beta}$ is most sparse with a small magnitude of nonzero element in $\boldsymbol{\beta}$. To further examine the performance of these tests, in case (vii) with the sample size 200, we set the single nonzero $\boldsymbol{\beta}$ coefficient at $0.2 + \delta$ with δ varying from 0 to 0.50 by an increment of 0.02 to illustrate the power increase pattern. Figure 2.2 shows all three power curves increase to 1 when the size δ in the alternative hypothesis becomes further distant from the null hypothesis. Our LR test is more powerful than the other competing tests. Empirically, these three tests are all shown to be consistent as their power rises to 1 when the deviation from the null tends to infinity. Taking all other settings in the simulation study into account, overall, we conclude that our LR test shows higher power in comparison with two existing PT-N and PT-NP tests, especially in the cases of small or moderate sample sizes.

2.6 Data Application

We apply the proposed LR test to analyze a real world data example from a pediatric cohort study consisting of 203 children, with 96 boys and 107 girls, age 8.1 to 14.4 years old. We consider two exposure variables X of macronutrient intakes calculated as the energy adjusted carbohydrate and fat. They are termed as carbohydrate intake and fat intake, respectively, obtained from the food frequency questionnaires (*Willett et al.*, 1997). The outcome variable Y is a HOMA-CP score defined by *Li et al.* (2004), which measures insulin resistance using the C-peptide biomarker produced by pancreas. A higher HOMA-CP score means more insulin resistant, leading potentially to a higher risk of developing diabetes in adulthood years.

In this analysis, we focus on studying a cluster of seven metabolites of glucose metabolites and acetylamino acids annotated by our collaborator Dr. Labarre (*LaBarre et al.*, 2020) at the University of Michigan Research Core of Metabolomics. One metabolite in this cluster is `N-acetylglycine`, which have been found in the literature to be positively associated with dietary fiber intake (*Lustgarten et al.*, 2014) and negatively associated with metabolic risk score (*Perng et al.*, 2017). The goal of central interest is to test if a cohesive cluster containing `N-acetylglycine` is involved as a group in a mediation pathway from dietary intakes to HOMA-CP score. This scientific question pertains to a hypothesis that food intakes may change metabolites and then further alter function of pancreas, so to elevate the risk of developing diabetes during later life time.

With the consultation with our collaborator, we choose a set of confounding variables, including age, gender, and puberty onset. We begin the data analysis to assess the total effect by fitting a linear model with outcome HOMA-CP on exposures to carbohydrate intake and fat intake, respectively, as well as the confounders. The estimated total effects are -0.015 with a p-value of 0.174 for fat intake and 0.004 with a p-value of 0.252 for carbohydrate intake. In the following analysis, we focus on

the pathway mediation analysis from exposures to outcome.We then calculate the p-values for the null hypothesis $H_0 : \boldsymbol{\alpha}^T\boldsymbol{\beta} = \mathbf{0}$ with $Q = 7$ using three methods of LR, PT-N and PT-NP. We perform the testing for the group-level mediation effect with exposure of fat intake, and obtain p-values equal to 0.01 (LR), 0.02 (PT-N), and 0.02 (PT-NP). Likewise, with exposure of carbohydrate intake, we obtain p-values 0.03 (LR), 0.04 (PT-N) and 0.04 (PT-N). All three methods reach an agreement that with 95% confidence this cluster of seven metabolites exhibits a significant group-level mediation effect on the associations between dietary intakes and HOMA-CP score. With no surprise, the LR test appears to have smaller p-values in both cases, being consistent with the findings in the simulation studies.

Taking a closer look at individually each of the seven metabolites in the cluster, we report in Table 2.3 estimates of the individual model parameters in $\boldsymbol{\alpha}$, $\boldsymbol{\beta}$ and $\boldsymbol{\alpha} \circ \boldsymbol{\beta}$, where $\circ$ is the element-wise product. The group-level mediation effects of fat and carbohydrate intakes through the seven metabolites are -0.012 and 0.003, respectively. For fat intakes, the negative mediation effect indicates that more fat intakes help reduce the insulin resistance through metabolites, where `N-acetyglicine` contributes most to the reduction of the insulin resistance score. In contrast, carbohydrate intakes increase the insulin resistance through metabolites, where again `N-acetyglicine` contributes most.

Table 2.3: Estimated coefficients for a cluster of seven metabolites.

Metabolite	Fat			Carbohydrate		
	$\boldsymbol{\alpha}$	$\boldsymbol{\beta}$	$\boldsymbol{\alpha} \circ \boldsymbol{\beta}$	$\boldsymbol{\alpha}$	$\boldsymbol{\beta}$	$\boldsymbol{\alpha} \circ \boldsymbol{\beta}$
L-histidine	-0.0019	0.334	-0.0006	0.0008	0.334	0.0003
N-acetyl-D-glucosamine	-0.0046	0.197	-0.0009	0.0009	0.200	0.0002
N-acetyl-DL-serine	0.0055	0.206	0.0011	-0.0017	0.204	-0.0004
3,4-hydroxyphenyl-lactate	0.0014	0.114	0.0002	-0.0006	0.114	-0.0001
2-deoxy-D-glucose	0.0041	-0.356	-0.0015	-0.0013	-0.356	0.0005
N-acetylglycine	0.0101	-0.840	-0.0085	-0.0030	-0.842	0.0025
D-lyxose	-0.0050	0.291	-0.0015	0.0016	0.294	0.0005

2.7 Concluding Remarks

This chapter studied a likelihood ratio approach to testing a group-level mediation effect with multiple mediators. We were able to overcome a key technical challenge arising from the constrained maximum likelihood estimation under irregular parameter spaces. In particular, the Lagrange Multiplier method was developed to carry out the constrained optimization via an efficient block coordinate decent algorithm, which was required to implement our LR test statistic. We established the asymptotic distributions of the proposed LR test statistic, in which a theoretical guarantee was given for a proper control of the type I error. Through both simulation studies and a data application, our LR method has showed less conservative and higher power than two existing methods, PT-N test and PT-NP test, especially when the sample size is moderate or small.

To apply our LR approach to testing for a cluster of high-dimensional potential mediators, one needs to first divide them into subgroups according to prior scientific knowledge or certain clustering techniques, and then carry out the test for a group-level mediation effect, each for one subgroup of mediator. A future work of interest would be to extend the current framework to the case of high-dimensional mediators with no need of dividing them into subgroups.

All test methods, including our LR test, have appeared to be conservative for the null case of $\boldsymbol{\alpha} = \boldsymbol{\beta} = \mathbf{0}$. This is an open problem in the theory of statistical inference for mediation effect, even in the setting of one single mediator. Some better solutions to overcome such conservatism are worth future exploration. In addition, extending the normal linear structural equation model to a more general model such as the family of generalized linear models and Cox proportional hazards model is appealing to deal with a broader range of data types and practical problems. Developing LR tests for group-level mediation effects beyond the linear structural equation model is an important research direction of great interest.

CHAPTER III

Generalized Structural Equation Models for Mediation Analysis with Data of Mixed Types

3.1 Introduction

In many biomedical studies, mediation analysis has received considerable attention as the choice of method to investigate a hypothesized mechanistic causal pathway that involves two or more potential causal factors. The mechanistic pathway of interest is typically organized as a form of direct acyclic graph (DAG) shown in Figure 3.1 (a), which enables us to understand how the effect of exposure X on outcome Y may be mediated through some intermediate variable M, called *mediator*. In addition to the direct causal path from exposure X to outcome Y denoted by $(X \to Y)$, there exists another causal path where exposure X causes mediator M denoted by $(X \to M)$, and then M affects outcome Y denoted by $(M \to Y)$. As a result, the mediation pathway has a route $X \to M \to Y$. These causal effects may be parameterized by α, β and γ in the DAG shown in Figure 3.1(a), corresponding to, respectively, the exposure-mediator, mediator-outcome and exposure-outcome relationships, and these directed edges or associations may be adjusted for a set of confounders $\mathbf{W}$ shown in Figure 3.1(b). The primary objective of a mediation analysis is twofold: (i) to evaluate the effect of exposure on outcome when the mediator is held constant, $X \to Y$, termed

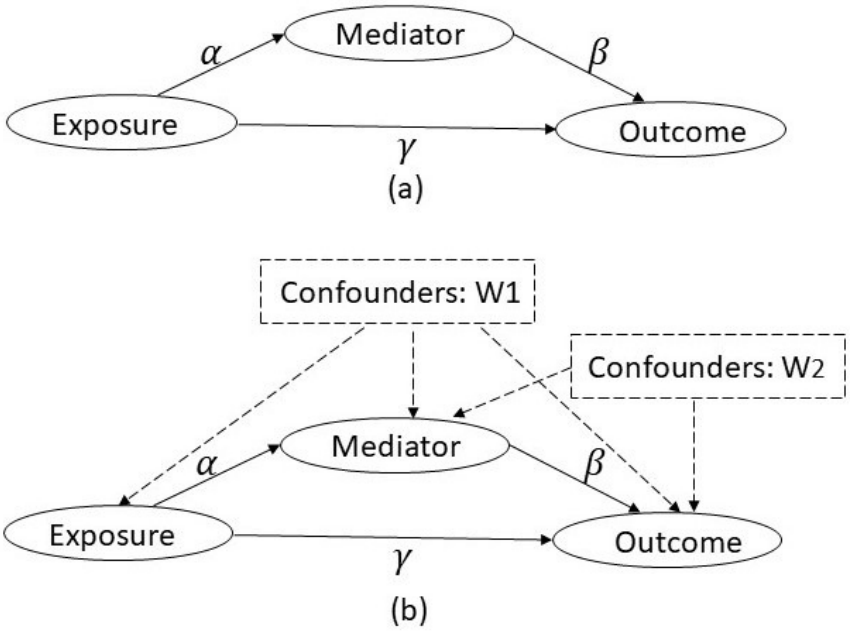

Figure 3.1: A DAG involving exposure, mediator and outcome: (a) randomization without confounders; (b) non-experimental study

as *direct effect* (DE); and (ii) to evaluate the effect of exposure on outcome through the mediator, $X \to M \to Y$, termed as *mediation effect or indirect effect* (IE).

Two types of modeling approaches are widely utilized in mediation analyses. The first type refers to the classical approach originally developed by *Baron and Kenny* (1986) through a system of linear regression models, the so-called structural equation modeling (SEM), which has been regarded as a standard methodology routinely used in practice. The second type emerges recently as a causal inference approach, which is formulated within the potential or counterfactual outcome framework (*Robins and Greenland*, 1992; *Pearl*, 2001). The latter has attracted considerable attention in the recent literature because of its role in making causal inference with two potential causal factors, an important extension from the causal inference literature with a signal causal factor. In the counterfactual framework, both DE and IE are later generalized to the so-called natural direct effect (NDE) and natural indirect effect

(NIE), in which some essential assumptions required to identify these natural causal effects have been extensively discussed in the literature (*Imai et al.*, 2010b; *Coffman and Zhong*, 2012; *Preacher*, 2015). Unfortunately, these identifiability assumptions cannot be checked in practical studies, imposing a great challenge in the use of the second type of mediation analysis approach. For example, "sequential ignorability" (*Imai et al.*, 2010b) is one of such assumptions, which requires the absence of unmeasured confounding in the DAG. Although it is hard to justify this condition, some sensitivity analyses may be performed to assess reproducibility of results by carefully designed variations of confounding scenarios.

A noticeable analytic limitation in contrast to the popularity of the mediation analysis methodology lies in the fact that mediator and/or outcome are often assumed to be continuous and normally distributed in the current literature. As seen in our motivating example in Section 3.2, mediator and/or outcome can be continuous, categorical or count variable. There are some scattered works concerning *ad hoc* cases, such as categorical mediator and/or outcome, or continuous but nonnormal mediator and/or outcome (*VanderWeele and Vansteelandt*, 2014, 2010; *Albert and Nelson*, 2011; *Huang et al.*, 2004; *Tingley et al.*, 2014). Usually, stronger model assumptions are imposed when data are of mixed types. Moreover, as becoming evident throughout this chapter, causal interpretations appear to be model-specific. When the joint normality is absent, there lacks of a unified framework of mediation analysis to handle data of mixed types, including modeling approach, statistical estimation and inference, software and interpretations. This motivates us to develop a flexible class of statistical models and analytics suitable for mediation analysis with data of mixed types. We propose to invoke copula dependence models to accommodate a broad range of data types.

In a similar spirit to the copula regression with data of mixed types developed by *Song et al.* (2009), we propose a class of generalized structural equation mod-

els (GSEMs) through Gaussian copulas to perform mediation analyses in that data may be of mixed types. This class of GSEMs provides three new methodological advantages: first, we develop a unified approach to analyzing exposure, mediator and outcome that may have different types, either categorical, discrete or continuous. Adopting the tool of copula dependence models, GSEMs gives rise to a general framework along with the generalized linear models (GLMs). Second, we jointly model exposure, mediator and outcome, and derive the joint likelihood function under different data scenarios, which is more flexible and rigorous than the approaches based only on the mean model for $\mathrm{E}(Y \mid M, X, \mathbf{W})$. Third, utilizing the hierarchical model specification, we develop a two-stage estimation procedure, where we handle the confounding factors in the first stage, allowing us to properly adjust for complex confounding scenarios without complicating the estimation of directed associations or causal effects in the second stage.

The remainder of the chapter is organized as follows: Section 3.2 introduces the motivating ELEMENT study. Section 3.3 describes GSEMs specified by a hierarchical copula mediation model approaches. Section 3.4 discusses the estimations of parameter and causal effects under eight scenarios of mixed data types, where exposure, mediator and outcome are continuous or discrete. Section 3.5 presents three examples to illustrate the proposed GSEM methodology. Section 3.6 reports extensive simulations to assess the performance of the GSEM methodology under three settings. Section 3.7 demonstrates an application of mediation analysis in the ELEMENT study. Section 3.8 concludes the chapter with discussions on the advantages and limitations of the proposed methodology. Detailed technical derivations are included in Appendix B.

3.2 ELEMENT Study

The data motivating our new methodology development comes from the "Early Life Exposures in Mexico to ENvironmental Toxicants" (ELEMENT) cohort study. ELEMENT recruited three birth cohorts of 1,643 mother-child pairs during pregnancy or delivery from maternity hospitals in Mexico city between 1993 and 2004 (*Hu et al.*, 2006; *Perng et al.*, 2019). The main objective is to study how environmental toxicants, such as phthalates, affect maternal and child health outcomes. Phthalates are a group of chemicals mostly used in plastics, and high levels of these chemicals during pregnancy have shown substantial adverse health effects on both mother and child (*Qian et al.*, 2020). Previous study has demonstrated third trimester maternal phthalate exposures of MEHHP, MEOHP and MIBP are linked to delayed infancy Body Mass Index (BMI) peak (*Zhou et al.*, 2021). Another literature has found that later age at infancy BMI peak has been associated with higher cardiometabolic risk biomarkers measured from children during peripuberty (*Perng et al.*, 2018). It is of great interest to understand potential causal mediation pathways among prenatal phthalate exposure, child' growth in early life, and child's cardiometabolic risk later in life. Thus, we hypothesize that the association between phthalate exposures during mothers' second and third trimesters and health outcomes of children in peripuberty may be mediated by the timing and tempo of children reaching their infancy BMI peak (labeled as on time or delayed).

The motivating data consists of 205 mother-child pairs, where prenatal exposures were measured between 1997 and 2005, while the child's cardiometabolic risk biomarkers were measured between 2008 to 2012. There are 97 boys and 108 girls. The mean ages of mothers at birth, and children at measurement during peripuberty are 27.7 yrs and 10.1 yrs, respectively.

The potential mediator, the timing of infancy BMI peak is given by estimated age (in month) when a child reaches his/her BMI peak between birth to 36 months.

Infancy BMI peak is an important milestone for the early-life growth, as well as an indicator for later obesity development for children (*Jensen et al.*, 2015; *Börnhorst et al.*, 2017). The estimated age is calculated from child's growth trajectory obtained by fitting eight serial anthropometry measurements in the Newton's Growth Models (*Baek et al.*, 2019). If the BMI peak was not reached before 36 months, it was categorized as delayed ($M = 1$), otherwise as on time ($M = 0$). There are 27.8% of children who were delayed in reaching their infancy BMI peaks.

The exposure variables include six urinary concentrations of phthalate metabolites (MEHHP, MEOHP, and MIBP) measured at the second trimester among 177 women and third trimester among 202 women. These three phthalate exposures at the third trimester were previously found to be positively associated with delayed infancy BMI peak (*Zhou et al.*, 2021). The outcome variables are three cardiometabolic biomarkers measured from children during peripuberty, including fasting glucose z-score, C-peptide z-score and a metabolic syndrome risk z-score (MetS z-score). Fasting glucose is a screening tool for diabetes. C-peptide is a substance released from pancreas when producing insulin. MetS z-score was calculated as the average of five z-scores for waist circumference, fasting glucose, C-peptide, ratio of triglycerides and High-density lipoprotein cholesterol, and average of systolic and diastolic blood pressures (*Perng et al.*, 2018).

Confounders $\mathbf{W}_1$ including mom's age at birth, education, and marital status are factors that occur prior to exposures, and they would influence exposures to toxicants, mediator and outcomes. Confounders $\mathbf{W}_2$ including breastfeeding duration, parity, child's gender, gestational age, and birth weight are factors that occur after exposures to toxicants, and would influence mediator and outcomes. Note that $\mathbf{W}_2$ are assumed to be not affected by exposures, a key assumption for identification of natural causal effects. Figure 3.2 describes hypothetical associations among exposures to phthalates (X), BMI infancy peak M, child's health outcomes Y as well as confounders $\mathbf{W}_1, \mathbf{W}_2$.

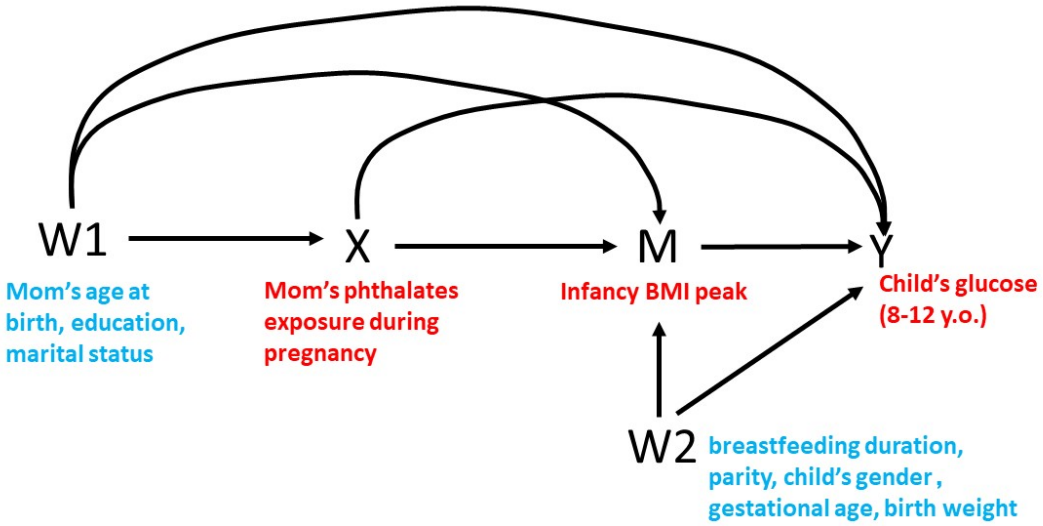

Figure 3.2: Association between phthalate exposures and health outcomes mediated by whether child's timing of reaching BMI peak is delayed or not.

3.3 Generalized Structural Equation Models

3.3.1 Framework

We begin with a discussion on our general modeling approach to constructing a joint distribution for three variables (X, M, Y) of mixed types under the graphic topology of acyclic direct graph (DAG) shown in Figure 3.1(a). Such joint distribution is specified by a hierarchical modeling approach, which enables us to define a class of generalized structural equations models (GSEMs) to address practical needs in mediation analyses with data of mixed types. The key feature in the classical linear normal SEM is that the covariance matrix of (X, M, Y), denoted by, Γ, takes a special

form arising from a DAG, given as follow:

$$\Gamma = (\mathbf{I} - \Theta)^{-1} \, \mathbf{\Sigma} \, (\mathbf{I} - \Theta)^{-\top}, \tag{3.1}$$

where $\mathbf{I}$ is an identity matrix, $\mathbf{\Sigma} = \text{diag}(\sigma_x^2, \sigma_m^2, \sigma_y^2)$ is a diagonal matrix of marginal variances, each for one variable. More specifically, similar to the classical SEM to ensure the acyclicity of the DAG, Θ is specified as a lower triangular matrix of the form:

$$\Theta = \begin{pmatrix} 0 & 0 & 0 \\ \alpha & 0 & 0 \\ \gamma & \beta & 0 \end{pmatrix}. \tag{3.2}$$

To incorporate the specification (3.1) in the formulation of a valid joint probability distribution for X, M, Y of mixed data types, we need to generalize the classical linear normal SEM to cases of nonlinear non-normal SEM, referred to as generalized structural equation models (GSEMs) in this chapter. We propose to construct a GSEM with the following modeling elements:

(i) The constructed joint distribution has to be so flexible that it can allow different types of marginal distributions because variables X, M, Y may be continuous, discrete or categorical in practical studies.

(ii) The proposed joint distribution accommodates an explicit form of the covariance or correlation matrix of a given form similar to that in (3.1) and (3.2). Such matrix form is deemed critical importance to model the topology of a DAG and to interpret parameters pertinent to mediation pathways.

(iii) Confounding factors do not enter the hierarchy of covariance matrix as shown in (3.1) and (3.2); rather, they are adjusted at the hierarchy of marginal distributions of individual variables, respectively. This consideration essentially suggests a hierarchical model specification where the marginal parameters and

dependence parameters are handled by different hierarchies of GSEMs.

(iv) The classical linear normal SEM is a special case of the proposed generalization.

These modeling characteristics above, as shown in the chapter, can be incorporated simultaneously in the proposed joint distribution by the copula modeling approach; see for example, *Song et al.* (2009) where a class of vector GLMs (VGLMs) is constructed with data of mixed of types. Here, we consider a more complicated scenario than VGLMs in which the dependence matrix takes a restricted form (3.1) and confounding factors are only allowed to enter the model in the hierarchies of marginal distributions. The Fréchet's theory of constructing multivariate distribution by bivariate distributions is the theoretical basis to ensure the validity of our proposed hierarchical models (*Joe*, 2014).

Copula models provide a natural hierarchical modeling framework to satisfy the above modeling requirements. According to Sklar's Theorem (*Sklar*, 1959), when (X, M, Y) are continuous random variables, their joint distribution can be expressed as

$$F(x, m, y) = C\{F_x(x; \cdot), F_m(m; \cdot), F_y(y; \cdot); \Gamma\}. \tag{3.3}$$

where $F_j(\cdot)$ and $f_j(\cdot)$ are the cumulative distribution function and density function, respectively, and $C(\cdot)$ is a suitable copula function that is independent of marginal parameters. This expression provides a hierarchical modeling framework, as desired. To incorporate the covariance structure given in (3.1) in the copula function, in this chapter we choose the Gaussian copula (*Song*, 2000) given as follows:

$$C(u_1, u_2, u_3) = \Phi_3\{\Phi^{-1}(u_1), \Phi^{-1}(u_2), \Phi^{-1}(u_3); \Gamma\}, \tag{3.4}$$

where $\Phi_3(\cdot; \Gamma)$ is a trivariate Gaussian distribution function with mean zero and correlation matrix Γ, and $\Phi(\cdot)$ is the standard univariate Gaussian distribution function and $\Phi^{-1}(\cdot)$ is the corresponding quantile function. Since all marginal parameters, in-

clude the variance parameters, are exclusively included in the marginal distributions, the dependence matrix Γ does not contain any marginal variance parameters but only correlation parameters. All parameters in Γ represent rank-based correlations, more general dependencies than Pearson correlations.

Following *Song et al.* (2009), we propose to extend the copula model (3.3) with a Gaussian copula by allowing distributions of both continuous and discrete random variables and Γ satisfying DAG model (3.1). In this way, the resulting joint distribution can be used to handle variables of mixed types, and to perform mediation analyses. We will focus on our discussion in two important settings: GSEMs with no confounders in Section 3.3.2 under double randomized trials, and GSEMs in observational studies in Section 3.3.3 in the presence of confounding factors.

3.3.2 GSEMs under No Confounders

We first consider the most favorable scenario: there are no confounders involved in the relationships among exposure, mediator and outcome. One such case can be achieved by randomization of both X and M (*Preacher*, 2015). This is the simplest setting in the mediation analysis for us to introduce the copula model. We consider three generalized linear models (GLM) that the marginal distributions of X, M and Y are elements of exponential dispersion family distribution (ED) models (*Jørgensen*, 1987) denoted by

$$X \sim ED(\mu_x, \phi_x), M \sim ED(\mu_m, \phi_m), Y \sim ED(\mu_y, \phi_y), \tag{3.5}$$

where (μ_x, μ_m, μ_y),and (ϕ_x, ϕ_m, ϕ_y) are the respective mean parameters and dispersion parameters. Note that these mean parameters are not modeled by covariates. Denote by $F_x(x;\cdot)$, $F_m(m;\cdot)$ and $F_y(y;\cdot)$ the cumulative distribution functions of X, M and Y, respectively. As a result, we are able to obtain a joint distribution of (X, M, Y)

via a copula function $C(u_1, u_2, u_3; \Gamma)$ in (3.3). It can be shown that (3.3) under assumptions (3.1), (3.2) and (3.4) is equivalent to the Gaussian copula model (3.6) based on latent variable representation. Let Z_x, Z_m and Z_y be three respective latent variable. Then (3.1)-(3.4) may be equivalently represented by,

$$\begin{aligned} Z_x &\sim N(0,1), \\ Z_m \mid Z_x &\sim N(\alpha Z_x, 1), \\ Z_y \mid Z_x, Z_m &\sim N(\gamma Z_x + \beta Z_m, 1). \end{aligned} \tag{3.6}$$

We can link observed variables with the latent variables via marginal quantile transformations: $X = F_x^{-1}\{\Phi(Z_x)\}$, $M = F_m^{-1}\{\Phi(Z_m^*)\}$ and $Y = F_y^{-1}\{\Phi(Z_y^*)\}$, where $Z_m^* = Z_m/\tau_m$, $Z_y^* = Z_y/\tau_y$, $\tau_m = \sqrt{\alpha^2+1}$, and $\tau_y = \sqrt{(\gamma+\alpha\beta)^2+\beta^2+1}$. The marginal distributions of Z_x, Z_m^* and Z_y^* all follow standard normal distribution. It is easy to see that the joint distribution of Z_x, Z_m^* and Z_y^* is trivariate normal given as follows. The detailed derivation of (3.7) can be found in Appendix B.1,

$$\begin{pmatrix} Z_x \\ Z_m^* \\ Z_y^* \end{pmatrix} \sim N\left\{ \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & \frac{\alpha}{\tau_m} & \frac{\eta}{\tau_y} \\ \frac{\alpha}{\tau_m} & 1 & \frac{\alpha\eta+\beta}{\tau_m\tau_y} \\ \frac{\eta}{\tau_y} & \frac{\alpha\eta+\beta}{\tau_m\tau_y} & 1 \end{pmatrix} \right\}. \tag{3.7}$$

If X is a continuous variable, $X = F_x^{-1}\{\Phi(Z_x)\}$ becomes $Z_x = \Phi^{-1}(F_x(X))$, namely the quantile of the standard normal distribution. If X is a discrete variable with non-zero point probability mass at $0, 1, 2, \cdots$, then $X = \sum_{x=0}^{\infty} xI(F_x(x-1) \leq \Phi(Z_x) < F_x(x))$. This implies that the event $\{X = x\}$ is equivalently to the event $\{\Phi^{-1}(F_x(x-1)) \leq Z_x < \Phi^{-1}(F_x(x))\}$. For simplicity, denote the lower and upper bounds of the normal quantiles as $\Phi^{-1}(F_x(x-1)) = l_x$, and $\Phi^{-1}(F_x(x)) = u_x$ in the rest of this chapter. Similar results hold for latent variable Z_m^* with respect to mediator M, with lower and upper quantile bounds denoted by l_m and u_m, Z_y^* with

respect to outcome Y, with lower and upper quantile bounds denoted by l_y and u_y. X, M and Y may be either continuous or discrete, we have a total of eight data type scenarios. For example, "CCC", "CDC" and "CCD" represent X, M, and Y are continuous; X and Y are continuous, M is discrete; X and M are continuous, Y is discrete, respectively. Table 3.1 presents the expressions of the corresponding likelihood functions, each for one scenario. The detailed derivation of $\pi(X, M, Y)$ is summarized in Appendix B.2.

Table 3.1: Joint distribution of X, M and Y, $\pi(X, M, Y)$ and expectation of potential outcome $\mathrm{E}\{Y(X_a, M(X_b))\}$ under eight scenarios, where X, M and Y are either continuous or discrete.

X	M	Y	$\pi(X, M, Y)$
C	C	C	$\pi(Z_x, Z_m^*, Z_y^*)$
C	C	D	$\pi(Z_x, Z_m^*)P\Big(Z_y^* \in (l_y, u_y)\vert Z_x, Z_m^*\Big)$
C	D	C	$\pi(Z_x, Z_y^*)P\Big(Z_m^* \in (l_m, u_m)\vert Z_x, Z_y^*\Big)$
C	D	D	$\pi(Z_x)P\Big(Z_m^* \in (l_m, u_m), Z_y^* \in (l_y, u_y)\vert Z_x\Big)$
D	C	C	$\pi(Z_m^*, Z_y^*)P\Big(Z_x \in (l_x, u_x)\vert Z_m^*, Z_y^*\Big)$
D	C	D	$\pi(Z_m^*)P\Big(Z_x \in (l_x, u_x), Z_y^* \in (l_y, u_y)\vert Z_m^*\Big)$
D	D	C	$\pi(Z_y^*)P\Big(Z_x \in (l_x, u_x), Z_m^* \in (l_m, u_m)\vert Z_y^*\Big)$
D	D	D	$P\Big(Z_x \in (l_x, u_x), Z_m^* \in (l_m, u_m), Z_y^* \in (l_y, u_y)\Big)$
X	M	Y	$\mathrm{E}\{Y(X_a, M(X_b))\}$
C	C	C	$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} F_y^{-1}(\Phi(Z_y^*))\pi(Z_y^*\vert Z_x = z_{x_a}, Z_m^* = z_m^*)\pi(Z_m^*\vert Z_x = z_{x_b})dZ_y^*dZ_m^*$
C	C	D	$\int_{-\infty}^{\infty}\sum_{y=1}^{\infty} yP(l_y \le Z_y^* < u_y\vert Z_x = z_{x_a}, Z_m^* = z_m^*)\pi(Z_m^*\vert Z_x = z_{x_b})dZ_m^*$
C	D	C	$\sum_{m=0}^{\infty} P(l_m \le Z_m^* < u_m\vert Z_x = z_{x_b})\int_{-\infty}^{\infty} F_y^{-1}(\Phi(Z_y^*))\frac{\int_{l_m}^{u_m}\pi(Z_y^*, Z_m^*\vert Z_x = z_{x_a})dZ_m^*}{P(l_m \le Z_m^* < u_m\vert Z_x = z_{x_a})}dZ_y^*$
C	D	D	$\sum_{m=0}^{\infty}\sum_{y=0}^{\infty} yP(l_y \le Z_y^* < u_y\vert Z_x = z_{x_a}, l_m \le Z_m^* < u_m)P(l_m \le Z_m^* < u_m\vert Z_x = z_{x_b})$
D	C	C	$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} F_y^{-1}(\Phi(Z_y^*))\pi(Z_y^*\vert l_{x_a} \le Z_x < u_{x_a}, Z_m = z_m)dZ_y\pi(Z_m\vert l_{x_b} \le Z_x < u_{x_b})dZ_m$
D	C	D	$\int_{-\infty}^{\infty}\sum_{y=0}^{\infty} yP(l_y \le Z_y^* < u_y\vert l_{x_a} \le Z_x < u_{x_a}, Z_m = z_m)\pi(Z_m\vert l_{x_b} \le Z_x < u_{x_b})dZ_m$
D	D	C	$\sum_{m=0}^{\infty}\int_{-\infty}^{\infty} F_y^{-1}(\Phi(Z_y^*))\pi(Z_y^*\vert l_{x_a} \le Z_x < u_{x_a}, l_m \le Z_m^* < u_m)dZ_y^*P(l_m \le Z_m^* < u_m\vert l_{x_b} \le Z_x < u_{x_b})$
D	D	D	$\sum_{m=0}^{\infty}\sum_{y=0}^{\infty} yP(l_y \le Z_y^* < u_y\vert l_{x_a} \le Z_x < u_{x_a}, l_m \le Z_m^* < u_m)P(l_m \le Z_m^* < u_m\vert l_{x_b} \le Z_x < u_{x_b})$

C: Continuous variable. D: Discrete variable.

Now let us study causal effects under the above model (3.1)-(3.4). To analyze natural direct effect (NDE) and natural indirect effect (NIE), we adopt the potential outcome framework from the causal inference literature (*Splawa-Neyman et al.*, 1990; *Rosenbaum and Rubin*, 1983; *Pearl*, 2001; *Robins and Greenland*, 1992). Following existing the counterfactual notions (*VanderWeele and Vansteelandt*, 2009), let $Y(x, m)$ denote the counterfactual outcome that would have been observed for a

subject had the exposure X been set to value x and mediator M to the value m; let $M(x)$ be the counterfactual value of mediator had the exposure X been set to x. Then $\mathrm{E}\{Y(x_a, M(x_b))\}$ is the expected outcome of Y had the exposure been set to x_a and mediator been set to $M(x_b)$, namely

$$\mathrm{E}\{Y(x_a, M(x_b))\} = \mathrm{E}_M[\mathrm{E}_Y\{Y|M, X = x_a\}|X = x_b].$$

Both NDE and NIE for a change of X from x_0 to x_1 are given by, respectively

$$\begin{aligned} \mathrm{NDE}(x_0, x_1) &= \mathrm{E}\{Y(x_1, M(x_0))\} - \mathrm{E}\{Y(x_0, M(x_0))\}, \\ \mathrm{NIE}(x_0, x_1) &= \mathrm{E}\{Y(x_1, M(x_1))\} - \mathrm{E}\{Y(x_1, M(x_0))\}. \end{aligned} \tag{3.8}$$

These are primarily determined by the hierarchies of causal effects dependencies among X, M and Y.

3.3.3 GSEMs in Observational Studies

Next we consider the setting of a non-experimental design, such as an observational studies, where confounding factors are used to adjust for selection bias. Let $\mathbf{W} = (\mathbf{W}_1^\top, \mathbf{W}_2^\top)^\top$ denote a matrix of all the confounding factors available in the dataset, where $\mathbf{W}_1 = (W_{1,1}, \ldots, W_{1,p_1})^\top$ influences X, and $\mathbf{W}$ influences M and Y. For simplicity, the first element of $\mathbf{W}$ is set to 1 for the intercept. For the identification of NDE and NIE, we impose the same assumptions discussed in the literature (*VanderWeele and Vansteelandt*, 2009; *VanderWeele*, 2015): (i) there are no unmeasured confounders of the associations between X and M, between M and Y, and between X and Y, respectively; (ii) no confounders of M and Y that are influenced by X. Then the hierarchy of marginal models, adjusting for confounders, is specified

by a set of three conditional density functions of X, M and Y given $\mathbf{W}_1$ and $\mathbf{W}$:

$$\begin{aligned} X \mid \mathbf{W}_1 &\sim ED(\mu_x, \phi_x), & g_x(\mu_x) &= \mathbf{W}_1\boldsymbol{\beta}_x, \\ M \mid \mathbf{W} &\sim ED(\mu_m, \phi_m), & g_m(\mu_m) &= \mathbf{W}\boldsymbol{\beta}_m, \\ Y \mid \mathbf{W} &\sim ED(\mu_y, \phi_y), & g_y(\mu_y) &= \mathbf{W}\boldsymbol{\beta}_y, \end{aligned} \tag{3.9}$$

where $\boldsymbol{\beta}_x, \boldsymbol{\beta}_m$ and $\boldsymbol{\beta}_y$ are vectors of regression coefficients; and $\phi(x)$, $\phi(m)$ and $\phi(y)$ are the dispersion parameters. In a similar way, we can establish the latent variable representation in the mediation analysis, as done for model (3.6).

When confounders are present, both NDE and NIE are given similarly as (3.8), with the inclusion of confounders in the marginal mean parameters, precisely, we would consider conditional NDE and NIE given $\mathbf{W} = \mathbf{w}$ as follows,

$$\begin{aligned} \text{NDE}(x_0, x_1; \mathbf{w}) &= \text{E}\{Y(x_1, M(x_0)) \mid \mathbf{W} = \mathbf{w}\} - \text{E}\{Y(x_0, M(x_0)) \mid \mathbf{W} = \mathbf{w}\}, \\ \text{NIE}(x_0, x_1; \mathbf{w}) &= \text{E}\{Y(x_1, M(x_1)) \mid \mathbf{W} = \mathbf{w}\} - \text{E}\{Y(x_1, M(x_0)) \mid \mathbf{W} = \mathbf{w}\}. \end{aligned} \tag{3.10}$$

This is natural as DAG is specified in the hierarchy adjusted by confounders in the marginal distributions. Compared with the conventional structural equation modeling approach to mediation analysis, one distinction for the GSEM methodology is that instead of conditioning on the exposure variable X, we propose to hierarchically model the marginal distribution of X and DAG for causal effects, where covariates $\mathbf{W}_1$ are used to adjust nodes not edges of DAG in Figure 3.1(b). One advantage of modeling X: scientifically, it is often the case that covariates $\mathbf{W}_1$ affect exposure X, M and Y, so the change of X may be due to the change of $\mathbf{W}_1$, however the conventional structural equation models ignore the dependence between X and $\mathbf{W}_1$, when they define the causal effects based on the change of X. Our proposed GSEMs models the change of X conditional on $\mathbf{W}$ to define the causal effects.

3.4 Parameter and Effect Estimations

3.4.1 Estimation of Model Parameters

We consider a general GSEM under non-experimental design given in Section 3.3.3, includes the setting with no confounders in Section 3.3.2 as a special case. Suppose we have a dataset of n observations, $(X_i, M_i, Y_i), i = 1, \cdots, n$. Although the exact likelihood fuction is available , optimization is computationally challenged. For consideration of practical usefulness, we adopt a computationally efficient method. Following the methodology of inference function for margins (IFM) (*Xu*, 1996; *Joe*, 2005; *Shih and Louis*, 1995; *Ferreira and Louzada*, 2014; *Ko and Hjort*, 2019), we develop a two-stage profile likelihood estimation procedure. The first stage involves fitting GLMs respectively on three univariate margins, X_i, M_i and Y_i to obtain the regression coefficients $\boldsymbol{\beta}_x$, $\boldsymbol{\beta}_m$ and $\boldsymbol{\beta}_y$ and dispersion parameters; and the second stage involves searching the dependence parameters α, β and γ given the estimates from the first stage.

Step I fits X_i marginally on covariates $\mathbf{W}_1$ via a GLM and obtain estimated regression coefficients $\hat{\boldsymbol{\beta}}_x$, as well the subject-specific estimates $\hat{\mu}_{x_i}$ and $\hat{\phi}_{x_i}$, where dispersion parameter $\hat{\phi}_{x_i}$ may be assumed the same across the subjects, i.e., $\hat{\phi}_x$. Then for each subject i, we calculate Z_{x_i} from $\hat{\mu}_{x_i}$ and $\hat{\phi}_x$. Z_{x_i} could be either a unique value if X is continuous or a range if X is discrete with lower and upper bounds. Repeat the same analysis for variables M and Y to obtain the $Z^*_{m_i}$ and $Z^*_{y_i}$. Step II plugs in Z_{x_i}, $Z^*_{m_i}$ and $Z^*_{y_i}$ to one log-likelihood function $\sum_{i=1}^n \log \pi(X_i, M_i, Y_i)$ in Table 3.1, which is a function of three parameters α, β and γ. We call "optim", the R function to search for estimates $\hat{\alpha}$, $\hat{\beta}$ and $\hat{\gamma}$ that minimize the negative log likelihood function. In particular, we use algorithm "Nelder-Mead" (*Nelder and Mead*, 1965), available in "optim", which only uses function values for optimization. The initial values used to start the search are set at 0 for α, β and γ.

3.4.2 Estimation of Causal Effects

To estimate the conditional NDE and NIE, we need to calculate $\mathrm{E}\{Y(x_a, M(x_b)\}$ for given x_a and x_b. Expectation for outcome $\mathrm{E}\{Y(x_a, M(x_b)\}$ under eight scenarios are listed in Table 3.1. The detailed derivations are available in Appendix B.3. Notably, these causal effects are calculated for a representative individual with mean covariate values $\mathbf{W} = \mathbf{w}$ when covariates are continuous, or a stratum $\mathbf{W} = \mathbf{w}$ when covariates are categorical. The trivariate Gaussian copula model (3.1)-(3.4) enables us to obtain some of the eight expectations in the closed forms. In the cases where the closed-forms expectations are unavailable, we invoke numerical techniques such as "Sparse grid" or "Monte Carlo" to approximate the integrals in the estimating of causal effects.

3.4.3 Bootstrap for Confidence Interval

Since our estimation procedure is a two-stage profile likelihood estimates and the expectation of conditional NDE/NIE appear complicated, it is difficult to derive the standard error of these causal effects using the Delta method (*Dorfman*, 1938). Therefore, we propose to adopt the parametric and nonparametric bootstrap approaches (*Efron*, 1987), to obtain the 95% confidence intervals (CI) for parameters α, β and γ as well as the NDE and NIE. In implementation, we generate 500 bootstrap samples, where we obtain 500 estimates of α, β and γ and NDE and NIE. Then respectively, a 95% CI is constructed by the 2.5 percentiles and 97.5 percentiles of these 500 empirical estimates. In the parametric bootstrap, 500 bootstrap datasets are generated from the estimates $\hat{\alpha}$, $\hat{\gamma}$ and $\hat{\gamma}$. In the non-parametric bootstrap, these bootstrap datasets are generated from randomly drawing of the observations with replacement.

3.5 Three Examples

In this section, we will illustrate how to estimate the conditional NDE and NIE via the proposed method through three examples with mixed data types. They are, (i) X, M and Y all follow normal distributions (CCC); (ii) X and Y follow normal distributions, while M follows Bernoulli distribution (CDC); and (iii) X and M follow normal distributions, while Y follows Bernoulli distribution (CCD). The proposed GSEM allows us to obtain the analytic forms of the conditional NDE and NIE for the first two examples (i) and (ii), and for the third case (iii), we use the Monte Carlo method to evaluate the integral when the closed forms of causal effects are unavailable. These three data types illustrated here are commonly encountered in practice. We present numerical results for these examples in Simulation Studies in Section 3.6.

3.5.1 Example CCC

In Table 3.1, when X, M and Y are all normally distributed (i.e., as "CCC" case), conditional NDE and NIE are given below in Proposition III.1.

Proposition III.1. *When X and M and Y are normally distributed, the conditional causal effects for a change of X from x_0 to x_1 are given by,*

$$\text{NDE}(x_0, x_1; \mathbf{w}) = \frac{\sigma_y \gamma (z_{x_1} - z_{x_0})}{\tau_y}, \ \text{NIE}(x_0, x_1; \mathbf{w}) = \frac{\sigma_y \alpha \beta (z_{x_1} - z_{x_0})}{\tau_y},$$

where $z_x = \frac{x - \mathbf{w}_x^\top \boldsymbol{\beta}_x}{\sigma_x}$. Furthermore, we have $\text{NDE}(x_0, x_1; \mathbf{w}) = 0$ if and only if $\gamma = 0$. $\text{NIE}(x_0, x_1; \mathbf{w}) = 0$ if and only if $\alpha = 0$ or $\beta = 0$. This is identical to the causal effect representation under the classical SEM.

Note that the causal effects only depend on parameters α, γ and β, but not on any confounders $\mathbf{W}$.

3.5.2 Example CDC

In Table 3.1, when X and Y are continuous, and M is dichotomous (i.e., as "CDC" case), conditional NDE and NIE are given below in Proposition III.2.

Proposition III.2. *When X and Y are normally distributed, and M is binary, the conditional causal effects for a change of X from x_0 to x_1 are given by,*

$$\text{NDE}(x_0, x_1; \mathbf{w}) = \frac{\sigma_y(\gamma + \alpha\beta)(z_{x_1} - z_{x_0})}{\tau_y} - \frac{\sigma_y \beta}{\tau_y}\left(\frac{p_{z_{x_0}}}{p_{z_{x_1}}} - \frac{1 - p_{z_{x_0}}}{1 - p_{z_{x_1}}}\right) d_{z_{x_1}}(\mathbf{w}),$$
$$\text{NIE}(x_0, x_1; \mathbf{w}) = \frac{\sigma_y \beta}{\tau_y}\left(\frac{p_{z_{x_0}}}{p_{z_{x_1}}} - \frac{1 - p_{z_{x_0}}}{1 - p_{z_{x_1}}}\right) d_{z_{x_1}}(\mathbf{w}),$$

where $u_m(\mathbf{w}) = \Phi^{-1}(F_m(0)) = -\Phi^{-1}\left(logit^{-1}(\mathbf{w}^\top \boldsymbol{\beta})\right)$, $p_x(\mathbf{w}) = \Phi(\tau_m u_m(\mathbf{w}) - \alpha x)$, $d_x(\mathbf{w}) = \phi(\tau_m u_m(\mathbf{w}) - \alpha x)$. *This implies that the sufficient conditions for* $\text{NDE}(x_0, x_1; \mathbf{w}) = 0$ *are* $\gamma = \beta = 0$ *or* $\gamma = \alpha = 0$*;* $\text{NIE}(x_0, x_1; \mathbf{w}) = 0$ *if* $\alpha = 0$ *or* $\beta = 0$.

Note that either α or β must be zero in order for the NDE being zero. This is because when M is discrete, Z_m and M are no longer one-to-one linked. As obvious in Figure 3.3, when only γ is zero, Z_x can change Z_m without changing M, Z_m can then change Z_y, which leads to a non-zero NDE. To obtain a zero NDE, we need either $\gamma = \alpha = 0$ or $\gamma = \beta = 0$.

3.5.3 Example CCD

In Table 3.1, when X and M are continuous, and Y is discrete (i.e., "CCD" case), NDE and NIE are given below in Proposition III.3.

Proposition III.3. *When X and Y are normally distributed, and M is binary, the*

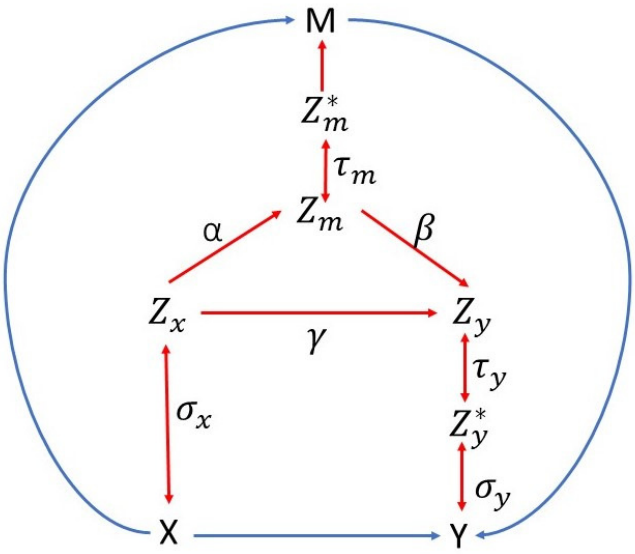

Figure 3.3: GSEM when X, Y are continuous and M is discrete.

conditional causal effects for a change of X *from* x_0 *to* x_1 *are given by,*

$$\text{NDE}(x_0, x_1; \mathbf{w}) = \int_{-\infty}^{\infty} (\Phi_1(z_{x_0}) - \Phi_1(z_{x_1}))\pi(Z_m^*|Z_x = z_{x_0})dZ_m^*,$$

$$\text{NIE}(x_0, x_1; \mathbf{w}) = \int_{-\infty}^{\infty} \Phi_1(z_{x_1})\pi(Z_m^*|Z_x = z_{x_0})dZ_m^* - \int_{-\infty}^{\infty} \Phi_1(z_{x_1})\pi(Z_m^*|Z_x = z_{x_1})dZ_m^*,$$

where $\Phi_1(z_{x_a}) = \Phi(\tau_m l_y(\mathbf{w}) - \gamma z_{x_a} - \tau_m \beta z_m^*)$, $l_y(\mathbf{w}) = \Phi^{-1}(F_y(0)) = -\Phi^{-1}\left(logit^{-1}(\mathbf{w}^\top \boldsymbol{\beta}_y)\right)$. $\text{NDE}(x_0, x_1; \mathbf{w}) = 0$ *if and only if* $\gamma = 0$. *This implies that the sufficient conditions for* $\text{NIE}(x_0, x_1; \mathbf{w}) = 0$ *are* $\alpha = 0$ *or* $\beta = 0$.

Since both conditional NDE and NIE do not have closed forms, we will resort to the Monte Carlo method to evaluate the intergrals by the following steps: (i) sampling Z_m^* from $N\left(\frac{\alpha}{\sqrt{\theta_{mx}^2+1}} z_{x_b}, \frac{1}{\sqrt{\theta_{mx}^2+1}}\right)$; (ii) plug Z_m^* into $\Phi_1(z_{x_a})$; and (iii) take average over all the values of $\Phi_1(z_{x_a})$.

Both conditional $\text{NDE}(x_0, x_1; \mathbf{w})$ and $\text{NIE}(x_0, x_1; \mathbf{w})$ may be interpreted as risk differences in the case of binary outcome Y. Alternatively, one may interpret these effects in terms of odds. According to *VanderWeele and Vansteelandt* (2010), we

calculate the odds ratios for the conditional NDE and NIE as follows, respectively

$$OR^{\text{NDE}}(x_0, x_1; \mathbf{w}) = \frac{A_{x_0,x_0}/(1 - A_{x_0,x_0})}{A_{x_1,x_0}/(1 - A_{x_1,x_0})}, \tag{3.11}$$
$$OR^{\text{NIE}}(x_0, x_1; \mathbf{w}) = \frac{A_{x_1,x_0}/(1 - A_{x_1,x_0})}{A_{x_1,x_1}/(1 - A_{x_1,x_1})},$$

where $A_{x_a,x_b} = \int_{-\infty}^{\infty} \Phi_1(z_{x_a})\pi(Z_m^*|Z_x = z_{x_b})dZ_m^*$ for $a = 0, 1$ and $b = 0, 1$.

3.6 Simulation Studies

We carry out four simulation studies to evaluate the performance of the proposed GSEM methodology. The first study assesses bias, mean squared error (MSE) and coverage probability of confident interval for causal effects and model parameters. The second study compares bias, MSE and coverage with an existing method Quasi-Bayesian Monte Carlo method (QBMC) (*Tingley et al.*, 2014). Both studies were performed to three settings discussed in Section 3.5. The third study concerns the setting where the outcome is binary and compares the odds ratios calculated from GSEM and an approximation approach proposed by *VanderWeele and Vansteelandt* (2010). The fourth study compares the efficiency of GSEM with the full maximum likelihood estimation (MLE) procedure, under the setting of CDC with no confounders.

3.6.1 Assessment of GSEM

Consider the three settings: (i) CCC: X, M and Y all follow Normal distributions; (ii) CDC: X and Y follow Normal distribution, and M follows Bernoulli distribution; (iii) CCD: X and M follow Normal distribution, and Y follows Bernoulli distribution. For each setting, we calculate average bias, MSE, and empirical coverage probability via both parametric bootstrap and non-parametric bootstrap for conditional NDE, NIE, α, γ and β. The sample size n varies over 200, 500, and 1000. For each sample size, we run 1,000 replicates. Coverage probability of confidence interval is obtained

by 500 bootstrap replicates.

The GSEM model is set up according to (3.6) and (3.9), while covariates $\mathbf{W} = (W_1, W_2, W_3, W_4)$, and $\mathbf{W}_1 = (W_1, W_2, W_3)$, with W_1 being a column of all ones for the intercept. Covariates W_2, W_3, W_4 are assumed to follow a multivariate normal distribution with mean zero and compound symmetry correlation $\rho_w = 0.2$ and standard deviation $\sigma_w = 0.3$. Set $\boldsymbol{\beta}_x = (0.5, 0.2, 0.2)^\top$, $\boldsymbol{\beta}_m = (0.8, 0.3, 0.3, 0.4)^\top$, $\boldsymbol{\beta}_y = (-0.2, 0.4, -0.2, 0.7)^\top$, and $\sigma_x = \sigma_m = \sigma_y = 0.3$. For the settings (i), (ii) and (iii), (α, γ, β) take the following values: $(0.20, 0.10, 0.20)$, $(0.15, 0.10, 0.75)$ and $(0.70, 0.10, 0.18)$.

First, we generate Z_x, Z_m, Z_m^*, Z_y and Z_y^* according to (3.6), and then given true parameters and simulated covariates, we generate exposure $X = \mathbf{W}_1^\top \boldsymbol{\beta}_x + \sigma_x Z_x$, and mediator M and outcome Y below. (i) CCC: $M = \mathbf{W}^\top \boldsymbol{\beta}_m + \sigma_m Z_m^*$, $Y = \mathbf{W}^\top \boldsymbol{\beta}_y + \sigma_y Z_y^*$; (ii) CDC: $M = I\{\Phi(Z_m^*) > \frac{1}{exp(\mathbf{W}^\top \boldsymbol{\beta}_m)+1}\}$, $Y = \mathbf{W}^\top \boldsymbol{\beta}_y + \sigma_y Z_y^*$; and (iii) CCD: $M = \mathbf{W}^\top \boldsymbol{\beta}_m + \sigma_m Z_m^*$, $Y = I\{\Phi(Z_y^*) > \frac{1}{exp(\mathbf{W}\boldsymbol{\beta}_y)+1}\}$.

We apply the GSEM methodology to first obtain $\hat{\boldsymbol{\beta}}_x$, $\hat{\boldsymbol{\beta}}_m$, $\hat{\boldsymbol{\beta}}_y$, $\hat{\sigma}_x$, $\hat{\sigma}_y$, $\hat{Z}_x$, $\hat{Z}_m^*$ (or $\hat{l}_m$ and $\hat{u}_m$), and $\hat{Z}_y^*$ (or $\hat{l}_y$ and $\hat{u}_y$) through the standard GLM regression, then we search $\hat{\alpha}$, $\hat{\beta}$, $\hat{\gamma}$ by minimizing the negative observed likelihood function. Lastly, we calculate the estimated NDE and NIE based on parameter estimates $\hat{\alpha}$, $\hat{\beta}$, $\hat{\gamma}$.

The simulation results are summarized in Table 3.2. Under the three settings, the magnitude of average bias for estimation of parameters and causal effects decrease as the sample size increases. The MSE for $n = 1000$ is at most one fifth of that for $n = 200$ in each setting. The 95% confidence interval have showed coverage rates for $n = 200, 500, 1000$ all close to the 95% nominal level using either parametric or non-parametric bootstrap. The computation cost in the simulation studies is between 30 to 60 minutes for a dataset with 500 sample size under three different settings of "CCC", "CDC" and "CDC".

Table 3.2: True value, bias, MSE, 95% coverage by parametric bootstrap (PB), and non-parametric bootstrap (NB) for NDE, NIE, α, γ and β under three settings. The sample size varies over 200, 500 and 1,000, with 1,000 data replicates for each sample size. The confidence interval to determine the coverage is obtained by 500 bootstrap replicates.

			Bias ($\times 10^{-3}$)			MSE ($\times 10^{-3}$)			Coverage (PB, %)			Coverage (NB, %)		
Setting		True	200	500	1000	200	500	1000	200	500	1000	200	500	1000
CCC	NDE	0.10	3.77	1.76	-0.40	4.99	1.97	0.90	94.2	95.0	95.9	94.6	95.1	96.0
	NIE	0.04	0.88	0.79	0.71	0.45	0.16	0.08	94.4	94.6	95.0	93.7	94.5	94.6
	α	0.20	5.29	2.34	0.54	5.72	2.13	0.98	93.1	94.7	95.6	93.4	94.4	95.1
	γ	0.10	4.93	2.05	-0.33	5.41	2.11	0.98	94.0	95.1	95.9	94.5	95.3	96.2
	β	0.20	2.66	2.39	3.25	5.75	2.06	1.11	93.5	94.0	93.8	93.7	94.5	93.8
CDC	NDE	0.11	0.98	0.97	-0.32	3.96	1.61	0.71	94.6	95.2	95.7	95.3	95.4	96.0
	NIE	0.05	3.53	1.53	0.46	1.46	0.55	0.24	93.3	93.4	94.3	93.7	93.0	94.4
	α	0.15	8.31	2.59	0.67	10.23	3.84	1.68	93.3	94.1	95.0	93.8	93.3	94.9
	γ	0.10	0.69	0.96	-0.52	8.01	3.15	1.32	94.7	94.1	95.2	93.9	94.2	95.4
	β	0.75	22.84	13.96	4.04	20.07	6.74	3.43	92.8	94.7	94.4	93.6	95.1	94.5
CCD	NDE	0.13	5.40	0.80	0.59	21.03	8.01	4.00	93.7	94.2	94.6	94.1	94.1	94.2
	NIE	0.14	3.70	1.64	1.25	5.45	2.17	1.05	94.7	93.6	94.4	94.6	94.4	94.5
	α	0.70	9.22	5.64	3.54	7.56	3.09	1.65	94.7	94.8	93.6	93.3	94.5	94.0
	γ	0.10	5.36	0.89	0.67	14.15	5.21	2.57	93.8	94.0	94.8	94.4	93.9	94.4
	β	0.18	7.24	2.53	1.92	9.33	3.48	1.71	93.8	93.7	94.2	93.5	94.6	94.5

3.6.2 Comparison to QBMC

In the second simulation study, we compare the coverage probability and MSE between our proposed GSEM method and the "mediation" R package (*Tingley et al.*, 2014; *Imai et al.*, 2010a), a quasi-Bayesion Monte Carlo simulation-based method (in short, QBMC). The comparison is conducted under two scenarios: one under datasets being generated from the GSEM, and the other from the QBMC. We focus on assessing their performance for conditional NDE and NIE equal to zero, and not zero.

The GSEM model specification is the same as discussed in Section 3.6.1. Let $\boldsymbol{\beta}_x = (0.5, 0.2, 0.2)^\top$, $\boldsymbol{\beta}_m = (0.8, 0.3, 0.3, 0.4)^\top$, $\boldsymbol{\beta}_y = (-0.2, 0.4, -0.2, 0.7)^\top$, and $\sigma_x = \sigma_m = \sigma_y = 0.3$. For the null effects, (α, γ, β) takes the following values: $(0.30, 0, 0)^\top$, $(0.15, 0, 0)^\top$ and $(0.70, 0, 0)^\top$. According to the propositions discussed in Section 3.5, the conditional NDE and NIE are both zero in these three settings. For the non-zero causal effects, (α, γ, β) takes values $(0.20, 0.10, 0.20)^\top$, $(0.15, 0.10, 0.75)^\top$ and $(0.70, 0.10, 0.18)^\top$.

The QBMC model is set up as follows: X is simulated from $N(\mathbf{W}_1^\top \boldsymbol{\beta}_x, \sigma_X)$, then M and Y are generated for the following three settings: (i) CCC: $M = \beta_{xm}X + \mathbf{W}^\top \boldsymbol{\beta}_m + \epsilon_M$, $Y = \beta_{xy}X + \beta_{my}M + \mathbf{W}^\top \boldsymbol{\beta}_y + \epsilon_Y$; (ii) CDC: $logit(P(M = 1)) = \beta_{xm}X + \mathbf{W}^\top \boldsymbol{\beta}_m$, $Y = \beta_{xy}X + \beta_{my}M + \mathbf{W}^\top \boldsymbol{\beta}_y + \epsilon_Y$; and (iii) CCD: $M = \beta_{xm}X + \mathbf{W}^\top \boldsymbol{\beta}_m + \epsilon_M$, $logit(P(Y = 1)) = \beta_{xy}X + \beta_{my}M + \mathbf{W}^\top \boldsymbol{\beta}_y$, where $\sigma_X = \sigma_M = \sigma_Y = 0.3$, $\boldsymbol{\beta}_x = (0.5, 0.2, 0.2)^\top$, $\boldsymbol{\beta}_m = (0.5, 0.2, 0.2, 0.4)^\top$, $\boldsymbol{\beta}_y = (-0.8, 0.2, -0.5, 0.4)^\top$. For settings (i) through (iii), $(\beta_{xm}, \beta_{xy}, \beta_{my})^\top$ take the following values $(0.30, 0, 0)^\top$, $(0.30, 0, 0)^\top$ and $(0.30, 0, 0)^\top$ for null effects, and $(0.20, 0.10, 0.20)^\top$, $(0.40, 0.10, 0.50)^\top$ and $(0.70, 0.50, 0.70)^\top$ for non-zero effects.

Table 3.3 summarizes the comparison results between GSEM and QBMC under the criteria of empirical coverage probability, average bias and MSE. When the data are generated from the GSEM, the coverage rates of GSEM are closer to the nominal 95% level in most cases. For average bias and MSE, the GSEM has shown advantages in some cases, such as CCD where the GSEM consistently provides smaller bias and MSE for NIE, no matter the true effects are zeros or not, compared to QBMC. In the CDC case, the GSEM provides smaller bias for NIE and NDE compared to QBMC when sample size $n = 500$ and the true effects are not zero. For the other cases, the results between the two methods are comparable.

When the data are generated from QBMC and effects are zero, two methods are comparable. When the effects are not zero, since the QBMC method does not provide formula to calculate the true effects, the true effects are estimated from a dataset with a very large sample $n = 100,000$, under which the estimated effects from the R package are then used as true effects. So the comparison under the non-zero effect setting needs to be interpreted cautiously.

Table 3.3: Coverage via the parametric bootstrap, bias and MSE comparison of GSEM and QBMC. When data are generated from GSEM, $(\alpha, \gamma, \beta)^\top$ takes values $(0.30, 0, 0)^\top$, $(0.15, 0, 0)^\top$ and $(0.70, 0, 0)^\top$ under zero NDE and NIE; and $(0.20, 0.10, 0.20)^\top$, $(0.15, 0.10, 0.75)^\top$ and $(0.70, 0.10, 0.18)^\top$ under non-zero NDE and NIE. Similarly, when data are generated from SEM, $(\beta_{xm}, \beta_{xy}, \beta_{my})^\top$ takes values $(0.30, 0, 0)^\top$, $(0.30, 0, 0)^\top$ and $(0.30, 0, 0)^\top$ under zero NDE and NIE; and $(0.20, 0.10, 0.20)^\top$, $(0.40, 0.10, 0.50)^\top$ and $(0.70, 0.50, 0.70)^\top$ under non-zero NDE and NIE. The sample size varies over 200 and 500, with 1,000 data replicates for each sample size.

					Coverage (%)		Bias ($\times 10^{-3}$)		MSE ($\times 10^{-3}$)	
Data	Setting	Effect	True	Method	200	500	200	500	200	500
GSEM	CCC	NDE	0	GSEM	95.0	94.6	-1.70	3.62	5.82	2.31
				QBMC	95.4	94.4	-1.68	3.59	5.67	2.27
		NIE	0	GSEM	92.6	94.3	-0.30	-0.37	0.58	0.20
				QBMC	93.2	94.0	-0.26	-0.36	0.54	0.20
		NDE	0.10	GSEM	95.3	94.8	-0.56	3.89	5.13	2.06
				QBMC	95.4	94.0	-1.74	3.37	5.03	2.03
		NIE	0.04	GSEM	94.0	94.6	1.28	-0.12	0.47	0.16
				QBMC	94.0	94.7	-0.03	-0.65	0.44	0.16
	CDC	NDE	0	GSEM	95.1	94.0	-1.97	3.40	5.24	2.12
				QBMC	95.1	93.8	-1.96	3.35	5.07	2.09
		NIE	0	GSEM	97.8	96.0	-0.02	-0.15	0.12	0.04
				QBMC	98.7	96.9	0.02	-0.12	0.10	0.03
		NDE	0.11	GSEM	93.9	94.2	-1.87	2.23	4.18	1.64
				QBMC	94.2	94.3	0.51	4.30	4.14	1.66
		NIE	0.05	GSEM	93.9	95.0	3.01	0.05	1.37	0.49
				QBMC	96.0	95.5	-2.77	-3.33	1.08	0.42
	CCD	NDE	0	GSEM	96.0	95.4	9.14	1.03	19.94	7.70
				QBMC	96.6	95.9	9.68	1.86	17.23	7.13
		NIE	0	GSEM	94.0	94.3	-6.63	-0.12	4.82	1.82
				QBMC	95.4	95.0	-8.28	-0.83	5.82	2.44
		NDE	0.13	GSEM	95.4	94.6	6.42	1.63	19.35	7.84
				QBMC	95.5	94.9	1.39	-1.96	16.74	7.20
		NIE	0.14	GSEM	93.6	95.0	-0.90	0.87	5.51	2.10
				QBMC	95.3	93.8	7.09	17.24	5.83	2.73
SEM	CCC	NDE	0	GSEM	95.0	94.6	-1.70	3.62	5.82	2.31
				QBMC	95.5	94.5	-1.68	3.59	5.67	2.27
		NIE	0	GSEM	92.6	94.3	-0.30	-0.37	0.58	0.20
				QBMC	93.4	93.9	-0.26	-0.36	0.54	0.20
		NDE	0.10	GSEM	95.3	94.0	3.80	8.38	5.45	2.24
				QBMC	95.1	94.6	2.59	7.85	5.33	2.20
		NIE	0.04	GSEM	93.6	94.2	-0.26	-1.70	0.49	0.17
				QBMC	93.7	93.4	-1.61	-2.25	0.47	0.17
	CDC	NDE	0	GSEM	94.0	93.9	-4.06	-0.98	5.16	2.06
				QBMC	95.2	94.1	-4.14	-0.96	5.02	2.03
		NIE	0	GSEM	99.5	99.3	0.34	-0.02	0.05	0.01
				QBMC	99.9	99.6	0.34	-0.02	0.04	0.01
		NDE	0.09	GSEM	93.6	94.0	5.24	7.49	5.28	2.17
				QBMC	95.7	93.5	1.95	5.13	5.01	2.05
		NIE	0.04	GSEM	93.5	93.6	5.41	3.76	3.32	1.23
				QBMC	95.7	96.7	5.41	4.83	3.31	1.24
	CCD	NDE	0	GSEM	94.3	94.2	1.08	3.93	13.72	5.42
				QBMC	95.6	94.5	1.91	4.10	12.72	5.21
		NIE	0	GSEM	93.7	95.1	-2.71	-0.79	1.17	0.40
				QBMC	95.2	94.8	-3.41	-1.21	1.10	0.42
		NDE	0.12	GSEM	94.3	95.3	11.24	7.20	21.25	8.26
				QBMC	95.0	95.5	5.36	3.70	18.25	7.63
		NIE	0.12	GSEM	94.2	93.2	-21.14	-18.14	5.46	2.19
				QBMC	95.7	96.5	-12.30	-4.60	5.75	2.31

3.6.3 Odds Ratio comparison for Binary Outcome

In the third simulation study, we consider the "CCD" case where X and M are continuous and Y is binary. The NDE and NIE on the odds ratio scale are compared by GSEM according to (3.11), and an alternative approximation approach proposed by *VanderWeele and Vansteelandt* (2010), which is termed as "VV" method in short. In particular, we consider two scenarios: rare outcome (around 4% of outcomes being "1") and abundant outcome (around 45% of outcomes being "1").

For both types of outcomes, data are generated from GSEM, with $\boldsymbol{\beta}_x = (\{0.5, 0.2, 0.2)^\top$, $\boldsymbol{\beta}_m = (0.8, 0.3, 0.3, 0.4)^\top$, $\alpha = 0.70$, $\gamma = 0.10$ and $\beta = 0.18$. For the case of rare outcome, $\boldsymbol{\beta}_y = (-0.2, 0.4, -0.2, 0.7)^\top$, and for the case of abundant outcome, $\boldsymbol{\beta}_y = (-3.0, 0.4, -0.2, 0.7)^\top$. Covariates $\mathbf{W}$ and $\mathbf{W}_1$ and variance parameters are set the same as those in the first and second simulation studies.

Table 3.4: True value, mean bias, MSE comparison of OR^{NDE} and OR^{NIE} for methods GSEM and "VV". Data are generated from GSEM, sample size varies over 500, 1,000, and 2,000 with 1000 data replicates for each sample size.

				Bias			MSE		
Setting	Odds Ratio	True	Method	500	1000	2000	500	1000	2000
Abundant	OR^{NDE}	1.704	GSEM	0.116	0.098	0.022	0.573	0.250	0.091
			VV	0.134	0.114	0.035	0.600	0.264	0.095
	OR^{NIE}	1.762	GSEM	0.070	0.026	0.017	0.136	0.056	0.027
			VV	0.315	0.251	0.238	0.349	0.166	0.105
Rare	OR^{NDE}	2.169	GSEM	1.135	0.662	0.290	22.062	5.109	1.629
			VV	0.618	0.350	0.078	10.039	2.870	1.055
	OR^{NIE}	2.113	GSEM	0.502	0.157	0.087	2.662	0.660	0.280
			VV	0.830	0.490	0.421	3.610	1.194	0.591

As shown in Table 3.4, for the case of abundant events, "VV" has a larger magnitude of average bias and MSE than GSEM for the odds ratio of NDE and NIE. For the case of rare events, "VV" has smaller bias and MSE for the odds ratio of NDE, but larger bias and MSE for the odds ratio of NIE. One explanation for the poor performance of GSEM on NDE might be that for rare events, the GLM may generate large bias due to the instability of the numerical algorithm. Nevertheless, GSEM still provides a better performance compared to "VV" method on the odds ratio of NIE.

For abundant events, GSEM clearly demonstrates the advantage over "VV".

3.6.4 Efficiency comparison with MLE

Given that GSEM is a framework based on inference functions for margins (IFM), a two-stage likelihood estimation method, in this simulation study we are interested in comparing GSEM with the full maximum likelihood estimation approach to evaluate the relative efficiency. For simplicity, we consider the CDC setting without confounding factors. The dataset is generated under a GSEM model with $\alpha = 0.25$, $\gamma = 0.50$, $\beta = 0.75$; $\beta_x = 0.50$ and $\beta_y = -0.20$ for mean parameters of X and Y; $\beta_m = 0.80$ for log odds ratio of M; and $\sigma_x = \sigma_y = 0.30$ for variance parameters of X and Y. The full MLE procedure is performed with the R function "optim" to simultaneously search for the estimates of the aforementioned eight parameters. Table 3.5 reports the comparison results of empirical variance, relative efficiency defined by $\frac{\text{Variance of GSEM}}{\text{Variance of MLE}}$, mean bias and MSE for α, γ and β from 1,000 simulated datasets. GSEM has slightly larger variance for α and γ then MLE, leading to less than one percent of efficiency loss. For bias and MSE, GSEM in general provides larger bias and MSE compared to the full MLE approach.

Table 3.5: Variance, relative efficiency, bias and MSE comparison of GSEM and MLE under the setting of CDC. When data are generated from GSEM. Relative efficiency (Rela. Efficiency)=$\frac{\text{Variance of GSEM}}{\text{Variance of MLE}}$. The sample size varies over 200 and 500, with 1,000 data replicates for each sample size.

			Variance ($\times 10^{-3}$)		Rela. Efficiency		Bias ($\times 10^{-3}$)		MSE ($\times 10^{-3}$)	
Parameter	True	Method	200	500	200	500	200	500	200	500
α	0.25	GSEM	9.986	3.767	99.66%	99.82%	10.005	2.929	10.076	3.772
		MLE	9.951	3.761	-	-	8.578	2.372	10.015	3.763
γ	0.50	GSEM	9.719	3.375	99.46%	99.82%	3.474	1.432	9.721	3.373
		MLE	9.666	3.369	-	-	1.682	0.717	9.659	3.366
β	0.75	GSEM	17.884	6.626	100.00%	100.00%	18.527	0.994	18.209	6.621
		MLE	17.887	6.630	-	-	17.848	0.707	18.188	6.624

3.7 Data Application

In this section, we present a data analysis of the motivating data example from the ELEMENT cohort study, see the detail of the data example in Section 3.2. We hypothesize the association between phthalate exposures during mothers' second and third trimesters and cardiometabolic outcomes of children in peripuberty may be mediated by the timing of children reaching their infancy BMI peaks on time ($M = 0$) or delayed ($M = 1$).

Exposures X include six maternal urinary phthalate concentrations of MEHHP, MEOHP and MIBP measured at second trimester (T2) and third trimester (T3). Exposure X appears right skewed, so a logorithm transformation of X is taken, and $\log(X)$ is appearing normally distributed. Outcomes Y include three standardized z-scores of health outcomes measured during adolescence, including fasting glucose z-score, C-peptide z-score, and MetS z-score. The outcome Y appears normally distributed. The mediator of reaching infancy BMI peak on time or not is binary where 27.8% of children have delayed infancy BMI peak time. Confounders $\mathbf{W}_1$ include mother's age at birth, education, and marital status, which affect the exposures, mediator and outcomes. Confounders $\mathbf{W}_2$ include breastfeeding duration, parity, child's gender, gestational age, and birth weight, a set of factors that only affect the mediator and outcomes.

We fit the dataset by GSEM discussed in Section 3.5.2, where $\mathbf{W}_1$ are confounders for X, and $\{\mathbf{W}_1, \mathbf{W}_2\}$ are confounders for M and Y in our model (3.9). We first estimate the model parameters and causal effects, then obtain 95% CI for conditional NDE, NIE and TE through the parametric bootstrap. From the results summarized in Table 3.6, we found that a positive NIE with an estimate of 0.015 and 95% CI (0.001, 0.026) for MEHHP during the second trimester on MetS z-score, suggesting that one unit increase in exposure to MEHHP during the second trimester would result in an increase of 0.015 in MetS z-score, which is mediated through the BMI infancy

peak. Similarly, a positive NIE with an estimate of 0.017 and 95% CI (0.001, 0.032) was identified for MEOHP during the second trimester on MetS z-score. However, both NDE and TE are not statistically significant for these two exposures on the MetS z-score. We will interpret the significant NIE results as the pathway mediation analysis.

Table 3.6: NDE, NIE and TE estimates and 95 % CI obtained from GSEM for ELEMENT study.

Exposure	Outcome	NDE	95% CI	NIE	95% CI	TE	95% CI
MEHHP T2	glucose	-0.013	(-0.121, 0.105)	0.016	(-0.001, 0.035)	0.003	(-0.109, 0.116)
	C-peptide	-0.098	(-0.219, 0.031)	0.014	(-0.003, 0.032)	-0.084	(-0.205, 0.038)
	MetS	-0.067	(-0.133, 0.004)	0.015	(0.001, 0.026)	-0.051	(-0.120, 0.016)
MEOHP T2	glucose	-0.011	(-0.130, 0.100)	0.017	(-0.002, 0.038)	0.006	(-0.115, 0.116)
	C-peptide	-0.105	(-0.236, 0.016)	0.015	(-0.003, 0.034)	-0.090	(-0.227, 0.032)
	MetS	-0.071	(-0.134, 0.005)	0.017	(0.001, 0.032)	-0.054	(-0.118, 0.022)
MIBP T2	glucose	0.038	(-0.078, 0.150)	0.014	(-0.007, 0.044)	0.052	(-0.053, 0.168)
	C-peptide	-0.019	(-0.128, 0.097)	0.011	(-0.007, 0.038)	-0.008	(-0.117, 0.108)
	MetS	-0.020	(-0.084, 0.053)	0.013	(-0.005, 0.041)	-0.007	(-0.073, 0.069)
MEHHP T3	glucose	0.053	(-0.069, 0.163)	0.006	(-0.013, 0.022)	0.059	(-0.066, 0.168)
	C-peptide	0.017	(-0.104, 0.137)	0.005	(-0.014, 0.020)	0.022	(-0.098, 0.142)
	MetS	0.005	(-0.069, 0.077)	0.005	(-0.016, 0.017)	0.011	(-0.066, 0.081)
MEOHP T3	glucose	0.055	(-0.071, 0.184)	0.003	(-0.022, 0.022)	0.058	(-0.067, 0.188)
	C-peptide	0.033	(-0.093, 0.173)	0.003	(-0.017, 0.017)	0.035	(-0.089, 0.175)
	MetS	0.009	(-0.065, 0.091)	0.003	(-0.020, 0.018)	0.013	(-0.062, 0.091)
MIBP T3	glucose	0.106	(-0.027, 0.257)	0.013	(-0.008, 0.039)	0.118	(-0.012, 0.266)
	C-peptide	0.020	(-0.126, 0.179)	0.011	(-0.009, 0.039)	0.030	(-0.110, 0.184)
	MetS	-0.014	(-0.099, 0.087)	0.012	(-0.008, 0.032)	-0.001	(-0.093, 0.102)

For sensitivity analysis, we perform the same analysis stratified on gender. For boys, the sample sizes for the second trimester and the third trimester exposures are 79 and 96; for girls, the same sample sizes are 98 and 106. The stratified analyses were presented in Table B.2 for boys and Table B.3 for girls in Appendix B.4. For boys, none of the effects are found to be significant, likely due to the small sample size. For girls, a significant NIE is identified for MIBP exposure during the second trimester on MetS z-score.

Another sensitivity analysis is performed to examine the effect of breastfeeding duration, a continuous variable about how many months that a mom had been breastfeeding her child. In previous analyses, we assume that breastfeeding duration is a confounder for mediator and outcome and the breastfeeding duration is not affected

by exposures X. This is one of the key assumptions for the identifications of natural effects. However, it is possible that breastfeeding could be influenced by exposures and then influences the BMI infancy peak, making it a potential mediator. Therefore, we perform the same analysis again without including the breastfeeding as a confounder to examine how different the results are with and without including breastfeeding. Table B.1 in Appendix B.4 summarizes the results for the sensitivity analysis excluding breastfeeding. Compare it with Table 3.6, we can see that the effect estimates are very similar, and the NIE remains significant for second trimester exposures to MEOHP and MIBP on MetS z score.

3.8 Concluding Remarks

This chapter developed a new framework for mediation analysis, GSEM, when exposure X, mediator M and outcome Y are of mixed types. We proposed a unified framework of hierarchical GSEMs based on the Gaussian copula models, which allows us to specify the joint density of X, M and Y using a hierarchical modeling approach. The GSEMs provide a flexible modeling of causal effects for mixed data types. The proposed GSEM characterizes interpretable parameters α, β and γ. We illustrated the GSEM methodology through three specific examples of practical importance. Through extensive simulation studies, the proposed profile likelihood estimation can accurately make statistical inference on NDE and NIE under the model assumptions.

Modeling X, M and Y respectively given covariates $\mathbf{W}$ allows us to potentially handle the high dimensional covarites easily and efficiently. In contrast, the conventional structural equation modeling involves both the parameters for causal effects and high-dimensional nuisance parameters for the confounding effects, leading to a complicated inference procedure. The ability to handle the high-dimensional covariates easily is especially critical as one of the core assumptions under the potential outcome framework is "sequential ignorability" meaning that we must include all the

possible confounders in order to identify the conditional causal effects (*VanderWeele and Vansteelandt*, 2009).

The current procedure focuses mainly on the estimation of the causal effects. To make inference about the causal effects, we need to develop formal statistical testing procedures to address the issue of composite null hypothesis. This is a challenging as it would be a composite task, which is our future work. We plan to extend the current framework of single mediator to multi-mediators of mixed types. There are some methods developed for continuous multi-mediators. However, there lack methodological developments for multi-mediators of mixed types. The flexibility of the copula dependent model allowing the different types of marginal distributions can provide a useful tool to handle the multi-mediators of mixed data types. Another potential future work is to adopt the current GSEM framework and extend it to accommodate multiple mediators with mixed data types. As seen in our real data application, the BMI infancy peak is a binary mediator and breastfeeding duration is potential a continuous mediator, and currently there is a lack of methods to analyze multiple mediators of mixed types. The flexibility of copula dependence models that handle mixed data types gives rise to a possible solution to modeling mediators of different types.

CHAPTER IV

Mediation Pathway Analysis with Categorical Exposure Variable

4.1 Introduction

Causal mediation analysis via the structural equation modeling (SEM) approach has been widely used in many scientific disciplines, such as social sciences, epidemiology, environmental health sciences and so on. *Baron and Kenny* (1986) first proposed the mediation analysis under the linear regression framework, and more recent literature has focused on the potential outcome framework (*Robins and Greenland*, 1992; *Pearl*, 2001), in which the concepts of direct and indirect effects have been extended to natural direct effect (NDE) and natural indirect effect (NIE).

In the SEM setting as shown in Figure 1.1, when both mediator M and outcome Y are continuous and normally distributed, the assessment of the NDE and NIE are well studied via the decomposition TE=NDE+NIE, see more details in (*VanderWeele and Vansteelandt*, 2009). Since exposure X is a dependent variable in both regression models of SEM, X can be either continuous or categorical. In this chapter, we focus on the case where exposure X is binary, such as a treatment variable with active drug and placebo. In this case, we are interested in assessing the causal effects of the treatment on an outcome of interest.

In this chapter, we focus on two important cases of binary exposure. The first case concerns the scenarios of drug absorption rate, an issue of effective drug dose. Our motivating example is a birth cohort study, "The Early Life Exposures in Mexico to ENvironmental Toxicants" (ELEMENT), which recruited three mother-child pairs from Mexico City during mid-90s. The overarching goal of this cohort study is to understand how environmental toxicants, such as phthalates, phenols and metals affect mother and child health outcomes, such as birth weight. Of those toxicants, exposure to lead is especially concerning as in-utero lead exposure may lead to low birth weight and developmental delay in offspring (*Zhu et al.*, 2010). A randomized control trial in the ELEMENT cohort was conducted to investigate whether maternal calcium supplementation helps suppress the release of the bone lead to the blood, which in turn may reduce offspring's exposure to lead during pregnancy (*Perng et al.*, 2019). In this chapter we plan to examine whether, and to what extent, the effect of calcium supplementation on birth weight is mediated by the mother's blood lead concentration during third trimester of pregnancy.

In this motivation example, we argue that even though mothers are allocated in the same treatment group, their individual absorption of calcium capacity is very likely to vary from person to person. It is practically unrealistic to obtain direct measurements of absorption. Thus, using binary exposure in the SEM is an approximation , ignoring the individual-level calcium dose. To fill this gap, we propose a new model, termed as *effective dose model*, a latent exposure variable Z is introduced to capture the effective dose of the treatment in the analysis of effective degree of those causal effects. Figure 4.1 describes a DAG with observed treatment variable (X), latent absorbed dose of calcium (Z), mediator maternal blood lead (M), and outcome birth weight (Y).

As a scientific premise, we hypothesize that for mothers allocated to the control group ($X = 0$), their effective dose of calcium $Z = 0$; for mothers allocated to calcium group

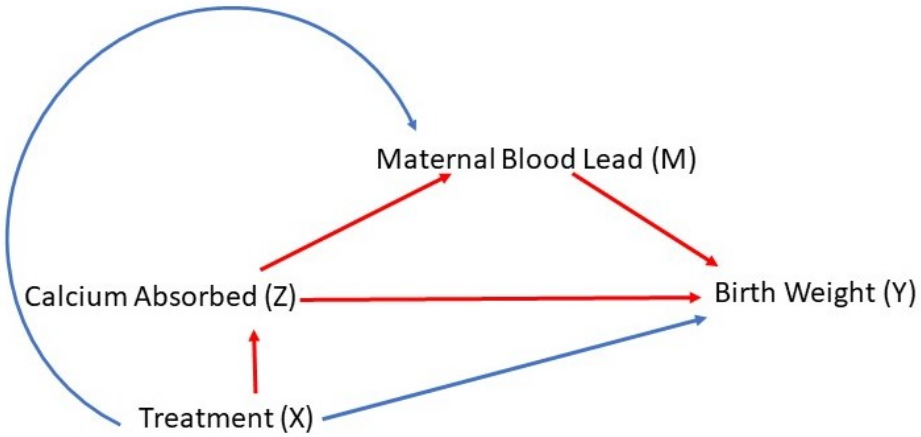

Figure 4.1: Hypothesized causal associations among calcium absorbed, maternal blood lead and birth weight.

$(X = 1)$, their absorbed dose of calcium Z follows a truncated normal distribution, denoted by $TN(a, 1)$, which only takes positive values and $a > 0$. Here, we see $Z = (1 - X)I[Z = 0] + XTN_{(0,+\infty)}(a, 1)$. The known value a describes the mean absorption rate for an individual. The higher the value a, the higher the subject's absorption of calcium. The subject specific a_i is assumed to have been obtained from external data, such as Phase I clinical trials that measure the pharmacokinetic dynamics of the drug. a_i can be modeled by statistical modeling characterizing the individual specific absorption rate. We believe this latent dose variable of calcium can more accurately describe effective exposure than a binary treatment variable. The conventional binary treatment variable X can be viewed as an approximation to the effective dose variable Z. While using a binary exposure X in a mediation analysis gives rise to certain simplicity, it indeed poses an assumption of the same dosage for individuals in the calcium group in the evaluation of causal effects. Specially, this assumption implies that the treatment effect on the outcome is linear, and ignores the individual variability in the absorption of calcium. Adopting Z instead of X in the mediation analysis allows us to obtain better examination of causal mediation pathways.

Let us discuss another important case of binary exposure. In some observational studies, often we only observe a binary exposure which arises from a continuous underlying unobserved exposure variable, by the means of dichotomization. For example, in electronic health records data, patient's diabetes diagnosis is commonly recorded as a binary variable of yes or no, but his/her fasting glucose levels measured by the lab are not available. In fact, the diagnosis of diabetes ($X = 1$) means his/her fasting glucose level is 126 mg/dl or higher. For all the diabetes patients, their glucose levels can vary substantially, and the information of severity provides more critical medical condition than a dichotomous disease status. The aforementioned effective dose scenario is not the same this type of binary exposure. Therefore, we propose another dichotomization model, in which a latent exposure variable is introduced to determines an observed binary exposure variable X. In the diabetes example, conceptually we can map the glucose level to a standardized latent z-score variable Z, $X = I(Z > 0)$.

Under the two proposed binary exposure models, we investigate both NDE and NIE in the causal mediation pathways. Along this line, we propose an extension under the latent variable model, leading to the new concepts of generalized natural direct/indirect effects (GNDE/GNIE). We demonstrate that GNDE and GNIE are more desirable to quantify and interpret causal mediation effects than the conventional NDE and NIE (*Pearl*, 2001).

The remainder of the chapter is organized as follows. Section 4.2 introduces two new frameworks of effective dose model and latent exposure model. Section 4.3 presents the formulation of NDE and NIE, which is then extended into the new concepts to GNDE and GNIE. Section 4.4 discusses the maximum likelihood estimation procedure for the model parameters and effects, as well as their confidence intervals. Section 4.5 presents simulation studies to examine the numerical behaviours of the two proposed models. Section 4.6 is devoted to the data application of the calcium

supplementation trial from ELEMENT study. Section 4.7 includes some concluding remarks on advantages and limitations of the proposed methodologies. Technical details can be found in Appendix C.

4.2 Method

This section presents two new models to describe the potential data generating mechanism of binary exposure. As a result, we generalize the definitions of the classical natural direct effect and natural indirect effect. These new concepts can provide better understanding and interpretation of the underlying intrinsic causal mechanism among binary exposure X, mediator M and outcome Y.

4.2.1 Effective Dose Model

We now present the effective dose model. Consider a randomized trial in that the allocation of treatment or exposure X follows a Bernoulli distribution with probability $p = 0.5$. Assume Z is a latent variable that represents the effective dose of the treatment that a subject would absorb. For a subject in the control arm ($X = 0$), the dose Z variable is naturally equal to 0; for a subject in the active drug arm ($X = 1$), the dose variable Z is assumed to follow a truncated normal distribution with mean a ($a > 0$) and variance 1 due to the fact that it takes positive values on the interval $(0, \infty)$, denoted as $TN_{(0,\infty)}(a, 1)$. Hence, the marginal distribution of Z is a mixture of point mass function at 0 and truncated normal distribution taking positive values; namely $Z \mid X = (1 - X)I[Z = 0] + XTN_{(0,+\infty)}(a, 1)$. To form a structured equation model (SEM) in a mediation analysis, we assume that continuous mediator M is normally distributed with mean αZ and variance σ_M^2, and that a continuous outcome Y is normally distributed with mean $\gamma Z + \beta M$ and variance σ_Y^2. Apparently, the

effective dose model is used to expand the commonly used SEM as follows:

$$X \sim Bern(p = 0.5), \qquad [Z \mid X] \sim (1 - X)I[Z = 0] + XTN_{(0,\infty)}(a, 1),$$
$$[M \mid Z\ ] \sim N(\alpha Z, \sigma_M^2), \qquad [Y \mid Z, M\ ] \sim N(\gamma Z + \beta M, \sigma_Y^2), \tag{4.1}$$

where a is pre-specified mean absorption rate of calcium.

4.2.2 Latent Exposure Model

We now turn to the latent exposure model for binary exposure X that arises from dichotomization. Assuming that a latent continuous exposure variable Z follows a normal distribution with mean μ and variance 1, we only observe a binary exposure variable X that is dichotomized from continuum Z. Exposure X may be interpreted as an observed categorical variable of the latent Z. Similar to effective dose model above, there exists a latent variable Z of the underlying exposure. The latent exposure model takes the following form:

$$Z \sim N(\mu, 1), \qquad [X \mid Z\ ] = I(Z > 0),$$
$$[M \mid Z\ ] \sim N(\alpha Z, \sigma_M^2), \qquad [Y \mid Z, M\ ] \sim N(\gamma Z + \beta M, \sigma_Y^2). \tag{4.2}$$

4.3 Mediation Pathway Analysis

Given two models (4.1) and (4.2), we now derive the natural direct effect (NDE) and natural indirect effect (NIE). The details of the derivations can be found in Appendix C.1. As a result, we are able to generalize these classical concepts; we define the generalized natural direct effect (GNDE) and generalized natural indirect effect (GNIE) based on the latent exposure models (4.1) and (4.2). We discuss some connections of the new concepts to the classical NDE and NIE.

4.3.1 Natural Direct and Indirect Effects

Let us begin with the potential outcome framework from the causal inference literature (*Robins and Greenland*, 1992; *Pearl*, 2013; *VanderWeele and Vansteelandt*, 2009). Adopting the counterfactual notions (*VanderWeele and Vansteelandt*, 2009), let $Y(x, m)$ denote the counterfactual outcome that would have been observed for the subject had the exposure X been set to value x and mediator M to value m; let $M(x)$ be the counterfactual value of mediator had the exposure X been set to value x. By definition, $\mathrm{E}\{Y(x_a, M(x_b))\}$ is the expected outcome of Y had the exposure been set to value x_a and mediator been set to value $M(x_b)$, and moreover we have

$$\mathrm{E}\{Y(x_a, M(x_b))\} = \mathrm{E}_M[\mathrm{E}_Y\{Y|M, X = x_a\}|X = x_b].$$

According to *VanderWeele and Vansteelandt* (2009), the natural direct effect (NDE), the natural indirect effect (NIE), and the total effect (TE) for a change of X from x_0 to x_1 are given by, respectively,

$$\begin{aligned}
\mathrm{NDE}(x_0, x_1) &= \mathrm{E}\{Y(x_1, M(x_0))\} - \mathrm{E}\{Y(x_0, M(x_0))\}, \\
\mathrm{NIE}(x_0, x_1) &= \mathrm{E}\{Y(x_1, M(x_1))\} - \mathrm{E}\{Y(x_1, M(x_0))\}, \\
\mathrm{TE}(x_0, x_1) &= \mathrm{E}\{Y(x_1, M(x_1))\} - \mathrm{E}\{Y(x_0, M(x_0))\}.
\end{aligned}$$

Proposition IV.1. *Under the* ***Effective Dose Model*** (4.1)*, the causal effects for a change of X from level 0 to level 1 are given by, respectively*

$$\text{NDE} = \left\{ \frac{\sigma_M^2 a}{\sigma_M^2 + \alpha^2} + \frac{\sigma_M}{\sqrt{\sigma_M^2 + \alpha^2}} \text{E}_M(A_1 \mid X = 0) \right\} \gamma,$$
$$\text{NIE} = \alpha \left(a + \frac{\phi(a)}{\Phi(a)} \right) \left(\beta + \frac{\alpha\gamma}{\sigma_M^2 + \alpha^2} \right) + \frac{\gamma\sigma_M}{\sqrt{\sigma_M^2 + \alpha^2}} \times$$
$$\{\text{E}_M(A_1 \mid X = 1) - \text{E}_M(A_1 \mid X = 0)\},$$

where $A_1 = \frac{\phi\left(\frac{\alpha M + \sigma_M^2 a}{\sigma_M \sqrt{\alpha^2 + \sigma_M^2}} \right)}{\Phi\left(\frac{\alpha M + \sigma_M^2 a}{\sigma_M \sqrt{\alpha^2 + \sigma_M^2}} \right)}$. *Moreover, we have the following sufficient conditions: if* $\gamma = 0$, *then* NDE $= 0$; *if* $\alpha = 0$ *or* $\gamma = \beta = 0$, *then* NIE $= 0$.

Proposition IV.2. *Under the* ***Latent Exposure Model*** (4.2), *the causal effects for a change of* X *from level 0 to level 1 are given by, respectively*

$$\text{NDE} = \frac{\gamma\sigma_M}{\sqrt{\sigma_M^2 + \alpha^2}} \text{E}_M(A_2 + A_3 \mid X = 0),$$
$$\text{NIE} = \frac{\alpha\phi(\mu)}{\Phi(\mu)(1 - \Phi(\mu))} \left(\beta + \frac{\alpha\gamma}{\sigma_M^2 + \alpha^2} \right) + \frac{\gamma\sigma_M}{\sqrt{\sigma_M^2 + \alpha^2}} \times$$
$$\{\text{E}_M(A_2 \mid X = 1) - \text{E}_M(A_2 \mid X = 0)\},$$

where $A_2 = \frac{\phi\left(\frac{\alpha M + \sigma_M^2 \mu}{\sigma_M \sqrt{\alpha^2 + \sigma_M^2}} \right)}{\Phi\left(\frac{\alpha M + \sigma_M^2 \mu}{\sigma_M \sqrt{\alpha^2 + \sigma_M^2}} \right)}$ *and* $A_3 = \frac{\phi\left(\frac{\alpha M + \sigma_M^2 \mu}{\sigma_M \sqrt{\alpha^2 + \sigma_M^2}} \right)}{1 - \Phi\left(\frac{\alpha M + \sigma_M^2 \mu}{\sigma_M \sqrt{\alpha^2 + \sigma_M^2}} \right)}$. *Moreover, we have the following sufficient conditions: if* $\gamma = 0$, *then* NDE $= 0$; *if* $\alpha = 0$ *or* $\gamma = \beta = 0$, *then* NIE $= 0$.

4.3.2 Generalized Natural Direct and Indirect Effects

Alternatively, we focus on causal effects under latent exposure models (4.1) and (4.2), with a change of Z from z_a to z_b:

$$\text{E}\{Y(Z_a, M(Z_b))\} = \text{E}_M[\text{E}_Y\{Y|M, Z = z_a\}|Z = z_b].$$

We term the resulting NDE/NIE/GTE as Generalized Natural Direct/Indirect/Total Effects. Precisely, for a change of Z from z_0 to z_1, GNDE, GNIE and GTE defined given by, respectively

$$\begin{aligned}
\text{GNDE}(z_0, z_1) &= \text{E}[Y(z_1, M(z_0)] - \text{E}[Y(z_0, M(z_0)], \\
\text{GNIE}(z_0, z_1) &= \text{E}[Y(z_1, M(z_1)] - \text{E}[Y(z_1, M(z_0)], \\
\text{GTE}(z_0, z_1) &= \text{E}[Y(z_1, M(z_1)] - \text{E}[Y(z_0, M(z_0)].
\end{aligned}$$

Since the causal effects are to make a comparison between treatment group and control group, let z_0 represent a possible latent value for a subject in the control group ($X = 0$), and let z_1 represent a latent value for a subject in the treatment group ($X = 1$). This means that for the effective dose model (4.1), all z_0 values equal to 0.

Proposition IV.3. *Under the* ***Effective Dose Model*** (4.1)*, the generalized causal effects for a change of Z from $z_0 = 0$ to z_1 are given by, respectively*

$$\text{GNDE}(z_0, z_1) = \gamma(z_1 - z_0) = \gamma z_1, \qquad \text{GNIE}(z_0, z_1) = \alpha\beta(z_1 - z_0) = \alpha\beta z_1,$$

where $z_1 = \text{E}(Z \mid X = 1) = a + \frac{\phi(a)}{\Phi(a)}$. Moreover, we have the sufficient and necessary conditions: GNDE $= 0$ *if and only if $\gamma = 0$;* GNIE $= 0$ *if and only if $\alpha = 0$ or $\beta = 0$.*

Proposition IV.4. *Under the* ***Latent Exposure Model*** (4.2)*, the generalized causal effects for a change of Z from z_0 to z_1 are given by, respectively*

$$\text{GNDE}(z_0, z_1) = \gamma(z_1 - z_0), \qquad \text{GNIE}(z_0, z_1) = \alpha\beta(z_1 - z_0),$$

where $z_1 - z_0 = \text{E}(Z \mid X = 1) - \text{E}(Z \mid X = 0) = \frac{\phi(\mu)}{\Phi(\mu)(1-\Phi(\mu))}$. Moreover, we have the sufficient and necessary conditions: GNDE $= 0$ *if and only if $\gamma = 0$;* GNIE $= 0$ *if and only if $\alpha = 0$ or $\beta = 0$.*

For both latent models for binary exposure (4.1) and (4.2), the parameter γ characterizes GNDE, while α and β characterize GNIE. Their interpretation are identical to those given in the classical SEM setting. In contrast to the sufficient conditions for NIE $= 0$ in Propositions IV.1 and IV.2, parameter γ determines the NIE. These are not necessary conditions. The new concepts of GNDE and GNIE apparently provide better interpretations of causal effects for the roles of parameters α, β and γ.

4.4 Estimation

This section presents the estimation of parameters and causal effects under both models. Due to the involvement of the latent exposure variable, the key step is to derive the observed data likelihood by integrating out the latent exposure variable Z. Then, by minimizing the negative log-likelihood function, we obtain parameter estimates, using the R function "optim". In particular, we use the algorithm "L-BFGS-B" (*Byrd et al.*, 1995) for the optimization. This algorithm allows us to specify the lower and/or upper bounds of the parameters, such as the constraints $\sigma_M^2 > 0$ and $\sigma_Y^2 > 0$.

Standard errors of the model parameter estimates are computed from the Fisher Information matrix, and standard errors of the generalized causal effects are obtained via the delta method (*Dorfman*, 1938). Consequently, a 95% confidence interval (CI) is constructed under the standard large-sample theory of maximum likelihood estimation (MLE), with point estimates and their standard errors.

4.4.1 Observed Likelihood in Effective Dose Model

Some key steps required by the calculation of the observed log likelihood function are given in this section. Consider a dataset consist of n observations (X_i, M_i, Y_i), as well as the latent variable Z_i, defined in model (4.1). The complete data likelihood for

subject i is given by

$$\pi(Y_i, M_i, Z_i \mid X_i) = \pi(Y_i \mid M_i, Z_i)\pi(M_i \mid Z_i)\pi(Z_i \mid X_i),$$

where $\pi(\cdot)$ is a generic notation of a density function.

When $X_i = 0$,

$$\pi(Y_i \mid M_i, Z_i)\pi(M_i \mid Z_i)\pi(Z_i \mid X_i) \propto \frac{1}{\sqrt{\sigma_M^2\sigma_Y^2}} exp\left\{-\frac{(Y_i - \beta M_i)^2}{2\sigma_Y^2} - \frac{M_i^2}{2\sigma_M^2}\right\}. \quad (4.3)$$

When $X_i = 1$,

$$\begin{aligned}
&\int_0^{+\infty} \pi(Y_i \mid M_i, Z_i)\pi(M_i \mid Z_i)\pi(Z_i \mid X_i = 1)dZ_i \\
\propto &\int_0^{+\infty} \frac{1}{\sqrt{\sigma_M^2\sigma_Y^2}} exp\left\{-\frac{(Y_i - \gamma Z_i - \beta M_i)^2}{2\sigma_Y^2} - \frac{(M_i - \alpha Z_i)^2}{2\sigma_M^2}\right\} \frac{\phi(Z_i - a_i)}{\Phi(a_i)} dZ_i \\
= &\frac{\Phi\left(\frac{\frac{\gamma(Y_i - \beta M_i)}{\sigma_Y^2} + \frac{\alpha M_i}{\sigma_M^2} + a_i}{\sqrt{\frac{\gamma^2}{\sigma_Y^2} + \frac{\alpha^2}{\sigma_M^2} + 1}}\right)}{\sqrt{\sigma_M^2\sigma_Y^2\left(\frac{\gamma^2}{\sigma_Y^2} + \frac{\alpha^2}{\sigma_M^2} + 1\right)}} exp\left\{-\frac{(Y_i - \beta M_i)^2}{2\sigma_Y^2} - \frac{M_i^2}{2\sigma_M^2} + \frac{\left(\frac{\gamma(Y_i - \beta M_i)}{\sigma_Y^2} + \frac{\alpha M_i}{\sigma_M^2} + a_i\right)^2}{2\left(\frac{\gamma^2}{\sigma_Y^2} + \frac{\alpha^2}{\sigma_M^2} + 1\right)}\right\}.
\end{aligned} \quad (4.4)$$

According to (4.3) and (4.4), for a random sample of n subjects, we have the log-likelihood function,

$$\begin{aligned}
\ell(\boldsymbol{\theta}) = &-\frac{n}{2}(\log\sigma_M^2 + \log\sigma_Y^2) - \frac{\sum_{i=1}^n (Y_i - \beta M_i)^2}{2\sigma_Y^2} - \frac{\sum_{i=1}^n M_i^2}{2\sigma_M^2} - \frac{n_1}{2}\log\left(\frac{\gamma^2}{\sigma_Y^2} + \frac{\alpha^2}{\sigma_M^2} + 1\right) \\
&+ \frac{\sum_{i:X_i=1}\left(\frac{\gamma(Y_i - \beta M_i)}{\sigma_Y^2} + \frac{\alpha M_i}{\sigma_M^2} + a_i\right)^2}{2\left(\frac{\gamma^2}{\sigma_Y^2} + \frac{\alpha^2}{\sigma_M^2} + 1\right)} + \sum_{i:X_i=1}\log\Phi\left(\frac{\frac{\gamma(Y_i - \beta M_i)}{\sigma_Y^2} + \frac{\alpha M_i}{\sigma_M^2} + a_i}{\sqrt{\frac{\gamma^2}{\sigma_Y^2} + \frac{\alpha^2}{\sigma_M^2} + 1}}\right),
\end{aligned}$$

where $n_1 = \sum_{i=1}^{n} I(X_i = 1)$ is the total number of subjects in the treatment group. We minimize the negative log-likelihood function $-\ell(\boldsymbol{\theta})$ by the algorithm "L-BFGS-B" with respect to model parameters α, γ, β, σ_M^2 and σ_Y^2, and obtain the parameter estimates. The initial values to start the search are set at $\hat{\alpha}^{(0)} = 0$, $\hat{\gamma}^{(0)} = 0$, $\hat{\beta}^{(0)} = 0$, $\hat{\sigma}_M^{2(0)} = \widehat{\text{var}}(M)$, $\hat{\sigma}_Y^{2(0)} = \widehat{\text{var}}(Y)$.

4.4.2 Observed Likelihood in Latent Exposure Model

The joint distribution of Z_i, M_i and Y_i may be expressed as,

$$\begin{pmatrix} Z_i \\ M_i \\ Y_i \end{pmatrix} \sim N \left\{ \begin{pmatrix} \mu \\ \alpha\mu \\ \eta\mu \end{pmatrix}, \begin{pmatrix} 1 & \alpha & \eta \\ \alpha & \alpha^2 + \sigma_M^2 & \alpha\eta + \beta\sigma_M^2 \\ \eta & \alpha\eta + \beta\sigma_M^2 & \eta^2 + \beta^2\sigma_M^2 + \sigma_Y^2 \end{pmatrix} \right\}, \tag{4.5}$$

where $\eta := \gamma + \beta\alpha$. It follows that the density of X_i, M_i and Y_i is

$$\begin{aligned} \pi(X_i, M_i, Y_i) =& \int_{Z_i} \pi(Z_i, M_i, Y_i) I[X_i = I(Z_i > 0)] dZ_i \\ =& \int_{Z_i} \pi(Z_i \mid M_i, Y_i) \pi(M_i, Y_i) I[X_i = I(Z_i > 0)] dZ_i \\ =& \pi(M_i, Y_i) P\left(Z_i < 0 \mid M_i, Y_i\right)^{1-X_i} P\left(Z_i > 0 \mid M_i, Y_i\right)^{X_i}. \end{aligned}$$

The observed data log-likelihood function is

$$\begin{aligned} \ell(\boldsymbol{\theta}) =& \sum_{i=1}^{n} \log \pi(M_i, Y_i) + \sum_{i=1}^{n} \log P\left(Z_i < 0 \mid M_i, Y_i\right) \\ & - \sum_{i=1}^{n} X_i \log \frac{P\left(Z_i < 0 \mid M_i, Y_i\right)}{1 - P\left(Z_i < 0 \mid M_i, Y_i\right)}, \end{aligned}$$

where densities $\pi(M_i, Y_i)$ and $\pi(Z_i \mid M_i, Y_i)$ can be straightforwardly obtained from (4.5). We then minimize the negative log-likelihood function $-\ell(\boldsymbol{\theta})$ by the algorithm "L-BFGS-B" in the R function optim with respect to model parameters α, γ, β, μ,

σ_M^2 and σ_Y^2 to obtain the parameter estimates. The initial values $\hat{\alpha}^{(0)}$ $\hat{\gamma}^{(0)}$, $\hat{\beta}^{(0)}$, $\hat{\mu}^{(0)}$, $\hat{\sigma}_M^{2(0)}$, $\hat{\sigma}_Y^{2(0)}$ are given by the moment estimates. In particular, $\hat{\mu}^{(0)}$ is estimated by the quantile of the standard normal distribution with the probability equal to the observed proportion of $X_i = 1$. Since $\mathrm{E}(Z_i \mid X_i = 0) = \mu - \frac{\phi(\mu)}{1-\Phi(\mu)}$ and $\mathrm{E}(Z_i \mid X_i = 1) = \mu + \frac{\phi(\mu)}{\Phi(\mu)}$, let $Z_i = \hat{\mu}^{(0)} - \frac{\phi(\hat{\mu}^{(0)})}{1-\Phi(\hat{\mu}^{(0)})}$ for $X_i = 0$, and $Z_i = \hat{\mu}^{(0)} + \frac{\phi(\hat{\mu}^{(0)})}{\Phi(\hat{\mu}^{(0)})}$ for $X_i = 1$. Then $\hat{\alpha}^{(0)}$ $\hat{\gamma}^{(0)}$, $\hat{\beta}^{(0)}$, $\hat{\sigma}_M^{2(0)}$, and $\hat{\sigma}_Y^{2(0)}$ are obtained by fitting the two regression models in (4.2).

4.5 Simulation

We carry out extensive simulations to evaluate the performance of the MLE in the two proposed models (4.1) and (4.2). In particular, we calculate average bias, mean square error (MSE), empirical coverage of 95% confidence interval, empirical standard error (ESE) of ML estimates, and average standard error (ASE) of estimates for both parameters and causal effects. Let $\hat{\theta}_j$ be estimate of a parameter θ under the j-th simulated dataset, where J denotes the total number of datasets. Let θ_0 be the true value. Then average bias is $\frac{1}{J}\sum_{j=1}^{J}(\hat{\theta}_j - \theta_0)$; MSE is $\frac{1}{J}\sum_{j=1}^{J}(\hat{\theta}_j - \theta_0)^2$; coverage probability is $\frac{1}{J}\sum_{j=1}^{J} I(\hat{\theta}_j^l \le \theta_0 \le \hat{\theta}_j^u)$, where $\hat{\theta}_j^l$ and $\hat{\theta}_j^u$ are 95% lower and upper limits; ESE is $\sqrt{\frac{1}{J}\sum_{j=1}^{J}(\hat{\theta}_j - \bar{\hat{\theta}})^2}$, where $\bar{\hat{\theta}} = \frac{1}{J}\sum_{j=1}^{J}\hat{\theta}_j$; and ASE is $\frac{1}{J}\sum_{j=1}^{J} se(\hat{\theta}_j)$, where $se(\hat{\theta}_j)$ is the standard error calculated from the Fisher Information matrix.

The sample size n varies over 200, 500 and 1000, and for each sample size, we generated $J = 10,000$ datasets.

4.5.1 Simulation with Effective Dose Model

Binary exposure variable X_i is first simulated from a Bernoulli distribution with probability 0.5. When $X_i = 0$ is generated, $Z_i = 0$; when $X_i = 1$, Z_i is drawn from a truncated normal distribution with mean $a_i = 0$ and variance 1 on interval $(0, \infty)$.

Then, we generate mediator and outcome by,

$$M_i = 0.3Z_i + \epsilon_{M_i}, \quad Y_i = 0.6Z_i + 0.5M_i + \epsilon_{Y_i}, \tag{4.6}$$

where $\epsilon_{M_i} \overset{\text{i.i.d.}}{\sim} N(0, 0.5)$, and $\epsilon_{Y_i} \overset{\text{i.i.d.}}{\sim} N(0, 0.5)$.

Table 4.1 reports the simulation results based on the above evaluation criteria. Average bias and MSE decrease as the sample size n increases from 200 to 1000, suggesting the accuracy of the estimates is improved with an increase of the sample size. The coverage probability for both model parameters and causal effects are all close to the nominal level 95.0%, ranging from 94.1% to 95.3%, indicating that the standard errors are well estimated. Both ESE and ASE are close and decrease as the sample size n increases, confirming the large-sample behaviours of the MLE. In summary, the simulation results demonstrate clearly that the MLE provides accurate estimation and appropriate statistical inference, even when the sample size is as low as 200.

In our simulation study setting, the initial values $\{\hat{\alpha}^{(0)} = 0, \hat{\gamma}^{(0)} = 0, \hat{\beta}^{(0)} = 0, \hat{\sigma}_M^{2(0)} = \widehat{\text{var}}(M), \hat{\sigma}_Y^{2(0)} = \widehat{\text{var}}(Y)\}$ seem to be reasonable for the "optim" function to search for the minimum. The "L-BFGS-B" algorithm does not fail even once to get a convergent solution over 30,000 datasets. The associated computation cost is very minimal and indeed negligible; a runtime of around 0.11 seconds is used to compute a dataset with 500 subjects.

4.5.2 Simulation with Latent Exposure Model

For the latent exposure model (4.2), we first generate latent variable Z_i from a normal distribution with mean 0.5 and variance 1, followed by dichotomizing Z_i into $X_i = 1$ if $Z_i > 0$, or $X_i = 0$ otherwise. Then we generate mediator M_i and outcome Y_i according to regression models in (4.6). The latent exposure model (4.2) has one

Table 4.1: Bias and MSE of parameter estimates and causal effects, and the coverage of the 95% confidence interval for the **effective dose model**; the sample size varies over 200, 500 and 1000, and the results are obtained by 10,000 replicates.

		Bias ($\times 10^{-3}$)			MSE ($\times 10^{-3}$)			Coverage (%)		
	True	200	500	1000	200	500	1000	200	500	1000
α	0.30	-3.11	-0.76	-0.32	7.54	3.01	1.50	95.0	95.0	95.3
γ	0.60	-2.31	-1.27	-0.32	9.69	3.95	1.94	94.9	94.7	95.0
β	0.50	1.81	0.52	0.09	5.93	2.36	1.19	95.0	94.9	94.9
σ_M^2	0.50	-2.15	-0.79	-0.94	2.68	1.07	0.54	94.2	94.5	94.5
σ_Y^2	0.50	-6.14	-1.74	-1.02	3.13	1.29	0.64	94.1	94.5	95.0
GNDE	0.48	-1.84	-1.01	-0.25	6.17	2.51	1.24	94.9	94.7	95.0
GNIE	0.12	-1.78	-0.55	-0.30	1.35	0.53	0.26	94.8	95.1	95.0
		ESE ($\times 10^{-2}$)			ASE ($\times 10^{-2}$)					
	True	200	500	1000	200	500	1000			
α	0.30	8.68	5.48	3.87	8.70	5.50	3.89			
γ	0.60	9.84	6.28	4.41	9.78	6.19	4.37			
β	0.50	7.70	4.86	3.45	7.67	4.85	3.43			
σ_M^2	0.50	5.18	3.28	2.33	5.17	3.28	2.32			
σ_Y^2	0.50	5.56	3.59	2.52	5.57	3.56	2.52			
GNDE	0.48	7.85	5.01	3.52	7.81	4.94	3.49			
GNIE	0.12	3.67	2.31	1.61	3.70	2.31	1.62			

more parameter than the effective dose model (4.1), which is mean parameter μ set at 0.5; the other five parameters are set the same as those given in Section 4.5.1.

Table 4.2 summarizes the simulation results based on 10,000 simulated datasets. Average bias, MSE, ESE and ASE decrease as the sample size n increases, and the coverage rate of 95% confident interval ranges from 93.8% to 95.4%, suggesting that the inference procedure is appropriate. Once again, our moment-based initial values appear reasonable to begin the "optim" function, as it is able to find the minimum in all 30,000 datasets. The computation cost is very minimal with a runtime of around 0.3 minutes for a dataset of 500 subjects on a standard personal computer.

4.6 Data Application

We present a data application from the ELEMENT study described in the Section of Introduction. The sample used in this mediation analysis is 368 mother-child

Table 4.2: Bias and MSE of parameter estimates and causal effects, and the coverage of the 95% confidence interval for the **latent exposure model**; the sample size varies over 200, 500 and 1000, and the results are obtained by 10,000 replicates.

		Bias ($\times 10^{-3}$)			MSE ($\times 10^{-3}$)			Coverage (%)		
	True	200	500	1000	200	500	1000	200	500	1000
α	0.30	-0.22	-0.71	-0.50	3.25	1.30	0.64	94.7	94.9	95.0
γ	0.60	-0.50	-0.05	-0.18	5.47	2.11	1.06	94.3	95.2	95.2
β	0.50	-2.63	-0.10	-0.16	8.04	3.19	1.57	94.8	94.9	94.8
μ	0.50	2.42	0.40	0.23	7.83	3.09	1.58	95.3	95.2	94.6
σ^2_M	0.50	-3.71	-0.92	-0.80	3.00	1.19	0.61	94.2	94.9	94.6
σ^2_Y	0.50	-8.38	-2.72	-1.21	4.65	1.86	0.94	93.8	94.3	94.7
GNDE	0.99	0.17	0.23	-0.13	14.36	5.52	2.79	94.5	95.2	95.4
GNIE	0.25	-3.75	-1.52	-0.93	3.03	1.19	0.57	94.0	94.5	94.7
		ESE ($\times 10^{-2}$)			ASE ($\times 10^{-2}$)					
	True	200	500	1000	200	500	1000			
α	0.30	5.70	3.60	2.53	5.66	3.59	2.54			
γ	0.60	7.39	4.59	3.26	7.28	4.61	3.26			
β	0.50	8.97	5.65	3.96	8.84	5.57	3.94			
μ	0.50	8.85	5.56	3.97	8.81	5.57	3.93			
σ^2_m	0.50	5.47	3.45	2.47	5.49	3.48	2.46			
σ^2_y	0.50	6.76	4.30	3.06	6.76	4.32	3.06			
GNDE	0.99	11.98	7.43	5.28	11.90	7.52	5.32			
GNIE	0.25	5.49	3.45	2.39	5.51	3.43	2.41			

pairs. The study is a randomized control trial where 184 women are supplemented with calcium ($X = 1$) and 184 women are supplemented with placebo ($X = 0$) during pregnancy. The calcium supplement is administered to suppress the release of the bone lead into the blood circulation during pregnancy (*Perng et al.*, 2019). In this way, prenatal lead exposure to offspring could be mitigated for the treatment group. Young children absorb lead more easily than adults (*Boeckx*, 1986), and exposure to lead can pose significant risks to offspring, such as preterm labor (*Vigeh et al.*, 2011), low birth weight (*Zhu et al.*, 2010), and neurodevelopmental delays (*Vigeh et al.*, 2014). One study (*Scholl et al.*, 2014) examined the relationships of calcium metabolism and birth weight. Therefore we hypothesize that mitigating the blood lead level by calcium supplementation , would affect birth weight in our study population. In this analysis, exposure X is calcium supplement or not, mediator M is mother's blood lead level measured at third trimester, and outcome Y is offspring's

birth weight. The absorption rate value a_i is estimated for each subject, by $a_i = \exp\left(-\frac{\text{calcium}_i}{\max_i(\text{calcium}_i)}\right)$, where calcium_i is the calcium baseline level measured at first trimester. Therefore, the higher the baseline calcium level, the lower the absorption rate a_i during supplementation.

The blood lead level has three observations with negative values due to below detection limit, and their values are changed to 0. The maximum value for blood lead is 34.3 ug/dL with a mean of 4.3 ug/dL. The birth weight ranges from 1.9 kg to 4.2 kg, with a mean of 3.2 kg. The initial values are set at $\{\hat{\alpha}^{(0)} = 0, \hat{\gamma}^{(0)} = 0, \hat{\beta}^{(0)} = 0, \hat{\sigma}_M^{2(0)} = \widehat{\text{var}}(M) = 12.6, \hat{\sigma}_Y^{2(0)} = \widehat{\text{var}}(Y) = 0.16\}$. The mediator and outcome are first centered to their respective means before the mediation analysis. In this way, the intercepts are removed from the SEM.

Table 4.3 summarizes the estimates and 95% confidence intervals for both model parameter and generalized causal effects. As expected, parameter α is estimated to be negatively associated with the blood level, meaning that on average the treatment group have a lower blood lead level. The CIs for both GNDE and GNIE contain 0 so the causal effects are not significant. The estimates of γ and β are close to zero. In conclusion, we do not have strong evidence from the data to support that there is a significant generalized direct effect, nor a significant generalized indirect effect.

Table 4.3: Parameter and effect estimates and 95% confidence interval for the effect of calcium supplementation on birth weight with potential mediator of maternal blood lead at third trimester

	Estimate	SE	95% CI
α	-0.2227	0.2342	(-0.6818, 0.2364)
γ	0.0050	0.0268	(-0.0476, 0.0575)
β	0.0013	0.0058	(-0.0101, 0.0127)
σ_M^2	12.6468	0.9394	(10.8057, 14.4880)
σ_Y^2	0.1558	0.0115	(0.1333, 0.1784)
GNDE	0.0055	0.0298	(-0.0529, 0.0639)
GNIE	-0.0003	0.0015	(-0.0032, 0.0026)

4.7 Concluding Remarks

This chapter develops a new framework of causal mediation analysis under the structural equation models, in which mediator M and outcome Y are continuous and normally distributed, and exposure X is binary. We present a useful extension of the classical SEM framework by a latent exposure model for the binary exposure under the premise that the "true" underlying dose of exposure is continuous and latent. We propose two models: the effective dose model and latent exposure model to characterize the mechanism as to how the binary exposure arises from a latent continuum. The two models have different interpretations: the effective dose model is suitable for a setting of the randomized control trial, while the latent exposure model is applicable to the observational study, where binary exposure X is recorded by dichotomization. We reestablish the concepts of direct and indirect effects under the latent variable models and propose a generalization of NDE and NIE based on changes of latent exposure Z. We showed that the generalized concepts offer more interpretable causal effects. The simulation studies show that the MLE in both models can achieve accurate estimates of parameter and causal mediation effects and provide numerically stable results with fast computation.

CHAPTER V

Future Plan

Chapter II developed a likelihood ratio test to jointly test for the mediation pathway involving multiple correlated mediators. This is useful in practice to deal with a cluster of mediators that often arise from studies with omics data. It has been shown that the joint hypothesis testing method improves power over existing methods under properly controlled type I error. This is a test method, along with the R package, that we would like to recommended to practitioners.

Below are a few future directions of methodological development that further extends the joint test.

(i) An extension of the current framework to the case of mediation pathway analysis with binary outcome. When the logistic regression is used to a binary outcome, the classical structural equation model becomes more complicated due to the presence of nonlinear models (e.g. logistic model). Such deviation from the linearity gives rise to substantial technical challenges that call for innovative solutions. One major challenge pertains to the form of parametrization for causal pathway, which is no longer the same expression of $\alpha\beta$ in the classical linear structural equation model. This extension may be carried out through the generalized structural equation model (GSEM) proposed in Chapter III. We plan to first extend the GSEM to accommodate multiple mediators and then to

re-establish the likelihood ratio test under the constrained maximum likelihood estimation or profile maximum likelihood estimation.

(ii) We plan to investigate some remedies that can alleviate the conservatism of type I error for the case of $\boldsymbol{\alpha} = \boldsymbol{\beta} = 0$. To do so, the key is to label this scenario from the other two null scenarios, $\boldsymbol{\alpha} = 0, \boldsymbol{\beta} \neq 0$ and $\boldsymbol{\alpha} \neq 0, \boldsymbol{\beta} = 0$. The conservatism is caused by different convergence rates for the test statistic in these three null situations. One possible solution would be to invoke a certain parameter fusion technique to identify these parameter groups. This relates to a simultaneous operation of estimation and clustering, which may be solved by the means of mixed integer programming.

(iii) An extension to the current framework is to add the extra terms for exposure and mediator interactions, and solve the NIE under the null hypothesis. Other future directions include investigating the influence of model mis-specification, such as the effects of unobserved confounders.

Chapter III developed a unified framework of generalized structural equation models (GSEM) for mediation analysis with mixed data types. We established this framework under one-dimensional exposure, one-dimensional mediator and one-dimensional outcome. There are many future directions to extend this methodology. For example,

(i) an extended GSEM that accommodate multi-dimensional mediators. As pointed out in Chapter II, handling multiple correlated mediators is a routine task in practice, and there are no systematic methodologies to perform joint analyses of such data of mixed types in the current literature. This extension can provide a needed toolbox to practitioners with improved statistical power, so to obtain new scientific findings.

(ii) An alternative formulation of GSEM may be considered via vine copulas. In Chapter III, we adopted the Gaussian copula to formulate a DAG topology for

the causal relationships. There are many other more flexible copula models, such as vine copulas, that can improve the goodness-of-fit of GSEMs in the analysis of data of mixed types. Naturally, there is a need of model selection; which copula model to be chosen for the formulation of GSEMs.

(iii) Other useful extensions include allowing the interaction terms between exposure and mediator, which are often encountered in practice; the influences of unobserved latent confounding factors as well as the mis-specification of parametric models.

Chapter IV developed two models that aim to characterize the underlying variation of personal exposure via the invocation of a latent exposure Z when the observed exposure (treatment) is binary. We proposed new concepts of generalized NDE and generalized NIE based on the latent exposure variable, which indeed give rise to better interpretation of causal effects. For our short-term future work, we plan to continue our methodological development in the following areas.

(i) We will incorporate confounding variables into both effective dose model and latent exposure model, and update the maximum likelihood estimation procedure in these expanded models.

(ii) We will develop a systematic approach to assess how different choices of distributional assumption for Z may influence conclusions of statistical inference on causal effects.

(iii) We will extend these two models in the framework of GSEMs developed in Chapter III to perform mediation analysis with data of mixed types.

(iv) We will also allow the exposure-mediator interactions, and examine the influences for unobserved latent confounding factors and mis-specification of the models.

APPENDICES

APPENDIX A

Supplement for Chapter II

A.1 Information Matrix

$$\begin{aligned}\mathbf{I}(\boldsymbol{\theta}) &= -\mathrm{E}\left(\frac{1}{n}\frac{\partial^2 \ell(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}\boldsymbol{\theta}^\top}\right) \\ &= \begin{pmatrix} \frac{1}{n}\boldsymbol{\Sigma}_M^{-1} \otimes \mathbf{B}^\top\mathbf{B}_{(L+1)Q\times(L+1)Q} & \mathbf{0}_{(L+1)Q\times(Q+L+1)} \\ \mathbf{0}_{(Q+L+1)\times(L+1)Q} & \frac{1}{n\sigma_y^2}\mathrm{E}(\mathbf{W}^\top\mathbf{W})_{(Q+L+1)\times(Q+L+1)} \end{pmatrix},\end{aligned}$$

where

$$\mathrm{E}(\mathbf{W}^\top\mathbf{W}) = \begin{pmatrix} \bar{\boldsymbol{\alpha}}^\top\mathbf{B}^\top\mathbf{B}\bar{\boldsymbol{\alpha}} + n\boldsymbol{\Sigma}_M & \bar{\boldsymbol{\alpha}}^\top\mathbf{B}^\top\mathbf{V} \\ \mathbf{V}^\top\mathbf{B}\bar{\boldsymbol{\alpha}} & \mathbf{V}^\top\mathbf{V} \end{pmatrix},$$

and $\mathbf{V}_{n\times(L+1)} = (\mathbf{Z}_1, \ldots, \mathbf{Z}_L, \mathbf{X})$.

A.2 Proof of Lemma II.2

First, we prove the part (i) of Lemma II.2. Recall that

$$\mathbf{H}(\boldsymbol{\theta}) = \nabla_{\boldsymbol{\theta}}\dot{h}(\boldsymbol{\theta}) = \begin{pmatrix} \mathbf{0}_{(L+1)Q\times(L+1)Q} & \tilde{\mathbf{H}}_{(L+1)Q\times(Q+L+1)} \\ \tilde{\mathbf{H}}^\top_{(Q+L+1)\times(L+1)Q} & \mathbf{0}_{(Q+L+1)\times(Q+L+1)} \end{pmatrix},$$

where

$$\tilde{\mathbf{H}}_{(L+1)Q\times(Q+L+1)} = \begin{pmatrix} \mathbf{I}_Q & \mathbf{0}_{Q\times(L+1)} \\ \mathbf{0}_{LQ\times Q} & \mathbf{0}_{LQ\times(L+1)} \end{pmatrix}.$$

Then, we have $\mathbf{H}^2(\boldsymbol{\theta}) = \text{Block-diag}\left(\tilde{\mathbf{H}}\tilde{\mathbf{H}}^\top, \tilde{\mathbf{H}}^\top\tilde{\mathbf{H}}\right)$.

Since $\mathbf{H}^2(\boldsymbol{\theta})$ is a diagonal matrix, and it has $2Q$ 1's and $(LQ+L+1)$ 0's on diagonal, implying that $\mathbf{H}^2(\boldsymbol{\theta})$ has $2Q$ nonzero eigenvalues equal to 1, and $(LQ+L+1)$ zero eigenvalues. This shows that $\mathbf{H}(\boldsymbol{\theta})$ has $2Q$ nonzero eigenvalues with their absolute values being 1. Note that $tr(\mathbf{H}(\boldsymbol{\theta})) = 0$, implying $h_1 = \cdots = h_Q = 1$, $h_{Q+1} = \cdots = h_{2Q} = -1$.

Now let us prove the part (ii) of Lemma II.2. According to Theorem 1.4 in (*Lu and Pearce*, 2000), matrix $\mathbf{A}(\boldsymbol{\theta}) = \mathbf{I}(\boldsymbol{\theta})^{-\frac{1}{2}}\mathbf{H}(\boldsymbol{\theta})\mathbf{I}(\boldsymbol{\theta})^{-\frac{1}{2}}$ has Q positive eigenvalues, Q negative eigenvalues and the rest eigenvalues are zero, due to the fact that the eigenvalues of $\mathbf{I}(\boldsymbol{\theta})^{-\frac{1}{2}}$ are all positive. Thus, the $2Q$ nonzero eigenvalues of $\mathbf{A}(\boldsymbol{\theta})$, $\upsilon_1 \geq \upsilon_2 \geq \cdots \geq \upsilon_Q > 0 > \upsilon_{Q+1} \geq \cdots \geq \upsilon_{2Q}$. Let $\mathbf{I}_{11} = \frac{1}{n}\boldsymbol{\Sigma}_M^{-1} \otimes \mathbf{B}^\top\mathbf{B}$ and $\mathbf{I}_{22} = \frac{1}{n\sigma_y^2}\text{E}(\mathbf{W}^\top\mathbf{W})$. Then, we write $\mathbf{I}(\boldsymbol{\theta}) = \text{Block-diag}\left(\mathbf{I}_{11}, \mathbf{I}_{22}\right)$. It follows that,

$$\begin{aligned}
\mathbf{A}(\boldsymbol{\theta}) &= \mathbf{I}(\boldsymbol{\theta})^{-\frac{1}{2}}\mathbf{H}(\boldsymbol{\theta})\mathbf{I}(\boldsymbol{\theta})^{-\frac{1}{2}} \\
&= \begin{pmatrix} \mathbf{I}_{11}^{-\frac{1}{2}} & \mathbf{0} \\ \mathbf{0} & \mathbf{I}_{22}^{-\frac{1}{2}} \end{pmatrix} \begin{pmatrix} \mathbf{0} & \tilde{\mathbf{H}} \\ \tilde{\mathbf{H}}^\top & \mathbf{0} \end{pmatrix} \begin{pmatrix} \mathbf{I}_{11}^{-\frac{1}{2}} & \mathbf{0} \\ \mathbf{0} & \mathbf{I}_{22}^{-\frac{1}{2}} \end{pmatrix}, \\
&= \begin{pmatrix} \mathbf{0} & \mathbf{I}_{11}^{-\frac{1}{2}}\tilde{\mathbf{H}}\mathbf{I}_{22}^{-\frac{1}{2}} \\ \mathbf{I}_{22}^{-\frac{1}{2}}\tilde{\mathbf{H}}^\top\mathbf{I}_{11}^{-\frac{1}{2}} & \mathbf{0} \end{pmatrix}.
\end{aligned}$$

Consequently, $tr(\mathbf{A}(\boldsymbol{\theta})) = 0$, and $(\mathbf{I}_{11}^{-\frac{1}{2}}\tilde{\mathbf{H}}\mathbf{I}_{22}^{-\frac{1}{2}})^\top = \mathbf{I}_{22}^{-\frac{1}{2}}\tilde{\mathbf{H}}^\top\mathbf{I}_{11}^{-\frac{1}{2}}$. Let $\mathbf{I}_{11}^{-\frac{1}{2}}\tilde{\mathbf{H}}\mathbf{I}_{22}^{-\frac{1}{2}} = \mathbf{C}$. We have $\mathbf{A}^2(\boldsymbol{\theta}) = \text{Block-diag}\left(\mathbf{C}\mathbf{C}^\top, \mathbf{C}^\top\mathbf{C}\right)$.

The eigenvalues of $\mathbf{A}^2(\boldsymbol{\theta})$ are $\lambda(\mathbf{A}^2(\boldsymbol{\theta})) = (\lambda(\mathbf{C}\mathbf{C}^\top), \lambda(\mathbf{C}^\top\mathbf{C}))$, where the nonzero eigenvalues of $\mathbf{C}\mathbf{C}^\top$ and $\mathbf{C}^\top\mathbf{C}$ are the same. This indicates $\upsilon_1^2 = \upsilon_{2Q}^2, \upsilon_2^2 = \upsilon_{2Q-1}^2, \ldots, \upsilon_Q^2 = \upsilon_{Q+1}^2$. In summary, $\mathbf{A}(\boldsymbol{\theta})$ has $2Q$ nonzero eigenvalues in a descending order $\upsilon_1 \geq \upsilon_2 \geq \cdots \geq \upsilon_Q > 0 > \upsilon_{Q+1} \geq \cdots \geq \upsilon_{2Q}$, satisfying $\sum_{i=1}^{2Q}\upsilon_i = tr(\mathbf{A}(\boldsymbol{\theta})) =$

0. This implies that $\upsilon_1 = -\upsilon_{2Q}, \upsilon_2 = -\upsilon_{2Q-1}, \ldots, \upsilon_Q = -\upsilon_{Q+1}$. The proof is completed.

A.3 Proof of Lemma II.3

Let $\mathbf{D} = \{\mathbf{Y}, \mathbf{W}, \mathbf{M}, \mathbf{B}\} = \{\mathbf{d}_i\}_{i=1}^n$ denote all the observed data where $\mathbf{d}_i$ represents the data from subject i. Let $\mathbf{u}(\boldsymbol{\theta}) = \sum_{i=1}^n \nabla_{\boldsymbol{\theta}} \ell(\boldsymbol{\theta}; \mathbf{d}_i)$ denote the score function of length $2Q + p$, where $p = LQ + L + 1$, and let $\mathbf{U}(\boldsymbol{\theta}) = \nabla_{\boldsymbol{\theta}} \mathbf{u}(\boldsymbol{\theta})$ be the Hessian matrix. Under the regularity conditions, by the Central Limit Theorem, $\frac{1}{\sqrt{n}}\mathbf{u}(\boldsymbol{\theta}_0) \xrightarrow{d} \mathrm{N}\{0, \mathbf{I}(\boldsymbol{\theta}_0)\}$. Moreover, by the Law of Large Number, $-\frac{1}{n}\mathbf{U}(\boldsymbol{\theta}_0) \xrightarrow{p} \mathbf{I}(\boldsymbol{\theta}_0)$. Let $\{\tilde{\boldsymbol{\theta}}, \tilde{\lambda}\}$ be the solution of the Lagrange multiplier equation (2.3). Then, they satisfy the following two equations:

$$\mathbf{u}(\boldsymbol{\theta}) + n\lambda \dot{h}(\boldsymbol{\theta}) = \mathbf{0}_{2Q+p}, \tag{A.1}$$

$$h(\boldsymbol{\theta}) = 0. \tag{A.2}$$

It is straightforward to show that the k-th order ($k \geq 3$) partial derivatives of $h(\boldsymbol{\theta})$ are all zero for any $\boldsymbol{\theta}$. Taking the Taylor expansion on $h(\tilde{\boldsymbol{\theta}})$ in (A.2) around $\boldsymbol{\theta}_0$,

$$h(\tilde{\boldsymbol{\theta}}) = h(\boldsymbol{\theta}_0) + \dot{h}(\boldsymbol{\theta}_0)^\top (\tilde{\boldsymbol{\theta}} - \boldsymbol{\theta}_0) + \frac{1}{2}(\tilde{\boldsymbol{\theta}} - \boldsymbol{\theta}_0)^\top \mathbf{H}(\boldsymbol{\theta}_0)(\tilde{\boldsymbol{\theta}} - \boldsymbol{\theta}_0).$$

Since $\boldsymbol{\alpha} = \boldsymbol{\beta} = \mathbf{0}$, $h(\boldsymbol{\theta}_0) = h(\tilde{\boldsymbol{\theta}}) = 0$ and $\dot{h}(\boldsymbol{\theta}_0) = \mathbf{0}_{2Q+p}$, then

$$(\tilde{\boldsymbol{\theta}} - \boldsymbol{\theta}_0)^\top \mathbf{H}(\boldsymbol{\theta}_0)(\tilde{\boldsymbol{\theta}} - \boldsymbol{\theta}_0) = 0. \tag{A.3}$$

On the other hand, taking the Taylor expansion of (A.1) around $\boldsymbol{\theta}_0$ gives, subject to a high order error term,

$$\begin{aligned}
\mathbf{u}(\boldsymbol{\theta}_0) + \mathbf{U}(\boldsymbol{\theta}_0)(\tilde{\boldsymbol{\theta}} - \boldsymbol{\theta}_0) + n\tilde{\lambda}\left\{\dot{h}(\boldsymbol{\theta}_0) + \mathbf{H}(\boldsymbol{\theta}_0)(\tilde{\boldsymbol{\theta}} - \boldsymbol{\theta}_0)\right\} &\approx \mathbf{0}_{2Q+p}, \\
\mathbf{u}(\boldsymbol{\theta}_0) + \mathbf{U}(\boldsymbol{\theta}_0)(\tilde{\boldsymbol{\theta}} - \boldsymbol{\theta}_0) + n\mathbf{H}(\boldsymbol{\theta}_0)\left[\tilde{\lambda}(\tilde{\boldsymbol{\theta}} - \boldsymbol{\theta}_0)\right] &\approx \mathbf{0}_{2Q+p}, \\
\{\mathbf{U}(\boldsymbol{\theta}_0) + n\tilde{\lambda}\mathbf{H}(\boldsymbol{\theta}_0)\}(\tilde{\boldsymbol{\theta}} - \boldsymbol{\theta}_0) &\approx -\mathbf{u}(\boldsymbol{\theta}_0).
\end{aligned}$$

Given that the matrix $\mathbf{U}(\boldsymbol{\theta}) + n\lambda\mathbf{H}(\boldsymbol{\theta})$ is invertible for $\{\boldsymbol{\theta}, \lambda\}$ in the small neighborhood of $\{\boldsymbol{\theta}_0, 0\}$, we have

$$\begin{aligned}
(\tilde{\boldsymbol{\theta}} - \boldsymbol{\theta}_0) &\approx -\{\mathbf{U}(\boldsymbol{\theta}_0) + n\tilde{\lambda}\mathbf{H}(\boldsymbol{\theta}_0)\}^{-1}\mathbf{u}(\boldsymbol{\theta}_0), \qquad \text{(A.4)} \\
\sqrt{n}(\tilde{\boldsymbol{\theta}} - \boldsymbol{\theta}_0) &\approx \frac{1}{\sqrt{n}}\{-\mathbf{U}(\boldsymbol{\theta}_0)/n - \tilde{\lambda}\mathbf{H}(\boldsymbol{\theta}_0)\}^{-1}\mathbf{u}(\boldsymbol{\theta}_0) \\
&\approx \{\mathbf{I}(\boldsymbol{\theta}_0) - \tilde{\lambda}\mathbf{H}(\boldsymbol{\theta}_0)\}^{-1}\frac{\mathbf{u}(\boldsymbol{\theta}_0)}{\sqrt{n}}.
\end{aligned}$$

This implies that for any $\lambda^* \in \mathbb{R}$, the conditional distribution of $\tilde{\boldsymbol{\theta}}$ given $\tilde{\lambda} = \lambda^*$ is

$$\left[\sqrt{n}(\tilde{\boldsymbol{\theta}} - \boldsymbol{\theta}_0) \mid \tilde{\lambda} = \lambda^*\right] \to \mathrm{N}\left(\mathbf{0}, \{\mathbf{I}(\boldsymbol{\theta}_0) - \lambda^*\mathbf{H}(\boldsymbol{\theta}_0)\}^{-1}\mathbf{I}(\boldsymbol{\theta}_0)\{\mathbf{I}(\boldsymbol{\theta}_0) - \lambda^*\mathbf{H}(\boldsymbol{\theta}_0)\}^{-1}\right).$$

By plugging (A.4) into (A.3), we define

$$f(\tilde{\lambda}) = \mathbf{u}(\boldsymbol{\theta}_0)^\top\{\mathbf{U}(\boldsymbol{\theta}_0) + n\tilde{\lambda}\mathbf{H}(\boldsymbol{\theta}_0)\}^{-1}\mathbf{H}(\boldsymbol{\theta}_0)\{\mathbf{U}(\boldsymbol{\theta}_0) + n\tilde{\lambda}\mathbf{H}(\boldsymbol{\theta}_0)\}^{-1}\mathbf{u}(\boldsymbol{\theta}_0).$$

Taking derivative of $f(\tilde{\lambda})$ in $\tilde{\lambda}$ yields,

$$\begin{aligned}
\frac{\partial f(\tilde{\lambda})}{\partial\tilde{\lambda}} = \dot{f}(\tilde{\lambda}) = -2n\mathbf{u}(\boldsymbol{\theta}_0)^\top\{\mathbf{U}(\boldsymbol{\theta}_0) + n\tilde{\lambda}\mathbf{H}(\boldsymbol{\theta}_0)\}^{-1}\mathbf{H}(\boldsymbol{\theta}_0)\{\mathbf{U}(\boldsymbol{\theta}_0) + n\tilde{\lambda}\mathbf{H}(\boldsymbol{\theta}_0)\}^{-1} \\
\mathbf{H}(\boldsymbol{\theta}_0)\{\mathbf{U}(\boldsymbol{\theta}_0) + n\tilde{\lambda}\mathbf{H}(\boldsymbol{\theta}_0)\}^{-1}\mathbf{u}(\boldsymbol{\theta}_0).
\end{aligned}$$

Note the fact that $f(\tilde{\lambda}) \approx f(0) + \dot{f}(0)\tilde{\lambda} = 0$. Then, we have

$$\begin{aligned} nf(0) &= \frac{\mathbf{u}(\boldsymbol{\theta}_0)^\top}{\sqrt{n}} \left\{ -\frac{\mathbf{U}(\boldsymbol{\theta}_0)}{n} \right\}^{-1} \mathbf{H}(\boldsymbol{\theta}_0) \left\{ -\frac{\mathbf{U}(\boldsymbol{\theta}_0)}{n} \right\}^{-1} \frac{\mathbf{u}(\boldsymbol{\theta}_0)}{\sqrt{n}} \\ &= \left[\frac{\mathbf{u}(\boldsymbol{\theta}_0)^\top}{\sqrt{n}} \left\{ -\frac{\mathbf{U}(\boldsymbol{\theta}_0)}{n} \right\}^{-\frac{1}{2}} \right] \left[\left\{ -\frac{\mathbf{U}(\boldsymbol{\theta}_0)}{n} \right\}^{-\frac{1}{2}} \mathbf{H}(\boldsymbol{\theta}_0) \left\{ -\frac{\mathbf{U}(\boldsymbol{\theta}_0)}{n} \right\}^{-\frac{1}{2}} \right] \\ &\quad \left[\left\{ -\frac{\mathbf{U}(\boldsymbol{\theta}_0)}{n} \right\}^{-\frac{1}{2}} \frac{\mathbf{u}(\boldsymbol{\theta}_0)}{\sqrt{n}} \right]. \end{aligned}$$

Since $\left\{ -\frac{\mathbf{U}(\boldsymbol{\theta}_0)}{n} \right\}^{-\frac{1}{2}} \xrightarrow{p} \mathbf{I}(\boldsymbol{\theta}_0)^{-\frac{1}{2}}$ and $\frac{\mathbf{u}(\boldsymbol{\theta}_0)}{\sqrt{n}} \xrightarrow{d} \mathrm{N}\{0, \mathbf{I}(\boldsymbol{\theta}_0)\}$, by Slutsky's Therorem, $\left\{ -\frac{\mathbf{U}(\boldsymbol{\theta}_0)}{n} \right\}^{-\frac{1}{2}} \frac{\mathbf{u}(\boldsymbol{\theta}_0)}{\sqrt{n}} \xrightarrow{d} \mathrm{N}\{0, \mathbf{I}\}$. Also $\left\{ -\frac{\mathbf{U}(\boldsymbol{\theta}_0)}{n} \right\}^{-\frac{1}{2}} \mathbf{H}(\boldsymbol{\theta}_0) \left\{ -\frac{\mathbf{U}(\boldsymbol{\theta}_0)}{n} \right\}^{-\frac{1}{2}} \xrightarrow{p} \mathbf{A}(\boldsymbol{\theta})$. It follows that, as $n \to \infty$,

$$nf(0) \xrightarrow{d} F_0, \text{ where } F_0 \overset{d}{\equiv} \sum_{q=1}^{2Q} \upsilon_q \xi_q = \sum_{q=1}^{Q} \upsilon_q (\xi_q - \xi_{q+Q}),$$

with $\xi_q \overset{i.i.d}{\sim} \chi_1^2, \quad q = 1, \cdots, 2Q$.

$$\begin{aligned} &n\dot{f}(0) \\ =&2\frac{\mathbf{u}(\boldsymbol{\theta}_0)^\top}{\sqrt{n}} \left\{ -\frac{\mathbf{U}(\boldsymbol{\theta}_0)}{n} \right\}^{-1} \mathbf{H}(\boldsymbol{\theta}_0) \left\{ -\frac{\mathbf{U}(\boldsymbol{\theta}_0)}{n} \right\}^{-1} \mathbf{H}(\boldsymbol{\theta}_0) \left\{ -\frac{\mathbf{U}(\boldsymbol{\theta}_0)}{n} \right\}^{-1} \frac{\mathbf{u}(\boldsymbol{\theta}_0)}{\sqrt{n}}, \end{aligned}$$

similarly, we have as $n \to \infty$,

$$n\dot{f}(0) \xrightarrow{d} G_0, \text{ where } G_0 \overset{d}{\equiv} 2\sum_{q=1}^{2Q} \upsilon_q^2 \xi_q = 2\sum_{q=1}^{Q} \upsilon_q^2 (\xi_q + \xi_{q+Q}),$$

with $\xi_q \overset{i.i.d.}{\sim} \chi_1^2, \quad q = 1, \cdots, 2Q$. In summary the asymptotic distribution of $\tilde{\lambda}$ is given as follows,

$$\tilde{\lambda} = -\frac{nf(0)}{n\dot{f}(0)} \xrightarrow{d} \Lambda_0, \text{ where } \Lambda_0 \overset{d}{\equiv} -\frac{\sum_{q=1}^{Q} \upsilon_q (\xi_q - \xi_{q+Q})}{2\sum_{q=1}^{Q} \upsilon_q^2 (\xi_q + \xi_{q+Q})},$$

with $\xi_q \overset{i.i.d.}{\sim} \chi^2_1, \quad q = 1, \cdots, 2Q$. The proof is completed.

A.4 Proof of Theorem II.4

When $\boldsymbol{\alpha} = \boldsymbol{\beta} = \mathbf{0}$, taking the Taylor expansion on $\{\mathbf{I}(\boldsymbol{\theta}_0) - \tilde{\lambda}\mathbf{H}(\boldsymbol{\theta}_0)\}^{-1}$ around a small neighborhood of $\tilde{\lambda} = 0$, we have, subject to a high order error term,

$$\{\mathbf{I}(\boldsymbol{\theta}_0) - \tilde{\lambda}\mathbf{H}(\boldsymbol{\theta}_0)\}^{-1} \approx \{\mathbf{I}(\boldsymbol{\theta}_0)\}^{-1} + \tilde{\lambda}\mathbf{I}(\boldsymbol{\theta}_0)^{-1}\mathbf{H}(\boldsymbol{\theta}_0)\mathbf{I}(\boldsymbol{\theta}_0)^{-1}.$$

It follows that

$$\begin{aligned}
\sqrt{n}(\tilde{\boldsymbol{\theta}} - \boldsymbol{\theta}_0) &\approx \left[\{\mathbf{I}(\boldsymbol{\theta}_0)\}^{-1} + \tilde{\lambda}\mathbf{I}(\boldsymbol{\theta}_0)^{-1}\mathbf{H}(\boldsymbol{\theta}_0)\mathbf{I}(\boldsymbol{\theta}_0)^{-1}\right]\frac{\mathbf{u}(\boldsymbol{\theta}_0)}{\sqrt{n}} \\
&= \{\mathbf{I}(\boldsymbol{\theta}_0)\}^{-1}\frac{1}{\sqrt{n}}\mathbf{u}(\boldsymbol{\theta}_0) + \tilde{\lambda}\mathbf{I}(\boldsymbol{\theta}_0)^{-1}\mathbf{H}(\boldsymbol{\theta}_0)\mathbf{I}(\boldsymbol{\theta}_0)^{-1}\frac{\mathbf{u}(\boldsymbol{\theta}_0)}{\sqrt{n}} \\
&\approx \sqrt{n}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0) + \tilde{\lambda}\mathbf{I}(\boldsymbol{\theta}_0)^{-1}\mathbf{H}(\boldsymbol{\theta}_0)\mathbf{I}(\boldsymbol{\theta}_0)^{-1}\frac{\mathbf{u}(\boldsymbol{\theta}_0)}{\sqrt{n}}.
\end{aligned}$$

Noting that $\sqrt{n}(\tilde{\boldsymbol{\theta}} - \hat{\boldsymbol{\theta}}) = \tilde{\lambda}\mathbf{I}(\boldsymbol{\theta}_0)^{-1}\mathbf{H}(\boldsymbol{\theta}_0)\mathbf{I}(\boldsymbol{\theta}_0)^{-1}\frac{\mathbf{u}(\boldsymbol{\theta}_0)}{\sqrt{n}}$, we have

$$\begin{aligned}
T_n &= -2\{\ell(\tilde{\boldsymbol{\theta}}) - \ell(\hat{\boldsymbol{\theta}})\} \\
&\approx \sqrt{n}(\tilde{\boldsymbol{\theta}} - \hat{\boldsymbol{\theta}})^\top \left\{-\frac{\mathbf{U}(\boldsymbol{\theta}_0)}{n}\right\}^{-1} \sqrt{n}(\tilde{\boldsymbol{\theta}} - \hat{\boldsymbol{\theta}}) \\
&\approx \tilde{\lambda}^2 \frac{\mathbf{u}(\boldsymbol{\theta}_0)}{\sqrt{n}}^\top \mathbf{I}(\boldsymbol{\theta}_0)^{-1}\mathbf{H}(\boldsymbol{\theta}_0)\mathbf{I}(\boldsymbol{\theta}_0)^{-1}\mathbf{H}(\boldsymbol{\theta}_0)\mathbf{I}(\boldsymbol{\theta}_0)^{-1}\frac{\mathbf{u}(\boldsymbol{\theta}_0)}{\sqrt{n}} \\
&= \tilde{\lambda}^2 \frac{\mathbf{u}(\boldsymbol{\theta}_0)}{\sqrt{n}}^\top \mathbf{I}(\boldsymbol{\theta}_0)^{-\frac{1}{2}}\mathbf{A}(\boldsymbol{\theta}_0)^2\mathbf{I}(\boldsymbol{\theta}_0)^{-\frac{1}{2}}\frac{\mathbf{u}(\boldsymbol{\theta}_0)}{\sqrt{n}}.
\end{aligned}$$

Note that

$$\tilde{\lambda} \xrightarrow{d} \Lambda_0, \text{ where } \Lambda_0 \overset{d}{\equiv} -\frac{\sum_{q=1}^{Q} \upsilon_q(\xi_q - \xi_{q+Q})}{2\sum_{q=1}^{Q} \upsilon_q^2(\xi_q + \xi_{q+Q})},$$

and

$$\frac{\mathbf{u}(\boldsymbol{\theta}_0)^\top}{\sqrt{n}}\mathbf{I}(\boldsymbol{\theta}_0)^{-\frac{1}{2}} \xrightarrow{d} \mathrm{N}(\mathbf{0}, \mathbf{I}_{2Q+p}).$$

Hence,

$$T_n \xrightarrow{d} \Lambda_1, \text{ where } \Lambda_1 \overset{d}{\equiv} \frac{\left[\sum_{q=1}^{Q} \upsilon_q(\xi_q - \xi_{q+Q})\right]^2}{4\sum_{q=1}^{Q} \upsilon_q^2(\xi_q + \xi_{q+Q})},$$

with $\xi_q \overset{i.i.d.}{\sim} \chi_1^2$ for $q = 1, \ldots, 2Q$. The proof is completed.

A.5 Additional Simulations

In this section, we include additional simulation scenarios, where exchangeable correlations of mediators are set to 0.25 and 0, separately. Moreover, we also consider a modified High-Dimensional Multiple Testing (HDMT) procedure for comparison(*Dai et al.*, 2020). HDMT was originally developed for a univariate screening of mediators with controlled false discovery rate in genome studies, which is an approach widely adopted in practice to avoid simultaneous inference. For the purpose of comparison, we slightly added a decision rule to the method in a testing problem involving multiple mediators, described as follows. First, we calculate the adjusted p-values for Q mediators using the HDMT method via the R package "HDMT". Then, under the α_0 significance level and α_1 false discovery rate, we propose to reject the null hypothesis if at least $\lceil Q\alpha_1 \rceil$ number of adjusted p-values are less than α_0. In this paper, we set $\alpha_0 = \alpha_1 = 0.05$ in all comparisons considered in this paper.

The univariate screening HDMT method involves choosing the tuning parameter $\lambda \in (0, 1)$ related to the control of false discovery rate, which is preferably close to 1(*Dai et al.*, 2020). Thus, in this simulation study, we choose the default value $\lambda = 0.5$ suggested in the R package "HDMT" and another value closer to 1, namely $\lambda = 0.9$. It is evident that the HDMT test with $\lambda = 0.5, 0.9$ cannot give a proper control of type I error. This may be due to the fact that the HDMT is developed to screen high-dimensional mediators by the Benjamini-Hochberg procedure of false discovery rate, which may not be suitable for a multi-dimensional simultaneous test in the Neyman-Pearson hypothesis testing framework. Because of poor control of

type I error by HDMT, this method will not be considered in the power comparison below. Tables A.1, A.2 and A.3 show the type I error and power comparison when the correlation among mediators are equal to 0.50, 0.25 and 0, respectively.

Table A.1: Empirical type I error under four null hypotheses, and power under four alternative hypotheses with 10,000 replicates. The sample size varies from 200, 500, and 1,000. The exchangeable correlation of mediators is set with correlation 0.5. Power increase (%) $= \dfrac{\text{power of LR test}}{\text{power of competing test}} - 1$.

n	Method	Null Hypothesis				Alternative Hypothesis				Percent of power increase			
		i	ii	iii	iv	v	vi	vii	viii	v	vi	vii	viii
200	LR	0.050	0.052	0.051	0.009	0.591	0.562	0.312	0.507	-	-	-	-
	PT-N	0.038	0.042	0.037	0.005	0.550	0.536	0.255	0.458	7.46%	4.78%	22.51%	10.72%
	PT-NP	0.032	0.038	0.028	0.001	0.507	0.497	0.242	0.426	16.55%	13.07%	28.72%	19.20%
	HDMT($\lambda = 0.5$)	0.439	0.910	0.115	0.001	0.039	0.062	0.225	0.817	-	-	-	-
	HDMT($\lambda = 0.9$)	0.934	1.000	0.787	0.018	0.251	0.344	0.633	0.991	-	-	-	-
500	LR	0.046	0.049	0.045	0.007	0.970	0.954	0.648	0.928	-	-	-	-
	PT-N	0.041	0.045	0.039	0.005	0.967	0.953	0.624	0.922	0.30%	0.12%	3.78%	0.66%
	PT-NP	0.038	0.044	0.035	0.001	0.962	0.947	0.620	0.916	0.78%	0.75%	4.5%	1.36%
	HDMT($\lambda = 0.5$)	0.961	1.000	0.767	0.002	0.059	0.066	0.599	1.000	-	-	-	-
	HDMT($\lambda = 0.9$)	1.000	1.000	1.000	0.019	0.332	0.361	0.928	1.000	-	-	-	-
1000	LR	0.051	0.049	0.046	0.006	1.000	1.000	0.917	0.999	-	-	-	-
	PT-N	0.048	0.048	0.043	0.004	1.000	1.000	0.911	0.999	0.00%	0.00%	0.67%	0.01%
	PT-NP	0.046	0.049	0.041	0.000	1.000	1.000	0.910	0.998	0.00%	0.00%	0.71%	0.05%
	HDMT($\lambda = 0.5$)	1.000	1.000	1.000	0.001	0.064	0.071	0.907	1.000	-	-	-	-
	HDMT($\lambda = 0.9$)	1.000	1.000	1.000	0.018	0.342	0.381	0.994	1.000	-	-	-	-

Table A.2: Empirical type I error under four null hypotheses, and power under four alternative hypotheses summarized over 10,000 replicates. The dimension of mediators Q is 30. The sample size varies from 200, 500, and 1,000. The exchangeable correlation among mediators is 0.25.

n	Method	Null Hypothesis				Alternative Hypothesis				Percent of power increase			
		i	ii	iii	iv	i	ii	iii	iv	i	ii	iii	iv
200	LR	0.045	0.048	0.049	0.009	0.573	0.519	0.402	0.475	-	-	-	-
	PT-N	0.034	0.038	0.037	0.005	0.536	0.497	0.340	0.433	6.84%	4.36%	18.18%	9.58%
	PT-NP	0.027	0.034	0.031	0.001	0.498	0.460	0.320	0.402	15.17%	12.95%	25.55%	18.23%
	HDMT($\lambda = 0.5$)	0.049	0.609	0.390	0.002	0.033	0.028	0.156	0.812	-	-	-	-
	HDMT($\lambda = 0.9$)	0.626	0.995	0.986	0.009	0.201	0.478	0.416	0.999	-	-	-	-
500	LR	0.043	0.049	0.044	0.007	0.959	0.932	0.807	0.905	-	-	-	-
	PT-N	0.037	0.045	0.040	0.005	0.956	0.929	0.786	0.898	0.26%	0.3%	2.62%	0.78%
	PT-NP	0.035	0.045	0.037	0.001	0.951	0.922	0.779	0.892	0.81%	1.07%	3.61%	1.4%
	HDMT($\lambda = 0.5$)	0.494	1.000	1.000	0.002	0.031	0.060	0.273	1.000	-	-	-	-
	HDMT($\lambda = 0.9$)	0.963	1.000	1.000	0.008	0.458	0.690	0.655	1.000	-	-	-	-
1000	LR	0.050	0.048	0.046	0.006	1.000	0.999	0.983	0.997	-	-	-	-
	PT-N	0.047	0.047	0.043	0.004	1.000	0.998	0.982	0.997	0.00%	0.01%	0.17%	0.01%
	PT-NP	0.045	0.047	0.042	0.000	1.000	0.998	0.980	0.997	0.00%	0.01%	0.32%	0.06%
	HDMT($\lambda = 0.5$)	0.946	1.000	1.000	0.002	0.048	0.093	0.420	1.000	-	-	-	-
	HDMT($\lambda = 0.9$)	0.999	1.000	1.000	0.010	0.634	0.741	0.854	1.000	-	-	-	-

Table A.3: Empirical type I error under four null hypotheses, and power under four alternative hypotheses summarized over 10,000 replicates. The dimension of mediators Q is 30. The sample size varies from 200, 500, and 1,000. The exchangeable correlation among mediators is 0.

n	Method	Null Hypothesis				Alternative Hypothesis				Percent of power increase			
		i	ii	iii	iv	i	ii	iii	iv	i	ii	iii	iv
200	LR	0.042	0.044	0.060	0.009	0.536	0.478	0.475	0.288	-	-	-	-
	PT-N	0.032	0.036	0.044	0.005	0.505	0.457	0.406	0.249	6.04%	4.48%	16.87%	15.32%
	PT-NP	0.023	0.031	0.040	0.001	0.469	0.426	0.379	0.242	14.13%	12.17%	25.16%	18.80%
	HDMT($\lambda = 0.5$)	0.023	0.872	0.785	0.004	0.113	0.130	0.137	0.939	-	-	-	-
	HDMT($\lambda = 0.9$)	0.033	0.964	1.000	0.004	0.138	0.235	0.261	1.000	-	-	-	-
500	LR	0.044	0.048	0.049	0.007	0.937	0.899	0.894	0.617	-	-	-	-
	PT-N	0.040	0.045	0.042	0.005	0.935	0.897	0.882	0.598	0.27%	0.30%	1.34%	3.13%
	PT-NP	0.035	0.041	0.042	0.001	0.928	0.891	0.875	0.594	0.94%	0.91%	2.22%	3.86%
	HDMT($\lambda = 0.5$)	0.027	1.000	1.000	0.004	0.112	0.170	0.174	1.000	-	-	-	-
	HDMT($\lambda = 0.9$)	0.039	1.000	1.000	0.005	0.138	0.316	0.316	1.000	-	-	-	-
1000	LR	0.050	0.046	0.046	0.006	0.999	0.996	0.997	0.898	-	-	-	-
	PT-N	0.048	0.044	0.044	0.004	0.999	0.996	0.996	0.894	0.00%	0.00%	0.06%	0.49%
	PT-NP	0.046	0.045	0.044	0.000	0.999	0.996	0.996	0.894	0.00%	0.04%	0.08%	0.45%
	HDMT($\lambda = 0.5$)	0.026	1.000	1.000	0.003	0.115	0.174	0.178	1.000	-	-	-	-
	HDMT($\lambda = 0.9$)	0.037	1.000	1.000	0.004	0.140	0.311	0.319	1.000	-	-	-	-

APPENDIX B

Supplement for Chapter III

B.1 Joint distribution of Z_x, Z_m^* and Z_y^*

We know that $Z_x \sim N(0,1)$, $Z_m \sim N(0, \alpha^2 + 1)$, and $Z_y \sim N(0, \eta^2 + \beta^2 + 1)$, where $\eta = \gamma + \beta\alpha$. then it's straightforward to obtain the covariances as follows,

$$\begin{aligned}
cov(Z_x, Z_m) &= cov(Z_x, \alpha Z_x + \epsilon_x) = \alpha \\
cov(Z_x, Z_y) &= cov(Z_x, \gamma Z_x + \beta Z_m + \epsilon_y) \\
&= cov(Z_x, \gamma Z_x + \beta(\alpha Z_x + \epsilon_x) + \epsilon_y) \\
&= \gamma + \beta\alpha \\
cov(Z_m, Z_y) &= cov(\alpha Z_x + \epsilon_x, \gamma Z_x + \beta Z_m + \epsilon_y) \\
&= cov(\alpha Z_x + \epsilon_x, \gamma Z_x + \beta(\alpha Z_x + \epsilon_x) + \epsilon_y) \\
&= \alpha(\gamma + \beta\alpha) + \beta
\end{aligned}$$

The joint distribution of Z_x, Z_m and Z_y is given by

$$\begin{pmatrix} Z_x \\ Z_m \\ Z_y \end{pmatrix} \sim N\left\{ \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & \alpha & \eta \\ \alpha & \alpha^2+1 & \alpha\eta+\beta \\ \eta & \alpha\eta+\beta & \eta^2+\beta^2+1 \end{pmatrix} \right\}$$

The joint distribution of Z_x, Z_m^* and Z_y^* is given by

$$\begin{pmatrix} Z_x \\ Z_m^* \\ Z_y^* \end{pmatrix} \sim N\left\{ \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & \frac{\alpha}{\sqrt{\alpha^2+1}} & \frac{\eta}{\sqrt{\eta^2+\beta^2+1}} \\ \frac{\alpha}{\sqrt{\alpha^2+1}} & 1 & \frac{\alpha\eta+\beta}{\sqrt{\alpha^2+1}\sqrt{\eta^2+\beta^2+1}} \\ \frac{\eta}{\sqrt{\eta^2+\beta^2+1}} & \frac{\alpha\eta+\beta}{\sqrt{\alpha^2+1}\sqrt{\eta^2+\beta^2+1}} & 1 \end{pmatrix} \right\}$$

where $\eta = \gamma + \beta\alpha$. The conditional distribution of any one or two variables from Z_x, Z_m^* and Z_y^* given the rest also follows the normal distribution. Next we will present the details for parameter estimation for a single subject under the eight scenarios where X, M and Y are discrete and continuous, respectively.

B.2 Joint density function of $\pi(X, M, Y)$

B.2.1 X, M and Y are all continuous

$$\begin{aligned}
&\pi(X, M, Y) \\
=&\int_{Z_x}\int_{Z_m^*}\int_{Z_y^*} \pi(X, M, Y, Z_x, Z_m^*, Z_y^*) dZ_x dZ_m^* dZ_y^* \\
=&\int_{Z_x}\int_{Z_m^*}\int_{Z_y^*} \pi(Z_x, Z_m^*, Z_y^*) I[X = F_x^{-1}\{\Phi(Z_x)\}] I[M = F_m^{-1}\{\Phi(Z_m^*)\}] \times \\
&\qquad I[Y = F_y^{-1}\{\Phi(Z_y^*)\}] dZ_x dZ_m^* dZ_y^* \\
=&\pi(Z_x, Z_m^*, Z_y^*)
\end{aligned}$$

where $Z_x = \Phi^{-1}\{F_x(X)\}$, $Z_m^* = \Phi^{-1}\{F_m(M)\}$, and $Z_y^* = \Phi^{-1}\{F_y(Y)\}$.

B.2.2 X, M are continuous, Y is discrete

$$
\begin{aligned}
&\pi(X, M, Y) \\
=&\int_{Z_x}\int_{Z_m^*}\int_{Z_y^*} \pi(X, M, Y, Z_x, Z_m^*, Z_y^*) dZ_x dZ_m^* dZ_y^* \\
=&\int_{Z_x}\int_{Z_m^*}\int_{Z_y^*} \pi(Z_x, Z_m^*, Z_y^*) I[X = F_x^{-1}\{\Phi(Z_x)\}] I[M = F_m^{-1}\{\Phi(Z_m^*)\}] \times \\
&\qquad\qquad I[Y = F_y^{-1}\{\Phi(Z_y^*)\}] dZ_x dZ_m^* dZ_y^* \\
=&\int_{Z_y^*} \pi(Z_y^*|Z_x, Z_m^*)\pi(Z_x, Z_m^*) I[\Phi^{-1}(F_y(Y-1)) \le Z_y^* < \Phi^{-1}(F_y(Y))] dZ_y^* \\
=&\pi(Z_x, Z_m^*) P\Big(Z_y^* \in (l_y, u_y) | Z_x, Z_m^*\Big)
\end{aligned}
$$

The joint distribution of Z_x and Z_m^* can be obtained from the joint distribution of Z_x, Z_m^* and Z_y^*. And the distribution of Z_y^* conditional on Z_x and Z_m^* is normally distributed with mean and variance given below.

$$
\begin{aligned}
&\mathrm{E}(Z_y^*|Z_x, Z_m^*) \\
&= \begin{pmatrix} \frac{\eta}{\sqrt{\eta^2+\beta^2+1}} \\ \frac{\alpha\eta+\beta}{\sqrt{\alpha^2+1}\sqrt{\eta^2+\beta^2+1}} \end{pmatrix}^\top \begin{pmatrix} 1 & \frac{\alpha}{\sqrt{\alpha^2+1}} \\ \frac{\alpha}{\sqrt{\alpha^2+1}} & 1 \end{pmatrix}^{-1} \begin{pmatrix} Z_x \\ Z_m \end{pmatrix} \\
&= \frac{\gamma Z_x + \sqrt{\alpha^2+1}\beta Z_m^*}{\sqrt{\eta^2+\beta^2+1}} \\
&\mathrm{var}(Z_y^*|Z_x, Z_m^*) \\
&= 1 - \begin{pmatrix} \frac{\eta}{\sqrt{\eta^2+\beta^2+1}} \\ \frac{\alpha\eta+\beta}{\sqrt{\alpha^2+1}\sqrt{\eta^2+\beta^2+1}} \end{pmatrix}^\top \begin{pmatrix} 1 & \frac{\alpha}{\sqrt{\alpha^2+1}} \\ \frac{\alpha}{\sqrt{\alpha^2+1}} & 1 \end{pmatrix}^{-1} \begin{pmatrix} \frac{\eta}{\sqrt{\eta^2+\beta^2+1}} \\ \frac{\alpha\eta+\beta}{\sqrt{\alpha^2+1}\sqrt{\eta^2+\beta^2+1}} \end{pmatrix} \\
&= \frac{1}{\eta^2+\beta^2+1}
\end{aligned}
$$

where $Z_x = \Phi^{-1}\{F_x(X)\}$, $Z_m^* = \Phi^{-1}\{F_m(M)\}$, $l_y = \Phi^{-1}(F_y(Y-1))$, and $u_y = \Phi^{-1}(F_y(Y))$.

B.2.3 X, Y are continuous, M is discrete

$$
\begin{aligned}
&\pi(X, M, Y) \\
=&\int_{Z_x}\int_{Z_m^*}\int_{Z_y^*} \pi(X, M, Y, Z_x, Z_m^*, Z_y^*) dZ_x dZ_m^* dZ_y^* \\
=&\int_{Z_x}\int_{Z_m^*}\int_{Z_y^*} \pi(Z_x, Z_m^*, Z_y^*) I[X = F_x^{-1}\{\Phi(Z_x)\}] I[M = F_m^{-1}\{\Phi(Z_m^*)\}] \times \\
&\qquad\qquad I[Y = F_y^{-1}\{\Phi(Z_y^*)\}] dZ_x dZ_m^* dZ_y^* \\
=&\int_{Z_m^*} \pi(Z_m^*|Z_x, Z_y^*)\pi(Z_x, Z_y^*) I[\Phi^{-1}(F_m(M-1)) \le Z_m^* < \Phi^{-1}(F_m(M))] dZ_m^* \\
=&\pi(Z_x, Z_y^*) P\Big(Z_m^* \in (l_m, u_m)|Z_x, Z_y^*\Big)
\end{aligned}
$$

The joint distribution of Z_x and Z_y^* can be obtained from the joint distribution of Z_x, Z_m^* and Z_y^*. And the distribution of Z_m^* conditional on Z_x and Z_y^* is normally

distributed with mean and variance given below.

$$
\begin{aligned}
&\mathrm{E}(Z_m^*|Z_x, Z_y^*) \\
&= \begin{pmatrix} \frac{\alpha}{\sqrt{\alpha^2+1}} \\ \frac{\alpha\eta+\beta}{\sqrt{\alpha^2+1}\sqrt{\eta^2+\beta^2+1}} \end{pmatrix}^\top \begin{pmatrix} 1 & \frac{\eta}{\sqrt{\eta^2+\beta^2+1}} \\ \frac{\eta}{\sqrt{\eta^2+\beta^2+1}} & 1 \end{pmatrix}^{-1} \begin{pmatrix} Z_x \\ Z_y^* \end{pmatrix} \\
&= \frac{(\alpha-\gamma\beta)Z_x + \beta\sqrt{\eta^2+\beta^2+1}Z_y^*}{(\beta^2+1)\sqrt{\alpha^2+1}} \\
&\mathrm{var}(Z_m^*|Z_x, Z_y^*) \\
&= 1 - \begin{pmatrix} \frac{\alpha}{\sqrt{\alpha^2+1}} \\ \frac{\alpha\eta+\beta}{\sqrt{\alpha^2+1}\sqrt{\eta^2+\beta^2+1}} \end{pmatrix}^\top \begin{pmatrix} 1 & \frac{\eta}{\sqrt{\eta^2+\beta^2+1}} \\ \frac{\eta}{\sqrt{\eta^2+\beta^2+1}} & 1 \end{pmatrix}^{-1} \begin{pmatrix} \frac{\alpha}{\sqrt{\alpha^2+1}} \\ \frac{\alpha\eta+\beta}{\sqrt{\alpha^2+1}\sqrt{\eta^2+\beta^2+1}} \end{pmatrix} \\
&= \frac{1}{(\beta^2+1)(\alpha^2+1)}
\end{aligned}
$$

where $Z_x = \Phi^{-1}\{F_x(X)\}$, $Z_y^* = \Phi^{-1}\{F_y(Y)\}$, $l_m = \Phi^{-1}(F_m(M-1))$, and $u_m = \Phi^{-1}(F_m(M))$.

B.2.4 X is continuous, M and Y are discrete

$$
\begin{aligned}
&\pi(X, M, Y) \\
=&\int_{Z_x}\int_{Z_m^*}\int_{Z_y^*} \pi(X, M, Y, Z_x, Z_m^*, Z_y^*) dZ_x dZ_m^* dZ_y^* \\
=&\int_{Z_x}\int_{Z_m^*}\int_{Z_y^*} \pi(Z_x, Z_m^*, Z_y^*) I[X = F_x^{-1}\{\Phi(Z_x)\}] I[M = F_m^{-1}\{\Phi(Z_m^*)\}] \times \\
&\qquad I[Y = F_y^{-1}\{\Phi(Z_y^*)\}] dZ_x dZ_m^* dZ_y^* \\
=&\int_{Z_m^*}\int_{Z_y^*} \pi(Z_m^*, Z_y^*|Z_x)\pi(Z_x) I[\Phi^{-1}(F_m(M-1)) \le Z_m^* < \Phi^{-1}(F_m(M))] \\
&\qquad I[\Phi^{-1}(F_y(Y-1)) \le Z_y^* < \Phi^{-1}(F_y(Y))] dZ_m dZ_y \\
=&\pi(Z_x) P\Big(Z_m^* \in (l_m, u_m), Z_y^* \in (l_y, u_y)|Z_x\Big)
\end{aligned}
$$

The joint distribution of Z_m^* and Z_y^* conditional on Z_x is normally distributed with mean and covariance given below.

$$
\begin{aligned}
&\mathrm{E}(Z_m^*, Z_y^* | Z_x) \\
&= \begin{pmatrix} \frac{\alpha}{\sqrt{\alpha^2+1}} \\ \frac{\eta}{\sqrt{\eta^2+\beta^2+1}} \end{pmatrix} Z_x \\
&\mathrm{var}(Z_m^*, Z_y^* | Z_x) \\
&= \begin{pmatrix} 1 & \frac{\alpha\eta+\beta}{\sqrt{\alpha^2+1}\sqrt{\eta^2+\beta^2+1}} \\ \frac{\alpha\eta+\beta}{\sqrt{\alpha^2+1}\sqrt{\eta^2+\beta^2+1}} & 1 \end{pmatrix} - \begin{pmatrix} \frac{\alpha}{\sqrt{\alpha^2+1}} \\ \frac{\eta}{\sqrt{\eta^2+\beta^2+1}} \end{pmatrix} \begin{pmatrix} \frac{\alpha}{\sqrt{\alpha^2+1}} \\ \frac{\eta}{\sqrt{\eta^2+\beta^2+1}} \end{pmatrix}^\top \\
&= \begin{pmatrix} \frac{1}{\alpha^2+1} & \frac{\beta}{\sqrt{\alpha^2+1}\sqrt{\eta^2+\beta^2+1}} \\ \frac{\beta}{\sqrt{\alpha^2+1}\sqrt{\eta^2+\beta^2+1}} & \frac{\beta^2+1}{\eta^2+\beta^2+1} \end{pmatrix}
\end{aligned}
$$

where $Z_x = \Phi^{-1}\{F_x(X)\}$, $l_m = \sqrt{\theta_{mx}^2+1}\Phi^{-1}(F_m(M-1))$, $u_m = \sqrt{\theta_{mx}^2+1}\Phi^{-1}(F_m(M))$, $l_y = \Phi^{-1}(F_y(Y-1))$, and $u_y = \Phi^{-1}(F_y(Y))$.

B.2.5 X is discrete, M and Y are continuous

$$
\begin{aligned}
&\pi(X, M, Y) \\
=&\int_{Z_x}\int_{Z_m^*}\int_{Z_y^*} \pi(X, M, Y, Z_x, Z_m^*, Z_y^*) dZ_x dZ_m^* dZ_y^* \\
=&\int_{Z_x}\int_{Z_m^*}\int_{Z_y^*} \pi(Z_x, Z_m^*, Z_y^*) I[X = F_x^{-1}\{\Phi(Z_x)\}] I[M = F_m^{-1}\{\Phi(Z_m^*)\}] \times \\
&\qquad I[Y = F_y^{-1}\{\Phi(Z_y^*)\}] dZ_x dZ_m^* dZ_y^* \\
=&\int_{Z_x} \pi(Z_x | Z_m^*, Z_y^*) \pi(Z_m^*, Z_y^*) I[\Phi^{-1}(F_x(X-1)) \leq Z_x < \Phi^{-1}(F_x(X))] dZ_x \\
=&\pi(Z_m^*, Z_y^*) P\Big(Z_x \in (l_x, u_x) | Z_m^*, Z_y^*\Big)
\end{aligned}
$$

The joint distribution of Z_m^* and Z_y^* can be obtained from the joint distribution of

Z_x, Z_m^* and Z_y^*. And the distribution of Z_x conditional on Z_m^* and Z_y^* is normally distributed with mean and covariance given below.

$$
\begin{aligned}
&\mathrm{E}(Z_x|Z_m^*, Z_y^*) \\
&= \begin{pmatrix} \frac{\alpha}{\sqrt{\alpha^2+1}} \\ \frac{\eta}{\sqrt{\eta^2+\beta^2+1}} \end{pmatrix}^\top \begin{pmatrix} 1 & \frac{\alpha\eta+\beta}{\sqrt{\alpha^2+1}\sqrt{\eta^2+\beta^2+1}} \\ \frac{\alpha\eta+\beta}{\sqrt{\alpha^2+1}\sqrt{\eta^2+\beta^2+1}} & 1 \end{pmatrix}^{-1} \begin{pmatrix} Z_m \\ Z_y \end{pmatrix} \\
&= \frac{(\alpha-\beta\gamma)\sqrt{\alpha^2+1}Z_m^*}{\alpha^2+\gamma^2+1} + \frac{\gamma\sqrt{\eta^2+\beta^2+1}Z_y^*}{\alpha^2+\gamma^2+1} \\
&\mathrm{var}(Z_x|Z_m^*, Z_y^*) \\
&= 1 - \begin{pmatrix} \frac{\alpha}{\sqrt{\alpha^2+1}} \\ \frac{\eta}{\sqrt{\eta^2+\beta^2+1}} \end{pmatrix}^\top \begin{pmatrix} 1 & \frac{\alpha\eta+\beta}{\sqrt{\alpha^2+1}\sqrt{\eta^2+\beta^2+1}} \\ \frac{\alpha\eta+\beta}{\sqrt{\alpha^2+1}\sqrt{\eta^2+\beta^2+1}} & 1 \end{pmatrix}^{-1} \begin{pmatrix} \frac{\alpha}{\sqrt{\alpha^2+1}} \\ \frac{\eta}{\sqrt{\eta^2+\beta^2+1}} \end{pmatrix} \\
&= \frac{1}{\alpha^2+\gamma^2+1}
\end{aligned}
$$

where $Z_m^* = \Phi^{-1}\{F_m(M)\}$, $Z_y^* = \Phi^{-1}\{F_y(Y)\}$, $l_x = \Phi^{-1}(F_x(X-1))$, and $u_x = \Phi^{-1}(F_x(X))$.

B.2.6 X and Y are discrete, M is continuous

$$
\begin{aligned}
&\pi(X, M, Y) \\
=&\int_{Z_x}\int_{Z_m^*}\int_{Z_y^*} \pi(X, M, Y, Z_x, Z_m^*, Z_y^*) dZ_x dZ_m^* dZ_y^* \\
=&\int_{Z_x}\int_{Z_m^*}\int_{Z_y^*} \pi(Z_x, Z_m^*, Z_y^*) I[X = F_x^{-1}\{\Phi(Z_x)\}] I[M = F_m^{-1}\{\Phi(Z_m^*)\}] \times \\
&\qquad\qquad I[Y = F_y^{-1}\{\Phi(Z_y^*)\}] dZ_x dZ_m^* dZ_y^* \\
=&\int_{Z_x}\int_{Z_y^*} \pi(Z_x, Z_y^*|Z_m^*)\pi(Z_m^*) I[\Phi^{-1}(F_x(X-1)) \le Z_x < \Phi^{-1}(F_x(X))] \\
&I[\Phi^{-1}(F_y(Y-1)) \le Z_y^* < \Phi^{-1}(F_y(Y))] dZ_x dZ_y^* \\
=&\pi(Z_m^*) P\Big(Z_x \in (l_x, u_x), Z_y^* \in (l_y, u_y) | Z_m^*\Big)
\end{aligned}
$$

The joint distribution of Z_x and Z_y^* conditional on Z_m^* is normally distributed with mean and covariance given below

$$
\begin{aligned}
&E(Z_x, Z_y^*|Z_m^*) \\
&\quad = \begin{pmatrix} \frac{\alpha}{\sqrt{\alpha^2+1}} \\ \frac{\alpha\eta+\beta}{\sqrt{\alpha^2+1}\sqrt{\eta^2+\beta^2+1}} \end{pmatrix} Z_m^* \\
&\mathrm{var}(Z_x, Z_y^*|Z_m^*) \\
&\quad = \begin{pmatrix} 1 & \frac{\eta}{\sqrt{\eta^2+\beta^2+1}} \\ \frac{\eta}{\sqrt{\eta^2+\beta^2+1}} & 1 \end{pmatrix} - \begin{pmatrix} \frac{\alpha}{\sqrt{\alpha^2+1}} \\ \frac{\alpha\eta+\beta}{\sqrt{\alpha^2+1}\sqrt{\eta^2+\beta^2+1}} \end{pmatrix} \begin{pmatrix} \frac{\alpha}{\sqrt{\alpha^2+1}} \\ \frac{\alpha\eta+\beta}{\sqrt{\alpha^2+1}\sqrt{\eta^2+\beta^2+1}} \end{pmatrix}^\top \\
&\quad = \frac{1}{\alpha^2+1} \begin{pmatrix} 1 & \frac{\gamma}{\sqrt{\eta^2+\beta^2+1}} \\ \frac{\gamma}{\sqrt{\eta^2+\beta^2+1}} & \frac{\gamma^2+\alpha^2+1}{\eta^2+\beta^2+1} \end{pmatrix}
\end{aligned}
$$

where $Z_m^* = \Phi^{-1}\{F_x(X)\}$, $l_x = \Phi^{-1}(F_x(X-1))$, $u_x = \Phi^{-1}(F_x(X))$, $l_y = \Phi^{-1}(F_y(Y-1))$, and $u_y = \Phi^{-1}(F_y(Y))$.

B.2.7 X and M are discrete, Y is continuous

$$
\begin{aligned}
&\pi(X, M, Y) \\
=&\int_{Z_x}\int_{Z_m^*}\int_{Z_y^*} \pi(X, M, Y, Z_x, Z_m^*, Z_y^*) dZ_x dZ_m^* dZ_y^* \\
=&\int_{Z_x}\int_{Z_m^*}\int_{Z_y^*} \pi(Z_x, Z_m^*, Z_y^*) I[X = F_x^{-1}\{\Phi(Z_x)\}] I[M = F_m^{-1}\{\Phi(Z_m^*)\}] \times \\
&\qquad\qquad I[Y = F_y^{-1}\{\Phi(Z_y^*)\}] dZ_x dZ_m^* dZ_y^* \\
=&\int_{Z_x}\int_{Z_m^*} \pi(Z_x, Z_m^*|Z_y^*)\pi(Z_y^*) I[\Phi^{-1}(F_x(X-1)) \le Z_x < \Phi^{-1}(F_x(X))] \\
&I[\Phi^{-1}(F_m(M-1)) \le Z_m^* < \Phi^{-1}(F_m(M))] dZ_x dZ_m^* \\
=&\pi(Z_y^*) P\Big(Z_x \in (l_x, u_x), Z_m^* \in (l_m, u_m)|Z_y^*\Big)
\end{aligned}
$$

The joint distribution of Z_x and Z_m^* conditional on Z_y^* is normally distributed with mean and covariance given below

$$
\begin{aligned}
&E(Z_x, Z_m^*|Z_y^*) \\
&\quad = \begin{pmatrix} \frac{\eta}{\sqrt{\eta^2+\beta^2+1}} \\ \frac{\alpha\eta+\beta}{\sqrt{\alpha^2+1}\sqrt{\eta^2+\beta^2+1}} \end{pmatrix} Z_y^* \\
&\text{var}(Z_x, Z_m^*|Z_y^*) \\
&\quad = \begin{pmatrix} 1 & \alpha \\ \alpha & \alpha^2+1 \end{pmatrix} - \frac{1}{\eta^2+\beta^2+1} \begin{pmatrix} \eta \\ \alpha\eta+\beta \end{pmatrix} \begin{pmatrix} \eta \\ \alpha\eta+\beta \end{pmatrix}^\top \\
&\quad = \frac{1}{\eta^2+\beta^2+1} \begin{pmatrix} \beta^2+1 & \frac{\alpha-\beta\gamma}{\sqrt{\alpha^2+1}} \\ \frac{\alpha-\beta\gamma}{\sqrt{\alpha^2+1}} & \frac{\gamma^2+\alpha^2+1}{\alpha^2+1} \end{pmatrix}
\end{aligned}
$$

where $Z_y^* = \Phi^{-1}\{F_y(Y)\}$, $l_x = \Phi^{-1}(F_x(X-1))$, $u_x = \Phi^{-1}(F_x(X))$, $l_m = \Phi^{-1}(F_m(M-1))$, and $u_m = \Phi^{-1}(F_m(M))$.

B.2.8 X, M and Y are discrete

$$
\begin{aligned}
&\pi(X, M, Y) \\
=&\int_{Z_x}\int_{Z_m^*}\int_{Z_y^*} \pi(X, M, Y, Z_x, Z_m^*, Z_y^*) dZ_x dZ_m^* dZ_y^* \\
=&\int_{Z_x}\int_{Z_m^*}\int_{Z_y^*} \pi(Z_x, Z_m^*, Z_y^*) I[X = F_x^{-1}\{\Phi(Z_x)\}] I[M = F_m^{-1}\{\Phi(Z_m^*)\}] \times \\
&\qquad\qquad I[Y = F_y^{-1}\{\Phi(Z_y^*)\}] dZ_x dZ_m^* dZ_y^* \\
=&\int_{Z_x}\int_{Z_m^*}\int_{Z_y^*} \pi(Z_x, Z_m^*, Z_y^*) I[\Phi^{-1}(F_x(X-1)) \leq Z_x < \Phi^{-1}(F_x(X))] \times \\
&\qquad\qquad I[\Phi^{-1}(F_m(M-1)) \leq Z_m^* < \Phi^{-1}(F_m(M))] \times \\
&\qquad\qquad I[\Phi^{-1}(F_y(Y-1)) \leq Z_y^* < \Phi^{-1}(F_y(Y))] dZ_x dZ_m^* dZ_y^* \\
=&P\Big(Z_x \in (l_x, u_x), Z_m^* \in (l_m, u_m), Z_y^* \in (l_y, u_y)\Big)
\end{aligned}
$$

where $l_x = \Phi^{-1}(F_x(X-1))$, $u_x = \Phi^{-1}(F_x(X))$, $l_m = \Phi^{-1}(F_m(M-1))$, $u_m = \Phi^{-1}(F_m(M))$, $l_y = \Phi^{-1}(F_y(Y-1))$, and $u_y = \Phi^{-1}(F_y(Y))$.

B.3 Effect calculation for $\mathrm{E}[Y(x_a, M(x_b)]$

B.3.1 X, M, and Y are continuous

$$
\begin{aligned}
&\mathrm{E}[Y(x_a, M(x_b)] \\
&= \int_{-\infty}^{\infty} \mathrm{E}(Y(x_a, M(x_b))|M(x_b) = m)\pi(m|X = x_b)dm \\
&= \int_{-\infty}^{\infty} \mathrm{E}(Y(x_a, m)|X = x_a, M = m)\pi(m|X = x_b)dm \\
&= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} y\pi(y|X = x_a, M = m)dy\pi(m|X = x_b)dm \\
&= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} F_y^{-1}(\Phi(Z_y^*))\pi(Z_y^*|Z_x = z_{x_a}, Z_m = z_m)\pi(Z_m^*|Z_x = z_{x_b})dZ_y^*dZ_m^*
\end{aligned}
$$

B.3.2 X and M are continuous, Y is discrete

$$
\begin{aligned}
&\mathrm{E}[Y(x_a, M(x_b)] \\
&= \int_{-\infty}^{\infty} \mathrm{E}(Y(x_a, M(x_b))|M(x_b) = m)\pi(m|X = x_b)dm \\
&= \int_{-\infty}^{\infty} \mathrm{E}(\sum_{y=0}^{\infty} yI(F_y(y-1) \leq \Phi(Z_y^*) < F_y(y))|Z_x = z_{x_a}, Z_m = z_m)\pi(Z_m|Z_x = z_{x_b})dZ_m \\
&= \int_{-\infty}^{\infty} \sum_{y=0}^{\infty} yP((F_y(y-1) \leq \Phi(Z_y^*) < F_y(y))|Z_x = z_{x_a}, Z_m = z_m)\pi(Z_m|Z_x = z_{x_b})dZ_m \\
&= \int_{-\infty}^{\infty} \sum_{y=1}^{\infty} yP(l_y \leq Z_y^* < u_y|Z_x = z_{x_a}, Z_m = z_m)\pi(Z_m|Z_x = z_{x_b})dZ_m
\end{aligned}
$$

B.3.3 X and Y are continuous, M is discrete

$$
\begin{aligned}
&\mathrm{E}[Y(x_a, M(x_b)] \\
&= \sum_{m=0}^{\infty} \mathrm{E}(Y(x_a, M(x_b))|M(x_b) = m)P(M(x_b) = m) \\
&= \sum_{m=0}^{\infty} \mathrm{E}(Y(x_a, m)|X = x_a, M(x_b) = m)P(M = m|X = x_b) \\
&= \sum_{m=0}^{\infty} \mathrm{E}(Y(x_a, m)|Z_x = z_{x_a}, l_m \leq Z_m^* < u_m)P(l_m \leq Z_m^* < u_m|Z_x = z_{x_b}) \\
&= \sum_{m=0}^{\infty} P(l_m \leq Z_m^* < u_m|Z_x = z_{x_b}) \int_{-\infty}^{\infty} F_y^{-1}(\Phi(Z_y^*))\pi(Z_y^*|Z_x = z_{x_a}, l_m \leq Z_m^* < u_m)dZ_y^* \\
&= \sum_{m=0}^{\infty} P(l_m \leq Z_m^* < u_m|Z_x = z_{x_b}) \int_{-\infty}^{\infty} F_y^{-1}(\Phi(Z_y^*)) \frac{\int_{l_m}^{u_m} \pi(Z_y^*, Z_m^*|Z_x = z_{x_a})dZ_m^*}{P(l_m \leq Z_m^* < u_m|Z_x = z_{x_a})} dZ_y^*
\end{aligned}
$$

B.3.4 X is continuous, M and Y are discrete

$$
\begin{aligned}
&\mathrm{E}[Y(x_a, M(x_b)] \\
&= \sum_{m=0}^{\infty} \mathrm{E}(Y(x_a, M(x_b))|M(x_b) = m)P(M(x_b) = m) \\
&= \sum_{m=0}^{\infty} \mathrm{E}(Y(x_a, m)|X = x_a, M(x_b) = m)P(M = m|X = x_b) \\
&= \sum_{m=0}^{\infty} \sum_{y=0}^{\infty} yP(l_y \leq Z_y^* < u_y|Z_x = z_{x_a}, l_m \leq Z_m^* < u_m)P(l_m \leq Z_m^* < u_m|Z_x = z_{x_b})
\end{aligned}
$$

B.3.5 X is discrete, M and Y are continuous

$$
\begin{aligned}
&\mathrm{E}[Y(x_a, M(x_b)] \\
&= \int_{-\infty}^{\infty} \mathrm{E}(Y(x_a, M(x_b))|M(x_b) = m)\pi(m|X = x_b)dm \\
&= \int_{-\infty}^{\infty} \mathrm{E}(Y(x_a, m)|X = x_a, M = m)\pi(m|X = x_b)dm \\
&= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} y\pi(y|X = x_a, M = m)dy\pi(m|X = x_b)dm \\
&= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} F_y^{-1}(\Phi(Z_y^*))\pi(Z_y^*|l_{x_a} \leq Z_x < u_{x_a}, Z_m^* = z_m^*)dZ_y^*\pi(z_m^*|l_{x_b} \leq Z_x < u_{x_b})dz_m^*
\end{aligned}
$$

B.3.6 X and Y are discrete, M is continuous

$$
\begin{aligned}
&\mathrm{E}[Y(x_a, M(x_b)] \\
&= \int_{-\infty}^{\infty} \mathrm{E}(Y(x_a, M(x_b))|M(x_b) = m)\pi(m|X = x_b)dm \\
&= \int_{-\infty}^{\infty} \mathrm{E}\left(\sum_{y=0}^{\infty} yI(F_y(y-1) \leq \Phi(Z_y^*) < F_y(y))|l_{x_a} \leq Z_x < u_{x_a}, Z_m = z_m\right) \times \\
&\qquad \pi(Z_m|l_{x_b} \leq Z_x < u_{x_b})dZ_m \\
&= \int_{-\infty}^{\infty} \sum_{y=0}^{\infty} yP((F_y(y-1) \leq \Phi(Z_y^*) < F_y(y))|l_{x_a} \leq Z_x < u_{x_a}, Z_m = z_m)\times \\
&\qquad \pi(Z_m|l_{x_b} \leq Z_x < u_{x_b})dZ_m \\
&= \int_{-\infty}^{\infty} \sum_{y=0}^{\infty} yP(l_y \leq Z_y^* < u_y|l_{x_a} \leq Z_x < u_{x_a}, Z_m = z_m)\pi(Z_m|l_{x_b} \leq Z_x < u_{x_b})dZ_m
\end{aligned}
$$

B.3.7 X and M are discrete, Y is continuous

$$
\begin{aligned}
&\mathrm{E}[Y(x_a, M(x_b)] \\
=&\sum_{m=0}^{\infty} \mathrm{E}(Y(x_a, M(x_b))|M(x_b) = m)P(M(x_b) = m) \\
=&\sum_{m=0}^{\infty} \mathrm{E}(Y(x_a, m)|X = x_a, M(x_b) = m)P(M = m|X = x_b) \\
=&\sum_{m=0}^{\infty} \mathrm{E}(Y(x_a, m)|l_{x_a} \leq Z_x < u_{x_a}, l_m \leq Z_m^* < u_m)P(l_m \leq Z_m^* < u_m|l_{x_b} \leq Z_x < u_{x_b}) \\
=&\sum_{m=0}^{\infty} \int_{-\infty}^{\infty} F_y^{-1}(\Phi(Z_y^*))\pi(Z_y^*|l_{x_a} \leq Z_x < u_{x_a}, l_m \leq Z_m^* < u_m)dZ_y^* \times \\
&\qquad P(l_m \leq Z_m^* < u_m|l_{x_b} \leq Z_x < u_{x_b})
\end{aligned}
$$

B.3.8 X, M and Y are discrete

$$
\begin{aligned}
&\mathrm{E}[Y(x_a, M(x_b)] \\
=&\sum_{m=0}^{\infty} \mathrm{E}(Y(x_a, M(x_b))|M(x_b) = m)P(M(x_b) = m) \\
=&\sum_{m=0}^{\infty} \mathrm{E}(Y(x_a, m)|X = x_a, M(x_b) = m)P(M = m|X = x_b) \\
=&\sum_{m=0}^{\infty} \sum_{y=0}^{\infty} yP(l_y \leq Z_y^* < u_y|l_{x_a} \leq Z_x < u_{x_a}, l_m \leq Z_m^* < u_m) \times \\
&\qquad P(l_m \leq Z_m^* < u_m|l_{x_b} \leq Z_x < u_{x_b})
\end{aligned}
$$

B.4 Real Data

Table B.1: NDE, NIE and TE estimates and 95 % CI obtained from GSEM for ELEMENT study. The breastfeeding duration was excluded from confounders.

Exposure	Outcome	NDE	95% CI	NIE	95% CI	TE	95% CI
	glucose	-0.013	(-0.121, 0.100)	0.017	(-0.002, 0.037)	0.004	(-0.107, 0.114)
MEHHP T2	C-peptide	-0.101	(-0.220, 0.029)	0.013	(-0.003, 0.032)	-0.088	(-0.212, 0.035)
	MetS	-0.065	(-0.132, 0.002)	0.016	(0.001, 0.026)	-0.049	(-0.118, 0.017)
	glucose	-0.010	(-0.127, 0.099)	0.018	(-0.000, 0.037)	0.008	(-0.110, 0.116)
MEOHP T2	C-peptide	-0.109	(-0.231, 0.012)	0.014	(-0.003, 0.033)	-0.094	(-0.233, 0.027)
	MetS	-0.070	(-0.132, 0.008)	0.017	(0.001, 0.033)	-0.052	(-0.115, 0.025)
	glucose	0.038	(-0.079, 0.147)	0.015	(-0.005, 0.047)	0.052	(-0.050, 0.168)
MIBP T2	C-peptide	-0.022	(-0.135, 0.090)	0.010	(-0.008, 0.038)	-0.012	(-0.121, 0.103)
	MetS	-0.019	(-0.086, 0.050)	0.013	(-0.005, 0.038)	-0.006	(-0.070, 0.069)
	glucose	0.055	(-0.066, 0.166)	0.008	(-0.011, 0.022)	0.063	(-0.063, 0.173)
MEHHP T3	C-peptide	0.009	(-0.112, 0.129)	0.006	(-0.012, 0.019)	0.015	(-0.106, 0.135)
	MetS	0.005	(-0.067, 0.075)	0.007	(-0.011, 0.018)	0.011	(-0.065, 0.081)
	glucose	0.055	(-0.068, 0.181)	0.005	(-0.019 0.024)	0.061	(-0.062, 0.189)
MEOHP T3	C-peptide	0.024	(-0.104, 0.166)	0.004	(-0.014, 0.018)	0.028	(-0.099, 0.167)
	MetS	0.008	(-0.069, 0.087)	0.005	(-0.016, 0.019)	0.013	(-0.064, 0.092)
	glucose	0.106	(-0.025, 0.253)	0.015	(-0.008, 0.045)	0.121	(-0.013, 0.269)
MIBP T3	C-peptide	0.013	(-0.132, 0.165)	0.011	(-0.007, 0.038)	0.024	(-0.120, 0.176)
	MetS	-0.014	(-0.100, 0.081)	0.013	(-0.007, 0.036)	-0.001	(-0.094, 0.103)

Table B.2: NDE, NIE and TE estimates and 95 % CI obtained from GSEM for boys in ELEMENT study.

Exposure	Outcome	NDE	95% CI	NIE	95% CI	TE	95% CI
	glucose	-0.038	(-0.213, 0.123)	0.004	(-0.069, 0.033)	-0.034	(-0.213, 0.124)
MEHHP T2	C-peptide	-0.051	(-0.237, 0.156)	0.004	(-0.065, 0.034)	-0.047	(-0.236, 0.152)
	MetS	-0.046	(-0.155, 0.084)	0.003	(-0.063, 0.024)	-0.043	(-0.163, 0.079)
	glucose	-0.036	(-0.199, 0.126)	0.006	(-0.075, 0.037)	-0.030	(-0.193, 0.132)
MEOHP T2	C-peptide	-0.032	(-0.224, 0.201)	0.005	(-0.051, 0.043)	-0.027	(-0.237, 0.191)
	MetS	-0.038	(-0.142, 0.076)	0.004	(-0.053, 0.033)	-0.034	(-0.153, 0.088)
	glucose	0.001	(-0.136, 0.124)	-0.008	(-0.045, 0.045)	-0.008	(-0.148, 0.118)
MIBP T2	C-peptide	0.030	(-0.130, 0.180)	-0.007	(-0.050, 0.036)	0.023	(-0.141, 0.175)
	MetS	-0.002	(-0.096, 0.090)	-0.007	(-0.040, 0.030)	-0.009	(-0.106, 0.084)
	glucose	0.030	(-0.092, 0.176)	0.005	(-0.065, 0.026)	0.035	(-0.099, 0.176)
MEHHP T3	C-peptide	-0.064	(-0.239, 0.102)	0.007	(-0.035, 0.023)	-0.057	(-0.238, 0.104)
	MetS	-0.029	(-0.114, 0.081)	0.009	(-0.030, 0.021)	-0.020	(-0.113, 0.084)
	glucose	0.025	(-0.107, 0.185)	-0.001	(-0.072, 0.025)	0.025	(-0.115, 0.176)
MEOHP T3	C-peptide	-0.030	(-0.189, 0.148)	0.003	(-0.056, 0.023)	-0.027	(-0.193, 0.152)
	MetS	-0.019	(-0.114, 0.086)	0.005	(-0.043, 0.023)	-0.014	(-0.118, 0.092)
	glucose	0.005	(-0.166, 0.174)	0.014	(-0.036, 0.062)	0.018	(-0.162, 0.182)
MIBP T3	C-peptide	-0.142	(-0.345, 0.075)	0.013	(-0.030, 0.060)	-0.129	(-0.329, 0.082)
	MetS	-0.096	(-0.212, 0.007)	0.018	(-0.029, 0.054)	-0.078	(-0.204, 0.023)

Table B.3: NDE, NIE and TE estimates and 95 % CI obtained from GSEM for girls in ELEMENT study.

Exposure	Outcome	NDE	95% CI	NIE	95% CI	TE	95% CI
MEHHP T2	glucose	0.042	(-0.127, 0.219)	0.017	(-0.007, 0.040)	0.059	(-0.117, 0.229)
	C-peptide	-0.092	(-0.263, 0.096)	0.023	(-0.007, 0.048)	-0.069	(-0.256, 0.119)
	MetS	-0.043	(-0.139, 0.075)	0.020	(-0.011, 0.033)	-0.023	(-0.126, 0.092)
MEOHP T2	glucose	0.048	(-0.142, 0.232)	0.018	(-0.016, 0.047)	0.066	(-0.131, 0.237)
	C-peptide	-0.115	(-0.298, 0.063)	0.025	(-0.011, 0.056)	-0.090	(-0.277, 0.104)
	MetS	-0.053	(-0.157, 0.062)	0.022	(-0.004, 0.042)	-0.031	(-0.142, 0.081)
MIBP T2	glucose	0.094	(-0.083, 0.272)	0.028	(-0.015, 0.106)	0.122	(-0.047, 0.295)
	C-peptide	-0.080	(-0.256, 0.102)	0.044	(-0.003, 0.118)	-0.035	(-0.210, 0.138)
	MetS	–0.030	(-0.132, 0.077)	0.037	(0.001, 0.092)	0.006	(-0.092, 0.105)
MEHHP T3	glucose	0.109	(-0.089, 0.325)	0.006	(-0.042, 0.029)	0.114	(-0.075, 0.326)
	C-peptide	0.148	(-0.017, 0.338)	0.007	(-0.046, 0.034)	0.155	(-0.015, 0.337)
	MetS	0.072	(-0.031, 0.187)	0.006	(-0.033, 0.025)	0.078	(-0.030, 0.196)
MEOHP T3	glucose	0.116	(-0.101, 0.333)	0.005	(-0.045, 0.036)	0.121	(-0.103, 0.341)
	C-peptide	0.143	(-0.064, 0.368)	0.006	(-0.049, 0.036)	0.149	(-0.061, 0.373)
	MetS	0.070	(-0.054, 0.196)	0.005	(-0.035, 0.031)	0.075	(-0.064, 0.203)
MIBP T3	glucose	0.204	(-0.013, 0.448)	0.012	(-0.020, 0.065)	0.216	(-0.002, 0.452)
	C-peptide	0.180	(-0.034, 0.414)	0.016	(-0.031, 0.085)	0.196	(-0.014, 0.422)
	MetS	0.048	(-0.069, 0.182)	0.015	(-0.029, 0.058)	0.063	(-0.078, 0.202)

APPENDIX C

Supplement for Chapter IV

C.1 NDE and NIE Derivations

C.1.1 Proposition IV.1

For Effective Dose Model (4.1), when $X = 1$ then $Z > 0$, then we have the following,

$$\begin{aligned}\pi(Z \mid M, X = 1) &\propto \pi(M \mid Z, X = 1)\pi(Z \mid X = 1) \\ &\propto exp\left\{-\frac{(M - \alpha Z)^2}{2\sigma_M^2}\right\} exp\left\{-\frac{(Z - a)^2}{2}\right\},\end{aligned}$$

This implies that $[Z \mid M, X = 1] \sim TN_{(0,\infty)}\left(\frac{\alpha M + \sigma_M^2 a}{\alpha^2 + \sigma_M^2}, \frac{\sigma_M^2}{\alpha^2 + \sigma_M^2}\right)$. For the derivations of NDE and NIE, let us denote $A_1 = \frac{\phi\left(\frac{\alpha M + \sigma_M^2 a}{\sigma_M \sqrt{\alpha^2 + \sigma_M^2}}\right)}{\Phi\left(\frac{\alpha M + \sigma_M^2 a}{\sigma_M \sqrt{\alpha^2 + \sigma_M^2}}\right)}$.

$$
\begin{aligned}
\mathrm{E}(Y \mid M, X=1) &= \int_0^\infty \mathrm{E}_Y(Y \mid M, Z, X=1)\pi(Z \mid M, X=1)dZ \\
&= \int_0^\infty (\gamma Z + \beta M)\pi(Z \mid M, X=1)dZ \\
&= \beta M + \gamma \mathrm{E}(Z \mid M, X=1) \\
&= \beta M + \gamma \left(\frac{\alpha M + \sigma_M^2 a}{\sigma_M^2 + \alpha^2} + \frac{\sigma_M}{\sqrt{\sigma_M^2 + \alpha^2}} A_1 \right), \\
\mathrm{E}(Y \mid M, X=0) &= \mathrm{E}(Y \mid M, Z=0) = \beta M, \\
\mathrm{E}(M \mid X=1) &= \int_0^\infty \mathrm{E}_Y(M \mid Z, X=1)\pi(Z \mid X=1)dZ \\
&= \int_0^\infty \alpha Z \pi(Z \mid X=1)dZ = \alpha \left(a + \frac{\phi(a)}{\Phi(a)} \right), \\
\mathrm{E}(M \mid X=0) &= 0.
\end{aligned}
$$

Therefore, the NDE and NIE are given below,

$$
\begin{aligned}
&\text{NDE} \\
=&\mathrm{E}\{Y(1, M(0))\} - \mathrm{E}\{Y(0, M(0))\} \\
=&\mathrm{E}_M\{\mathrm{E}_Y(Y \mid M, X=1) \mid X=0\} - \mathrm{E}_M\{\mathrm{E}_Y(Y \mid M, X=0) \mid X=0\} \\
=&\mathrm{E}_M \left\{ \beta M + \gamma \left(\frac{\alpha M + \sigma_M^2 a}{\sigma_M^2 + \alpha^2} + \frac{\sigma_M}{\sqrt{\sigma_M^2 + \alpha^2}} A_1 \right) \mid X=0 \right\} - \mathrm{E}_M(\beta M \mid X=0) \\
=&\left\{ \frac{\sigma_M^2 a}{\sigma_M^2 + \alpha^2} + \frac{\sigma_M}{\sqrt{\sigma_M^2 + \alpha^2}} \mathrm{E}_M(A_1 \mid X=0) \right\} \gamma,
\end{aligned}
$$

NIE

$$
\begin{aligned}
&= \mathrm{E}\{Y(1, M(1))\} - \mathrm{E}\{Y(1, M(0))\} \\
&= \mathrm{E}_M\{\mathrm{E}_Y(Y \mid M, X=1) \mid X=1\} - \mathrm{E}_M\{\mathrm{E}_Y(Y \mid M, X=1) \mid X=0\} \\
&= \mathrm{E}_M\left\{\beta M + \gamma\left(\frac{\alpha M + \sigma_M^2 a}{\sigma_M^2 + \alpha^2} + \frac{\sigma_M}{\sqrt{\sigma_M^2 + \alpha^2}} A_1\right) \mid X = 1\right\} - \mathrm{NDE} \\
&= \alpha\left(a + \frac{\phi(a)}{\Phi(a)}\right)\left(\beta + \frac{\alpha\gamma}{\sigma_M^2 + \alpha^2}\right) + \frac{\gamma\sigma_M}{\sqrt{\sigma_M^2 + \alpha^2}} \times \\
&\qquad \{\mathrm{E}_M(A_1 \mid X=1) - \mathrm{E}_M(A_1 \mid X=0)\}.
\end{aligned}
$$

C.1.2 Proposition IV.2

For Latent Exposure Model (4.2), when $Z > 0$, we have $X = 1$ and the following,

$$
\begin{aligned}
\pi(Z \mid M, X=1) &\propto \pi(M \mid Z)\pi(X \mid Z)\pi(Z) \\
&\propto exp\left\{-\frac{(M - \alpha Z)^2}{2\sigma_M^2}\right\} exp\left\{-\frac{(Z-\mu)^2}{2}\right\} I(Z > 0),
\end{aligned}
$$

This implies that $[Z \mid M, X=1] \sim TN_{(0,\infty)}\left(\frac{\alpha M + \sigma_M^2 \mu}{\alpha^2 + \sigma_M^2}, \frac{\sigma_M^2}{\alpha^2 + \sigma_M^2}\right)$. Similarly, when $Z < 0$, we have $[Z \mid M, X=0] \sim TN_{(-\infty,0)}\left(\frac{\alpha M + \sigma_M^2 \mu}{\alpha^2 + \sigma_M^2}, \frac{\sigma_M^2}{\alpha^2 + \sigma_M^2}\right)$. For the remaining derivations of NDE and NIE, we denote $A_2 = \frac{\phi\left(\frac{\alpha M + \sigma_M^2 \mu}{\sigma_M\sqrt{\alpha^2 + \sigma_M^2}}\right)}{\Phi\left(\frac{\alpha M + \sigma_M^2 \mu}{\sigma_M\sqrt{\alpha^2 + \sigma_M^2}}\right)}$ and $A_3 = \frac{\phi\left(\frac{\alpha M + \sigma_M^2 \mu}{\sigma_M\sqrt{\alpha^2 + \sigma_M^2}}\right)}{1 - \Phi\left(\frac{\alpha M + \sigma_M^2 \mu}{\sigma_M\sqrt{\alpha^2 + \sigma_M^2}}\right)}$.

$$
\begin{aligned}
\mathrm{E}(Y \mid M, X=1) &= \int_0^\infty \mathrm{E}_Y(Y \mid M, Z, X=1)\pi(Z \mid M, X=1)dZ \\
&= \int_0^\infty (\gamma Z + \beta M)\pi(Z \mid M, X=1)dZ \\
&= \beta M + \gamma \mathrm{E}(Z \mid M, X=1) \\
&= \beta M + \gamma \left(\frac{\alpha M + \sigma_M^2 \mu}{\sigma_M^2 + \alpha^2} + \frac{\sigma_M}{\sqrt{\sigma_M^2 + \alpha^2}} A_2 \right), \\
\mathrm{E}(Y \mid M, X=0) &= \mathrm{E}(Y \mid M, Z=0) \\
&= \beta M + \gamma \left(\frac{\alpha M + \sigma_M^2 \mu}{\sigma_M^2 + \alpha^2} - \frac{\sigma_M}{\sqrt{\sigma_M^2 + \alpha^2}} A_3 \right), \qquad \text{(C.1)} \\
\mathrm{E}(M \mid X=1) &= \int_0^\infty \mathrm{E}_Y(M \mid Z, X=1)\pi(Z \mid X=1)dZ \\
&= \int_0^\infty \alpha Z \pi(Z \mid X=1)dZ = \alpha \left(\mu + \frac{\phi(\mu)}{\Phi(\mu)} \right), \\
\mathrm{E}(M \mid X=0) &= \alpha \left(\mu - \frac{\phi(\mu)}{1-\Phi(\mu)} \right).
\end{aligned}
$$

Therefore, the NDE and NIE are given below,

$$
\begin{aligned}
&\text{NDE} \\
=&\mathrm{E}\{Y(1, M(0))\} - \mathrm{E}\{Y(0, M(0))\} \\
=&\mathrm{E}_M\{\mathrm{E}_Y(Y \mid M, X=1) \mid X=0\} - \mathrm{E}_M\{\mathrm{E}_Y(Y \mid M, X=0) \mid X=0\} \\
=&\mathrm{E}_M \left\{ \beta M + \gamma \left(\frac{\alpha M + \sigma_M^2 \mu}{\sigma_M^2 + \alpha^2} + \frac{\sigma_M}{\sqrt{\sigma_M^2 + \alpha^2}} A_2 \right) \mid X=0 \right\} - \\
&\quad \mathrm{E}_M \left\{ \beta M + \gamma \left(\frac{\alpha M + \sigma_M^2 \mu}{\sigma_M^2 + \alpha^2} - \frac{\sigma_M}{\sqrt{\sigma_M^2 + \alpha^2}} A_3 \right) \mid X=0 \right\} \\
=&\frac{\gamma \sigma_M}{\sqrt{\sigma_M^2 + \alpha^2}} \mathrm{E}_M(A_2 + A_3 \mid X=0),
\end{aligned}
$$

NIE

$$
\begin{aligned}
&= \mathrm{E}\{Y(1, M(1))\} - \mathrm{E}\{Y(1, M(0))\} \\
&= \mathrm{E}_M\{\mathrm{E}_Y(Y \mid M, X = 1) \mid X = 1\} - \mathrm{E}_M\{\mathrm{E}_Y(Y \mid M, X = 1) \mid X = 0\} \\
&= \mathrm{E}_M \left\{ \beta M + \gamma \left(\frac{\alpha M + \sigma_M^2 \mu}{\sigma_M^2 + \alpha^2} + \frac{\sigma_M}{\sqrt{\sigma_M^2 + \alpha^2}} A_2 \right) \mid X = 1 \right\} - \\
&\qquad \mathrm{E}_M \left\{ \beta M + \gamma \left(\frac{\alpha M + \sigma_M^2 \mu}{\sigma_M^2 + \alpha^2} + \frac{\sigma_M}{\sqrt{\sigma_M^2 + \alpha^2}} A_2 \right) \mid X = 0 \right\} \\
&= \frac{\alpha \phi(\mu)}{\Phi(\mu)(1 - \Phi(\mu))} \left(\beta + \frac{\alpha \gamma}{\sigma_M^2 + \alpha^2} \right) + \frac{\gamma \sigma_M}{\sqrt{\sigma_M^2 + \alpha^2}} \times \\
&\qquad \{\mathrm{E}_M(A_2 \mid X = 1) - \mathrm{E}_M(A_2 \mid X = 0)\}.
\end{aligned}
$$

BIBLIOGRAPHY

BIBLIOGRAPHY

Aitchison, J., S. Silvey, et al. (1958), Maximum-likelihood estimation of parameters subject to restraints, *The Annals of Mathematical Statistics*, *29*(3), 813–828.

Albert, J. M., and S. Nelson (2011), Generalized causal mediation analysis, *Biometrics*, *67*(3), 1028–1038.

Baek, J., B. Zhu, P. X. Song, et al. (2019), Bayesian analysis of infant's growth dynamics with in utero exposure to environmental toxicants, *The Annals of Applied Statistics*, *13*(1), 297–320.

Barfield, R., J. Shen, A. C. Just, P. S. Vokonas, J. Schwartz, A. A. Baccarelli, T. J. VanderWeele, and X. Lin (2017), Testing for the indirect effect under the null for genome-wide mediation analyses, *Genetic epidemiology*, *41*(8), 824–833.

Baron, R. M., and D. A. Kenny (1986), The moderator–mediator variable distinction in social psychological research: Conceptual, strategic, and statistical considerations., *Journal of personality and social psychology*, *51*(6), 1173.

Boeckx, R. L. (1986), Lead poisoning in children, *Analytical Chemistry*, *58*(2), 274A–288A.

Bollen, K. A., and R. Stine (1990), Direct and indirect effects: Classical and bootstrap estimates of variability, *Sociological methodology*, pp. 115–140.

Börnhorst, C., et al. (2017), Potential selection effects when estimating associations between the infancy peak or adiposity rebound and later body mass index in children, *International journal of obesity*, *41*(4), 518–526.

Byrd, R. H., P. Lu, J. Nocedal, and C. Zhu (1995), A limited memory algorithm for bound constrained optimization, *SIAM Journal on scientific computing*, *16*(5), 1190–1208.

Coffman, D. L., and W. Zhong (2012), Assessing mediation using marginal structural models in the presence of confounding and moderation., *Psychological methods*, *17*(4), 642.

Dai, J. Y., J. L. Stanford, and M. LeBlanc (2020), A multiple-testing procedure for high-dimensional mediation hypotheses, *Journal of the American Statistical Association*, (just-accepted), 1–39.

Daniel, R. M., B. L. De Stavola, S. Cousens, and S. Vansteelandt (2015), Causal mediation analysis with multiple mediators, *Biometrics*, *71*(1), 1–14.

Derkach, A., R. M. Pfeiffer, T.-H. Chen, and J. N. Sampson (2019), High dimensional mediation analysis with latent variables, *Biometrics*, *75*(3), 745–756.

Djordjilović, V., C. M. Page, J. M. Gran, T. H. Nøst, T. M. Sandanger, M. B. Veierød, and M. Thoresen (2019), Global test for high-dimensional mediation: Testing groups of potential mediators, *Statistics in medicine*, *38*(18), 3346–3360.

Dorfman, R. (1938), A note on the! d-method for finding variance formulae., *Biometric Bulletin*.

Efron, B. (1987), Better bootstrap confidence intervals, *Journal of the American statistical Association*, *82*(397), 171–185.

Ferreira, P. H., and F. Louzada (2014), A modified version of the inference function for margins and interval estimation for the bivariate clayton copula sur tobit model: An simulation approach, *arXiv preprint arXiv:1404.3287*.

Hu, H., et al. (2006), Fetal lead exposure at each stage of pregnancy as a predictor of infant mental development, *Environmental health perspectives*, *114*(11), 1730–1735.

Huang, B., S. Sivaganesan, P. Succop, and E. Goodman (2004), Statistical assessment of mediational effects for logistic mediational models, *Statistics in medicine*, *23*(17), 2713–2728.

Huang, Y.-T. (2019), Variance component tests of multivariate mediation effects under composite null hypotheses, *Biometrics*, *75*(4), 1191–1204.

Huang, Y.-T., and W.-C. Pan (2016), Hypothesis test of mediation effect in causal mediation model with high-dimensional continuous mediators, *Biometrics*, *72*(2), 402–413.

Huang, Y.-T., et al. (2018), Joint significance tests for mediation effects of socioeconomic adversity on adiposity via epigenetics, *The Annals of Applied Statistics*, *12*(3), 1535–1557.

Huang, Y.-T., et al. (2019), Genome-wide analyses of sparse mediation effects under composite null hypotheses, *The Annals of Applied Statistics*, *13*(1), 60–84.

Imai, K., L. Keele, and D. Tingley (2010a), A general approach to causal mediation analysis., *Psychological methods*, *15*(4), 309.

Imai, K., L. Keele, and T. Yamamoto (2010b), Identification, inference and sensitivity analysis for causal mediation effects, *Statistical science*, pp. 51–71.

Jensen, S. M., C. Ritz, K. T. Ejlerskov, C. Mølgaard, and K. F. Michaelsen (2015), Infant bmi peak, breastfeeding, and body composition at age 3 y, *The American journal of clinical nutrition*, *101*(2), 319–325.

Joe, H. (2005), Asymptotic efficiency of the two-stage estimation method for copula-based models, *Journal of multivariate Analysis*, *94*(2), 401–419.

Joe, H. (2014), *Dependence modeling with copulas*, CRC press.

Jørgensen, B. (1987), Exponential dispersion models, *Journal of the Royal Statistical Society: Series B (Methodological)*, *49*(2), 127–145.

Ko, V., and N. L. Hjort (2019), Model robust inference with two-stage maximum likelihood estimation for copulas, *Journal of Multivariate Analysis*, *171*, 362–381.

LaBarre, J. L., et al. (2020), Mitochondrial nutrient utilization underlying the association between metabolites and insulin resistance in adolescents, *The Journal of Clinical Endocrinology & Metabolism*, *105*(7), dgaa260.

Li, X., Z. Zhou, H. Qi, X. Chen, and G. Huang (2004), Replacement of insulin by fasting c-peptide in modified homeostasis model assessment to evaluate insulin resistance and islet beta cell function, *Zhong nan da xue xue bao. Yi xue ban= Journal of Central South University. Medical sciences*, *29*(4), 419–423.

Liu, Z., J. Shen, R. Barfield, J. Schwartz, A. A. Baccarelli, and X. Lin (2021), Large-scale hypothesis testing for causal mediation effects with applications in genome-wide epigenetic studies, *Journal of the American Statistical Association*, pp. 1–15.

Lu, L.-Z., and C. E. M. Pearce (2000), Some new bounds for singular values and eigenvalues of matrix products, *Annals of Operations Research*, *98*(1-4), 141–148.

Lustgarten, M. S., L. L. Price, A. Chalé, and R. A. Fielding (2014), Metabolites related to gut bacterial metabolism, peroxisome proliferator-activated receptor-alpha activation, and insulin sensitivity are associated with physical function in functionally-limited older adults, *Aging cell*, *13*(5), 918–925.

MacKinnon, D. P., C. M. Lockwood, J. M. Hoffman, S. G. West, and V. Sheets (2002), A comparison of methods to test mediation and other intervening variable effects., *Psychological methods*, *7*(1), 83.

Nelder, J. A., and R. Mead (1965), A simplex method for function minimization, *The computer journal*, *7*(4), 308–313.

Pearl, J. (2001), Direct and indirect effects, in *Proceedings of the Seventeenth conference on Uncertainty in artificial intelligence*, pp. 411–420.

Pearl, J. (2013), Direct and indirect effects, *arXiv preprint arXiv:1301.2300*.

Perng, W., E. C. Hector, P. X. Song, M. M. Tellez Rojo, S. Raskind, M. Kachman, A. Cantoral, C. F. Burant, and K. E. Peterson (2017), Metabolomic determinants of metabolic risk in mexican adolescents, *Obesity*, *25*(9), 1594–1602.

Perng, W., J. Baek, C. W. Zhou, A. Cantoral, M. M. Tellez-Rojo, P. X. Song, and K. E. Peterson (2018), Associations of the infancy body mass index peak with anthropometry and cardiometabolic risk in mexican adolescents, *Annals of human biology*, *45*(5), 386–394.

Perng, W., et al. (2019), Early life exposure in mexico to environmental toxicants (element) project, *BMJ open*, *9*(8), e030,427.

Preacher, K. J. (2015), Advances in mediation analysis: A survey and synthesis of new developments, *Annual review of psychology*, *66*.

Qian, Y., H. Shao, X. Ying, W. Huang, and Y. Hua (2020), The endocrine disruption of prenatal phthalate exposure in mother and offspring, *Frontiers in Public Health*, *8*.

Robins, J. M., and S. Greenland (1992), Identifiability and exchangeability for direct and indirect effects, *Epidemiology*, *3*(2), 143–155.

Rosenbaum, P. R., and D. B. Rubin (1983), The central role of the propensity score in observational studies for causal effects, *Biometrika*, *70*(1), 41–55.

Rubin, D. B. (1978), Bayesian inference for causal effects: The role of randomization, *The Annals of statistics*, pp. 34–58.

Scholl, T. O., X. Chen, and T. P. Stein (2014), Maternal calcium metabolic stress and fetal growth, *The American journal of clinical nutrition*, *99*(4), 918–925.

Shih, J. H., and T. A. Louis (1995), Inferences on the association parameter in copula models for bivariate survival data, *Biometrics*, pp. 1384–1399.

Sklar, M. (1959), Fonctions de repartition an dimensions et leurs marges, *Publ. inst. statist. univ. Paris*, *8*, 229–231.

Sobel, M. E. (1982), Asymptotic confidence intervals for indirect effects in structural equation models, *Sociological methodology*, *13*, 290–312.

Song, P. X. (2000), Multivariate dispersion models generated from gaussian copula, *Scandinavian Journal of Statistics*, *27*(2), 305–320.

Song, P. X.-K., M. Li, and Y. Yuan (2009), Joint regression analysis of correlated data using gaussian copulas, *Biometrics*, *65*(1), 60–68.

Song, Y., et al. (2020), Bayesian shrinkage estimation of high dimensional causal mediation effects in omics studies, *Biometrics*, *76*(3), 700–710.

Splawa-Neyman, J., D. M. Dabrowska, and T. Speed (1990), On the application of probability theory to agricultural experiments. essay on principles. section 9., *Statistical Science*, pp. 465–472.

Tingley, D., T. Yamamoto, K. Hirose, L. Keele, and K. Imai (2014), Mediation: R package for causal mediation analysis.

Ullman, J. B., and P. M. Bentler (2003), Structural equation modeling, *Handbook of psychology*, pp. 607–634.

Valeri, L., and T. J. VanderWeele (2013), Mediation analysis allowing for exposure–mediator interactions and causal interpretation: theoretical assumptions and implementation with sas and spss macros., *Psychological methods*, *18*(2), 137.

VanderWeele, T. (2015), *Explanation in Causal Inference: Methods for Mediation and Interaction*, Oxford University Press.

VanderWeele, T., and S. Vansteelandt (2014), Mediation analysis with multiple mediators, *Epidemiologic methods*, *2*(1), 95–115.

VanderWeele, T. J. (2011), Causal mediation analysis with survival data, *Epidemiology (Cambridge, Mass.)*, *22*(4), 582.

VanderWeele, T. J., and S. Vansteelandt (2009), Conceptual issues concerning mediation, interventions and composition, *Statistics and its Interface*, *2*(4), 457–468.

VanderWeele, T. J., and S. Vansteelandt (2010), Odds ratios for mediation analysis for a dichotomous outcome, *American journal of epidemiology*, *172*(12), 1339–1348.

Vansteelandt, S., and T. J. VanderWeele (2012), Natural direct and indirect effects on the exposed: effect decomposition under weaker assumptions, *Biometrics*, *68*(4), 1019–1027.

Vigeh, M., K. Yokoyama, Z. Seyedaghamiri, A. Shinohara, T. Matsukawa, M. Chiba, and M. Yunesian (2011), Blood lead at currently acceptable levels may cause preterm labour, *Occupational and environmental medicine*, *68*(3), 231–234.

Vigeh, M., K. Yokoyama, T. Matsukawa, A. Shinohara, and K. Ohtani (2014), Low level prenatal blood lead adversely affects early childhood mental development, *Journal of child neurology*, *29*(10), 1305–1311.

Willett, W. C., G. R. Howe, and L. H. Kushi (1997), Adjustment for total energy intake in epidemiologic studies, *The American journal of clinical nutrition*, *65*(4), 1220S–1228S.

Wolak, F. A. (1989), Local and global testing of linear and nonlinear inequality constraints in nonlinear econometric models, *Econometric Theory*, *5*(1), 1–35.

Xu, J. J. (1996), Statistical modelling and inference for multivariate and longitudinal discrete response data, Ph.D. thesis, University of British Columbia.

Zhao, Y., and X. Luo (2016), Pathway lasso: estimate and select sparse mediation pathways with high dimensional mediators, *arXiv preprint arXiv:1603.07749*.

Zhao, Y., M. A. Lindquist, and B. S. Caffo (2020), Sparse principal component based high-dimensional mediation analysis, *Computational Statistics & Data Analysis*, *142*, 106,835.

Zhou, C., W. Perng, J. Baek, D. Watkins, A. Cantoral, M. M. Tellez-Rojo, K. E. Peterson, and P. X. Song (2021), Association of prenatal phthalates exposure with time to infancy body mass index peak: a longitudinal study in mexico city, *Working Paper*.

Zhu, M., E. F. Fitzgerald, K. H. Gelberg, S. Lin, and C. M. Druschel (2010), Maternal low-level lead exposure and fetal growth, *Environmental health perspectives*, *118*(10), 1471–1475.