JOURNAL OF BUSINESS & ECONOMIC STATISTICS
2022, VOL. 40, NO. 4, 1759–1771
https://doi.org/10.1080/07350015.2021.1971089

Varying Coefficient Mediation Model and Application to Analysis of Behavioral Economics Data

Yujie Liao[a], Jingyuan Liu[b,c], Donna L. Coffman[d], and Runze Li[a]

[a]Department of Statistics, Pennsylvania State University, University Park, PA; [b]MOE Key Laboratory of Econometrics, Department of Statistics and Data Science, School of Economics, Wang Yanan Institute for Studies in Economics, Xiamen University, Xiamen, China; [c]Fujian Key Lab of Statistics, Xiamen University, Xiamen, China; [d]Department of Epidemiology and Biostatistics, Temple University, Philadelphia, PA

ABSTRACT
This article is concerned with causal mediation analysis with varying indirect and direct effects. We propose a varying coefficient mediation model, which can also be viewed as an extension of moderation analysis on a causal diagram. We develop a new estimation procedure for the direct and indirect effects based on B-splines. Under mild conditions, rates of convergence and asymptotic distributions of the resulting estimates are established. We further propose a *F*-type test for the direct effect. We conduct simulation study to examine the finite sample performance of the proposed methodology, and apply the new procedures for empirical analysis of behavioral economics data.

ARTICLE HISTORY
Received May 2021
Accepted July 2021

KEYWORDS
B-splines; Indirect effect; Mediation analysis; Varying coefficient models

1. Introduction

Since the seminal work of Baron and Kenny (1986), mediation models have been applied in many fields such as finance, psychology, communication research, genetics, and epidemiology (MacKinnon 2008; Huang et al. 2015; Huang, VanderWeele, and Lin 2014). Mediation occurs when an intervention or exposure has an effect on an outcome via a third, intermediate variable. The effect through the third variable is called the indirect effect. The effect of the exposure on the outcome that does not occur via the mediator is called the direct effect. Mediation analysis has received increasing attention in the recent literature. However, much of existing literature has focused on relatively simple parametric relationships (e.g., linear) among the variables although more recently significant advances have allowed for multiple mediators, "effect modifiers" (i.e., moderators), and mixed variable types (e.g., binary outcome or binary mediators). For example, path-specific effects were introduced to characterize the indirect effect of a pathway through multiple mediators (Avin, Shpitser, and Pearl 2005; Daniel et al. 2015). Generalized mediation models with mixed variable types have been proposed (Albert and Nelson 2011; Albert et al. 2019), and various methods of variable selection, dimension reduction, and decorrelation in the high-dimensional case have been introduced (Huang and Pan 2016; Zhang et al. 2016; Chén et al. 2018; Zhou, Wang, and Zhao 2020). Effect modifiers, or variables that modify the direct and indirect effects, have been incorporated into the mediation model albeit in a parametric form by, for example, using product terms between the effect modifier and the exposure or between the effect modifier and the mediator (Edwards and Lambert 2007; Preacher, Rucker, and Hayes 2007; Valeri and VanderWeele 2013). However, the indirect effect may vary in a nonparametric way across levels of the effect modifier.

Motivated by an empirical analysis of behavioral economic data presented in Section 4, this article aims to develop a varying coefficient mediation model (VCMM), which allows coefficients of exposure variables and mediators to vary as smooth functions of a continuous effect modifier variable. We provide a foundation for statistical inference for varying indirect effects under the regression analysis framework for mediation. As is known, compared with the potential outcomes framework for mediation analysis, the regression framework has stricter model assumptions, such as linearity. By introducing these varying coefficient functions into mediation regression models, we can relax some of these assumptions and add significant flexibility compared with the ordinary parametric mediation models. The indirect and direct effects can then be represented as curves of another variable, and estimated through polynomial spline techniques inspired by Huang, Wu, and Zhou (2002, 2004). In fact, when a coefficient function is set to be the effect modifier itself, the product of the varying coefficient function and the main effect is indeed an interaction term in linear mediation models. Thus, the VCMM can also be viewed as an extension of the classic moderation analysis in causal mediation models, where the interaction effects are of primary interest. In this article, we establish consistency and convergence rates of the resulting estimates. We further derive their asymptotic distributions. Although both varying coefficient models (Cleveland et al. 1992; Fan et al. 1999; Cai, Fan, and Li 2000; Fan and Zhang 2008) and linear mediation analysis have been systematically studied in many existing literatures, the derivation of the asymptotic properties in the proposed VCMM is not trivial partly because

CONTACT Jingyuan Liu *jingyuan@xmu.edu.cn* Department of Statistics, Xiamen University, D307 Economics Bldg, Xiamen, Fujian, 361005 China.

the mediators are random and typically unbounded, and therefore existing theory for varying coefficient models with bounded covariates are not applicable for the setting studied in this article. In addition, the point-wise confidence intervals and an F-type test are also proposed for possible inferential pursuit.

We further use the advocated methodology of VCMM to analyze small business owners' affective and behavioral changes, in response to economic pressure. The learned helplessness theory convincingly shows that people encountering an event out of their control are likely to have a sense of helplessness (Hiroto and Seligman 1975; Pryce et al. 2011), and increased depression would predict increased psychological intentions to withdraw (Pollack, Vanepps, and Hayes 2012). Thus, the mediating role of depression on the entrepreneurs' withdrawal intentions in response to economic stress is of practical interest. In addition, entrepreneurs with different backgrounds, such as length of tenures in their own companies, may perceive stress differently. Some may consider economic pressure as a challenge while others view it as a hindrance. Therefore, it is of interest to explore the buffering/promoting effect of individual backgrounds on the effect of economic stress on withdrawal intentions via depressed affect. The results of this analysis are given in Section 4.

The rest of this article is organized as follows. In Section 2, we propose a new varying coefficient mediation model and develop an estimation procedure for the direct and indirect effect coefficient functions. We further study theoretical properties of the proposed estimator. In Section 3, we conduct simulation studies to verify the finite sample performance of the proposed methods. Section 4 includes the empirical varying coefficient mediation analysis using a behavioral economics dataset. Some discussion and conclusion are given in Section 5. Technical details are provided in the appendix.

2. Varying Coefficient Mediation Model

2.1. Model Setup

We consider a varying coefficient mediation model that allows the direct and indirect coefficients to vary as smooth functions over values of another variable (i.e., effect modifier). Specifically, let y be the response, $\boldsymbol{x}$ be the q-dimensional exposure variable, u be the effect modifier, and $\boldsymbol{m}$ be the p-dimensional mediator. The VCMM is then presented as follows:

$$y = \boldsymbol{\alpha}_0(u)^T \boldsymbol{m} + \boldsymbol{\alpha}_1(u)^T \boldsymbol{x} + \epsilon_1 \tag{1}$$

$$\boldsymbol{m} = \boldsymbol{\Gamma}(u)^T \boldsymbol{x} + \boldsymbol{\epsilon}, \tag{2}$$

where ϵ_1 is an error term for model (1), with mean 0 and variance σ_1^2, and $\boldsymbol{\epsilon}$ is a p-dimensional error vector for model (2), with mean $\mathbf{0}$ and covariance matrix $\boldsymbol{\Sigma}_\epsilon$. The errors are independent of each other, as well as with $\boldsymbol{m}$, $\boldsymbol{x}$ and u. For any u in its support $\mathcal{U}$, $\boldsymbol{\alpha}_0(u) = (\alpha_{01}(u), \dots, \alpha_{0p}(u))^T$ and $\boldsymbol{\alpha}_1(u) = (\alpha_{11}(u), \dots, \alpha_{1q}(u))^T$ are the p- and q-dimensional coefficient vectors, respectively, and $\boldsymbol{\Gamma}(u) = (\boldsymbol{\Gamma}_1(u)^T, \dots, \boldsymbol{\Gamma}_q(u)^T)^T$ is the $q \times p$ coefficient matrix, where $\boldsymbol{\Gamma}_l(u) = (\Gamma_{l1}(u), \dots, \Gamma_{lp}(u))^T$ is the lth row of $\boldsymbol{\Gamma}(u)$, $l = 1, \dots, q$. Plugging model (2) into model (1) yields

$$\begin{aligned} y &= \{\boldsymbol{\Gamma}(u)\boldsymbol{\alpha}_0(u) + \boldsymbol{\alpha}_1(u)\}^T \boldsymbol{x} + \epsilon_1 + \boldsymbol{\alpha}_0(u)^T \boldsymbol{\epsilon} \\ &\triangleq \{\boldsymbol{\beta}(u) + \boldsymbol{\alpha}_1(u)\}^T \boldsymbol{x} + \epsilon_{\text{tot}} \\ &\triangleq \boldsymbol{\gamma}(u)^T \boldsymbol{x} + \epsilon_{\text{tot}}, \end{aligned} \tag{3}$$

where $\boldsymbol{\beta}(u) = \boldsymbol{\Gamma}(u)\boldsymbol{\alpha}_0(u)$ is called the indirect effect of exposure $\boldsymbol{x}$ on y mediated through $\boldsymbol{m}$, and is our main interest in this article. $\boldsymbol{\alpha}_1(u)$ is called the direct effect of $\boldsymbol{x}$, and $\boldsymbol{\gamma}(u) = \boldsymbol{\beta}(u) + \boldsymbol{\alpha}_1(u)$ is called the total effect. Moreover, $\epsilon_{\text{tot}} = \epsilon_1 + \boldsymbol{\alpha}_0(u)^T \boldsymbol{\epsilon}$ is the total error with mean 0 and variance σ_{tot}^2.

2.2. Estimating Procedures

In this section, we systematically study the estimation procedure of the direct effect function $\boldsymbol{\alpha}_1(u)$ and the indirect effect function $\boldsymbol{\beta}(u) = \boldsymbol{\Gamma}(u)\boldsymbol{\alpha}_0(u)$. Suppose $\{(\boldsymbol{x}_i, \boldsymbol{m}_i, u_i, y_i)\}_{i=1}^n$ is a random sample. Then the sample-level presentation of model (1) and (2) are

$$\begin{aligned} y_i &= \boldsymbol{\alpha}_0(u_i)^T \boldsymbol{m}_i + \boldsymbol{\alpha}_1(u_i)^T \boldsymbol{x}_i + \epsilon_{1i} \\ &= \sum_{j=1}^{p} \alpha_{0j}(u_i) m_{ij} + \sum_{l=1}^{q} \alpha_{1l}(u_i) x_{il} + \epsilon_{1i}, \end{aligned} \tag{4}$$

$$\boldsymbol{m}_i = \boldsymbol{\Gamma}(u_i)^T \boldsymbol{x}_i + \boldsymbol{\epsilon}_i = \sum_{l=1}^{q} \boldsymbol{\Gamma}_l(u_i) x_{il} + \boldsymbol{\epsilon}_i. \tag{5}$$

We apply a regression spline method to estimate the potential varying effects described by the corresponding coefficient functions. Specifically, using B-splines, we approximate $\{\alpha_{0j}(u_i), j = 1, \dots, p\}$ and $\{\alpha_{1l}(u_i), l = 1, \dots, q\}$ by

$$\alpha_{0j}(u_i) \approx \sum_{k=1}^{K_0} a_{0jk} b_{0jk}(u_i) \triangleq \boldsymbol{a}_{0j}^T \boldsymbol{b}_{0j}(u_i), \text{ and}$$

$$\alpha_{1l}(u_i) \approx \sum_{k=1}^{K_1} a_{1lk} b_{1lk}(u_i) \triangleq \boldsymbol{a}_{1l}^T \boldsymbol{b}_{1l}(u_i),$$

where for all $j = 1, \dots, p$ and $l = 1, \dots, q$, $\{b_{0jk}(\cdot), k = 1 \dots, K_0\}$ and $\{b_{1lk}(\cdot), k = 1 \dots, K_1\}$ are sets of B-spline bases in linear spaces $\mathbb{G}_{\alpha_{0j}}$ and $\mathbb{G}_{\alpha_{1l}}$ on $\mathcal{U}$, respectively, with fixed degrees and knots. K_0 and K_1 are the numbers of basis functions for $\alpha_{0j}(\cdot)$ and $\alpha_{1l}(\cdot)$, $l = 1, \dots, q$, $j = 1, \dots, p$, respectively, which are allowed to increase with the sample size n.

Since the lth element of $\hat{\boldsymbol{\beta}}(u)$ is only determined by the lth row of $\hat{\boldsymbol{\Gamma}}(u)$, we consider expanding $\boldsymbol{\Gamma}(u)$ row by row. Denote $\{d_{ljk}(\cdot), k = 1, \dots, K_m\}$ to be the B-spline basis in linear space $\mathbb{G}_{\Gamma_{lj}}$ on $\mathcal{U}$, where K_m is the number of basis functions correspondingly. For ease of presentation, set $d_{lk}(u) \equiv d_{ljk}(u)$, then for every $j = 1, \dots, p$, we can approximate $\{\boldsymbol{\Gamma}_l(u_i), l = 1, \dots q\}$ as

$$\boldsymbol{\Gamma}_l(u_i) \approx \sum_{k=1}^{K_m} \boldsymbol{c}_{lk} d_{lk}(u_i) \triangleq \boldsymbol{C}_l^T \boldsymbol{d}_l(u_i),$$

where $\boldsymbol{c}_{lk} = (c_{l1k}, \dots, c_{lpk})^T \in \mathbb{R}^{p \times 1}$ and $\boldsymbol{C}_l = (\boldsymbol{c}_{l1}, \dots, \boldsymbol{c}_{lK_m})^T \in \mathbb{R}^{K_m \times p}$.

Thus model (4) and (5) can be approximated by

$$y_i \approx \sum_{j=1}^{p} a_{0j}^T b_{0j}(u_i) m_{ij} + \sum_{l=1}^{q} a_{1l}^T b_{1l}(u_i) x_{il} + \epsilon_{1i}, \quad (6)$$

$$m_i \approx \sum_{l=1}^{q} C_l^T d_l(u_i) x_{il} + \epsilon_i. \quad (7)$$

Let $\alpha_0^* = (a_{01}^T, \ldots, a_{0p}^T)^T \in \mathbb{R}^{pK_0 \times 1}$, $m_i^* = (m_{i1} b_{01}(u_i)^T, \ldots, m_{ip} b_{0p}(u_i)^T)^T \in \mathbb{R}^{pK_0 \times 1}$, $\alpha_1^* = (a_{11}^T, \ldots, a_{1q}^T)^T \in \mathbb{R}^{qK_1 \times 1}$, $x_i^* = (x_{i1} b_{11}(u_i)^T, \ldots, x_{iq} b_{1q}(u_i)^T)^T \in \mathbb{R}^{qK_1 \times 1}$, $C = (C_1^T, \ldots, C_q^T)^T \in \mathbb{R}^{qK_m \times p}$ and $x_i^m = (x_{i1} d_1(u_i)^T, \ldots, x_{iq} d_q(u_i)^T)^T \in \mathbb{R}^{qK_m \times 1}$. Then Equations (6) and (7) become $y_i \approx \alpha_0^{*T} m_i^* + \alpha_1^{*T} x_i^* + \epsilon_{1i}$ and $m_i \approx C^T x_i^m + \epsilon_i$. In the matrix form,

$$Y \approx M^* \alpha_0^* + X^* \alpha_1^* + E_1 \quad \text{and} \quad M \approx X^m C + E,$$

where $Y = (y_1, \ldots, y_n)^T$, $M^* = (m_1^*, \ldots, m_n^*)^T$, $X^* = (x_1^*, \ldots, x_n^*)^T$, $E_1 = (\epsilon_{11}, \ldots, \epsilon_{1n})^T$, and $M = (m_1, \ldots, m_n)^T$, $X^m = (x_1^m, \ldots x_n^m)^T$ and $E = (\epsilon_1, \ldots, \epsilon_n)^T$. Therefore, estimating coefficient functions $\alpha_0(u)$, $\alpha_1(u)$ and $\Gamma(u)$ is converted to estimating coefficients a_0^*, a_1^* and C.

By minimizing $\ell_1(\alpha_0^*, \alpha_1^*) = \|Y - M^* \alpha_0^* - X^* \alpha_1^*\|^2$ with respect to α_0^* and α_1^*, we obtained the least-square estimates of α_0^* and α_1^*. The estimate of C can be obtained by minimizing $\ell_2(C) = \|M - X^m C\|^2$ with respect to C. Specifically,

$$\hat{\alpha}_0^* = \{M^{*T}(I_n - P_{X^*})M^*\}^{-1} M^{*T}(I_n - P_{X^*})Y,$$
$$\hat{\alpha}_1^* = \{X^{*T}(I_n - P_{M^*})X^*\}^{-1} X^{*T}(I_n - P_{M^*})Y,$$
$$\hat{C} = (X^{mT} X^m)^{-1} X^{mT} M,$$

where $P_{X^*} = X^*(X^{*T}X^*)^{-1}X^{*T}$ and $P_{M^*} = M^*(M^{*T}M^*)^{-1}M^{*T}$ are the corresponding projection matrices.

For sake of simplicity, we further set $b_{0j}(u_i) \equiv b_0(u_i)$, for all $j = 1, \ldots, p$, $b_{1l}(u_i) \equiv b_1(u_i)$, for all $l = 1, \ldots, q$, and $d_l(u_i) = d(u_i)$, for all $l = 1, \ldots q$. Thus,

$$\hat{\alpha}_0(u_i) = (b_0(u_i)^T \hat{a}_{01}, \ldots, b_0(u_i)^T \hat{a}_{0p})^T = \{I_p \otimes b_0(u_i)^T\} \hat{\alpha}_0^*, \quad (8)$$

$$\hat{\alpha}_1(u_i) = (b_1(u_i)^T \hat{a}_{11}, \ldots, b_1(u_i)^T \hat{a}_{1q})^T = \{I_q \otimes b_1(u_i)^T\} \hat{\alpha}_1^*, \quad (9)$$

$$\hat{\Gamma}(u_i) = (\hat{C}_1^T d(u_i), \ldots, \hat{C}_q^T d(u_i))^T = \{I_q \otimes d(u_i)^T\} \hat{C}, \quad (10)$$

with $\otimes$ denoting the Kronecker product. The estimates $\hat{\alpha}_0(u)$, $\hat{\alpha}_1(u)$ and $\hat{\Gamma}(u)$ at any point u can also be obtained. Thus, the direct effect $\alpha_1(u)$ and the indirect effect $\beta(u)$ are estimated by $\hat{\alpha}_1(u)$ and $\hat{\beta}(u) = \hat{\Gamma}(u)\hat{\alpha}_0(u)$, respectively.

Remark 1. To implement the proposed estimation procedure, one needs to determine K_0, K_1, and K_m, which control the model complexity of $\alpha_0(\cdot)$, $\alpha_1(\cdot)$ and $\Gamma(\cdot)$, respectively. Condition C6 (in the appendix) presents technical conditions on these tuning parameters to establish the sampling properties of the proposed estimation procedure. In practical implementation, one may use cross-validation to select K_0, K_1 and K_m. This leads an expensive computational cost. Thus, we would suggest setting K_0, K_1 and K_m be the same, chosen by 5- or 10-fold cross-validation. This is implemented in our real data analysis.

2.3. Hypothesis Tests

We in this section provide tests for the direct effect $\alpha_1(u)$ based on model (1), as well as the total effect $\gamma(u)$ based on model (3). Denote $\alpha_1(u) = (\alpha_{11}(u), \alpha_{12}(u), \ldots, \alpha_{1q}(u))^T$, where $\alpha_{11}(u)$ refers to the intercept. It is of interest to test

$$H_0 : \alpha_{1l}(u) = 0 \text{ for all } u \in \mathcal{U}, l = 2, \ldots, q,$$
$$\text{versus } H_1 : \alpha_{1l}(u) \neq 0 \text{ for some } u \in \mathcal{U}, l = 2, \ldots, q. \quad (11)$$

With the aid of the approximated model (6), we formulate this hypothesis problem as

$$H_0 : a_{1lk} = 0 \text{ for all } k = 1, \ldots, K_1, l = 2, \ldots, q,$$
$$\text{versus } H_1 : a_{1lk} \neq 0 \text{ for some } k = 1, \ldots, K_1, l = 2, \ldots, q, \quad (12)$$

which is essentially testing regression coefficients in linear regression models. Instead of testing many hypotheses on one-dimensional parameters $H_{0lk} : a_{1lk} = 0$ for $k = 1, \ldots, K_1$ and $l = 2, \ldots, q$, which leads to a large scale of multiple testings, we test H_0 directly. Generalized likelihood ratio test was proposed for the generalized varying coefficient models in Cai, Fan, and Li (2000). As it is well known, the traditional F-test is equivalent to the likelihood ratio test in the normal linear model. Thus, we adopt the following F-type test statistic for hypothesis (12)

$$T_n = \frac{(\text{RSS}_0 - \text{RSS})/(q-1)K_1}{\text{RSS}/(N - pK_0 - qK_1)},$$

where RSS_0 and RSS are the residual sum of squares under H_0 and H_1, respectively. In general, T_n approximately follows F distribution with degrees of freedom $(q-1)K_1$ and $N - pK_0 - qK_1$ when the sample size is large enough. In our simulation, we set the critical value of T_n to be the corresponding critical value of the F distribution, and find it performs well in terms of retaining Type I error rate.

Remark 2. In theory, as K_1 tends to ∞, the degrees of F-distribution tend to ∞. Thus, the traditional F distribution cannot be applied directly. Note that $\text{RSS}/(N - pK_0 - qK_1)$ converges to the variance of ϵ_1 in probability, and $\text{RSS}_0 - \text{RSS}$ tends to a χ^2 distribution with degrees of freedom $(q-1)K_1$ in distribution. Thus, if $(q-1)K_1$ is relatively large, we may consider the following standardized T_n

$$T_{sn} = \{(q-1)K_1\}\{T_n - 1\}/\sqrt{2(q-1)K_1}.$$

It can be shown that T_{sn} asymptotically follows a $N(0, 1)$ by using the properties of F-distribution and χ^2-distribution with diverging degrees of freedom.

2.4. Asymptotic Properties

We next study the asymptotic properties of the direct effect estimate, $\hat{\alpha}_1(u)$, and the indirect effect estimate, $\hat{\beta}(u) = \hat{\Gamma}(u)\hat{\alpha}_0(u)$. We begin by introducing notations. Suppose that $\|a\|_{L_2}$ denotes the L_2 norm of a square integrable function $a(\cdot)$ on $\mathcal{U}$, that is, $\|a(\cdot)\|_{L_2}^2 = \int_{u \in \mathcal{U}} |a(u)|^2 f_U(u) du$. Let $\text{dist}\{\alpha_{0j}, \mathbb{G}_{\alpha_{0j}}\} = \inf_{g_{0j} \in \mathbb{G}_{\alpha_{0j}}} \sup_{u \in \mathcal{U}} |\alpha_{0j}(u) - g_{0j}(u)|$ be the L_∞ distance between $\alpha_{0j}(\cdot)$ and $\mathbb{G}_{\alpha_{0j}}$, and take $\rho_{n0} = \max_{1 \le j \le p} \text{dist}\{\alpha_{0j}, \mathbb{G}_{\alpha_{0j}}\}$. Similarly, let

$\mathrm{dist}\{\alpha_{1l}, \mathbb{G}_{\alpha_{1l}}\} = \inf_{g_{1l}\in\mathbb{G}_{\alpha_{1l}}} \sup_{u\in\mathcal{U}} |\alpha_{1l}(u) - g_{1l}(u)|$ and take $\rho_{n1} = \max_{1\le l\le q} \mathrm{dist}\{\alpha_{1l}, \mathbb{G}_{\alpha_{1l}}\}$. Finally, let $\mathrm{dist}\{\Gamma_{lj}, \mathbb{G}_{\Gamma_{lj}}\} = \inf_{g_{lj}\in\mathbb{G}_{\Gamma_{lj}}} \sup_{u\in\mathcal{U}} |\Gamma_{lj}(u) - g_{lj}(u)|$, and take $\rho_{n2} = \max_{1\le j\le p} \max_{1\le l\le q} \mathrm{dist}\{\Gamma_{lj}, \mathbb{G}_{\Gamma_{lj}}\}$. The approximation rates ρ_{n0}, ρ_{n1} and ρ_{n2} depend on K_0, K_1, and K_m, respectively, under commonly used smoothing conditions on the elements of $\boldsymbol{\alpha}_0$, $\boldsymbol{\alpha}_1$, and $\boldsymbol{\Gamma}$, respectively (Schumaker 2007). Under Conditions C4 and C5 (in the appendix), it has been shown that $\rho_{n0} = O(K_0^{-2})$, $\rho_{n1} = O(K_1^{-2})$ and $\rho_{n2} = O(K_m^{-2})$.

First, we establish the asymptotic normality of $\hat{\boldsymbol{\xi}}^* = (\hat{\boldsymbol{\alpha}}_0^{*T}, \hat{\boldsymbol{\alpha}}_1^{*T})^T$, and then derive the asymptotic distribution of $\hat{\boldsymbol{\alpha}}_0(u)$ and the estimated direct effect, $\hat{\boldsymbol{\alpha}}_1(u)$.

Let $z_i^* = [\boldsymbol{m}_i^{*T}, \boldsymbol{x}_i^{*T}]^T$ and $Z^* = [M^*, X^*]$, then $\ell_1(\boldsymbol{\alpha}_0^*, \boldsymbol{\alpha}_1^*)$ has a unique minimizer $\hat{\boldsymbol{\xi}}^* = (\hat{\boldsymbol{\alpha}}_0^{*T}, \hat{\boldsymbol{\alpha}}_1^{*T})^T = (\boldsymbol{Z}^{*T}\boldsymbol{Z}^*)^{-1}\boldsymbol{Z}^{*T}\boldsymbol{Y}$ when $\boldsymbol{Z}^*$ is of full rank. Set $\tilde{y}_i = \boldsymbol{\alpha}_0(u_i)^T\boldsymbol{m}_i + \boldsymbol{\alpha}_1(u_i)^T\boldsymbol{x}_i$, for any $i = 1, \dots, n$, and $\tilde{Y} = (\tilde{y}_1, \dots, \tilde{y}_n)^T$. Define $\tilde{\boldsymbol{\xi}}^* = (\tilde{\boldsymbol{\alpha}}_0^{*T}, \tilde{\boldsymbol{\alpha}}_1^{*T})^T = (\boldsymbol{Z}^{*T}\boldsymbol{Z}^*)^{-1}\boldsymbol{Z}^{*T}\tilde{\boldsymbol{Y}}$. Then, it follows that $\mathrm{E}(\hat{\boldsymbol{\xi}}^*) = \tilde{\boldsymbol{\xi}}^*$, where the expectation is taken conditional on $\mathcal{X} = \{(\boldsymbol{x}_i, u_i) : i = 1, \dots, n\}$. Specifically, $\mathrm{E}(\hat{\boldsymbol{\alpha}}_0^*) = \tilde{\boldsymbol{\alpha}}_0^*$ and $\mathrm{E}(\hat{\boldsymbol{\alpha}}_1^*) = \tilde{\boldsymbol{\alpha}}_1^*$. Moreover, $\mathrm{E}\{\hat{\boldsymbol{\alpha}}_0(u_i)\} = \tilde{\boldsymbol{\alpha}}_0(u_i) \triangleq \{\boldsymbol{I}_p \otimes \boldsymbol{b}_0(u_i)^T\}\tilde{\boldsymbol{\alpha}}_0^*$ and $\mathrm{E}\{\hat{\boldsymbol{\alpha}}_1(u_i)\} = \tilde{\boldsymbol{\alpha}}_1(u_i) \triangleq \{\boldsymbol{I}_q \otimes \boldsymbol{b}_1(u_i)^T\}\tilde{\boldsymbol{\alpha}}_1^*$.

Since the length of vector $\hat{\boldsymbol{\xi}}^*$ depends on the number of knots, which tends to infinity as $n \to \infty$, we establish the asymptotic normality of $\hat{\boldsymbol{\xi}}^*$ by deriving the asymptotic normality of $\boldsymbol{g}^T\hat{\boldsymbol{\xi}}^*$ for any nonzero constant vector $\boldsymbol{g}$ in the following theorem (see Appendix for proof).

Theorem 1. Under Conditions C1–C7 in the appendix, for any nonzero constant vector $\boldsymbol{g} \in \mathbb{R}^{(pK_0+qK_1)\times 1}$, the estimate $\hat{\boldsymbol{\xi}}^*$ satisfies

$$\{\mathrm{var}_a(\boldsymbol{g}^T\hat{\boldsymbol{\xi}}^*)\}^{-1/2}\boldsymbol{g}^T(\hat{\boldsymbol{\xi}}^* - \tilde{\boldsymbol{\xi}}^*) \xrightarrow{D} N(0, 1),$$

where $\mathrm{var}_a(\boldsymbol{g}^T\hat{\boldsymbol{\xi}}^*) = \frac{1}{n}\sigma_1^2\boldsymbol{g}\boldsymbol{\Sigma}_{z^*z^*}^{-1}\boldsymbol{g}^T$ is the asymptotic variance.

Theorem 1 implies that $\sqrt{n}(\hat{\boldsymbol{\xi}}^* - \tilde{\boldsymbol{\xi}}^*)$ asymptotically follows a multivariate normal distribution with mean $\boldsymbol{0}$ and covariance matrix $\sigma_1^2\boldsymbol{\Sigma}_{z^*z^*}^{-1}$. Furthermore, for any nonzero constant matrix $\boldsymbol{G}$, $\sqrt{n}\boldsymbol{G}(\hat{\boldsymbol{\xi}}^* - \tilde{\boldsymbol{\xi}}^*)$ asymptotically follows a multivariate normal distribution with mean $\boldsymbol{0}$ and covariance matrix $\sigma_1^2\boldsymbol{G}\boldsymbol{\Sigma}_{z^*z^*}^{-1}\boldsymbol{G}^T$. In particular, to obtain the asymptotic normality of $\hat{\boldsymbol{\alpha}}_0^*$ and $\hat{\boldsymbol{\alpha}}_1^*$, note that

$$\boldsymbol{\Sigma}_{z^*z^*} = \mathrm{E}(z_i^*z_i^{*T}) = \mathrm{E}\begin{pmatrix} \boldsymbol{m}_i^*\boldsymbol{m}_i^{*T} & \boldsymbol{m}_i^*\boldsymbol{x}_i^{*T} \\ \boldsymbol{x}_i^*\boldsymbol{m}_i^{*T} & \boldsymbol{x}_i^*\boldsymbol{x}_i^{*T} \end{pmatrix} \triangleq \begin{pmatrix} \boldsymbol{\Sigma}_{m^*m^*} & \boldsymbol{\Sigma}_{m^*x^*} \\ \boldsymbol{\Sigma}_{x^*m^*} & \boldsymbol{\Sigma}_{x^*x^*} \end{pmatrix}.$$

Denote $\boldsymbol{\Sigma}_{x^*x^*.m^*} = \boldsymbol{\Sigma}_{x^*x^*} - \boldsymbol{\Sigma}_{x^*m^*}\boldsymbol{\Sigma}_{m^*m^*}^{-1}\boldsymbol{\Sigma}_{m^*x^*}$ and $\boldsymbol{\Sigma}_{m^*m^*.x^*} = \boldsymbol{\Sigma}_{m^*m^*} - \boldsymbol{\Sigma}_{m^*x^*}\boldsymbol{\Sigma}_{x^*x^*}^{-1}\boldsymbol{\Sigma}_{x^*m^*}$. By the formula for the inverse of a block matrix, it follows that

$$\boldsymbol{\Sigma}_{z^*z^*}^{-1} = \begin{pmatrix} \boldsymbol{\Sigma}_{m^*m^*.x^*}^{-1} & -\boldsymbol{\Sigma}_{m^*m^*.x^*}^{-1}\boldsymbol{\Sigma}_{m^*x^*}\boldsymbol{\Sigma}_{x^*x^*}^{-1} \\ -\boldsymbol{\Sigma}_{x^*x^*.m^*}^{-1}\boldsymbol{\Sigma}_{x^*m^*}\boldsymbol{\Sigma}_{m^*m^*}^{-1} & \boldsymbol{\Sigma}_{x^*x^*.m^*}^{-1} \end{pmatrix}.$$

Thus, $\sqrt{n}(\hat{\boldsymbol{\alpha}}_0^* - \tilde{\boldsymbol{\alpha}}_0^*)$ asymptotically follows a multivariate normal distribution with mean $\boldsymbol{0}$ and covariance matrix $\sigma_1^2\boldsymbol{\Sigma}_{m^*m^*.x^*}^{-1}$, and $\sqrt{n}(\hat{\boldsymbol{\alpha}}_1^* - \tilde{\boldsymbol{\alpha}}_1^*)$ asymptotically follows a multivariate normal distribution with mean $\boldsymbol{0}$ and covariance matrix $\sigma_1^2\boldsymbol{\Sigma}_{x^*x^*.m^*}^{-1}$. Therefore, we may obtain the asymptotical normality of $\hat{\boldsymbol{\alpha}}_0(u)$ and $\hat{\boldsymbol{\alpha}}_1(u)$ as follows:

$$\hat{\boldsymbol{\alpha}}_0(u) \overset{a}{\sim} N(\tilde{\boldsymbol{\alpha}}_0(u), \frac{1}{n}\sigma_1^2\{\boldsymbol{I}_p \otimes \boldsymbol{b}_0(u)^T\}\boldsymbol{\Sigma}_{m^*m^*.x^*}^{-1}\{\boldsymbol{I}_p \otimes \boldsymbol{b}_0(u)\}), \tag{13}$$

$$\hat{\boldsymbol{\alpha}}_1(u) \overset{a}{\sim} N(\tilde{\boldsymbol{\alpha}}_1(u), \frac{1}{n}\sigma_1^2\{\boldsymbol{I}_q \otimes \boldsymbol{b}_1(u)^T\}\boldsymbol{\Sigma}_{x^*x^*.m^*}^{-1}\{\boldsymbol{I}_q \otimes \boldsymbol{b}_1(u)\}). \tag{14}$$

Remark 3. Under Conditions C1–C6, it can be shown by using related techniques in the proof of Theorem 2 that for $\omega = 0, 1$, $\|\tilde{\alpha}_{\omega j} - \alpha_{\omega j}\|_{L_2}^2 = O_p(\rho_{n\omega}^2) = O_p(K_\omega^{-4})$. Furthermore, it can be shown that $\|\hat{\alpha}_{\omega j} - \tilde{\alpha}_{\omega j}\|_{L_2}^2 = O_p(1/n + K_\omega/n)$. Thus, $\|\hat{\alpha}_{\omega j} - \alpha_{\omega j}\|_{L_2}^2 = O_p(K_\omega/n + K_\omega^{-4})$. This provides us the rate of convergence of $\hat{\alpha}_{\omega j}$. Since $E\{\hat{\alpha}_{\omega j}\} = \tilde{\alpha}_{\omega j}$, the asymptotic bias of $\hat{\alpha}_{\omega j}$ is $\tilde{\alpha}_{\omega j} - \alpha_{\omega j}$, which is controlled by $\rho_{n\omega} = O(K_\omega^{-2})$. In practice, we take $K_\omega = O(n^{1/5})$. Thus, the asymptotic bias of $\hat{\alpha}_{\omega j}$ is of order $n^{-2/5}$ and from Equations (13) and (14), the asymptotic variance of $\hat{\alpha}_{\omega j}$ is of order $n^{-4/5}$.

We next study the asymptotic properties of $\hat{\boldsymbol{\beta}}(\cdot)$. Similar to the asymptotic normality of $\hat{\boldsymbol{\xi}}^*$, we establish the asymptotic normality of $\hat{C}$ and further establish the asymptotic normality of $\hat{\boldsymbol{\Gamma}}(u)$ in the appendix. This enables us to derive the rate of convergence of $\hat{\beta}_l(\cdot)$ in the following theorem, whose proof is given in the Appendix.

Theorem 2 (Convergence Rate). Under Conditions C1–C6 in the appendix, if $K_0 \to \infty$ and $K_m \to \infty$ as $n \to \infty$,

$$\left\|\hat{\beta}_l(\cdot) - \beta_l(\cdot)\right\|_{L_2}^2 = O_p\{1/n + K_0/n + K_m/n + \rho_{n0}^2 + \rho_{n2}^2\}.$$

Let $\tilde{C} = (X^{mT}X^m)^{-1}X^{mT}\tilde{M}$, where $\tilde{M} = (\tilde{\boldsymbol{m}}_1, \dots, \tilde{\boldsymbol{m}}_n)^T$ with $\tilde{\boldsymbol{m}}_i = \boldsymbol{\Gamma}(u_i)^T\boldsymbol{x}_i$ for $i = 1, \dots, n$. Define $\tilde{\boldsymbol{\Gamma}}(u_i) = \{\boldsymbol{I}_q \otimes \boldsymbol{d}(u_i)^T\}\tilde{C}$. We can obtain that $\mathrm{E}(\hat{C}) = \tilde{C}$ and $\mathrm{E}\{\hat{\boldsymbol{\Gamma}}(u_i)\} = \tilde{\boldsymbol{\Gamma}}(u_i)$ conditioning on $\mathcal{X}$. Define $\tilde{\boldsymbol{\beta}}(u) = \tilde{\boldsymbol{\Gamma}}(u)\tilde{\boldsymbol{\alpha}}_0(u)$. Then the difference between $\hat{\beta}_l(u)$ and $\beta_l(u)$ can be decomposed into two parts. That is, $\hat{\beta}_l(u) - \beta_l(u) = \{\hat{\beta}_l(u) - \tilde{\beta}_l(u)\} + \{\tilde{\beta}_l(u) - \beta_l(u)\}$, where $\hat{\beta}_l(u) - \tilde{\beta}_l(u)$ contributes to the variance of estimation and $\tilde{\beta}_l(u) - \beta_l(u)$ to the bias. We can show that $\left\|\tilde{\beta}_l - \beta_l\right\|_{L_2}^2 = O_p\{\rho_{n0}^2 + \rho_{n2}^2\}$, and $\left\|\hat{\beta}_l - \tilde{\beta}_l\right\|_{L_2}^2 = O_p\{1/n + K_0/n + K_m/n\}$. Furthermore, we can establish the asymptotic normality of $\hat{\boldsymbol{\beta}}(u)$ in the following theorem, whose proof is given in the Appendix.

Theorem 3. (a) (Asymptotic Normality). Under Conditions C1–C7 in the Appendix, it follows that

$$[\mathrm{cov}_a\{\hat{\boldsymbol{\beta}}(u)\}]^{-1/2}\{\hat{\boldsymbol{\beta}}(u) - \tilde{\boldsymbol{\beta}}(u)\} \xrightarrow{D} N(\boldsymbol{0}, \boldsymbol{I}),$$

where $\mathrm{cov}_a\{\hat{\boldsymbol{\beta}}(u)\} = \frac{1}{n}[\tilde{\boldsymbol{\alpha}}_0(u)^T\boldsymbol{\Sigma}_\epsilon\tilde{\boldsymbol{\alpha}}_0(u)\{\boldsymbol{I}_q \otimes \boldsymbol{d}(u)^T\}\boldsymbol{\Sigma}_{x^m x^m}^{-1}\{\boldsymbol{I}_q \otimes \boldsymbol{d}(u)\} + \sigma_1^2\tilde{\boldsymbol{\Gamma}}(u)\{\boldsymbol{I}_p \otimes \boldsymbol{b}_0(u)^T\}\boldsymbol{\Sigma}_{m*m*.x*}^{-1}\{\boldsymbol{I}_p \otimes \boldsymbol{b}_0(u)\}\tilde{\boldsymbol{\Gamma}}(u)^T]$ is the asymptotic covariance matrix.

(b) (Bias) Suppose Conditions C1–C6 hold, if $K_0 \to \infty$ and $K_m \to \infty$ as $n \to \infty$, $\sup_{u\in\mathcal{U}} |\tilde{\beta}_l(u) - \beta_l(u)| = O_p(\rho_{n0} + \rho_{n2})$.

The bias term in Theorem 3(b) follows naturally from (Huang, Wu, and Zhou 2004, theor. 4). In Huang, Wu, and Zhou (2004), covariates are assumed to be bounded. However, under model (1) and (2), the mediator $\boldsymbol{m}$ is not bounded. Fortunately, this assumption can be relaxed based on Condition C2. The bias term is negligible compared with the variance term.

3. Simulation Studies

In this section, we discuss the implementation of the VCMM estimation procedures and evaluate the performance of the proposed method via Monte Carlo simulation studies.

We generate the random sample $(\boldsymbol{x}_i, y_i, \boldsymbol{m}_i, u_i)$, $i = 1, \ldots n$ as follows. The dimension of $\boldsymbol{x}_i$ is set to be $q = 3$, and the first element of $\boldsymbol{x}_i$ is 1 to allow existence of the intercept term. The remaining elements of $\boldsymbol{x}_i$, denoted $\boldsymbol{x}_{-1,i}$, as well as u_i are generated from

$$(u_i^*, \boldsymbol{x}_{-1,i}) \sim N_3(\mathbf{0}, \boldsymbol{\Sigma}), \quad u_i = \Phi(u_i^*),$$

where the (k_1, k_2)-element of $\boldsymbol{\Sigma}$ is set to be $\rho^{|k_1-k_2|}$ with $\rho = 0.5$, and $\Phi(\cdot)$ is the cumulative distribution function of the standard normal distribution. Therefore, u_i is uniformly distributed on $[0, 1]$ and correlated with $\boldsymbol{x}_{-1,i}$.

Example 1. The coefficient functions $\boldsymbol{\alpha}_0(u)$, $\boldsymbol{\alpha}_1(u)$, and $\boldsymbol{\Gamma}(u)$ are taken to be

$$\boldsymbol{\alpha}_0(u) = (\alpha_{01}(u), \alpha_{02}(u))^T,$$
$$\boldsymbol{\alpha}_1(u) = (\alpha_{11}(u), \alpha_{12}(u), \alpha_{13}(u))^T,$$

and

$$\boldsymbol{\Gamma}(u) = \begin{pmatrix} \Gamma_{11}(u) & \Gamma_{12}(u) \\ \Gamma_{21}(u) & \Gamma_{22}(u) \\ \Gamma_{31}(u) & \Gamma_{32}(u) \end{pmatrix},$$

where

$$\alpha_{01}(u) = 0.5\cos(\frac{\pi}{2}u) + 0.5, \quad \alpha_{02}(u) = 0.5\sin(\frac{\pi}{2}u) + 0.5,$$
$$\alpha_{11}(u) = 0.5 + 2\xi(u), \quad \alpha_{12}(u) = 1 + 1.5u^2,$$
$$\alpha_{13}(u) = 1 + \sin^2(\pi u),$$
$$\Gamma_{11}(u) = 0.5u^3, \quad \Gamma_{12(u)} = 2\cos^2(\pi u - \frac{\pi}{2}) + 0.5,$$
$$\Gamma_{21}(u) = 2.5\sin^2(\pi u) + 0.5, \quad \Gamma_{22}(u) = u^2 - 0.5,$$
$$\Gamma_{31}(u) = 3(0.5 - u)^2 - 1.5, \quad \Gamma_{32}(u) = -u + 2u^3 - 0.5,$$

with $\xi(\cdot)$ being the cumulative distribution function of Gamma(0.5,1). Then $\boldsymbol{m}_i$ and y_i's are generated using model (1) and (2), where $\boldsymbol{\epsilon}_i \sim N(0, \boldsymbol{I}_2)$, with $\boldsymbol{I}_2$ being a 2×2 identity matrix, and $\epsilon_{1i} \sim N(0, 0.5)$.

To apply B-spline approximations to the varying coefficients, we set the degree of splines to be 3, and the number of interior knots to be 3. Moreover, since u is uniformly generated from $[0, 1]$, the knots are reasonably taken with equal spacing. For simplicity, we set the basis splines for approximating $\boldsymbol{\alpha}_0(u)$, $\boldsymbol{\alpha}_1(u)$ and $\boldsymbol{\Gamma}(u)$ to be the same, that is, $\boldsymbol{b}_0(u_i) = \boldsymbol{b}_1(u_i) = \boldsymbol{d}(u_i)$, $i = 1, \ldots, n$. In practice, one may choose different B-spline basis functions for different coefficients based on prior knowledge or professional experience. For instance, we may choose different locations of knots for the B-spline bases, and knots are placed at the locations at which curvatures of coefficients are changing. Furthermore, we may choose different degrees of B-spline for different coefficients when we see the potential different smoothness of coefficients.

In our simulation, we set $n = 100$, 200, and 400, and conduct 1000 replications for each case. The estimate $\hat{\boldsymbol{\alpha}}_1(u)$ and $\hat{\boldsymbol{\beta}}(u) = \hat{\boldsymbol{\Gamma}}(u)\hat{\boldsymbol{\alpha}}_0(u)$ are evaluated at a set of grid points $\{u_k, k = 1, \ldots, n_{\text{grid}}\}$ evenly distributing over [0,1] with $n_{\text{grid}} = 500$ points. Figures 1 and 2 depict the estimated coefficient functions with different sample sizes. In these figures, the means of the estimated coefficient functions are plotted as green dashed curves, while the true curves are in solid black. Red dashed lines are the 95% point-wise confidence bands. It can be seen from Figures 1 and 2 that the proposed estimation procedure performs well. The means of estimated functions and their true curves are so close that the bias indeed is invisible. The true curves fall within the corresponding 95% point-wise confidence bands. As expected, the confidence bands become narrower as n increases.

We next examine the accuracy of the proposed standard error estimation procedure by comparing the estimated standard error with true standard deviation of the proposed estimated coefficient functions at three representative points $u = 1/4$, $1/2$, and $3/4$. Tables 1 and 2 summarize the estimation results for the direct effect estimates $\hat{\alpha}_{12}(u)$, $\hat{\alpha}_{13}(u)$ and the indirect effect estimates $\hat{\boldsymbol{\beta}}(u)$. In Tables 1 and 2, *Mean(sd)* denotes the mean and standard deviation of $\hat{\alpha}_{12}(u), \hat{\alpha}_{13}(u), \hat{\beta}_2(u)$, and $\hat{\beta}_3(u)$ over the 1000 replications. The standard deviations of estimates over the 1000 replications can be viewed as the true deviations of these estimates. *SE(sd)* in the second column describes the average and standard deviation of the 1000 estimated standard errors calculated from the variance formula, by directly plugging in the estimated $\tilde{\boldsymbol{\alpha}}_0(u)$ and $\tilde{\boldsymbol{\Gamma}}(u)$ with B-spline basis functions at the grid points. In our simulation, we estimate σ_1^2 and $\boldsymbol{\Sigma}$ by $\hat{\sigma}_1^2 = ||Y - M^*\hat{\boldsymbol{\alpha}}_0^{*T} - X^*\hat{\boldsymbol{\alpha}}_1^{*T}||^2/(n - pK_0 - qK_1)$ and $\hat{\boldsymbol{\Sigma}}_\epsilon = \frac{1}{n-q}\sum_{i=1}^n \hat{\boldsymbol{\epsilon}}_i \hat{\boldsymbol{\epsilon}}_i^T$, respectively, where $\hat{\boldsymbol{\epsilon}}_i = \boldsymbol{m}_i - \hat{\boldsymbol{\Gamma}}(u_i)^T \boldsymbol{x}_i$. As can be seen in Tables 1 and 2, the standard error estimates, calculated based on Theorem 3, are very close to the standard deviations for the 1000 replications. Indeed, all differences between the average estimated standard errors and the standard deviations in Tables 1 and 2 are less than one standard deviation of the 1000 standard errors. This implies that the standard error estimates based on Theorem 3 provide us an accurate estimation of the standard error. *Coverage probability (CP)* is the proportion that the true values are covered by the corresponding 95% point-wise confidence intervals over 1000 simulations. The Monte Carlo errors for 1000 simulations is 1.35% for confidence level 95%. Most CP values in Tables 1 and 2 with $n = 200$ and $n = 400$ fall in $95\% \pm 1.35\%$. This implies that the point-wise confidence intervals based on Theorem 3 are valid. The CP values in Table 2 with $n = 100$ are slightly less 95%−1.35%=93.65%, and the phenomenon disappears when n increases to 200 and 400. This may imply that the estimation of indirect effect needs more sample to achieve desired accuracy.

We next evaluate the testing procedure for the direct effects. We keep the setting of $\boldsymbol{\alpha}_0(u)$ and $\boldsymbol{\Gamma}(u)$ unchanged but change $\boldsymbol{\alpha}_1(u)$ to be

$$\boldsymbol{\alpha}_1(u) = \{0.5 + 2\xi(u), w(1 + 1.5u^2), w(1 + \sin^2(\pi u)\},$$

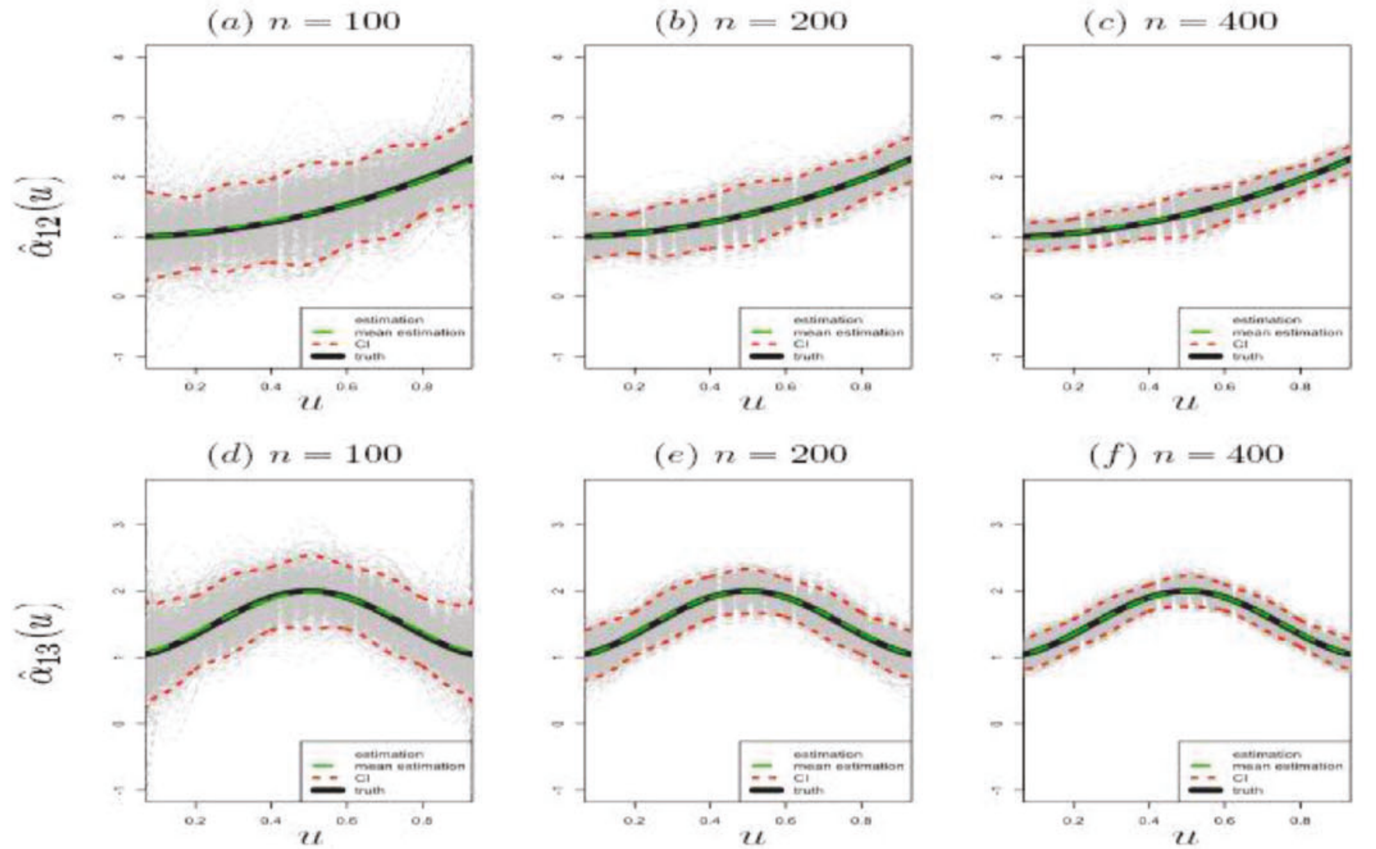

Figure 1. The mean of the estimated direct effect coefficient functions. The green and black curves are the estimated values and true values, respectively. The red dashed curves are the estimated function plus/minus 1.96 times the standard errors from 1000 replications.

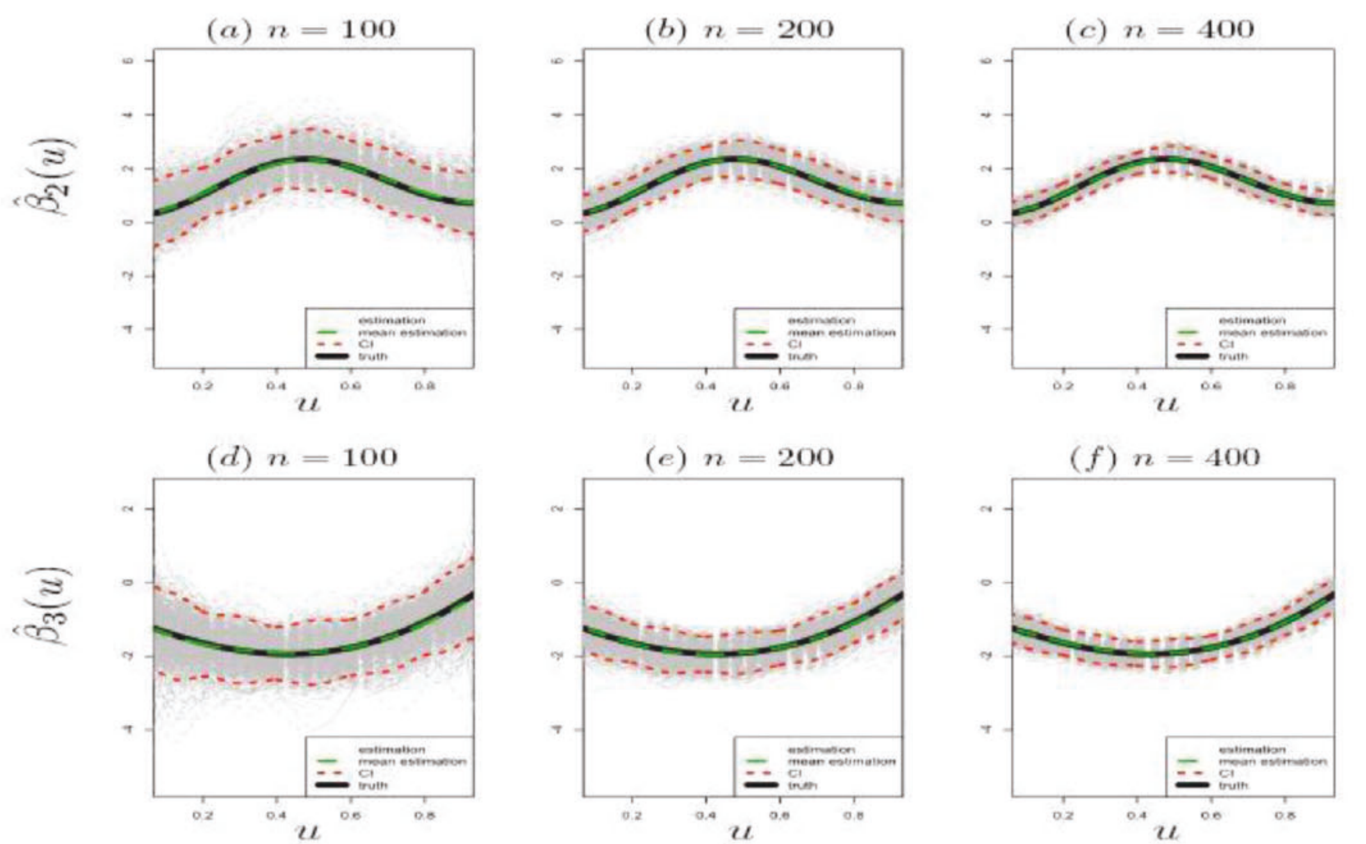

Figure 2. The Mean of the estimated indirect effect coefficient functions $\hat{\beta}_2(u)$ and $\hat{\beta}_3(u)$. Caption is the same as that of Figure 1.

Table 1. Estimation of direct effects.

		$\hat{\alpha}_{12}(u)$			$\hat{\alpha}_{13}(u)$		
n	u	Mean(sd)	SE(sd)	CP	Mean(sd)	SE(sd)	CP
100	1/4	1.113(0.353)	0.335(0.083)	0.942	1.517(0.292)	0.272(0.060)	0.937
	1/2	1.384(0.430)	0.417(0.085)	0.940	1.984(0.282)	0.275(0.060)	0.947
	3/4	1.827(0.363)	0.336(0.082)	0.931	1.531(0.280)	0.264(0.058)	0.938
200	1/4	1.087(0.213)	0.209(0.033)	0.941	1.516(0.185)	0.170(0.024)	0.929
	1/2	1.374(0.265)	0.270(0.036)	0.950	1.995(0.170)	0.177(0.025)	0.956
	3/4	1.842(0.212)	0.209(0.033)	0.948	1.515(0.179)	0.164(0.023)	0.921
400	1/4	1.107(0.139)	0.139(0.015)	0.959	1.500(0.126)	0.114(0.011)	0.922
	1/2	1.376(0.186)	0.183(0.017)	0.951	2.000(0.129)	0.120(0.012)	0.956
	3/4	1.842(0.139)	0.139(0.014)	0.951	1.509(0.115)	0.110(0.010)	0.945

Table 2. Estimation of Indirect Effects.

		$\hat{\beta}_2(u)$			$\hat{\beta}_3(u)$		
n	u	mean(sd)	SE(sd)	CP	mean(sd)	SE(sd)	CP
100	1/4	1.394(0.515)	0.463(0.100)	0.917	−1.766(0.446)	0.408(0.084)	0.919
	1/2	2.335(0.584)	0.554(0.107)	0.946	−1.910(0.445)	0.412(0.084)	0.929
	3/4	1.311(0.551)	0.467(0.100)	0.916	−1.321(0.452)	0.405(0.086)	0.922
200	1/4	1.390(0.345)	0.306(0.042)	0.929	−1.774(0.288)	0.270(0.037)	0.932
	1/2	2.347(0.362)	0.369(0.047)	0.958	−1.919(0.286)	0.278(0.037)	0.946
	3/4	1.306(0.332)	0.308(0.044)	0.936	−1.317(0.293)	0.268(0.037)	0.926
400	1/4	1.371(0.227)	0.210(0.020)	0.939	−1.744(0.188)	0.186(0.017)	0.946
	1/2	2.349(0.254)	0.254(0.023)	0.953	−1.920(0.196)	0.193(0.018)	0.951
	3/4	1.280(0.222)	0.209(0.020)	0.930	−1.303(0.191)	0.184(0.017)	0.948

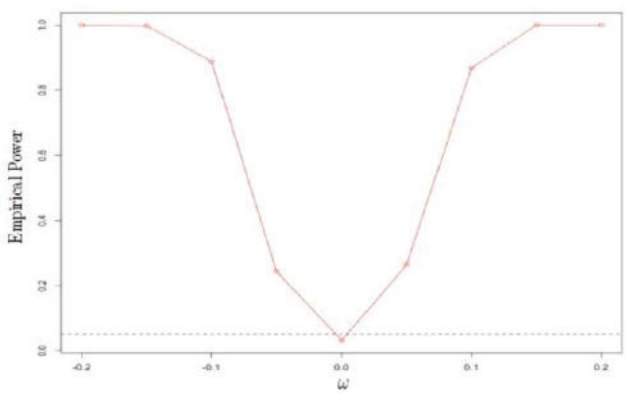

Figure 3. Empirical size and power for the direct effect test at the significance level $\alpha = 0.05$.

for a sequence of $w \in \{-0.2, -0.15, -0.1, -0.05, 0, 0.05, 0.1, 0.15, 0.2\}$. The value $w = 0$ corresponds to the null hypothesis, so that we can examine how the proposed test retains Type I error rate. Figure 3 depicts the empirical size and power at the significance level $\alpha = 0.05$ over 500 replications for testing the direct effect. As can be seen in the figure, the proposed test T_n retains Type I error well, and the power increases to 1 quickly when $|w|$ increases from 0 to 0.15.

4. Real Data Example

We illustrate in this section our proposed procedure via an application to the behavioral economics dataset introduced previously. According to the learned helplessness theory (Hiroto and Seligman 1975; Pryce et al. 2011), entrepreneurs under economic stress are compelled to change their standard behavioral patterns and more likely to develop a sense of depression, isolation, and helplessness. These feelings, in turn, lead to greater job-related withdrawal intentions.

Pollack, Vanepps, and Hayes (2012) conducted a linear mediation analysis and suggested that social ties, a variable that measures the number of daily contacts an entrepreneur has with members in his or her own networking group about work-related matters, buffers the impact of economic burden on depressed affect, and reduces the indirect effect on entrepreneurs' withdrawal intentions. However, the moderating role of social ties was only analyzed based on interaction terms in this article. Since the social ties variable is highly skewed in the collected dataset, the interaction effect of the social ties and main effect cannot be captured well and generalized for different levels of social ties. Moreover, other factors such as tenure can also play a crucial role in moderating the withdrawal intentions of small business owners in response to economic stress. Individuals with shorter tenures might have tremendous enthusiasm but fewer back-up options, so they are more likely to insist on business even under stress, while entrepreneurs with longer tenures are likely to have a sound economic basis but possibly tired of the business. This hypothesis motivates us to examine the buffering/promoting effect of entrepreneurs' tenures in the aforementioned mediation relationship. Rather than conducting standard moderation analysis that only involves the interaction terms, we explore the indirect and direct effects of economic stress as flexible functions of tenure so that effects at different tenures can be thoroughly examined.

The data are available in Pollack, Vanepps, and Hayes (2012). Small business owners ($n = 300$) were recruited into the study and a total of 262 participants (162 male, 100 female) provided complete data. Economic stress was measured on a scale ranging from 1 to 7 in the last year, where a larger number indicates higher economic stress. Depressed affect was measured on a scale from 1 to 5, where larger values reflect more depressed affect. Entrepreneurs' intentions to withdraw from entrepreneurship over the next year was measured on a scale from 1 to 7, with higher scores indicative of greater intentions. Other covariates include entrepreneurial self-efficacy, social competence, age, tenure (how long a person has worked in the company), social ties, and gender (1 = male, 0 = female).

In this analysis, we take the economic stress variable to be the exposure variable x, the depressed affect variable to be the mediator m, with $p = 1$, and the withdrawal intention to be the response y. Since tenure is also skewed to the right and some values are close to 0, u is defined as $\log(1 + \text{tenure})$. In this analysis, we also include an intercept term (z_1) and covariates in both models (2) and (3). The covariates consist of entrepreneurial self-efficacy (z_2), social competence (z_3), age (z_4), the log-transformed social ties variable (z_5) since the social ties variable itself is very skewed, and gender (z_6). This leads to

$$y = \alpha_0(u)m + \alpha_1(u)x + \sum_{j=1}^{6} \alpha_{j+1}(u)z_j + \epsilon_1, \qquad (15)$$

$$m = \gamma_1(u)x + \sum_{j=1}^{6} \gamma_{j+1}(u)z_j + \epsilon. \qquad (16)$$

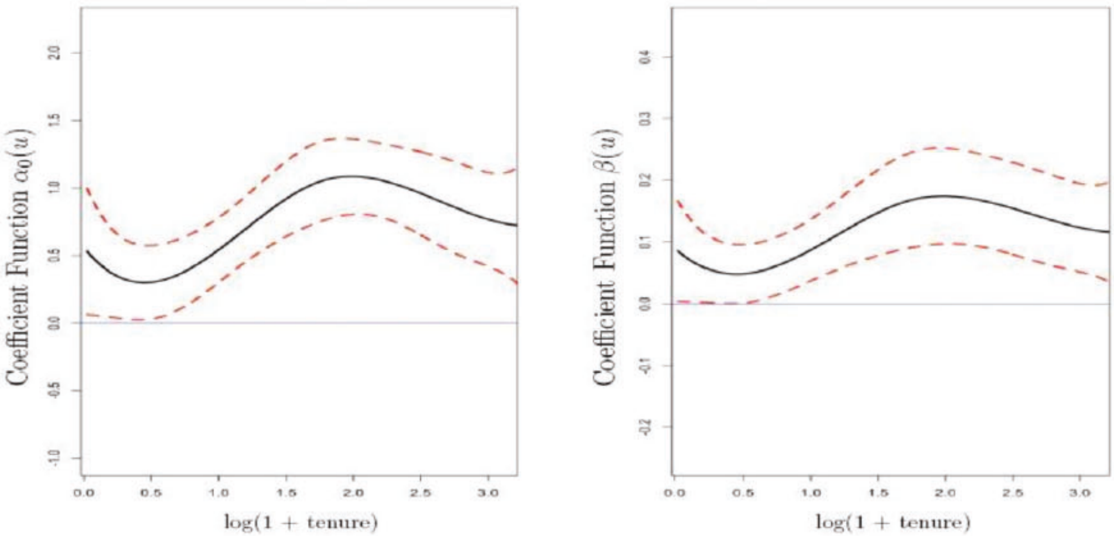

Figure 4. The estimated coefficient functions for $\alpha_0(u)$ and the indirect effect $\beta(u)$ for Economic Stress. The dashed curves are the estimated functions plus/minus 1.96 times the estimated standard errors.

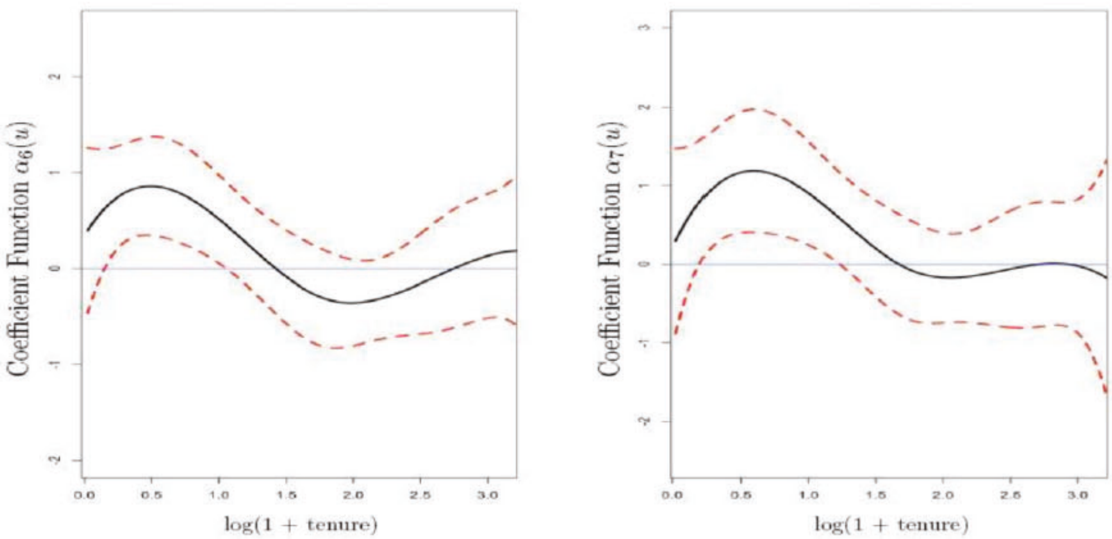

Figure 5. The estimated coefficient functions for $\alpha_6(u)$ associated with gender and $\alpha_7(u)$ associated with social ties. The dashed curves are the estimated functions plus/minus 1.96 times the estimated standard errors.

We rewrite models (15) and (16) as the form of models (2) and (3). Thus, the proposed procedures in Section 2 can be directly applied for both models (15) and (16). We use cubic splines with the number of knots chosen to be 2 using 10-fold cross-validation. We first test the significance of the varying pattern based on the procedures similar to Cai, Fan, and Li (2000). The p-value for the null hypothesis that γ_j's are constant is 0.41, indicating a favor of constant effects. Furthermore, given varying $\alpha_0(\cdot)$, $\alpha_6(\cdot)$, and $\alpha_7(\cdot)$, the p-value for the null hypothesis that $\alpha_k(\cdot)$ is constant for all $k = 1, \ldots, 5$, is 0.47. Thus, we set α_k's, $l = 1, \ldots, 5$, and all γ_j's to be constant in our analysis, and allow the other coefficients to be varying over u. That is, only effects of depressed affect, social ties, and gender on the response are varying with u. Therefore, we consider the following model for subsequent analysis.

$$y_i = \alpha_0(u_i)m_i + \alpha_1 x_i + \alpha_2 z_{i1} + \alpha_3 z_{i2} + \alpha_4 z_{i3} + \alpha_5 z_{i4} + \alpha_6(u_i) z_{i5} + \alpha_7(u_i) z_{i6} + \epsilon_{1i},$$

$$m_i = \gamma_1 x_i + \sum_{j=1}^{6} \gamma_{j+1} z_{ij} + \epsilon_i.$$

The left panel of Figure 4 illustrates the effect of depression on withdrawal intentions, which is positive and indeed changes over tenure. The two dashed lines are the 95% pointwise confidence band. As can be seen, the effect of increased depressed affect on withdrawal intentions becomes stronger with longer tenure ($\geq$ 2 years). This probably is because young businessmen are usually full of energy and vigor, and less likely to withdraw from entrepreneurship in response to depression, while individuals with longer tenure might be more aware of the difficulty they meet with, and have back-up options. The estimated coefficient γ_1 associated with economic stress is 0.16, and significant at the $\alpha = 0.05$ level. The indirect effect $\beta(\cdot) = \gamma_1 \alpha_0(\cdot)$ of economic stress is still positive and shares a pattern similar to that of $\alpha_0(\cdot)$, as indicated by the right panel of Figure 4.

The direct effect α_1 of economic stress on withdrawal intentions is not significant.

The coefficient functions associated with gender and social ties are presented in Figure 5. Interestingly, compared with females, males are more likely to withdraw when tenure is short ($\leq$ 3 years), implied by the positive estimates of $\alpha_6(u)$ for short tenure. When tenure is longer, there are no significant differences between males and females. This might be explained by more alternative working options for males. Finally, for entrepreneurs with shorter tenure, an increase in the work-related social ties triggers greater withdrawal intentions, while the effect of work-related social ties on withdrawal intentions is not statistically significant for entrepreneurs with longer tenure.

For the purpose of comparison, we also fit the data with constant-coefficient linear mediation model. The estimated effect of depressed affect on withdrawal intentions α_0 and the coefficient γ_1 are 0.74 and 0.16, respectively. The indirect effect of economic stress $\beta = \gamma_1\alpha_0$ is then estimated to be 0.12, with a 95% confidence interval $(0.07, 0.17)$. However, the dynamic patterns of α_0 and β over u cannot be captured. All coefficients are considered as constants and the promoting effect of tenure is ignored. The direct effect of economic stress on withdrawal intentions is not significant, in accordance with the VCMM. In addition, the coefficients associated with gender and social ties in the constant linear mediation model are not significant at the level $\alpha = 0.05$, while the VCMM illustrates that both of them have positive effect when tenure falls in a certain interval.

5. Conclusion

In this article we propose the VCMM, which is distinguished from the linear mediation model in that all the direct and indirect effects are varying with a variable of particular interest. The model is flexible in the sense that both varying and invariant coefficients can be included in the same model. We propose an estimation procedure based on the polynomial spline based method (Huang, Wu, and Zhou 2002, 2004), and establish asymptotic normality of the resulting estimates. A F-type test are proposed for testing of the direct effects. Simulation studies are conducted to verify them empirically. Finally, in the real data analysis, we evaluate the effect of economic stress on small business owners' intentions to withdraw from entrepreneurship via depressed effect, and illustrate that the indirect effect indeed varies as a function of tenure.

Acknowledgments

The authors would like to thank the AE and reviewers for their constructive comments, which lead to a significant improvement of this work. All authors equally contribute to the article, and the authors are listed according to seniority. Jingyuan Liu is corresponding author.

Funding

Liao and Li's research was supported by National Science Foundation (NSF) grants DMS 1820702, DMS 1953196, and DMS 2015539, and National Institutes of Health (NIH) grant R01CA229542 funded by the National Cancer Institute (NCI). Liu's research was supported by National Natural Science Foundation of China 11771361 and 11871409. Coffman's research was supported by the NIH grant R01CA229542. The content is solely the responsibility of the authors and does not necessarily represent the official views of NSF, NCI or NIH.

Appendix. Lemmas and Technical Proofs

Suppose that given sequences of positive numbers a_n and b_n, $a_n \lesssim b_n$ and $b_n \gtrsim a_n$ mean a_n/b_n is bounded, and $a_n \asymp b_n$ means both $a_n \lesssim b_n$ and $a_n \gtrsim b_n$ hold.

A.1. Technical Conditions

The following technical conditions along with some notations are imposed to facilitate the technical proofs, although they may not be the weakest ones.

Technical Conditions:

C1. The points $\{u_i, i = 1, \ldots, n\}$ are independently distributed with distribution F_U on a bounded and compact support $\mathcal{U}$ and Lebesgue density $f_U(u)$ which is bounded away from 0 and infinity uniformly over $u \in \mathcal{U}$.

C2. For some constant $\delta > 0$, all elements of $\mathbf{x}$ and $\mathbf{m}$ have finite $(2+\delta)$th moments. Specifically, there are constants N_1 and N_2 such that $\mathrm{E}(|x_l|^{2+\delta}) \leq N_1$ for $l = 1, \ldots, q$, and $\mathrm{E}(|m_j|^{2+\delta}) \leq N_2$ for $j = 1, \ldots, q$.

C3. Both ϵ_1 and $\boldsymbol{\epsilon}$ have finite $(2+\delta)$th moments. In other words, there exist constants N_3 and N_3' such that $\mathrm{E}(|\epsilon_1|^{2+\delta}) \leq N_3 < \infty$ and $\mathrm{E}(\|\boldsymbol{\epsilon}\|^{2+\delta}) < N_3' < \infty$. All eigenvalues of $\boldsymbol{\Sigma}_\epsilon$ are bounded away from 0 and infinity.

C4. The second-order derivatives of coefficient functions $\boldsymbol{\Gamma}(u)$, $\boldsymbol{\alpha}_0(u)$ and $\boldsymbol{\alpha}_1(u)$ are assumed to be continuous over $\mathcal{U}$. Thus, they and their second-order derivatives are bounded. Denote M_Γ, M_0 and M_1 to be the bounds of $\boldsymbol{\Gamma}(u)$, $\boldsymbol{\alpha}_0(u)$ and $\boldsymbol{\alpha}_1(u)$ over $u \in \mathcal{U}$, respectively.

C5. $\limsup_{n\to\infty}\left(\frac{\max_\omega K_\omega}{\min_\omega K_\omega}\right) < \infty, \omega = 0, 1, m.$

C6. Let $K_{01} = \max\{K_0, K_1\}$. Assume $\lim_{n\to\infty} n^{-\delta} K_{01}^{(2+\delta)(\nu+1)} = 0$ for some $\nu \in (0, 1.5]$, and, $\lim_{n\to\infty} n^{-\delta} K_m^{(2+\delta)(\nu_2+1)} = 0$ for some $\nu_2 \in (0, 1.5]$.

C7. Let $\boldsymbol{\Sigma}_{z^*z^*} = \mathrm{E}(z_i^* z_i^{*T})$ and $\hat{\boldsymbol{\Sigma}}_{z^*z^*} = \frac{1}{n}\sum_{i=1}^n z_i^* z_i^{*T}$. Assume $\lambda_{\max}(\boldsymbol{\Sigma}_{z^*z^*}^{-1}) = O_p(K_{01}^{\nu})$, $\lambda_{\max}(\boldsymbol{\Sigma}_{z^*z^*}^{-1/2}\hat{\boldsymbol{\Sigma}}_{z^*z^*}\boldsymbol{\Sigma}_{z^*z^*}^{-1/2}) < N_4$ for some constant N_4 when n is large, and $\lambda_{\max}\{\boldsymbol{\Sigma}_{z^*z^*}^{-1/2} \cdot (\boldsymbol{\Sigma}_{z^*z^*}^{1/2}\hat{\boldsymbol{\Sigma}}_{z^*z^*}^{-1}\boldsymbol{\Sigma}_{z^*z^*}^{1/2} - I)^2\boldsymbol{\Sigma}_{z^*z^*}^{-1/2}\} \to 0$ as $n \to \infty$, where $\lambda_{\max}(\cdot)$ denotes the largest eigenvalue.

Similarly, let $\boldsymbol{\Sigma}_{x^m x^m} = \mathrm{E}(\mathbf{x}_i^m \mathbf{x}_i^{mT})$ and $\hat{\boldsymbol{\Sigma}}_{x^m x^m} = \frac{1}{n}\sum_{i=1}^n \mathbf{x}_i^m \mathbf{x}_i^{mT}$. Then $\lambda_{\max}(\boldsymbol{\Sigma}_{x^m x^m}^{-1}) = O_p(K_m^{\nu_2})$, $\lambda_{\max}(\boldsymbol{\Sigma}_{x^m x^m}^{-1/2}\hat{\boldsymbol{\Sigma}}_{x^m x^m}\boldsymbol{\Sigma}_{x^m x^m}^{-1/2}) < N_4'$ for some constant N_4' when n is large, and $\lambda_{\max}\{\boldsymbol{\Sigma}_{x^m x^m}^{-1/2}(\boldsymbol{\Sigma}_{x^m x^m}^{1/2}\hat{\boldsymbol{\Sigma}}_{x^m x^m}^{-1}\boldsymbol{\Sigma}_{x^m x^m}^{1/2} - I)^2\boldsymbol{\Sigma}_{x^m x^m}^{-1/2}\} \to 0$ as $n \to \infty$.

Condition C1 guarantees that the observations u_i, $i = 1, \ldots, n$, are randomly scattered. Such a condition can be weakened when $\{u_i\}$ are deterministic under other mild conditions (Huang, Wu, and Zhou 2004). Conditions C2 and C3 are implied by non-singularity of the covariance matrices. Condition C4 is needed to derive the convergence rate of $\hat{\boldsymbol{\Gamma}}(u)$, $\hat{\boldsymbol{\alpha}}_0(u)$, and $\hat{\boldsymbol{\beta}}(u)$. Conditions C5 and C6 are adapted from Huang, Wu, and Zhou (2004). And Conditions C5–C7 are only purely technical and serve to facilitate theoretical proofs of the proposed estimation procedure. These are some mild conditions that can be satisfied in many practical scenarios.

A.2. Proof of Theorem 1

Since $\hat{\xi}^* = (\hat{\alpha}_0^{*T}, \hat{\alpha}_1^{*T})^T = (Z^{*T}Z^*)^{-1}Z^{*T}Y$ and $\tilde{\xi}^* = (\tilde{\alpha}_0^{*T}, \tilde{\alpha}_1^{*T})^T = \{Z^{*T}Z^*\}^{-1}Z^{*T}\tilde{Y}$, it follows by the definition of $\tilde{Y}$ that $\hat{\xi}^* - \tilde{\xi}^* = (Z^{*T}Z^*)^{-1}Z^{*T}E_1$. Let $K = pK_0 + qK_1$. For any nonzero constant vector $g \in \mathbb{R}^{K\times 1}$, we want to establish the asymptotic normality of

$$\begin{aligned}\sqrt{n}g^T(\hat{\xi}^* - \tilde{\xi}^*) &= \sqrt{n}g^T(Z^{*T}Z^*)^{-1}Z^{*T}E_1\\ &= \frac{1}{\sqrt{n}}g^T(\frac{1}{n}Z^{*T}Z^*)^{-1}Z^{*T}E_1\\ &= \frac{1}{\sqrt{n}}g^T\mathbf{\Sigma}_{z^*z^*}^{-1}Z^{*T}E_1 + \frac{1}{\sqrt{n}}g^T(\hat{\mathbf{\Sigma}}_{z^*z^*}^{-1} - \mathbf{\Sigma}_{z^*z^*}^{-1})Z^{*T}E_1.\end{aligned}$$

Without loss of generality, we assume $\|g\| = 1$. We consider the term $g^T\mathbf{\Sigma}_{z^*z^*}^{-1}Z^{*T}E_1$ at first by checking Lyapounov condition, where $g^T\mathbf{\Sigma}_{z^*z^*}^{-1}Z^{*T}E_1 = \sum_{i=1}^n g^T\mathbf{\Sigma}_{z^*z^*}^{-1}z_i^*\epsilon_{1i}$.

Let $S_n^2 = \text{var}(\sum_{i=1}^n g^T\mathbf{\Sigma}_{z^*z^*}^{-1}z_i^*\epsilon_{1i}) = n\sigma_1^2 g^T\mathbf{\Sigma}_{z^*z^*}^{-1}g$. Thus, for δ in Conditions C2 and C3,

$$\begin{aligned}&\frac{1}{S_n^{2+\delta}}\sum_{i=1}^n \text{E}(|g^T\mathbf{\Sigma}_{z^*z^*}^{-1}z_i^*\epsilon_{1i}|^{2+\delta})\\ &= \frac{1}{S_n^{2+\delta}}\sum_{i=1}^n\left\{\text{E}\left(\left|\tilde{g}^Tz_i^*\right|^{2+\delta}\right)\cdot \text{E}(|\epsilon_{1i}|^{2+\delta})\right\}\\ &\quad\text{by independence and } \tilde{g} \stackrel{\Delta}{=} \mathbf{\Sigma}_{z^*z^*}^{-1}g\\ &= \frac{1}{S_n^{2+\delta}}\sum_{i=1}^n\left\{\text{E}\left(\left|\tilde{g}^T\mathbf{\Sigma}_{z^*z^*}^{1/2}\mathbf{\Sigma}_{z^*z^*}^{-1/2}z_i^*\right|^{2+\delta}\right)\cdot N_3\right\}\\ &\quad\text{by condition } C3\\ &\le \frac{1}{S_n^{2+\delta}}\sum_{i=1}^n\left\{\text{E}\left(\left\|\tilde{g}^T\mathbf{\Sigma}_{z^*z^*}^{1/2}\right\|^{2+\delta}\cdot\left\|\mathbf{\Sigma}_{z^*z^*}^{-1/2}z_i^*\right\|^{2+\delta}\right)\cdot N_3\right\}\\ &\quad\text{by Cauchy–Swartz Inequality}\\ &= \frac{1}{S_n^{2+\delta}}\sum_{i=1}^n\left[\left\|g^T\mathbf{\Sigma}_{z^*z^*}^{-1}g\right\|^{\frac{2+\delta}{2}}\cdot \text{E}\left\{\left(z_i^{*T}\mathbf{\Sigma}_{z^*z^*}^{-1}z_i^*\right)^{\frac{2+\delta}{2}}\right\}\cdot N_3\right]\end{aligned}$$

Note that

$$\text{E}\left\{\left(z_i^{*T}\mathbf{\Sigma}_{z^*z^*}^{-1}z_i^*\right)^{\frac{2+\delta}{2}}\right\} \le \lambda_{\max}(\mathbf{\Sigma}_{z^*z^*}^{-1})^{\frac{2+\delta}{2}}\cdot \text{E}\{(z_i^{*T}z_i^*)^{\frac{2+\delta}{2}}\}.$$

Since $\text{E}\{(z_i^{*T}z_i^*)^{\frac{2+\delta}{2}}\} = \text{E}\{(\sum_{i=1}^n |z_{ik}^*|^2)^{\frac{2+\delta}{2}}\} \le K^{(2+\delta)/2}\frac{1}{K}\sum_{k=1}^K \text{E}(|z_{ik}^*|^{2+\delta}) \lesssim K^{(2+\delta)/2} \lesssim K_{01}^{(2+\delta)/2}$ by Jensen's Inequality and Condition C2, it follows by Condition C7 that

$$\begin{aligned}\text{E}\left\{\left(z_i^{*T}\mathbf{\Sigma}_{z^*z^*}^{-1}z_i^*\right)^{\frac{2+\delta}{2}}\right\} &\lesssim K_{01}^{\nu(2+\delta)/2}\cdot \text{E}\{(z_i^{*T}z_i^*)^{\frac{2+\delta}{2}}\}\\ &\lesssim K_{01}^{(\nu+1)(2+\delta)/2}.\end{aligned}$$

Thus, it follows that

$$\begin{aligned}\frac{1}{S_n^{2+\delta}}\sum_{i=1}^n \text{E}(|g^T\mathbf{\Sigma}_{z^*z^*}^{-1}z_i^*\epsilon_{1i}|^{2+\delta}) &\lesssim \frac{n}{n^{\frac{2+\delta}{2}}}\cdot K_{01}^{(\nu+1)(2+\delta)/2}\\ &= \left(n^{-\delta}K_{01}^{(2+\delta)(\nu+1)}\right)^{1/2},\end{aligned}$$

which tends to 0 as $n\to\infty$ by Condition C6.

Thus,

$$(n\sigma_1^2g^T\mathbf{\Sigma}_{z^*z^*}^{-1}g)^{-1/2}g^T\mathbf{\Sigma}_{z^*z^*}^{-1}Z^{*T}E_1 \xrightarrow{D} N(0,1).$$

We next deal with the second term $R_n = \frac{1}{\sqrt{n}}g^T(\hat{\mathbf{\Sigma}}_{z^*z^*}^{-1} - \mathbf{\Sigma}_{z^*z^*}^{-1})Z^{*T}E_1$. Since $\text{E}(R_n|Z^*) = 0$, $\text{E}(R_n) = 0$. Thus, $R_n = o_p(1)$ and is negligible if we can show that

$$\begin{aligned}\text{var}(R_n) &= \text{E}(R_n^2)\\ &= \sigma_1^2\text{E}\{g^T(\hat{\mathbf{\Sigma}}_{z^*z^*}^{-1} - \mathbf{\Sigma}_{z^*z^*}^{-1})\hat{\mathbf{\Sigma}}_{z^*z^*}(\hat{\mathbf{\Sigma}}_{z^*z^*}^{-1} - \mathbf{\Sigma}_{z^*z^*}^{-1})g\}\end{aligned}$$

tends to 0 as $n\to\infty$.

Note that

$$\begin{aligned}&g^T(\hat{\mathbf{\Sigma}}_{z^*z^*}^{-1} - \mathbf{\Sigma}_{z^*z^*}^{-1})\hat{\mathbf{\Sigma}}_{z^*z^*}(\hat{\mathbf{\Sigma}}_{z^*z^*}^{-1} - \mathbf{\Sigma}_{z^*z^*}^{-1})g\\ &= g^T\mathbf{\Sigma}_{z^*z^*}^{-1/2}(\mathbf{\Sigma}_{z^*z^*}^{1/2}\hat{\mathbf{\Sigma}}_{z^*z^*}^{-1}\mathbf{\Sigma}_{z^*z^*}^{1/2} - I)\mathbf{\Sigma}_{z^*z^*}^{-1/2}\cdot\hat{\mathbf{\Sigma}}_{z^*z^*}\\ &\quad\cdot\mathbf{\Sigma}_{z^*z^*}^{-1/2}(\mathbf{\Sigma}_{z^*z^*}^{1/2}\hat{\mathbf{\Sigma}}_{z^*z^*}^{-1}\mathbf{\Sigma}_{z^*z^*}^{1/2} - I)\mathbf{\Sigma}_{z^*z^*}^{-1/2}g\\ &\le \lambda_{\max}(\mathbf{\Sigma}_{z^*z^*}^{-1/2}\hat{\mathbf{\Sigma}}_{z^*z^*}\mathbf{\Sigma}_{z^*z^*}^{-1/2})\\ &\quad\cdot g^T\mathbf{\Sigma}_{z^*z^*}^{-1/2}(\mathbf{\Sigma}_{z^*z^*}^{1/2}\hat{\mathbf{\Sigma}}_{z^*z^*}^{-1}\mathbf{\Sigma}_{z^*z^*}^{1/2} - I)^2\mathbf{\Sigma}_{z^*z^*}^{-1/2}g\\ &\le \lambda_{\max}(\mathbf{\Sigma}_{z^*z^*}^{-1/2}\hat{\mathbf{\Sigma}}_{z^*z^*}\mathbf{\Sigma}_{z^*z^*}^{-1/2})\lambda_{\max}\{\mathbf{\Sigma}_{z^*z^*}^{-1/2}\\ &\quad(\mathbf{\Sigma}_{z^*z^*}^{1/2}\hat{\mathbf{\Sigma}}_{z^*z^*}^{-1}\mathbf{\Sigma}_{z^*z^*}^{1/2} - I)^2\mathbf{\Sigma}_{z^*z^*}^{-1/2}\}\cdot\|g\|^2,\end{aligned}$$

which tends to 0 by Condition C7. Thus, $\text{var}(R_n)$ tends to 0 as $n\to\infty$ and R_n is negligible.

Therefore, for any $K\times 1$ nonzero constant vector g with norm 1,

$$\{\text{Cov}_a(g^T\hat{\xi}^*)\}^{-1/2}g^T(\hat{\xi}^* - \tilde{\xi}^*)\xrightarrow{D}N(0,1),$$

where $\text{cov}_a(g^T\hat{\xi}^*) = \frac{1}{n}\sigma_1^2 g\mathbf{\Sigma}_{z^*z^*}^{-1}g^T$.

A.3. Asymptotic Normality of $\hat{\Gamma}(u)$

To derive the asymptotic normality of $\hat{\beta}(u)$, we first derive the asymptotic distribution of $\hat{\Gamma}(u)$. Suppose that X^mX^{mT} is invertible, then $\ell_2(C)$ has a unique minimizer $\hat{C} = (X^{mT}X^m)^{-1}X^{mT}M$. Set $\tilde{m}_i = \Gamma(u_i)^Tx_i$ for any $i = 1,\dots,n$, $\tilde{M} = (\tilde{m}_1,\dots,\tilde{m}_n)^T$, and $\tilde{C} = (X^{mT}X^m)^{-1}X^{mT}\tilde{M}$. We can obtain that $\text{E}(\hat{C}) = \tilde{C}$ and $\text{E}\{\hat{\Gamma}(u_i)\} = \tilde{\Gamma}(u_i) \stackrel{\Delta}{=} \{I_q\otimes d(u_i)^T\}\tilde{C}$ conditioning on $\mathcal{X}$.

Lemma A.1. Under Conditions C1–C7, for any nonzero constant vector $h\in\mathbb{R}^{qK_mp\times 1}$, the estimate $\hat{C}$ satisfies

$$\left[\text{cov}_a\{h^T\text{vec}(\hat{C}^T)\}\right]^{-1/2}h^T\text{vec}\{(\hat{C}-\tilde{C})^T\}\xrightarrow{D}N(0,1),$$

where $\text{cov}_a\{h^T\text{vec}(\hat{C}^T)\} = \frac{1}{n}h^T(\mathbf{\Sigma}_{x^mx^m}^{-1}\otimes\mathbf{\Sigma}_\epsilon)h$.

Proof. Since $\hat{C} = (X^{mT}X^m)^{-1}X^{mT}M$ and $\tilde{C} = (X^{mT}X^m)^{-1}X^{mT}\tilde{M}$, we can obtain that

$$\hat{C}-\tilde{C} = (X^{mT}X^m)^{-1}X^{mT}E.$$

To derive the asymptotic distribution of $\hat{C}$, let us first define the covariance of a matrix. Let B be a matrix of dimension $m\times n$, where $B = (b_1,\dots,b_n)$. Then $\text{vec}(B) = (b_1^T,\dots,b_n^T)^T$, and $\text{cov}(B) = \text{cov}\{\text{vec}(B^T)\}$.

To find the asymptotic distribution of $\sqrt{n}(\hat{C}-\tilde{C})$, suppose for any nonzero constant vector $h\in\mathbb{R}^{qK_mp\times 1}$, we have

$$\begin{aligned}&\sqrt{n}h^T[\text{vec}\{(\hat{C}-\tilde{C})^T\}]\\ &= \frac{1}{\sqrt{n}}h^T[\text{vec}\{E^TX^m(\frac{1}{n}X^{mT}X^m)^{-1}\}]\\ &= \frac{1}{\sqrt{n}}h^T\text{vec}(E^TX^m\mathbf{\Sigma}_{x^mx^m}^{-1}) + \frac{1}{\sqrt{n}}h^T\text{vec}\{E^TX^m(\hat{\mathbf{\Sigma}}_{x^mx^m}^{-1} - \mathbf{\Sigma}_{x^mx^m}^{-1})\}.\end{aligned}$$

Without loss of generality, we assume $\|\boldsymbol{h}\| = 1$. Similar to the proof of Theorem 1, we want to derive the asymptotic distribution of $\frac{1}{\sqrt{n}}\boldsymbol{h}^T\text{vec}(\boldsymbol{E}^T\boldsymbol{X}^m\boldsymbol{\Sigma}_{x^mx^m}^{-1})$ and show that $\frac{1}{\sqrt{n}}\boldsymbol{h}^T\text{vec}\{\boldsymbol{E}^T\boldsymbol{X}^m(\hat{\boldsymbol{\Sigma}}_{x^mx^m}^{-1} - \boldsymbol{\Sigma}_{x^mx^m}^{-1})\}$ is negligible.

First, we show the asymptotic normality of $\boldsymbol{h}^T\text{vec}(\boldsymbol{E}^T\boldsymbol{X}^m\boldsymbol{\Sigma}_{x^mx^m}^{-1}) = \sum_{i=1}^n \boldsymbol{h}^T\text{vec}(\boldsymbol{\epsilon}_i\boldsymbol{x}_i^{mT}\boldsymbol{\Sigma}_{x^m}^{-1})$ by checking Lyapounov condition, where $\boldsymbol{\epsilon}_i \in \mathbb{R}^{p\times 1}$ denotes the ith row of $\boldsymbol{E}$, $i = 1,\dots,n$. By direct calculation, we can show that

$$\mathcal{S}_n^2 = \text{var}\{\boldsymbol{h}^T\text{vec}(\boldsymbol{E}^T\boldsymbol{X}^m\boldsymbol{\Sigma}_{x^mx^m}^{-1})\} = n\boldsymbol{h}^T(\boldsymbol{\Sigma}_{x^mx^m}^{-1}\otimes\boldsymbol{\Sigma}_\epsilon)\boldsymbol{h}.$$

Since $\boldsymbol{h}^T\text{vec}(\boldsymbol{\epsilon}_i\boldsymbol{x}_i^{mT}\boldsymbol{\Sigma}_{x^mx^m}^{-1}) = \boldsymbol{h}^T\{(\boldsymbol{\Sigma}_{x^mx^m}^{-1}\boldsymbol{x}_i^m)\otimes\boldsymbol{I}_p\}\boldsymbol{\epsilon}_i \leq \left\|\boldsymbol{h}^T\{(\boldsymbol{\Sigma}_{x^mx^m}^{-1}\boldsymbol{x}_i^m)\otimes\boldsymbol{I}_p\}\right\|\cdot\|\boldsymbol{\epsilon}_i\|$ by Cauchy-Schwartz Inequality, we can obtain that for some constant $\delta > 0$,

$$\begin{aligned}
&\frac{1}{\mathcal{S}_n^{2+\delta}}\sum_{i=1}^n \text{E}\{|\boldsymbol{h}^T\text{vec}(\boldsymbol{\epsilon}_i\boldsymbol{x}_i^{mT}\boldsymbol{\Sigma}_{x^m}^{-1})|^{2+\delta}\}\\
&\leq \frac{1}{\mathcal{S}_n^{2+\delta}}\sum_{i=1}^n\left(\text{E}[\left\|\boldsymbol{h}^T\{(\boldsymbol{\Sigma}_{x^mx^m}^{-1}\boldsymbol{x}_i^m)\otimes\boldsymbol{I}_p\}\right\|^{2+\delta}]\cdot\text{E}(\|\boldsymbol{\epsilon}_i\|^{2+\delta})\right)\\
&\qquad\text{by independence}\\
&\leq \frac{1}{\mathcal{S}_n^{2+\delta}}\sum_{i=1}^n\left(\text{E}[\left\|\boldsymbol{h}^T\{(\boldsymbol{\Sigma}_{x^mx^m}^{-1}\boldsymbol{x}_i^m)\otimes\boldsymbol{I}_p\}\right\|^{2+\delta}]\cdot N_3'\right)\\
&\qquad\text{by Condition C3}\\
&\leq \frac{1}{\mathcal{S}_n^{2+\delta}}\sum_{i=1}^n\Big(\text{E}\Big[\Big\|\boldsymbol{h}^T(\boldsymbol{\Sigma}_{x^mx^m}^{-1}\\
&\quad\otimes\boldsymbol{\Sigma}_\epsilon)^{1/2}(\boldsymbol{\Sigma}_{x^mx^m}^{-1}\otimes\boldsymbol{\Sigma}_\epsilon)^{-1/2}\cdot\{(\boldsymbol{\Sigma}_{x^mx^m}^{-1}\boldsymbol{x}_i^m)\otimes\boldsymbol{I}_p\}\Big\|^{2+\delta}\Big]\cdot N_3'\Big)\\
&\leq \frac{1}{\mathcal{S}_n^{2+\delta}}\sum_{i=1}^n\Big(\left\|\boldsymbol{h}^T(\boldsymbol{\Sigma}_{x^mx^m}^{-1}\otimes\boldsymbol{\Sigma}_\epsilon)^{1/2}\right\|^{2+\delta}\\
&\quad\cdot\text{E}\left[\left\|(\boldsymbol{\Sigma}_{x^mx^m}^{-1}\otimes\boldsymbol{\Sigma}_\epsilon)^{-1/2}\cdot\{(\boldsymbol{\Sigma}_{x^mx^m}^{-1}\boldsymbol{x}_i^m)\otimes\boldsymbol{I}_p\}\right\|_F^{2+\delta}\right]\cdot N_3'\Big)\\
&= \frac{1}{\mathcal{S}_n^{2+\delta}}\sum_{i=1}^n\Big(\left\|\boldsymbol{h}^T(\boldsymbol{\Sigma}_{x^mx^m}^{-1}\otimes\boldsymbol{\Sigma}_\epsilon)\boldsymbol{h}\right\|^{(2+\delta)/2}\\
&\quad\cdot\text{E}\left[\left\|(\boldsymbol{\Sigma}_{x^mx^m}^{-1}\otimes\boldsymbol{\Sigma}_\epsilon)^{-1/2}\cdot\{(\boldsymbol{\Sigma}_{x^mx^m}^{-1}\boldsymbol{x}_i^m)\otimes\boldsymbol{I}_p\}\right\|_F^{2+\delta}\right]\cdot N_3'\Big)\\
&\lesssim \frac{1}{n^{(2+\delta)/2}}\sum_{i=1}^n\text{E}\Big[\Big\|(\boldsymbol{\Sigma}_{x^mx^m}^{-1}\otimes\boldsymbol{\Sigma}_\epsilon)^{-1/2}\\
&\quad\cdot\{(\boldsymbol{\Sigma}_{x^mx^m}^{-1}\boldsymbol{x}_i^m)\otimes\boldsymbol{I}_p\}\Big\|_F^{2+\delta}\Big].
\end{aligned}$$

Note that

$$\begin{aligned}
&\text{E}\left[\left\|(\boldsymbol{\Sigma}_{x^mx^m}^{-1}\otimes\boldsymbol{\Sigma}_\epsilon)^{-1/2}\cdot\{(\boldsymbol{\Sigma}_{x^mx^m}^{-1}\boldsymbol{x}_i^m)\otimes\boldsymbol{I}_p\}\right\|_F^{2+\delta}\right]\\
&= \text{E}\Big[\text{tr}\{(\boldsymbol{x}_i^{mT}\boldsymbol{\Sigma}_{x^mx^m}^{-1})\\
&\quad\otimes\boldsymbol{I}_p\}(\boldsymbol{\Sigma}_{x^mx^m}^{-1}\otimes\boldsymbol{\Sigma}_\epsilon)^{-1}\{(\boldsymbol{\Sigma}_{x^mx^m}^{-1}\boldsymbol{x}_i^m)\otimes\boldsymbol{I}_p\}\Big]^{(2+\delta)/2}\\
&= \text{E}\left[\text{tr}\{(\boldsymbol{x}_i^{mT}\boldsymbol{\Sigma}_{x^mx^m}^{-1}\boldsymbol{x}_i^m)\otimes\boldsymbol{\Sigma}_\epsilon^{-1}\}\right]^{(2+\delta)/2}\\
&\leq \text{E}\left[\text{tr}\{\lambda_{\max}(\boldsymbol{\Sigma}_{x^mx^m}^{-1})\left\|\boldsymbol{x}_i^m\right\|^2\boldsymbol{\Sigma}_\epsilon^{-1}\}\right]^{(2+\delta)/2}\\
&= \{\lambda_{\max}(\boldsymbol{\Sigma}_{x^mx^m}^{-1})\}^{(2+\delta)/2}\cdot\text{E}(\left\|\boldsymbol{x}_i^m\right\|^{2+\delta})\{\text{tr}(\boldsymbol{\Sigma}_\epsilon^{-1})\}^{(2+\delta)/2}\\
&\lesssim K_m^{\nu_2(2+\delta)/2}\cdot\text{E}(\left\|\boldsymbol{x}_i^m\right\|^{2+\delta})\cdot\{\text{tr}(\boldsymbol{\Sigma}_\epsilon^{-1})\}^{(2+\delta)/2}\\
&\qquad\text{by Conditions C7 and C3}\\
&\lesssim K_m^{(\nu_2+1)(2+\delta)/2},
\end{aligned}$$

since $\text{E}(\left\|\boldsymbol{x}_i^m\right\|^{2+\delta}) = \text{E}\{(\sum_{k=1}^{qK_m}|x_{ik}^m|^2)^{(2+\delta)/2}\} \leq \text{E}\{(qK_m)^{\delta/2}\sum_{k=1}^{qK_m}|x_{ik}^m|^{(2+\delta)}\} \lesssim K_m^{(\delta+2)/2}$ by Condition C2 and Jensen's Inequality. Thus,

$$\frac{1}{\mathcal{S}_n^{2+\delta}}\sum_{i=1}^n\text{E}\{|\boldsymbol{h}^T\text{vec}(\boldsymbol{\epsilon}_i\boldsymbol{x}_i^{mT}\boldsymbol{\Sigma}_{x^m}^{-1})|^{2+\delta}\} \lesssim \frac{1}{n^{\delta/2}}\cdot K_m^{(\nu_2+1)(2+\delta)/2},$$

which tends to 0 by Condition C6. Therefore,

$$\{n\boldsymbol{h}^T(\boldsymbol{\Sigma}_{x^mx^m}^{-1}\otimes\boldsymbol{\Sigma}_\epsilon)\boldsymbol{h}\}^{-1/2}\boldsymbol{h}^T\text{vec}(\boldsymbol{E}^T\boldsymbol{X}^m\boldsymbol{\Sigma}_{x^mx^m}^{-1}) \xrightarrow{D} N(0,1).$$

Second, we want to show that $\mathcal{R}_n = \frac{1}{\sqrt{n}}\boldsymbol{h}^T\text{vec}\{\boldsymbol{E}^T\boldsymbol{X}^m(\hat{\boldsymbol{\Sigma}}_{x^mx^m}^{-1} - \boldsymbol{\Sigma}_{x^mx^m}^{-1})\}$ is negligible. Note that $\mathcal{R}_n = \frac{1}{\sqrt{n}}\boldsymbol{h}^T\{(\hat{\boldsymbol{\Sigma}}_{x^mx^m}^{-1} - \boldsymbol{\Sigma}_{x^mx^m}^{-1})\boldsymbol{X}^{mT}\otimes\boldsymbol{I}_p\}\text{vec}(\boldsymbol{E}^T)$. Since $\text{E}(\mathcal{R}_n|\boldsymbol{X}^m) = 0$, $\text{E}(\mathcal{R}_n) = 0$. We next show $\mathcal{R}_n = o_P(1)$ by showing that $\text{E}(\mathcal{R}_n^2) = o_P(1)$. Note that

$$\begin{aligned}
&\text{E}(\mathcal{R}_n^2)\\
&= \text{E}\Big[\frac{1}{n}\boldsymbol{h}^T\{(\hat{\boldsymbol{\Sigma}}_{x^mx^m}^{-1} - \boldsymbol{\Sigma}_{x^mx^m}^{-1})\boldsymbol{X}^{mT}\otimes\boldsymbol{I}_p\}\text{vec}(\boldsymbol{E}^T)\text{vec}(\boldsymbol{E}^T)^T\\
&\quad\{\boldsymbol{X}^m(\hat{\boldsymbol{\Sigma}}_{x^mx^m}^{-1} - \boldsymbol{\Sigma}_{x^mx^m}^{-1})\otimes\boldsymbol{I}_p\}\boldsymbol{h}\Big]\\
&= \text{E}\Big[\frac{1}{n}\boldsymbol{h}^T\{(\hat{\boldsymbol{\Sigma}}_{x^mx^m}^{-1} - \boldsymbol{\Sigma}_{x^mx^m}^{-1})\boldsymbol{X}^{mT}\otimes\boldsymbol{I}_p\}\text{E}\{\text{vec}(\boldsymbol{E}^T)\text{vec}(\boldsymbol{E}^T)^T|\boldsymbol{X}^m\}\\
&\quad\{\boldsymbol{X}^m(\hat{\boldsymbol{\Sigma}}_{x^mx^m}^{-1} - \boldsymbol{\Sigma}_{x^mx^m}^{-1})\otimes\boldsymbol{I}_p\}\boldsymbol{h}\Big]\\
&= \text{E}\Big[\frac{1}{n}\boldsymbol{h}^T\{(\hat{\boldsymbol{\Sigma}}_{x^mx^m}^{-1} - \boldsymbol{\Sigma}_{x^mx^m}^{-1})\boldsymbol{X}^{mT}\otimes\boldsymbol{I}_p\}(\boldsymbol{I}_n\otimes\boldsymbol{\Sigma}_\epsilon)\\
&\quad\{\boldsymbol{X}^m(\hat{\boldsymbol{\Sigma}}_{x^mx^m}^{-1} - \boldsymbol{\Sigma}_{x^mx^m}^{-1})\otimes\boldsymbol{I}_p\}\boldsymbol{h}\Big]\\
&= \text{E}\Big(\boldsymbol{h}^T\Big[\{(\hat{\boldsymbol{\Sigma}}_{x^mx^m}^{-1} - \boldsymbol{\Sigma}_{x^mx^m}^{-1})\hat{\boldsymbol{\Sigma}}_{x^mx^m}\\
&\quad(\hat{\boldsymbol{\Sigma}}_{x^mx^m}^{-1} - \boldsymbol{\Sigma}_{x^mx^m}^{-1})\}\otimes\boldsymbol{\Sigma}_\epsilon\Big]\boldsymbol{h}\Big),
\end{aligned}$$

and

$$\begin{aligned}
&\boldsymbol{h}^T\left[\{(\hat{\boldsymbol{\Sigma}}_{x^mx^m}^{-1} - \boldsymbol{\Sigma}_{x^mx^m}^{-1})\hat{\boldsymbol{\Sigma}}_{x^mx^m}(\hat{\boldsymbol{\Sigma}}_{x^mx^m}^{-1} - \boldsymbol{\Sigma}_{x^mx^m}^{-1})\}\otimes\boldsymbol{\Sigma}_\epsilon\right]\boldsymbol{h}\\
&\leq \lambda_{\max}\Big[\{(\hat{\boldsymbol{\Sigma}}_{x^mx^m}^{-1} - \boldsymbol{\Sigma}_{x^mx^m}^{-1})\hat{\boldsymbol{\Sigma}}_{x^mx^m}\\
&\quad(\hat{\boldsymbol{\Sigma}}_{x^mx^m}^{-1} - \boldsymbol{\Sigma}_{x^mx^m}^{-1})\}\otimes\boldsymbol{\Sigma}_\epsilon\Big]\cdot\|\boldsymbol{h}\|^2\\
&\leq \lambda_{\max}\{(\hat{\boldsymbol{\Sigma}}_{x^mx^m}^{-1} - \boldsymbol{\Sigma}_{x^mx^m}^{-1})\hat{\boldsymbol{\Sigma}}_{x^mx^m}(\hat{\boldsymbol{\Sigma}}_{x^mx^m}^{-1} - \boldsymbol{\Sigma}_{x^mx^m}^{-1})\}\\
&\quad\lambda_{\max}(\boldsymbol{\Sigma}_\epsilon)\cdot\|\boldsymbol{h}\|^2\\
&= \lambda_{\max}\{\boldsymbol{\Sigma}_{x^mx^m}^{-1/2}(\boldsymbol{\Sigma}_{x^mx^m}^{1/2}\hat{\boldsymbol{\Sigma}}_{x^mx^m}^{-1}\boldsymbol{\Sigma}_{x^mx^m}^{1/2} - \boldsymbol{I})\\
&\quad\boldsymbol{\Sigma}_{x^mx^m}^{-1/2}\hat{\boldsymbol{\Sigma}}_{x^mx^m}^{-1}\boldsymbol{\Sigma}_{x^mx^m}^{-1/2}(\boldsymbol{\Sigma}_{x^mx^m}^{1/2}\hat{\boldsymbol{\Sigma}}_{x^mx^m}^{-1}\boldsymbol{\Sigma}_{x^mx^m}^{1/2} - \boldsymbol{I})\boldsymbol{\Sigma}_{x^mx^m}^{-1/2}\}\\
&\quad\cdot\lambda_{\max}(\boldsymbol{\Sigma}_\epsilon)\cdot\|\boldsymbol{h}\|^2\\
&\leq \lambda_{\max}(\boldsymbol{\Sigma}_{x^mx^m}^{-1/2}\hat{\boldsymbol{\Sigma}}_{x^mx^m}^{-1}\boldsymbol{\Sigma}_{x^mx^m}^{-1/2})\cdot\lambda\{\boldsymbol{\Sigma}_{x^mx^m}^{-1/2}(\boldsymbol{\Sigma}_{x^mx^m}^{1/2}\\
&\quad\hat{\boldsymbol{\Sigma}}_{x^mx^m}^{-1}\boldsymbol{\Sigma}_{x^mx^m}^{1/2} - \boldsymbol{I})^2\boldsymbol{\Sigma}_{x^mx^m}^{-1/2}\}\cdot\lambda_{\max}(\boldsymbol{\Sigma}_\epsilon)\cdot\|\boldsymbol{h}\|^2,
\end{aligned}$$

which tends to 0 by Conditions C3 and C7. Thus, $\text{E}(\mathcal{R}_n^2) = o_P(1)$ and $\mathcal{R}_n$ is negligible.

As a result, it follows that

$$\left\{\frac{1}{n}\boldsymbol{h}^T(\boldsymbol{\Sigma}_{x^mx^m}^{-1}\otimes\boldsymbol{\Sigma}_\epsilon)\boldsymbol{h}\right\}^{-1/2}\boldsymbol{h}^T[\text{vec}\{(\hat{\boldsymbol{C}} - \tilde{\boldsymbol{C}})^T\}] \xrightarrow{D} N(0,1).$$

This completes the proof of this lemma.

Lemma A.1 implies that $\sqrt{n}(\hat{\boldsymbol{C}} - \tilde{\boldsymbol{C}})$ asymptotically follows a matrix normal distribution with mean $\boldsymbol{0}$ and covariance matrix $(\boldsymbol{\Sigma}_{x^mx^m}^{-1}\otimes\boldsymbol{\Sigma}_\epsilon)$. Based on this,

$$\hat{\boldsymbol{\Gamma}}(u) \overset{a}{\sim} N(\tilde{\boldsymbol{\Gamma}}(u), \frac{1}{n}\left[\{\boldsymbol{I}_q\otimes\boldsymbol{d}(u)^T\}\boldsymbol{\Sigma}_{x^mx^m}^{-1}\{\boldsymbol{I}_q\otimes\boldsymbol{d}(u)\}\right]\otimes\boldsymbol{\Sigma}_\epsilon).$$

A.4. Proof of Theorem 2

Under Conditions C1–C6, it follows by Huang, Wu, and Zhou (2004) that $\|\hat{\alpha}_{0j} - \tilde{\alpha}_{0j}\|_{L_2}^2 = O_P(1/n + K_0/n)$ and $\left\|\hat{\Gamma}_{lj} - \tilde{\Gamma}_{lj}\right\|_{L_2}^2 = O_P(1/n + K_m/n)$. Moreover, $\|\tilde{\alpha}_{0j} - \alpha_{0j}\|_{L_2}^2 = O_P(\rho_{n0}^2)$ for any $j = 1, \ldots, p$, and $\left\|\tilde{\Gamma}_{lj} - \Gamma_{lj}\right\|_{L_2}^2 = O_P(\rho_{n2}^2)$ for any $j = 1, \ldots, p$, $l = 1, \ldots, q$. The results of convergence rates for longitudinal data in Huang, Wu, and Zhou (2004) can be applied to this article under the assumption that the number of observation times for each individual i is $n_i = 1$.

It is noteworthy that in Huang, Wu, and Zhou (2004), covariates are assumed to be bounded. However, under model (1) and (2), the mediator $\boldsymbol{m}_i$ is no longer bounded. Some lemmas in Huang, Wu, and Zhou (2004) are slightly affected. For example, in the proof of Lemma A.4, to obtain the order of $\mathrm{E}\{M^2(t)B^2(t)\}$, we cannot apply the boundness condition of covariate M and properties of B-splines directly. Instead, we need to check $\mathrm{E}\{M^2(t)B^2(t)\} \le \mathrm{E}(|M^2(t)|_1) \cdot \left\|B^2(t)\right\|_\infty$ explicitly and then obtain the order. Another example is in the proof of Lemma A.7 in Huang, Wu, and Zhou (2004). Instead of bound the covariates directly, we can control the order of covariates by Condition C2. The only lemma influenced substantially by this additional constraint is Lemma A.2 in Huang, Wu, and Zhou (2004). However, we can still show that it is applicable here. The reason is as follows.

Let $z_i = (m_{i1}, \ldots, m_{ip}, x_{i1}, \ldots, x_{iq})^T = (z_1, \ldots, z_h, \ldots, z_{p+q})^T$, where $h = 1, \ldots, p+q$. For simplicity, assume $b_0(u) = b_1(u) = b(u)$ for any $u \in \mathcal{U}$. Let

$$w_i = w_{i,h,h',k,k'} = \{z_{ih}b_{hk}(u_i)z_{ih'}b_{h'k'}(u_i)\} - \mathrm{E}\{z_{ih}b_{hk}(u_i)z_{ih'}b_{h'k'}(u_i)\},$$

where $k = 1, \ldots, K$, and $K = K_1 = K_0$. We want to show that $P(|\sum_{i=1}^n w_i| > ns)$, $s > 0$, can be bounded by the same constant derived from Bernstein's inequality in Huang, Wu, and Zhou (2004). Instead of obtaining the bound of $P\left(\left|\sum_{i=1}^n w_i\right| > ns\right)$ directly, we decompose it into P_1, P_2 and P_3 as follows.

$$\begin{aligned}
&P\left(\left|\sum_{i=1}^n w_i\right| > ns\right)\\
=&P\left(\sum_{i=1}^n w_i > ns, \max|z_{ih}| \le M_n, \max|z_{ih'}| \le M_n\right)\\
&+ P\left(\sum_{i=1}^n w_i < -ns, \max|z_{ih}| \le M_n, \max|z_{ih'}| \le M_n\right)\\
&+ P\left(\left|\sum_{i=1}^n w_i\right| > ns, \max|z_{ih}| \right.\\
&\qquad \left. \ge M_n \text{ or } \max|z_{ih'}| \ge M_n \text{ or both, for some} i\right)\\
\triangleq& P_1 + P_2 + P_3
\end{aligned}$$

It is easy to show that P_1 and P_2 can be controlled using the same procedures in (Huang, Wu, and Zhou 2004, lem. A.2). In addition, P_3 can be controlled by $nP(|z| \ge M_n)$ based on countable subadditivity. Since z satisfies the sub-exponential tail probability in $p+q$, that is, there exists $s_0 > 0$, such that for $0 \le s < s_0$, $\max_{1 \le h \le p+q} E\left\{\exp\left(sz_h^2|u\right)\right\} < \infty$. By Lemma S3 in Liu, Li, and Wu (2014), there exist m_1, $m_2 > 0$, such that $P(|z| \ge M_n) \le m_1 exp\{-m_2 M_n\}$. Thus, $P_3 \le cnm_1\exp\{-m_2 M_n\} \le c(1-\epsilon)^n$ for some constant $c > 0$, by taking $M_n = m_2^{-1}\log\{\frac{nm_1}{(1-\epsilon)^n}\}$, and P_3 tends to 0 as n increases.

Therefore, the results of convergence rates in Huang, Wu, and Zhou (2004) naturally apply to this article. Then, $\|\hat{\alpha}_{0j} - \alpha_{0j}\|_{L_2}^2 = O_P(1/n + K_0/n + \rho_{n0}^2)$ and $\left\|\hat{\Gamma}_{lj} - \Gamma_{lj}\right\|_{L_2}^2 = O_P(1/n + K_m/n + \rho_{n2}^2)$. Since

$$\begin{aligned}
\left\|\hat{\beta}_l - \beta_l\right\|_{L_2}^2 &= \int_{u\in\mathcal{U}} |\hat{\beta}_l(u) - \beta_l(u)|^2 f_U(u)du\\
&= \int_{u\in\mathcal{U}} \left|\sum_{j=1}^p \{\hat{\Gamma}_{lj}(u)\hat{\alpha}_{0j}(u) - \Gamma_{lj}(u)\alpha_{0j}(u)\}\right|^2 f_U(u)du\\
&\le p \cdot \sum_{j=1}^p \int_{u\in\mathcal{U}} \left|\hat{\Gamma}_{lj}(u)\hat{\alpha}_{0j}(u) - \Gamma_{lj}(u)\alpha_{0j}(u)\right|^2 f_U(u)du
\end{aligned}$$

by the Cauchy–Schwartz inequality, and

$$\begin{aligned}
&\int_{u\in\mathcal{U}} \left|\hat{\Gamma}_{lj}(u)\hat{\alpha}_{0j}(u) - \Gamma_{lj}(u)\alpha_{0j}(u)\right|^2 f_U(u)du\\
=&\int_{u\in\mathcal{U}} \left|\hat{\Gamma}_{lj}(u)\{\hat{\alpha}_{0j}(u) - \alpha_{0j}(u)\} + \alpha_{0j}(u)\{\hat{\Gamma}_{lj}(u) - \Gamma_{lj}(u)\}\right|^2 f_U(u)du\\
\le& 2\int_{u\in\mathcal{U}} \left|\hat{\Gamma}_{lj}(u)\{\hat{\alpha}_{0j}(u) - \alpha_{0j}(u)\}\right|^2 f_U(u)du + 2\\
&\quad \int_{u\in\mathcal{U}} \left|\alpha_{0j}(u)\{\hat{\Gamma}_{lj}(u) - \Gamma_{lj}(u)\}\right|^2 f_U(u)du\\
\le& 2M_\Gamma^2 \int_{u\in\mathcal{U}} \left|\hat{\alpha}_{0j}(u) - \alpha_{0j}(u)\right|^2 f_U(u)du + 2M_0^2\\
&\quad \int_{u\in\mathcal{U}} \left|\hat{\Gamma}_{lj}(u) - \Gamma_{lj}(u)\right|^2 f_U(u)du\\
=& 2M_\Gamma^2 \left\|\hat{\alpha}_{0j} - \alpha_{0j}\right\|_{L_2}^2 + 2M_0^2 \left\|\hat{\Gamma}_{lj} - \Gamma_{lj}\right\|_{L_2}^2
\end{aligned}$$

by Condition C4, we can derive that

$$\begin{aligned}
&\int_{u\in\mathcal{U}} \left|\hat{\Gamma}_{lj}(u)\hat{\alpha}_{0j}(u) - \Gamma_{lj}(u)\alpha_{0j}(u)\right|^2 f_U(u)du\\
&\quad = O_P\{1/n + K_0/n + K_m/n + \rho_{n0}^2 + \rho_{n2}^2\}.
\end{aligned}$$

Thus, $\left\|\hat{\beta}_l - \beta_l\right\|_{L_2}^2 = O_P\{1/n + K_0/n + K_m/n + \rho_{n0}^2 + \rho_{n2}^2\}$.

A.5. Proof of Theorem 3 (a)

First, because ϵ_1 and $\boldsymbol{\epsilon}$ are independent, the joint log-likelihood function

$$\sum_{i=1}^n \log\{f(y_i|\boldsymbol{m}_i^*, \boldsymbol{x}_i^*, \boldsymbol{\alpha}_0^*, \boldsymbol{\alpha}_1^*)\} + \sum_{i=1}^n \log\{f(\boldsymbol{m}_i|\boldsymbol{x}_i^m, \boldsymbol{C})\}$$

implies that $\hat{\boldsymbol{C}}$ and $\hat{\boldsymbol{\alpha}}_0^*$ are independent, and therefore, $\hat{\boldsymbol{\Gamma}}(u)$ and $\hat{\boldsymbol{\alpha}}_0(u)$ are independent.

The asymptotic distribution of $\hat{\boldsymbol{\beta}}(u)$ can be obtained as follows. For any $\boldsymbol{a}_{q\times 1}$, we consider Cramer's Device. Since

$$\begin{aligned}
&\mathrm{cov}_a\{\mathrm{vec}(\hat{\boldsymbol{\Gamma}}(u)^T - \tilde{\boldsymbol{\Gamma}}(u)^T)\}\\
&\quad = \frac{1}{n}\left[\{\boldsymbol{I}_q \otimes \boldsymbol{d}(u)^T\}\boldsymbol{\Sigma}_{x^m x^m}^{-1}\{\boldsymbol{I}_q \otimes \boldsymbol{d}(u)\}\right] \otimes \boldsymbol{\Sigma}_\epsilon,
\end{aligned}$$

$\mathrm{vec}(\boldsymbol{I}_p\hat{\boldsymbol{\Gamma}}(u)^T\boldsymbol{a}) = (\boldsymbol{a}^T \otimes \boldsymbol{I}_p)\mathrm{vec}(\hat{\boldsymbol{\Gamma}}(u)^T)$, and $\mathrm{vec}(\hat{\boldsymbol{\Gamma}}(u)^T\boldsymbol{a} - \tilde{\boldsymbol{\Gamma}}(u)^T\boldsymbol{a}) = (\boldsymbol{a}^T \otimes \boldsymbol{I}_p)\{\mathrm{vec}(\hat{\boldsymbol{\Gamma}}(u)^T) - \mathrm{vec}(\tilde{\boldsymbol{\Gamma}}(u)^T)\}$, we can get that

$$\begin{aligned}
&\mathrm{cov}_a\{\mathrm{vec}(\hat{\boldsymbol{\Gamma}}(u)^T\boldsymbol{a} - \tilde{\boldsymbol{\Gamma}}(u)^T\boldsymbol{a})\}\\
=&\mathrm{cov}_a[(\boldsymbol{a}^T \otimes \boldsymbol{I}_p)\{\mathrm{vec}(\hat{\boldsymbol{\Gamma}}(u)^T - \tilde{\boldsymbol{\Gamma}}(u)^T)\}]\\
=&(\boldsymbol{a}^T \otimes \boldsymbol{I}_p)\left(\frac{1}{n}\left[\{\boldsymbol{I}_q \otimes \boldsymbol{d}(u)^T\}\boldsymbol{\Sigma}_{x^m x^m}^{-1}\{\boldsymbol{I}_q \otimes \boldsymbol{d}(u)\}\right] \otimes \boldsymbol{\Sigma}_\epsilon\right)(\boldsymbol{a} \otimes \boldsymbol{I}_p)\\
=&\frac{1}{n}\left[\boldsymbol{a}^T\{\boldsymbol{I}_q \otimes \boldsymbol{d}(u)^T\}\boldsymbol{\Sigma}_{x^m x^m}^{-1}\{\boldsymbol{I}_q \otimes \boldsymbol{d}(u)\}\boldsymbol{a}\right] \otimes \boldsymbol{\Sigma}_\epsilon.
\end{aligned}$$

If we let $\tilde{\theta}(u) = \tilde{\boldsymbol{\Gamma}}(u)^T \boldsymbol{a}$ and $\hat{\theta}(u) = \hat{\boldsymbol{\Gamma}}(u)^T \boldsymbol{a}$, then

$$[\mathrm{cov}_a\{\hat{\theta}(u)\}]^{-1/2}(\hat{\theta}(u)^T - \tilde{\theta}(u)^T) \xrightarrow{D} N(\mathbf{0}, \boldsymbol{I}),$$

where $\mathrm{cov}_a\{\hat{\theta}(u)\} = \frac{1}{n}\big[\boldsymbol{a}^T\{\boldsymbol{I}_q \otimes \boldsymbol{d}(u)^T\}\boldsymbol{\Sigma}_{x^m x^m}^{-1}\{\boldsymbol{I}_q \otimes \boldsymbol{d}(u)\}\boldsymbol{a}\big]\boldsymbol{\Sigma}_{\epsilon}$.

To obtain the asymptotic distribution of $\hat{\boldsymbol{\beta}}(u) = \hat{\boldsymbol{\Gamma}}(u)\hat{\boldsymbol{\alpha}}_0(u)$ by the delta method, let $\boldsymbol{\Sigma}_{\theta(u)} = \mathrm{cov}_a\{\hat{\theta}(u)\}$ and $\boldsymbol{\Sigma}_{\alpha_0(u)} = \mathrm{cov}_a\{\hat{\boldsymbol{\alpha}}_0(u)\} = \frac{1}{n}\sigma_1^2\{\boldsymbol{I}_p \otimes \boldsymbol{b}_0(u)^T\}\boldsymbol{\Sigma}_{m*m*.x*}^{-1}\{\boldsymbol{I}_p \otimes \boldsymbol{b}_0(u)\}$ for simplicity.

Since $\hat{\boldsymbol{\alpha}}_0(u)$ and $\hat{\boldsymbol{\Gamma}}(u)$ are independent and

$$\begin{pmatrix} \boldsymbol{\Sigma}_{\theta(u)} & \mathbf{0} \\ \mathbf{0} & \boldsymbol{\Sigma}_{\alpha_0(u)} \end{pmatrix}^{-1/2} \left\{ \begin{pmatrix} \hat{\theta}(u)^T \\ \hat{\boldsymbol{\alpha}}_0(u) \end{pmatrix} - \begin{pmatrix} \tilde{\theta}(u)^T \\ \tilde{\boldsymbol{\alpha}}_0(u) \end{pmatrix} \right\} \xrightarrow{D} N(\mathbf{0}, \boldsymbol{I}),$$

we have

$$\{\tilde{\boldsymbol{\alpha}}_0(u)^T \boldsymbol{\Sigma}_{\theta(u)}\tilde{\boldsymbol{\alpha}}_0(u) + \tilde{\theta}(u)^T \boldsymbol{\Sigma}_{\alpha_0(u)}\tilde{\theta}(u)\}^{-1/2}\{\hat{\theta}(u)^T\hat{\boldsymbol{\alpha}}_0(u) - \tilde{\theta}(u)^T\tilde{\boldsymbol{\alpha}}_0(u)\} \xrightarrow{D} N(\mathbf{0}, \boldsymbol{I}),$$

where $\tilde{\boldsymbol{\alpha}}(u)^T\boldsymbol{\Sigma}_{\theta(u)}\tilde{\boldsymbol{\alpha}}_0(u) = \frac{1}{n}\tilde{\boldsymbol{\alpha}}_0(u)^T\boldsymbol{\Sigma}_{\epsilon}\tilde{\boldsymbol{\alpha}}_0(u)\boldsymbol{a}^T\{\boldsymbol{I}_q \otimes \boldsymbol{d}(u)^T\}\boldsymbol{\Sigma}_{x^m x^m}^{-1}\{\boldsymbol{I}_q \otimes \boldsymbol{d}(u)\}\boldsymbol{a}$, and $\tilde{\theta}(u)^T \cdot \boldsymbol{\Sigma}_{\alpha_0(u)}\tilde{\theta}(u) = \frac{1}{n}\sigma_1^2\boldsymbol{a}^T\tilde{\boldsymbol{\Gamma}}(u)\{\boldsymbol{I}_p \otimes \boldsymbol{b}_0(u)^T\}\boldsymbol{\Sigma}_{m*m*.x*}^{-1}\{\boldsymbol{I}_p \otimes \boldsymbol{b}_0(u)\}\tilde{\boldsymbol{\Gamma}}(u)^T\boldsymbol{a}$. Thus,

$$\begin{aligned}
&\tilde{\boldsymbol{\alpha}}_0(u)^T\boldsymbol{\Sigma}_{\theta(u)}\tilde{\boldsymbol{\alpha}}_0(u) + \tilde{\theta}(u)^T\boldsymbol{\Sigma}_{\alpha_0(u)}\tilde{\theta}(u) \\
=&\frac{1}{n}\boldsymbol{a}^T\big[\tilde{\boldsymbol{\alpha}}_0(u)^T\boldsymbol{\Sigma}_{\epsilon}\tilde{\boldsymbol{\alpha}}_0(u)\{\boldsymbol{I}_q \otimes \boldsymbol{d}(u)^T\}\boldsymbol{\Sigma}_{x^m x^m}^{-1}\{\boldsymbol{I}_q \otimes \boldsymbol{d}(u)\} \\
&+ \sigma_1^2\tilde{\boldsymbol{\Gamma}}(u)\{\boldsymbol{I}_p \otimes \boldsymbol{b}_0(u)^T\}\boldsymbol{\Sigma}_{m*m*.x*}^{-1}\{\boldsymbol{I}_p \otimes \boldsymbol{b}_0(u)\}\tilde{\boldsymbol{\Gamma}}(u)^T\big]\boldsymbol{a}.
\end{aligned}$$

Since $\{\hat{\theta}(u)^T\hat{\boldsymbol{\alpha}}_0(u) - \tilde{\theta}(u)^T\tilde{\boldsymbol{\alpha}}_0(u)\} = \boldsymbol{a}^T\{\hat{\boldsymbol{\Gamma}}(u)\hat{\boldsymbol{\alpha}}_0(u) - \tilde{\boldsymbol{\Gamma}}(u)\tilde{\boldsymbol{\alpha}}_0(u)\}$, it follows that

$$\begin{aligned}
&[\mathrm{cov}_a\{\hat{\boldsymbol{\beta}}(u)\}]^{-1/2}\{\hat{\boldsymbol{\Gamma}}(u)\hat{\boldsymbol{\alpha}}_0(u) - \tilde{\boldsymbol{\Gamma}}(u)\tilde{\boldsymbol{\alpha}}_0(u)\} \\
=&[\mathrm{cov}_a\{\hat{\boldsymbol{\beta}}(u)\}]^{-1/2}\{\hat{\boldsymbol{\beta}}(u) - \tilde{\boldsymbol{\beta}}(u)\} \\
&\xrightarrow{D} N\{\mathbf{0}, \boldsymbol{I}\},
\end{aligned}$$

where $\mathrm{cov}_a\{\hat{\boldsymbol{\beta}}(u)\} = \frac{1}{n}[\tilde{\boldsymbol{\alpha}}_0(u)^T\boldsymbol{\Sigma}_{\epsilon}\tilde{\boldsymbol{\alpha}}_0(u)\{\boldsymbol{I}_q \otimes \boldsymbol{d}(u)^T\}\boldsymbol{\Sigma}_{x^m x^m}^{-1}\{\boldsymbol{I}_q \otimes \boldsymbol{d}(u)\} + \sigma_1^2\tilde{\boldsymbol{\Gamma}}(u)\{\boldsymbol{I}_p \otimes \boldsymbol{b}_0(u)^T\} \cdot \boldsymbol{\Sigma}_{m*m*.x*}^{-1}\{\boldsymbol{I}_p \otimes \boldsymbol{b}_0(u)\}\tilde{\boldsymbol{\Gamma}}(u)^T]$.

References

Albert, J. M., Cho, J. I., Liu, Y., and Nelson, S. (2019), "Generalized Causal Mediation and Path Analysis: Extensions and Practical Considerations," *Statistical Methods in Medical Research*, 28, 1793–1807. [1759]

Albert, J. M., and Nelson, S. (2011), "Generalized Causal Mediation Analysis," *Biometrics*, 67, 1028–1038. [1759]

Avin, C., Shpitser, I., and Pearl, J. (2005), "Identifiability of Path-Specific Effects," in *Proceedings of the 19th International Joint Conference on Artificial Intelligence*, pp. 357–363. [1759]

Baron, R. M., and Kenny, D. A. (1986), "The Moderator–Mediator Variable Distinction in Social Psychological Research: Conceptual, Strategic, and Statistical Considerations," *Journal of Personality and Social Psychology* 51, 1173. [1759]

Cai, Z., Fan, J., and Li, R. (2000), "Efficient Estimation and Inferences for Varying-Coefficient Models," *Journal of the American Statistical Association*, 95, 888–902. [1759,1761,1766]

Chén, O. Y., Crainiceanu, C. M., Ogburn, E. L., Caffo, B. S., Wager, T. D., and Lindquist, M. A. (2018), "High-Dimensional Multivariate Mediation: With Application to Neuroimaging Data," *Biostatistics*, 19, 121–136. [1759]

Cleveland, W. S., Grosse, E., Shyu, W. M., Chambers, J. M., and Hastie, T. (1992), "Local Regression Models," in *Statistical Models in S*, eds. J. M. Chambers and T. Hastie, Boca Raton, FL: Wadsworth & Brooks/Cole, pp. 305–373. [1759]

Daniel, R., De Stavola, B., Cousens, S., and Vansteelandt, S. (2015), "Causal Mediation Analysis With Multiple Mediators," *Biometrics*, 71, 1–14. [1759]

Edwards, J. R., and Lambert, L. S. (2007), "Methods for Integrating Moderation and Mediation: A General Analytical Framework Using Moderated Path Analysis," *Psychological Methods*, 12, 1. [1759]

Fan, J., and Zhang, W. (2008), "Statistical Methods With Varying Coefficient Models," *Statistics and its Interface*, 1, 179. [1759]

Fan, J., and Zhang, W. (1999), "Statistical Estimation in Varying Coefficient Models," *The Annals of Statistics*, 27, 1491–1518. [1759]

Hiroto, D. S., and Seligman, M. E. (1975), "Generality of Learned Helplessness in Man," *Journal of Personality and Social Psychology*, 31, 311. [1760,1765]

Huang, J. Z., Wu, C. O., and Zhou, L. (2002), "Varying-Coefficient Models and Basis Function Approximations for the Analysis of Repeated Measurements," *Biometrika*, 89, 111–128. [1759,1767]

Huang, J. Z., Wu, C. O., and Zhou, L. (2004), "Polynomial Spline Estimation and Inference for Varying Coefficient Models With Longitudinal Data," *Statistica Sinica*, 14, 763–788. [1759,1763,1767,1770]

Huang, Y. T., Liang, L., Moffatt, M. F., Cookson, W. O., and Lin, X. (2015), "igwas: Integrative Genome-Wide Association Studies of Genetic and Genomic Data for Disease Susceptibility Using Mediation Analysis," *Genetic Epidemiology*, 39, 347–356. [1759]

Huang, Y. T., and Pan, W. C. (2016), "Hypothesis Test of Mediation Effect in Causal Mediation Model With High-Dimensional Continuous Mediators," *Biometrics*, 72, 402–413. [1759]

Huang, Y. T., VanderWeele, T. J., and Lin, X. (2014), "Joint Analysis of SNP and Gene Expression Data in Genetic Association Studies of Complex Diseases," *The Annals of Applied Statistics*, 8, 352. [1759]

Liu, J., Li, R., and Wu, R. (2014), "Feature Selection for Varying Coefficient Models With Ultrahigh-Dimensional Covariates," *Journal of the American Statistical Association*, 109, 266–274. [1770]

MacKinnon, D. P. (2008), *Introduction to Statistical Mediation Analysis*, New York: Routledge. [1759]

Pollack, J. M., Vanepps, E. M., and Hayes, A. F. (2012), "The Moderating Role of Social Ties on Entrepreneurs' Depressed Affect and Withdrawal Intentions in Response to Economic Stress," *Journal of Organizational Behavior*, 33, 789–810. [1760,1765]

Preacher, K. J., Rucker, D. D., and Hayes, A. F. (2007), "Assessing Moderated Mediation Hypotheses: Theory, Methods, and Prescriptions," *Multivariate Behavioral Research*, 42, 185–227. [1759]

Pryce, C. R., Azzinnari, D., Spinelli, S., Seifritz, E., Tegethoff, M., and Meinlschmidt, G. (2011), "Helplessness: A Systematic Translational Review of Theory and Evidence for Its Relevance to Understanding and Treating Depression," *Pharmacology & Therapeutics*, 132, 242–267. [1760,1765]

Schumaker, L. (2007), *Spline Functions: Basic Theory*, Cambridge, MA: Cambridge University Press. [1762]

Valeri, L., and VanderWeele, T. J. (2013), "Mediation Analysis Allowing for Exposure–Mediator Interactions and Causal Interpretation: Theoretical Assumptions and Implementation With SAS and SPSS Macros," *Psychological Methods*, 18, 137. [1759]

Zhang, H., Zheng, Y., Zhang, Z., Gao, T., Joyce, B., Yoon, G., Zhang, W., Schwartz, J., Just, A., Colicino, E., Vokonas, P., Zhao, L., Lv, J., Baccarelli, A., Hou, L., and Liu, L. (2016), "Estimating and Testing High-Dimensional Mediation Effects in Epigenetic Studies," *Bioinformatics*, 32, 3150–3154. [1759]

Zhou, R., Wang, L., and Zhao, S. D. (2020), "Estimation and Inference for the Indirect Effect in High-Dimensional Linear Mediation Model," *Biometrika*, 107, 575–589. [1759]