

Statistica Sinica Preprint No: SS-2021-0032	
Title	A Simultaneous Likelihood Test for Joint Mediation Effects of Multiple Mediators
Manuscript ID	SS-2021-0032
URL	http://www.stat.sinica.edu.tw/statistica/
DOI	10.5705/ss.202021.0032
Complete List of Authors	Wei Hao and Peter X.-K. Song
Corresponding Author	Peter X.-K. Song
E-mail	pxsong@umich.edu
Notice: Accepted version subject to English editing.	

A SIMULTANEOUS LIKELIHOOD TEST FOR JOINT MEDIATION EFFECTS OF MULTIPLE MEDIATORS

Wei Hao¹ and Peter X.-K. Song¹

Department of Biostatistics, University of Michigan, MI, USA¹

Abstract: Mediation analysis via structural equation models has become a widely used tool to study whether the effect of an exposure on an outcome is mediated by some intermediate factors. When multiple mediators are present, statistical inference on the joint mediation effect is challenging due to the involvement of composite null hypotheses with a large number of parameter configurations. We propose a simultaneous likelihood ratio test in which a block coordinate descent algorithm is invoked to solve the constrained likelihood under the irregular null parameter space using the Lagrange Multiplier approach. We establish the asymptotic null distribution, and examine the performance of the proposed joint test statistic via extensive simulations with a comparison to existing tests. The simulation results show that our method controls type I error properly and in general provides better power than existing test methods. We apply our method to examine whether a group of glucose metabolites and acetylamino acids mediate the effect of nutrient intakes on insulin resistance.

Key words and phrases: Constrained maximum likelihood, directed acyclic graph, Lagrange multiplier, multi-dimensional mediators, structural equation model.

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, 1974; 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 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.

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

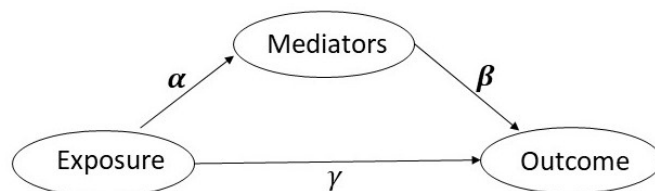


Figure 1: A DAG involving exposure, mediators and outcome.

method (Bollen and Stine, 1990), joint significant test (MacKinnon et al., 2002). Recently, with available omics data, testing for mediation effects had received much attentions, especially handling a group of multiple or even high-dimensional mediators. Several methods have been developed, such as multiple testing approaches for genome-wide association analysis proposed for simultaneous single mediator tests with multiple comparison correction (Huang, 2019; Huang 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 important 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 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 paper, 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 , $q = 1, \dots, Q$, where Q is a fixed number of mediators. One example of possible combination is $\boldsymbol{\alpha} = \boldsymbol{\beta} = \mathbf{0}$, which is of great

interest in practice, and is well known for its overly conservative type I error control. More subtle cases may arise from cancellations among some individual products of $\alpha_q\beta_q$, $q = 1, \dots, Q$ 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 $\boldsymbol{\alpha} = \boldsymbol{\beta} = \mathbf{0}$. Bore with the fundamental Neyman-Pearson Lemma, the likelihood ratio (LR) test is known as the uniformly most power test for a simple hypothesis testing problem under mild regularity conditions (Neyman and Pearson, 1933), and the Wilks' generalized LR test is one of top finite-sample performers in the literature. To bridge this gap, in this paper 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 paper 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 simulation studies and a data application, we demonstrate that our LR test can not only properly control type I error 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 paper is organized as follows: Section 2 introduces the linear structural equation model. Section 3 concerns the development of likelihood ratio test, including the Lagrange Multiplier and the asymptotic null distributions for the LR test statistic. Section 4 presents an implementation of the LR test. Section 5 shows the numerical performance of the LR test in terms of type I error rate and power, and comparisons to existing methods. Section 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 7 concludes the paper with discussions on both advantages and limitations of the proposed LR method.

Detailed technical derivations and proofs are included in the Appendix.

2. Framework

2.1 Structure Equation Model

Consider a data set of n observations, $(X_i, M_{i,j}, Y_i)$, $i = 1, \dots, n$, 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 paper, 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:

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

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

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 counterfac-

2.2 Unconstrained Parameter Estimation

tual 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 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 perform both unconstrained and constrained maximum likelihood estimations (MLE) under the null and alternative hypotheses. SEM (2.1) may be rewritten as a matrix form:

$$\mathbf{Y} = \mathbf{W}\bar{\boldsymbol{\beta}} + \boldsymbol{\epsilon}, \quad \mathbf{M} = \mathbf{B}\bar{\boldsymbol{\alpha}} + \mathbf{E}, \quad (2.2)$$

where $\bar{\boldsymbol{\beta}} = (\beta_1, \dots, \beta_Q, \eta_1, \dots, \eta_L, \gamma)^\top$, $\mathbf{Y}$ 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 $\mathbf{W}_i = (M_{i,1}, \dots, M_{i,Q}, Z_{i,1}, \dots, Z_{i,L}, X_i)^\top$, $i = 1, \dots, n$, 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 $\mathbf{B}_i = (X_i, Z_{i,1}, \dots, Z_{i,L})$, and $\mathbf{E} = (\mathbf{E}_1^\top, \dots, \mathbf{E}_n^\top)^\top$ with $\mathbf{E}_i \sim MVN(\mathbf{0}, \boldsymbol{\Sigma}_M)$. Here $\bar{\boldsymbol{\alpha}}$ is an $(L+1) \times Q$

2.3 Constrained Parameter Estimation

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

$$\begin{aligned} -2\ell(\boldsymbol{\theta}) = & n \log(\sigma_Y^2) + n \log(|\boldsymbol{\Sigma}_M|) + \sigma_Y^{-2}(\mathbf{Y} - \mathbf{W}\bar{\boldsymbol{\beta}})^\top (\mathbf{Y} - \mathbf{W}\bar{\boldsymbol{\beta}}) \\ & + \text{tr}\{(\mathbf{M} - \mathbf{B}\bar{\boldsymbol{\alpha}})\boldsymbol{\Sigma}_M^{-1}(\mathbf{M} - \mathbf{B}\bar{\boldsymbol{\alpha}})^\top\}. \end{aligned} \quad (2.3)$$

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

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

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 ,

$$g(\bar{\boldsymbol{\alpha}}, \bar{\boldsymbol{\beta}}, \sigma_Y^2, \boldsymbol{\Sigma}_M, \lambda) = -2\ell(\boldsymbol{\theta}) - 2\lambda \boldsymbol{\alpha}^\top \boldsymbol{\beta}. \quad (2.4)$$

2.3 Constrained Parameter Estimation10

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

$$\bar{\alpha} = (\mathbf{B}^\top \mathbf{B})^{-1} \mathbf{B}^\top \mathbf{M} + \lambda (\mathbf{B}^\top \mathbf{B})^{-1} \boldsymbol{\beta}^* \boldsymbol{\Sigma}_M = \hat{\alpha} + \lambda (\mathbf{B}^\top \mathbf{B})^{-1} \boldsymbol{\beta}^* \boldsymbol{\Sigma}_M, \quad (2.5)$$

$$\bar{\beta} = (\mathbf{W}^\top \mathbf{W})^{-1} \mathbf{W}^\top \mathbf{Y} + \lambda \sigma_Y^2 (\mathbf{W}^\top \mathbf{W})^{-1} \boldsymbol{\alpha}^* = \hat{\beta} + \lambda \sigma_Y^2 (\mathbf{W}^\top \mathbf{W})^{-1} \boldsymbol{\alpha}^*, \quad (2.6)$$

and the equations of variance parameters,

$$\sigma_y^2 = (\mathbf{Y} - \mathbf{W}\bar{\beta})^\top (\mathbf{Y} - \mathbf{W}\bar{\beta})/n, \text{ and } \boldsymbol{\Sigma}_M = (\mathbf{M} - \mathbf{B}\bar{\alpha})^\top (\mathbf{M} - \mathbf{B}\bar{\alpha})/n, \quad (2.7)$$

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 $\bar{\alpha}$, we denote the first row of $\hat{\alpha}$ 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 $\bar{\beta}$, denote the first Q rows of vector $\hat{\beta}$ 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 (2.8), and we shall choose the one that yields the higher log-likelihood,

$$\tilde{\lambda} = \frac{-(\mathbf{a}_1^\top \mathbf{b}_2 + \mathbf{b}_1^\top \mathbf{a}_2) \pm \sqrt{(\mathbf{a}_1^\top \mathbf{b}_2 + \mathbf{b}_1^\top \mathbf{a}_2)^2 - 4\mathbf{b}_1^\top \mathbf{b}_2 \mathbf{a}_1^\top \mathbf{a}_2}}{2\mathbf{b}_1^\top \mathbf{b}_2}. \quad (2.8)$$

Remark 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 (2.4). 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 $(\tilde{\boldsymbol{\theta}}, \tilde{\lambda})$.

3. Likelihood Ratio Test for Joint Mediation Effect

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 $\alpha_q \beta_q \neq 0, q = 1, \dots, Q$ 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:

$$T_n = -2 \left\{ \sup_{\boldsymbol{\theta} \in \boldsymbol{\Theta}: \boldsymbol{\alpha}^\top \boldsymbol{\beta} = 0} \ell(\boldsymbol{\theta}) - \sup_{\boldsymbol{\theta} \in \boldsymbol{\Theta}} \ell(\boldsymbol{\theta}) \right\} = -2 \{ \ell(\tilde{\boldsymbol{\theta}}) - \ell(\hat{\boldsymbol{\theta}}) \}, \quad (3.1)$$

3.2 Properties of the LR test

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 and 2.3.

3.2 Properties of the LR test

This section concerns the asymptotic distributions of the likelihood ratio statistic T_n in (3.1) 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

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

where

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

3.2 Properties of the LR test

The information matrix $\mathbf{I}(\boldsymbol{\theta}) = -\mathbb{E} \left(\frac{1}{n} \frac{\partial^2 \ell(\boldsymbol{\theta})}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^\top} \right)$ 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 1. *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 $h_1 \geq h_2 \geq \dots \geq h_{2Q}$, then $h_1 = \dots = h_Q = 1$, $h_{Q+1} = \dots = h_{2Q} = -1$.*
- (ii) *The matrix $\mathbf{A}(\boldsymbol{\theta})$ has $2Q$ nonzero eigenvalues. If nonzero eigenvalues are arranged in a descending order as of the form $v_1 \geq v_2 \geq \dots \geq v_Q > 0 > v_{Q+1} \geq \dots \geq v_{2Q}$, then they satisfy $\sum_{i=1}^{2Q} v_i = 0$, and $v_1 = -v_{2Q}, v_2 = -v_{2Q-1}, \dots, v_Q = -v_{Q+1}$.*

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 1 is presented in Appendix A.2.

Lemma 2. *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*

3.2 Properties of the LR test

$\tilde{\boldsymbol{\theta}}$ and $\tilde{\lambda}$ are given by, as $n \rightarrow \infty$,

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

with $\xi_q \stackrel{i.i.d.}{\sim} \chi_1^2$, $q = 1, \dots, 2Q$.

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

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

where $v_1, \dots, v_Q$ are Q positive eigenvalues of $\mathbf{A}(\boldsymbol{\theta}_0)$.

Lemma 2 leads to an asymptotic joint distribution of $\tilde{\boldsymbol{\theta}}$ and $\tilde{\lambda}$ due to the fact $[\tilde{\boldsymbol{\theta}}, \tilde{\lambda}] = [\tilde{\boldsymbol{\theta}} \mid \tilde{\lambda}][\tilde{\lambda}]$. 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 2 is presented in Appendix A.3.

Theorem 1. 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 $n \rightarrow \infty$, $T_n \xrightarrow{d} \chi_1^2$,

(ii) when $\boldsymbol{\alpha} = \boldsymbol{\beta} = \mathbf{0}$, as $n \rightarrow \infty$, $T_n \xrightarrow{d} \Lambda_1$ with $\Lambda_1 \triangleq \frac{\{\sum_{q=1}^Q v_q(\xi_q - \xi_{q+Q})\}^2}{4 \sum_{q=1}^Q v_q^2(\xi_q + \xi_{q+Q})}$,

where $\xi_q \stackrel{i.i.d.}{\sim} \chi_1^2$, $q = 1, \dots, 2Q$.

In this paper, we write $\Lambda_1 \sim \kappa_Q$ distribution. The proof of Theorem 1 involves deriving the asymptotic distributions of the constrained MLE.

3.2 Properties of the LR test

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 1, 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 2. The proof of Theorem 1 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_1^2 distributed variables $\xi_q, q = 1, \dots, 2Q$. We conduct a simulation study to confirm the validity of our theoretical derivations for Theorem 1 (ii). Our numerical assessment focuses on the tail probability of the distribution of the test statistic of T_n when $\boldsymbol{\alpha} = \boldsymbol{\beta} = \mathbf{0}$. See more details in the supplementary materials S1.

It follows from Theorem 1 that we propose a test for $H_0 : \boldsymbol{\alpha}^\top \boldsymbol{\beta} = 0$, termed as LR test, given by the decision function:

$$\phi_n = I[T_n > (\chi_{1,(1-\alpha)}^2 \vee \kappa_{Q,(1-\alpha)})], \quad (3.2)$$

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 1, and $\chi_{1,(1-\alpha)}^2$ is the $(1 - \alpha)$ quantile of the χ_1^2 distribution. When $\phi_n = 1$, we reject the null H_0 ; otherwise, accept the null H_0 . Section S5 in the supplementary material illustrates via a

simulation study that $\chi_{1,(1-\alpha)}^2$ overwhelmingly dominates $\kappa_{Q,(1-\alpha)}$, and such dominance can reach 100% with large sample sizes.

Theorem 2. *The LR test in (3.2) controls the type I error; that is*

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

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

Proof. Divide the parameter space under the H_0 , $\Theta = \{(\alpha, \beta) : \alpha^\top \beta = 0\}$ into two disjoint sub-spaces: $\Theta_1 = \{(0, 0)\}$ and $\Theta_2 = \Theta \setminus \Theta_1$. Then,

$$\begin{aligned} & \sup_{\theta \in \Theta: \alpha^\top \beta = 0} P_\theta(\phi_n = 1) \\ &= \sup_{\theta \in \Theta_1 \cup \Theta_2} P_\theta(T_n > \chi_{1,(1-\alpha)}^2 \vee \kappa_{Q,(1-\alpha)}) \\ &= \max\left\{ \sup_{\theta \in \Theta_1} P_\theta(T_n > \chi_{1,(1-\alpha)}^2 \vee \kappa_{Q,(1-\alpha)}), \sup_{\theta \in \Theta_2} P_\theta(T_n > \chi_{1,(1-\alpha)}^2 \vee \kappa_{Q,(1-\alpha)}) \right\} \\ &\leq \max\left\{ \sup_{\theta \in \Theta_1} P_\theta(T_n > \kappa_{Q,(1-\alpha)}), \sup_{\theta \in \Theta_2} P_\theta(T_n > \chi_{1,(1-\alpha)}^2) \right\} \\ &\leq \alpha. \end{aligned}$$

□

4. Implementation

In practice, to perform the LR test ϕ_n , we first compute two p -values of $p_1 = 1 - F_{\chi_1^2}(T_n)$ and $p_2 = 1 - F_{\kappa_Q}(T_n)$, where $F_{\chi_1^2}$ is the CDF of the χ_1^2 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: $\boldsymbol{\theta}_1 = \{\tilde{\boldsymbol{\alpha}}, \tilde{\boldsymbol{\beta}}\}$, and $\boldsymbol{\theta}_2 = \{\tilde{\sigma}_Y^2, \tilde{\boldsymbol{\Sigma}}_M\}$, as well as λ . The unconstrained MLE $\hat{\boldsymbol{\theta}} = \{\hat{\boldsymbol{\alpha}}, \hat{\boldsymbol{\beta}}, \hat{\sigma}_Y^2, \hat{\boldsymbol{\Sigma}}_M\}$ 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.

5. Simulation Studies

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; Huang et al., 2018). In addition, we consider a comparison to a recent method of High-Dimensional Multiple Testing (HDMT) proposed by Dai et al. (2020). HDMT was 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. We present

Algorithm 1 Search for constrained MLE

- 1: Compute the unconstrained MLE $\hat{\boldsymbol{\theta}} = \{\hat{\boldsymbol{\alpha}}, \hat{\boldsymbol{\beta}}, \hat{\sigma}_Y^2, \hat{\boldsymbol{\Sigma}}_M\}$, and evaluate the log-likelihood $\ell(\hat{\boldsymbol{\theta}})$. At the j th-iteration, let $\boldsymbol{\theta}_1^{(j)} = \{\tilde{\boldsymbol{\alpha}}^{(j)}, \tilde{\boldsymbol{\beta}}^{(j)}\}$, and let $\boldsymbol{\theta}_2^{(j)} = \{\tilde{\sigma}_Y^{2(j)}, \tilde{\boldsymbol{\Sigma}}_M^{(j)}\}$. Set $\boldsymbol{\theta}_1^{(0)} = \{\hat{\boldsymbol{\alpha}}, \hat{\boldsymbol{\beta}}\}$ and $\boldsymbol{\theta}_2^{(0)} = \{\hat{\sigma}_Y^2, \hat{\boldsymbol{\Sigma}}_M\}$ as the initial values.
 - 2: **for** $j = 0, 1, \dots, J$ **do**
 - 3: calculate $\lambda^{(j)} = \underset{\lambda}{\operatorname{argmax}}\{\ell(\boldsymbol{\theta}_1^{(j)}, \boldsymbol{\theta}_2^{(j)}, \lambda)\}$ from (2.8);
 - 4: calculate $\boldsymbol{\theta}_1^{(j+1)} = \underset{\boldsymbol{\theta}_1}{\operatorname{argmax}}\{\ell(\boldsymbol{\theta}_1, \boldsymbol{\theta}_2^{(j)}, \lambda^{(j)})\}$ from (2.5) and (2.6);
 - 5: calculate $\boldsymbol{\theta}_2^{(j+1)}$ from $\boldsymbol{\theta}_1^{(j+1)}$ based on (2.7);
 - 6: calculate $\delta = \|\boldsymbol{\theta}_1^{(j+1)} - \boldsymbol{\theta}_1^{(j)}\|$;
 - 7: **if** $|\delta| < tol$ **then** break
 - 8: **end if**
 - 9: **end for**
 - 10: Output $\tilde{\boldsymbol{\theta}} = \{\tilde{\boldsymbol{\alpha}}^{(j+1)}, \tilde{\boldsymbol{\beta}}^{(j+1)}, \tilde{\sigma}_Y^{2(j+1)}, \tilde{\boldsymbol{\Sigma}}_M^{(j+1)}\}$, and calculate the log-likelihood.
 - 11: 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 .
 - 12: 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 .
 - 13: Report $\max(p_1, p_2)$ as the final p -value.
-

some comparison results involving the HDMT method in the supplementary materials S3 (Tables S2-S5).

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 $N(0, I_2)$. Conditional on 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 $Q = 30$ or $Q = 60$, $\gamma = -2$, $\boldsymbol{\eta} = (2, -3, 2)^\top$, $\sigma_Y^2 = 1$. Here $\text{vec}(\boldsymbol{\zeta})$ consists of 18 repeated sequences of $(-2, 3, -3, 1, 1)$ for $Q = 30$, and 36 repeated sequences for $Q = 60$. A compound symmetry correlation with $\rho = 0.5$ is set for mediators. The sample size n varies over 200, 500, and 1000. For each sample size, we run 10,000 replicates. To evaluate the influence of Q and/or ρ on the performance of the LR test, we conduct additional simulations with $Q = 90$ and $\rho = 0, 0.25, 0.75$, and related results are summarized in Tables S1, S2, S3 and S5 in the supplementary materials.

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 $\boldsymbol{\alpha} =$

$\beta = \mathbf{0}$. Here sparsity refers to the number of zero parameters in α and/or β . For $Q = 30$, the detailed specifications of α and β can be found in Table 1; for $Q = 60$, the same patterns are repeated. We report in Table 2 the estimated empirical type I error rate as the proportion of rejections from the 10,000 replicates. For $Q = 30$ and 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. In the cases of small n (200) and $Q = 60$, the type I error of the LR test becomes slightly inflated. This is not surprising because a larger number of mediators implies a more complex model with more parameters, and thus a larger sample size is needed.

5.3 Power Comparison

We evaluate and compare power under the same basic model specifications above, in which α and β are specified in four sets of alternative scenarios different from the null hypothesis; see the detail in Table 1 for $Q = 30$. The design for the four alternative hypotheses corresponds the following scenarios of pathways: (v) both α and β are sparse; (vi) α is sparse and

[illegible]

5.3 Power Comparison22

Table 2: Empirical type I error under four null hypotheses, and power under four alternative hypotheses with 10,000 replicates for $Q = 30$ and $Q = 60$. The sample size n varies from 200, 500, and 1,000. The compound symmetry correlation of mediators is set with correlation 0.5. Power increase (%)

$$= \frac{\text{power of LR test}}{\text{power of competing test}} - 1.$$

Q	n	Method	Null Hypothesis				Alternative Hypothesis				Percent of power increase			
			i	ii	iii	iv	v	vi	vii	viii	v	vi	vii	viii
30	200	LR	0.048	0.052	0.051	0.010	0.603	0.561	0.306	0.512	-	-	-	-
		PT-N	0.036	0.043	0.037	0.006	0.557	0.536	0.252	0.463	8.33%	4.67%	21.53%	10.56%
		PT-NP	0.029	0.039	0.029	0.001	0.517	0.496	0.239	0.430	16.76%	13.08%	28.14%	19.13%
	500	LR	0.045	0.046	0.045	0.007	0.970	0.959	0.654	0.931	-	-	-	-
		PT-N	0.038	0.043	0.040	0.005	0.967	0.957	0.631	0.925	0.31%	0.17%	3.73%	0.64%
		PT-NP	0.036	0.043	0.036	0.001	0.963	0.951	0.627	0.918	0.74%	0.81%	4.32%	1.41%
	1000	LR	0.049	0.047	0.048	0.008	1.000	1.000	0.923	0.999	-	-	-	-
		PT-N	0.046	0.045	0.046	0.005	1.000	1.000	0.917	0.998	0.00%	0.00%	0.57%	0.04%
		PT-NP	0.045	0.046	0.044	0.001	1.000	1.000	0.916	0.999	0.01%	0.01%	0.76%	0.03%
60	200	LR	0.062	0.056	0.066	0.022	0.469	0.408	0.303	0.433	-	-	-	-
		PT-N	0.039	0.041	0.041	0.010	0.389	0.360	0.203	0.344	20.52%	13.27%	49.48%	25.87%
		PT-NP	0.035	0.038	0.033	0.004	0.327	0.305	0.187	0.293	43.29%	33.58%	61.95%	47.73%
	500	LR	0.051	0.055	0.054	0.010	0.932	0.880	0.630	0.897	-	-	-	-
		PT-N	0.043	0.049	0.043	0.007	0.923	0.875	0.583	0.884	0.88%	0.62%	8.17%	1.51%
		PT-NP	0.040	0.048	0.038	0.001	0.910	0.856	0.571	0.867	2.41%	2.78%	10.39%	3.44%
	1000	LR	0.053	0.052	0.050	0.008	0.999	0.995	0.912	0.998	-	-	-	-
		PT-N	0.050	0.050	0.045	0.006	0.999	0.995	0.902	0.998	0.02%	0.01%	1.09%	0.03%
		PT-NP	0.050	0.051	0.044	0.001	0.999	0.995	0.897	0.997	0.03%	-0.01%	1.60%	0.10%

5.3 Power Comparison²³

β is not sparse; (vii) α is not sparse and β is sparse; and (viii) both α and β are not sparse. Regardless of specific settings, the overall absolute group-level effect is fixed at 0.16, i.e. $|\alpha^\top \beta| = 0.16$. For $Q = 60$, we repeat the same patterns for α and β with a fixed size 0.16 in that scaling by $\sqrt{2}$ is applied on the parameters. Table 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 the underlying parameter configurations and sparsity. The power also decreases as Q increases in each setting of alternative hypothesis as individual signal strengths decrease by a factor of $1/\sqrt{2}$. Among these four cases, case (vii) appears to be the most challenging scenario, where β is most sparse with a small magnitude of nonzero element. To further examine the performance of these tests, in case (vii) with the sample size 200 and $Q = 30$, we set the single nonzero β coefficient at $0.2 + \delta$ with δ varying from 0 to 0.5 by an increment of 0.02 to illustrate the power increase pattern. Figure 2 shows all three power

5.3 Power Comparison²⁴

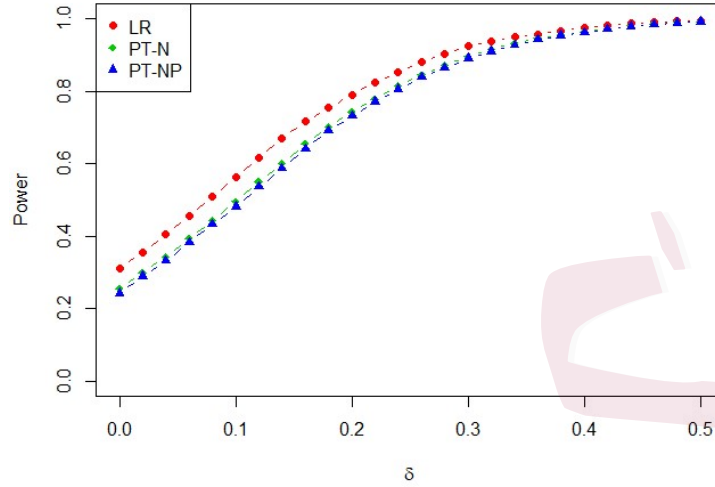


Figure 2: Power curves of three tests under the simulation case (vii).

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. In sum, these simulation results indicate that our LR test exhibits higher power than two existing PT-N and PT-NP tests, especially in the cases of small and moderate sample sizes.

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 LaBarre et al. (2020), 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 that all passed related data QC screening and 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. 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. First, 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 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-acetyglycine contributes most to the reduction of the insulin resistance score. In contrast, carbohydrate intakes increase the insulin resistance through metabolites, where again N-acetyglycine contributes most. In closing, we inspect some data quality issues such as truncation pattern due to limit of detection (LOD) and normality assumption using QQ plots of the residuals from the respective regressions of mediator-exposure (i.e. fat and carbohydrate). There are no truncation patterns on the lower part of the distributions, and all distributions look approximately normal. Refer to all details in the supplementary materials S4.

Table 3: Estimated coefficients for a cluster of seven metabolites.

Metabolite	Fat			Carbohydrate		
	α	β	$\alpha \circ \beta$	α	β	$\alpha \circ \beta$
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

7. Concluding Remarks

This paper 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. The associated computational cost is negligible, on average 0.15 seconds for a dataset of sample size 1000. The R package “MedLRT” implementing the LR test method is available at <https://github.com/haowei72/MedLRT>. 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.

This paper did not attempt to develop a solution to differentiate different null parameter configurations arising from the composite null hypothesis. The LR test approach attempted to take a step in the direction of solving this conservatism problem through the kappa distribution, which

was found as the limiting distribution of the Wilks' generalized likelihood ratio statistic for the null case of $\alpha = \beta = \mathbf{0}$, being different from the chi-square distribution under the other null cases. This technical contribution serves an important technical preparation; because once a new method enables us to differentiate null parameter configurations, the respective limiting distributions of the LR test statistic are ready to be applied to achieve the optimal solution (i.e. a desirable size alpha LR test).

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. In addition, to deal with the issue of dimensionality as well as complex patterns arising from the simultaneous testing setup (e.g. 243 possible null parameter configurations for $Q = 5$), alternative solutions such as an extension of Dai et al. (2020)'s approach is worth an exploration.

All test methods, including our LR test, have appeared to be conservative for the null case of $\alpha = \beta = \mathbf{0}$. This is an open problem in the theory of statistical inference for mediation effect, even in the setting of one single

mediator. The technical difficulty pertains to the presence of multiple null parameter configurations, each giving rise to a specific distribution for the test statistic, but the lack of knowledge which null configuration is the truth hinders us from obtaining a desirable size α in the type I error control. The κ_Q distribution is proposed to improve the overly conservative type I error control in that the κ_Q distribution has some chance to be selected in the decision making. However, as shown in the simulation study, under the case of $\alpha = \beta = \mathbf{0}$ this improvement is moderate and the type I error rate is still below 0.05. Some better solutions to overcome such conservatism are worth future exploration.

A. Appendix

A.1 Information Matrix

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

where

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

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

A.2 Proof of Lemma 1

First, we prove the part (i) of Lemma 1. Recall that

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

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

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 $\text{tr}(\mathbf{H}(\boldsymbol{\theta})) = 0$, implying $h_1 = \dots = h_Q = 1$, $h_{Q+1} = \dots = h_{2Q} = -1$.

Now we prove part (ii) of Lemma 1. From 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 since the eigenvalues of $\mathbf{I}(\boldsymbol{\theta})^{-\frac{1}{2}}$ are all positive. Thus, the $2Q$ nonzero eigenvalues of $\mathbf{A}(\boldsymbol{\theta})$, $v_1 \geq v_2 \geq \dots \geq v_Q > 0 > v_{Q+1} \geq \dots \geq v_{2Q}$. Let $\mathbf{I}_{11} = \frac{1}{n}\boldsymbol{\Sigma}_M^{-1} \otimes \mathbf{B}^{\top}\mathbf{B}$ and $\mathbf{I}_{22} = \frac{1}{n\sigma_y^2}\mathbf{E}(\mathbf{W}^{\top}\mathbf{W})$. Writing $\mathbf{I}(\boldsymbol{\theta}) = \text{Block-diag}(\mathbf{I}_{11}, \mathbf{I}_{22})$, we have

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

A.3 Proof of Lemma 232

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 $\mathbf{A}^2(\boldsymbol{\theta}) = \text{Block-diag}(\mathbf{C}\mathbf{C}^\top, \mathbf{C}^\top \mathbf{C})$. 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 non-zero eigenvalues of $\mathbf{C}\mathbf{C}^\top$ and $\mathbf{C}^\top \mathbf{C}$ are the same. This indicates $v_1^2 = v_{2Q}^2, v_2^2 = v_{2Q-1}^2, \dots, v_Q^2 = v_{Q+1}^2$. In summary, $\mathbf{A}(\boldsymbol{\theta})$ has $2Q$ nonzero eigenvalues in a descending order $v_1 \geq v_2 \geq \dots \geq v_Q > 0 > v_{Q+1} \geq \dots \geq v_{2Q}$, satisfying $\sum_{i=1}^{2Q} v_i = tr(\mathbf{A}(\boldsymbol{\theta})) = 0$. This implies that $v_1 = -v_{2Q}, v_2 = -v_{2Q-1}, \dots, v_Q = -v_{Q+1}$.

A.3 Proof of Lemma 2

Let $\mathbf{D} = \{\mathbf{Y}, \mathbf{W}, \mathbf{M}, \mathbf{B}\} = \{\mathbf{d}_i\}_{i=1}^n$ denote all observations 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 $p = LQ + L + 1$. 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, $\frac{1}{\sqrt{n}} \mathbf{u}(\boldsymbol{\theta}_0) \xrightarrow{d} N\{0, \mathbf{I}(\boldsymbol{\theta}_0)\}$. 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 (2.4). Then, they satisfy the following two equations:

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

It is easy 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 the 2nd equation of (A.1) around $\boldsymbol{\theta}_0$, $h(\tilde{\boldsymbol{\theta}}) = h(\boldsymbol{\theta}_0) + \dot{h}(\boldsymbol{\theta}_0)^\top (\tilde{\boldsymbol{\theta}} - \boldsymbol{\theta}_0) + \frac{1}{2} (\tilde{\boldsymbol{\theta}} - \boldsymbol{\theta}_0)^\top \mathbf{H}(\boldsymbol{\theta}_0) (\tilde{\boldsymbol{\theta}} - \boldsymbol{\theta}_0)$.

A.3 Proof of Lemma 233

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

$$(\tilde{\boldsymbol{\theta}} - \boldsymbol{\theta}_0)^\top \mathbf{H}(\boldsymbol{\theta}_0)(\tilde{\boldsymbol{\theta}} - \boldsymbol{\theta}_0) = 0. \quad (\text{A.2})$$

Similarly, taking the Taylor expansion of the first equation 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

$$\begin{aligned} (\tilde{\boldsymbol{\theta}} - \boldsymbol{\theta}_0) &\approx -\{\mathbf{U}(\boldsymbol{\theta}_0) + n\tilde{\lambda}\mathbf{H}(\boldsymbol{\theta}_0)\}^{-1}\mathbf{u}(\boldsymbol{\theta}_0), \\ \sqrt{n}(\tilde{\boldsymbol{\theta}} - \boldsymbol{\theta}_0) &\approx \frac{1}{\sqrt{n}}\{-\mathbf{U}(\boldsymbol{\theta}_0)/n - \tilde{\lambda}\mathbf{H}(\boldsymbol{\theta}_0)\}^{-1}\mathbf{u}(\boldsymbol{\theta}_0) \\ &\approx \{\mathbf{I}(\boldsymbol{\theta}_0) - \tilde{\lambda}\mathbf{H}(\boldsymbol{\theta}_0)\}^{-1}\frac{\mathbf{u}(\boldsymbol{\theta}_0)}{\sqrt{n}}. \end{aligned} \quad (\text{A.3})$$

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

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

By plugging (A.3) into (A.2), we define

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

A.3 Proof of Lemma 234

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

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

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 $\frac{\mathbf{u}(\boldsymbol{\theta}_0)}{\sqrt{n}} \xrightarrow{d} N\{0, \mathbf{I}(\boldsymbol{\theta}_0)\}$, by Slutsky's Theorem, $\left\{ -\frac{\mathbf{U}(\boldsymbol{\theta}_0)}{n} \right\}^{-\frac{1}{2}} \frac{\mathbf{u}(\boldsymbol{\theta}_0)}{\sqrt{n}} \xrightarrow{d} N\{0, \mathbf{I}\}$. 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 $n \rightarrow \infty$,

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

with $\xi_q \stackrel{i.i.d.}{\sim} \chi_1^2$, $q = 1, \dots, 2Q$.

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

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

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

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

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

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

A.4 Proof of Theorem 1

When $\alpha = \beta = \mathbf{0}$, taking the Taylor expansion on $\{\mathbf{I}(\theta_0) - \tilde{\lambda}\mathbf{H}(\theta_0)\}^{-1}$ around a small neighborhood of $\tilde{\lambda} = 0$, we have, subject to a high order error term, $\{\mathbf{I}(\theta_0) - \tilde{\lambda}\mathbf{H}(\theta_0)\}^{-1} \approx \{\mathbf{I}(\theta_0)\}^{-1} + \tilde{\lambda}\mathbf{I}(\theta_0)^{-1}\mathbf{H}(\theta_0)\mathbf{I}(\theta_0)^{-1}$. It follows that

$$\begin{aligned} \sqrt{n}(\tilde{\theta} - \theta_0) &\approx \left[\{\mathbf{I}(\theta_0)\}^{-1} + \tilde{\lambda}\mathbf{I}(\theta_0)^{-1}\mathbf{H}(\theta_0)\mathbf{I}(\theta_0)^{-1} \right] \frac{\mathbf{u}(\theta_0)}{\sqrt{n}} \\ &= \{\mathbf{I}(\theta_0)\}^{-1} \frac{1}{\sqrt{n}} \mathbf{u}(\theta_0) + \tilde{\lambda}\mathbf{I}(\theta_0)^{-1}\mathbf{H}(\theta_0)\mathbf{I}(\theta_0)^{-1} \frac{\mathbf{u}(\theta_0)}{\sqrt{n}} \\ &\approx \sqrt{n}(\hat{\theta} - \theta_0) + \tilde{\lambda}\mathbf{I}(\theta_0)^{-1}\mathbf{H}(\theta_0)\mathbf{I}(\theta_0)^{-1} \frac{\mathbf{u}(\theta_0)}{\sqrt{n}}. \end{aligned}$$

Noting that $\sqrt{n}(\tilde{\theta} - \hat{\theta}) = \tilde{\lambda}\mathbf{I}(\theta_0)^{-1}\mathbf{H}(\theta_0)\mathbf{I}(\theta_0)^{-1} \frac{\mathbf{u}(\theta_0)}{\sqrt{n}}$, we have

$$\begin{aligned} T_n &= -2\{\ell(\tilde{\theta}) - \ell(\hat{\theta})\} \\ &\approx \sqrt{n}(\tilde{\theta} - \hat{\theta})^\top \left\{ -\frac{\mathbf{U}(\theta_0)}{n} \right\}^{-1} \sqrt{n}(\tilde{\theta} - \hat{\theta}) \\ &\approx \tilde{\lambda}^2 \frac{\mathbf{u}(\theta_0)^\top}{\sqrt{n}} \mathbf{I}(\theta_0)^{-1} \mathbf{H}(\theta_0) \mathbf{I}(\theta_0)^{-1} \mathbf{H}(\theta_0) \mathbf{I}(\theta_0)^{-1} \frac{\mathbf{u}(\theta_0)}{\sqrt{n}} \\ &= \tilde{\lambda}^2 \frac{\mathbf{u}(\theta_0)^\top}{\sqrt{n}} \mathbf{I}(\theta_0)^{-\frac{1}{2}} \mathbf{A}(\theta_0)^2 \mathbf{I}(\theta_0)^{-\frac{1}{2}} \frac{\mathbf{u}(\theta_0)}{\sqrt{n}}. \end{aligned} \tag{A.4}$$

Note $\tilde{\lambda} \xrightarrow{d} \Lambda_0$, where $\Lambda_0 \stackrel{d}{=} -\frac{\sum_{q=1}^Q v_q(\xi_q - \xi_{q+Q})}{2 \sum_{q=1}^Q v_q^2(\xi_q + \xi_{q+Q})}$, and $\frac{\mathbf{u}(\boldsymbol{\theta}_0)^\top}{\sqrt{n}} \mathbf{I}(\boldsymbol{\theta}_0)^{-\frac{1}{2}} \xrightarrow{d} \mathbf{N}(\mathbf{0}, \mathbf{I}_{2Q+p})$.

Hence,

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

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

Acknowledgements

This research work was supported by grants NSF DMS1811734, DMS2113564 and NIH R24ES028502. The authors thank AE and three anonymous reviewers for constructive comments that led to a significant improvement of the paper. They also thank Drs. Jennifer LaBarre and Karen Peterson for constructive discussions on the metabolomics data application.

References

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

REFERENCES37

- Bollen, K. A. and R. Stine (1990). Direct and indirect effects: Classical and bootstrap estimates of variability. *Sociological methodology*, 115–140.
- 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.
- Huang, Y.-T. (2019). Variance component tests of multivariate mediation effects under composite null hypotheses. *Biometrics* 75(4), 1191–1204.
- 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.
- 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.
- LaBarre, J. L., K. E. Peterson, M. T. Kachman, W. Perng, L. Tang, W. Hao, L. Zhou, A. Karnovsky, A. Cantoral, M. M. Téllez-Rojo, et al.

REFERENCES₃₈

- (2020). Mitochondrial nutrient utilization underlying the association between metabolites and insulin resistance in adolescents. *The Journal of Clinical Endocrinology & Metabolism* 105(7), dgaa260.
- 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- α 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.
- Neyman, J. and E. S. Pearson (1933). On the problem of the most efficient tests of statistical hypotheses. *Philosophical Transactions of the Royal Society of London. Series A, Containing Papers of a Mathematical or Physical Character* 231(694-706), 289–337.

REFERENCES³⁹

- Pearl, J. (2001). Direct and indirect effects. In *Proceedings of the Seventeenth conference on Uncertainty in artificial intelligence*, pp. 411–420.
- 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.
- Robins, J. M. and S. Greenland (1992). Identifiability and exchangeability for direct and indirect effects. *Epidemiology* 3(2), 143–155.
- Rubin, D. B. (1974). Estimating causal effects of treatments in randomized and nonrandomized studies. *Journal of educational Psychology* 66(5), 688.
- Sobel, M. E. (1982). Asymptotic confidence intervals for indirect effects in structural equation models. *Sociological methodology* 13, 290–312.
- 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. and S. Vansteelandt (2009). Conceptual issues concerning mediation, interventions and composition. *Statistics and its Interface* 2(4), 457–468.

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.

Department of Biostatistics, School of Public Health, University of Michigan, MI, USA

E-mail: (weihao@umich.edu)

Department of Biostatistics, School of Public Health, University of Michigan, MI, USA

E-mail: (pxsong@umich.edu)