Statistica Sinica **33** (2023), 2305-2326
doi:https://doi.org/10.5705/ss.202021.0032

A SIMULTANEOUS LIKELIHOOD TEST FOR JOINT MEDIATION EFFECTS OF MULTIPLE MEDIATORS

Wei Hao and Peter X. K. Song

University of Michigan

Abstract: Mediation analysis using 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, a statistical inference on the joint mediation effect is challenging because of the composite null hypotheses with a large number of parameter configurations. We propose a simultaneous likelihood ratio test that uses a block coordinate descent algorithm 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 using extensive simulations and a comparison with existing tests. The simulation results show that our method controls the type I error properly and, in general, provides better power than that of 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 analyses provide a popular way of understanding whether or not the effect of an exposure on an outcome is mediated by some intermediate variables, called mediators. The mediation analysis approach, first proposed by Baron and Kenny (1986), has been applied extensively in many disciplines to perform pathway analyses. Using the counterfactual outcome framework in the causal inference literature (Rubin (1974); Robins and Greenland (1992); Pearl (2001)), the mediation approach has been extended to study causal mediation pathways using directed acyclic graphs (DAG) formed under a certain scientific hypothesis, as shown in Figure 1. With a few extra assumptions of causation, such an extension allows us to decompose the total causal effect into the sum of a direct effect and an indirect effect in the presence of interactions and nonlinearities (Pearl

Corresponding author: Peter X. K. Song, Department of Biostatistics, University of Michigan, Ann Arbor, MI 48109, USA. E-mail: pxsong@umich.edu.

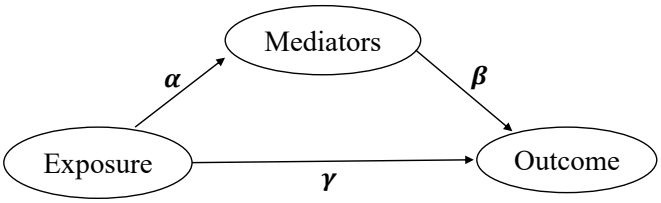

Figure 1. A DAG involving exposure, mediators, and an outcome.

(2001); VanderWeele and Vansteelandt (2009)). This new causal framework has received much attention in the literature.

Many methods have been developed to test the existence of a mediation effect (or the indirect effect) in the case of a single potential mediator, including Sobel's test (Sobel (1982)), the bootstrap method (Bollen and Stine (1990)), and the joint significant test (MacKinnon et al. (2002)). Recently, with the growing availability of omics data, testing for mediation effects has received much attention, especially for handling a group of multiple or even high-dimensional mediators. These methods include the multiple testing approaches for genome-wide association analyses proposed for simultaneous single mediator tests with a multiple comparison correction (Huang (2019b,a); Dai, Stanford and LeBlanc (2020); Liu et al. (2022)).

In such methods, the test for a causal mediation effect focuses on a single mediator, performing a univariate screening analysis of the mediators, while ignoring the dependence among multiple mediators. Although multiple testing corrections have been applied to identify potential mediators, the interpretation of the causal effect is still limited to each of the selected mediators, rather than being a simultaneous inference for the group-level mediation effect. However, when multiple correlated mediators exist, particularly a cohesive cluster of biologically relevant mediators, the group-level mediation effect is not simply a summation of the individual mediation effects, as pointed out by VanderWeele (2015). Therefore, a conclusion drawn from a univariate screening test with multiple comparison corrections does not necessarily produce a valid statistical inference for the group-level mediation effect. Although these univariate screening procedures are useful for discovering individual potential mediators, it is important to analyze a cluster of correlated multiple mediators jointly, which requires a test for their group-level mediation effect.

The mediation relationships specified by a DAG, shown in Figure 1, have been analyzed extensively using the linear normal structural equation model (SEM). When exposure-mediator interaction terms are absent from 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 the exposure-mediator association, and $\boldsymbol{\beta}$ is the vector of coefficients for the mediator-outcome association. We 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 when 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, for $q = 1, \ldots, Q$, where Q is a fixed number of mediators. One 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 individual products of $\alpha_q\beta_q$, for $q = 1, \ldots, Q$, to satisfy $\boldsymbol{\alpha}^\top\boldsymbol{\beta} = 0$. Two existing approaches to testing this group-level mediation effect include the product test based on the normal product distribution (PT-NP) (Huang and Pan (2016); Huang (2018)) and the product test based on normality (PT-N) (Huang and Pan (2016); Huang (2018)). Although these two methods have shown satisfactory numerical performance in simulation studies, few works provide rigorous theoretical justifications, such as the results of asymptotic distributions of such test statistics under the null, especially for the case $\boldsymbol{\alpha} = \boldsymbol{\beta} = \mathbf{0}$. With the fundamental Neyman–Pearson lemma, the likelihood ratio (LR) test is known as the uniformly most powerful test for a simple hypothesis testing problem under mild regularity conditions (Neyman and Pearson (1933)), and Wilks' generalized LR test is one of the top finite-sample performers in the literature. To bridge this gap, we investigate a simultaneous LR test for the joint group-level mediation effect under the null hypothesis $\boldsymbol{\alpha}^\top\boldsymbol{\beta} = 0$. We also establish the asymptotic distributions of the proposed test statistic and confirm the theoretical results using numerical analyses.

This study makes two methodological contributions. First, we develop a constrained optimization to compute the LR 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. Using numerical experiments, including simulation studies and a data application, we demonstrate that our LR test not only properly controls the type I error, but also improves the power over that of two existing tests, PT-NP and PT-N.

The remainder of the paper is organized as follows: Section 2 introduces the linear SEM. In Section 3, we develop the LR 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 the type I error rate and power, and compares it with that of existing methods. In Section 6, we test for a group-level mediation effect of a metabolite cluster on the association between dietary intake and insulin resistance. Section 7 concludes the paper. Detailed technical derivations and proofs are included in the Appendix.

2. Framework

2.1. SEM

Consider a data set of n observations, (X_i, M_i, Y_i), for $i = 1, \ldots, n$, randomly sampled from n subjects. For the ith 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. We consider the case when both Q and L are fixed and $Q + L + 1 < n$. A linear SEM takes the following form:

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

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

Denote the collection of distinct model parameters by $\boldsymbol{\theta} = \{\boldsymbol{\alpha}, \boldsymbol{\beta}, \gamma, \boldsymbol{\eta}, \boldsymbol{\zeta}, \boldsymbol{\Sigma}_M, \sigma_Y^2\}$, and let $\boldsymbol{\Theta}$ be the generic notation for a parameter space. In the counterfactual outcome paradigm (Robins and Greenland (1992); Pearl (2001)), under the fundamental assumptions of consistency and the absence of unmeasured confounders, VanderWeele and the colleague (VanderWeele and Vansteelandt (2014)) show that when the 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 the LR 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 (MLEs) under the null and alternative hypotheses. SEM

(2.1) may be rewritten in matrix form, as follows:

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

where $\bar{\boldsymbol{\beta}} = (\beta_1, \ldots, \beta_Q, \eta_1, \ldots, \eta_L, \gamma)^\top$, $\mathbf{Y}$ is an $n \times 1$ vector of outcomes, $\mathbf{W}$ is an $n \times (Q + L + 1)$ matrix of mediators, confounders, and the exposure variable, with $\mathbf{W}_i = (M_{i,1}, \ldots, M_{i,Q}, Z_{i,1}, \ldots, Z_{i,L}, X_i)^\top$, for $i = 1, \ldots, n$, and $\boldsymbol{\epsilon} \sim MVN(\mathbf{0}, \sigma_Y^2 \mathbf{I}_n)$. Similarly, $\mathbf{M}$ is an $n \times Q$ matrix of mediators, $\mathbf{B}$ is an $n \times (L+1)$ matrix of exposure and confounding variables with $\mathbf{B}_i = (X_i, Z_{i,1}, \ldots, Z_{i,L})$, and $\mathbf{E} = (\mathbf{E}_1^\top, \ldots, \mathbf{E}_n^\top)^\top$, with $\mathbf{E}_i \sim MVN(\mathbf{0}, \boldsymbol{\Sigma}_M)$. Here, $\bar{\boldsymbol{\alpha}}$ is an $(L+1) \times Q$ matrix of parameters, with the first row vector being $\boldsymbol{\alpha}^\top$ in model (2.1), and its remaining $L \times Q$ submatrix is the parameter matrix of $\boldsymbol{\zeta}$. It follows that the two times negative log-likelihood function is given by

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

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

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

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 we obtain using the Lagrange multiplier. Consider a Lagrange objective function of the following form, with tuning parameter $\lambda \geq 0$ ::

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

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

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

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

and the equations of the variance parameters

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

where $\boldsymbol{\beta}^*$ is an $(L+1) \times Q$ matrix with the first row being $\boldsymbol{\beta}^\top$ and the rest of the elements zeros, and $\boldsymbol{\alpha}^*$ is a $(Q+L+1) \times 1$ vector with the first Q elements being $\boldsymbol{\alpha}$ and the rest equal to zero. Given that $\boldsymbol{\alpha}^\top$ appears in the first row of $\bar{\boldsymbol{\alpha}}$, we denote the first row of $\hat{\bar{\boldsymbol{\alpha}}}$ by $\mathbf{a}_1^\top$, and the first row of $(\mathbf{B}^\top\mathbf{B})^{-1}\boldsymbol{\beta}^*\boldsymbol{\Sigma}_M$ by $\mathbf{b}_1^\top$. It follows that $\boldsymbol{\alpha}^\top = \mathbf{a}_1^\top + \lambda \mathbf{b}_1^\top$. Similarly, given that $\boldsymbol{\beta}$ is in the first Q rows of vector $\bar{\boldsymbol{\beta}}$, denote the first Q rows of vector $\hat{\bar{\boldsymbol{\beta}}}$ by $\mathbf{a}_2$, and the first Q rows of $(\mathbf{W}^\top\mathbf{W})^{-1}\boldsymbol{\alpha}^*$ by $\mathbf{b}_2$. Under the constraint $\boldsymbol{\alpha}^\top\boldsymbol{\beta} = 0$, we obtain $(\mathbf{a}_1^\top + \lambda\mathbf{b}_1^\top)(\mathbf{a}_2 + \lambda\mathbf{b}_2) = 0$. This leads to two possible solutions of λ given in (2.8), and we choose the one that yields the higher log-likelihood,

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

Remark 1. After we obtain the constrained MLE solutions $(\tilde{\boldsymbol{\theta}}, \tilde{\lambda})$ using the Lagrange multiplier, we evaluate the Hessian matrix of the function $g(\cdot)$ in (2.4). It is easy to show that in a 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. LR Test for the 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). Because the null hypothesis allows internal cancellation, it does not preclude the possibility of component-wise nonzero mediation effects in the sense that $\alpha_q\beta_q \neq 0$, for $q = 1, \ldots, Q$, but $\boldsymbol{\alpha}^\top\boldsymbol{\beta} = 0$. Following Wilks' theory of the LR test, we construct an LR test statistic of the form

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

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 examines the asymptotic distributions of the LR 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 properly controls the type I error with theoretical guarantees. Technical proofs for all lemmas and theorems presented in this section are given in the Appendix. We begin with some notation. For ease of exposition, we redefine $\boldsymbol{\theta} = (\boldsymbol{\alpha}^\top, \boldsymbol{\zeta}, \boldsymbol{\beta}^\top, \boldsymbol{\eta}^\top, \gamma)^\top$, where $\boldsymbol{\zeta}$ denotes the row vector of LQ elements vectorized from the matrix $\boldsymbol{\zeta}_{L\times Q}$. Define the constraint function by $h(\boldsymbol{\theta}) = \boldsymbol{\alpha}^\top \boldsymbol{\beta}$. It is easy to see that its gradient $\dot{h}(\boldsymbol{\theta}) = \nabla_{\boldsymbol{\theta}} h(\boldsymbol{\theta}) = (\boldsymbol{\beta}^\top, \mathbf{0}_{LQ}^\top, \boldsymbol{\alpha}^\top, \mathbf{0}_{L+1}^\top)^\top$. Let

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

where

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

The information matrix $\mathbf{I}(\boldsymbol{\theta}) = -\mathrm{E}[(1/n)\{\partial^2 \ell(\boldsymbol{\theta})/\partial\boldsymbol{\theta}\boldsymbol{\theta}^\top\}]$ has a closed-form expression, presented in Appendix A.1. Let $\mathbf{A}(\boldsymbol{\theta}) = \mathbf{I}(\boldsymbol{\theta})^{-1/2}\mathbf{H}(\boldsymbol{\theta})\mathbf{I}(\boldsymbol{\theta})^{-1/2}$. To derive the asymptotic properties, we first introduce a lemma that establishes the eigenvalue bounds of the 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 the nonzero eigenvalues are arranged in descending order $h_1 \geq h_2 \geq \cdots \geq h_{2Q}$, then $h_1 = \cdots = h_Q = 1$, $h_{Q+1} = \cdots = h_{2Q} = -1$.

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

The above properties for the eigenvalues of $\mathbf{A}(\boldsymbol{\theta})$ are used to establish the 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. The asymptotic distributions of the constrained MLE $\tilde{\boldsymbol{\theta}}$ and $\tilde{\lambda}$ are given by, as $n \to \infty$,*

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

with $\xi_q \overset{i.i.d.}{\sim} \chi_1^2, q = 1, \ldots, 2Q$ *where* $\upsilon_1, \ldots, \upsilon_Q$ *are* Q *positive eigenvalues of* $\mathbf{A}(\boldsymbol{\theta}_0)$. *For any* $\lambda^* \in \mathbb{R}$, *conditional on a value* $\tilde{\lambda} = \lambda^*$,

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

Lemma 2 leads to an asymptotic joint distribution of $\tilde{\boldsymbol{\theta}}$ and $\tilde{\lambda}$ because $[\tilde{\boldsymbol{\theta}}, \tilde{\lambda}] = [\tilde{\boldsymbol{\theta}}|\tilde{\lambda}][\tilde{\lambda}]$. Thus, we obtain the asymptotic distribution of the LR test statistic for $\boldsymbol{\alpha} = \boldsymbol{\beta} = \mathbf{0}$. 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 LR test statistic* T_n *are given as follows:*

(i) when $(\boldsymbol{\alpha}^\top, \boldsymbol{\beta}^\top)^\top \neq \mathbf{0}$, *as* $n \to \infty$, $T_n \xrightarrow{d} \chi_1^2$;

(ii) when $\boldsymbol{\alpha} = \boldsymbol{\beta} = \mathbf{0}$, *as* $n \to \infty$, $T_n \xrightarrow{d} \Lambda_1$ *with* $\Lambda_1 \overset{d}{\equiv} \{\sum_{q=1}^{Q} \upsilon_q(\xi_q - \xi_{q+Q})\}^2 / \{(4\sum_{q=1}^{Q} \upsilon_q^2(\xi_q + \xi_{q+Q})\}$, *where* $\xi_q \overset{i.i.d.}{\sim} \chi_1^2, q = 1, \ldots, 2Q$.

In this paper, we write $\Lambda_1 \sim \kappa_Q$ distribution. The proof of Theorem 1 involves deriving the asymptotic distribution of the constrained MLE. The classical large-sample work for the LR test, for example, Aitchison and Silvey (1958); Wolak (1989), may be applied directly to prove part (i) of Theorem 1. However, the proof of part (ii) is nontrivial and needs specific technical arguments and treatments to manipulate the 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 estimating both the matrix $\mathbf{A}(\boldsymbol{\theta})$ and its Q eigenvalues, we invoke a Monte Carlo simulation with a large number of draws (say 10,000) independently from $2Q$ χ_1^2 distributed variables ξ_q, for $q = 1, \ldots, 2Q$. 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 the Supplementary Material S1.

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

$$\phi_n = I\{T_n > (\chi^2_{1,(1-\alpha)} \vee \kappa_{Q,(1-\alpha)})\}, \tag{3.2}$$

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 $\chi^2_{1,(1-\alpha)}$ is the $(1-\alpha)$ quantile of the χ_1^2 dis-

tribution. When $\phi_n = 1$, we reject the null H_0; otherwise, we accept the null H_0. Section S2 in the Supplementary Material uses a simulation study to demonstrate that $\chi^2_{1,(1-\alpha)}$ overwhelmingly dominates $\kappa_{Q,(1-\alpha)}$, and that such dominance can reach 100% with large sample sizes.

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

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

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

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

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

4. Implementation

In practice, to perform the LR test ϕ_n, we first compute the two p-values $p_1 = 1 - F_{\chi^2_1}(T_n)$ and $p_2 = 1 - F_{\kappa_Q}(T_n)$, where $F_{\chi^2_1}$ is the CDF of the χ^2_1 distribution, and F_{κ_Q} is the CDF of the κ_Q distribution. Then, we reject the null hypothesis if $\max(p_1, p_2)$ is smaller than the significance level α.

To obtain the constrained MLE, we develop a block coordinate descent algorithm. We partition $\boldsymbol{\theta}$ into two sets, $\boldsymbol{\theta}_1 = \{\tilde{\boldsymbol{\alpha}}, \tilde{\boldsymbol{\beta}}\}$ and $\boldsymbol{\theta}_2 = \{\tilde{\sigma}^2_Y, \tilde{\boldsymbol{\Sigma}}_M\}$, as well as λ. The unconstrained MLE $\hat{\boldsymbol{\theta}} = \{\hat{\boldsymbol{\alpha}}, \hat{\boldsymbol{\beta}}, \hat{\sigma}^2_Y, \hat{\boldsymbol{\Sigma}}_M\}$ provides the initial values used 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; and given $\boldsymbol{\theta}_1$, update $\boldsymbol{\theta}_2$. See Algorithm 1, where the default number of Monte Carlo simulations is set at 10,000.

Algorithm 1 Search for constrained MLE

Compute the unconstrained MLE $\hat{\boldsymbol{\theta}} = \{\hat{\tilde{\boldsymbol{\alpha}}}, \hat{\tilde{\boldsymbol{\beta}}}, \hat{\sigma}_Y^2, \hat{\boldsymbol{\Sigma}}_M\}$ and evaluate the log-likelihood $\ell(\hat{\boldsymbol{\theta}})$. At the jth iteration, let $\boldsymbol{\theta}_1^{(j)} = \{\tilde{\boldsymbol{\alpha}}^{(j)}, \tilde{\tilde{\boldsymbol{\beta}}}^{(j)}\}$ and $\boldsymbol{\theta}_2^{(j)} = \{\tilde{\sigma}_Y^{2(j)}, \tilde{\boldsymbol{\Sigma}}_M^{(j)}\}$. Set $\boldsymbol{\theta}_1^{(0)} = \{\hat{\tilde{\boldsymbol{\alpha}}}, \hat{\tilde{\boldsymbol{\beta}}}\}$ and $\boldsymbol{\theta}_2^{(0)} = \{\hat{\sigma}_Y^2, \hat{\boldsymbol{\Sigma}}_M\}$ as the initial values.
for $j = 0, 1, \ldots, J$ **do**
  calculate $\lambda^{(j)} = \text{argmax}_\lambda\{\ell(\boldsymbol{\theta}_1^{(j)}, \boldsymbol{\theta}_2^{(j)}, \lambda)\}$ from (2.8);
  calculate $\boldsymbol{\theta}_1^{(j+1)} = \text{argmax}_{\boldsymbol{\theta}_1}\{\ell(\boldsymbol{\theta}_1, \boldsymbol{\theta}_2^{(j)}, \lambda^{(j)})\}$ from (2.5) and (2.6);
  calculate $\boldsymbol{\theta}_2^{(j+1)}$ from $\boldsymbol{\theta}_1^{(j+1)}$ based on (2.7);
  calculate $\delta = \| \boldsymbol{\theta}_1^{(j+1)} - \boldsymbol{\theta}_1^{(j)} \|$;
  if $|\delta| < tol$ **then** break
  end if
end for
Output $\tilde{\boldsymbol{\theta}} = \{\tilde{\tilde{\boldsymbol{\alpha}}}^{(j+1)}, \tilde{\tilde{\boldsymbol{\beta}}}^{(j+1)}, \tilde{\sigma}_Y^{2(j+1)}, \tilde{\boldsymbol{\Sigma}}_M^{(j+1)}\}$, and calculate the log-likelihood.
Calculate the test statistic $T = -2\{\ell(\tilde{\boldsymbol{\theta}}) - \ell(\hat{\boldsymbol{\theta}})\}$, and compute the p-value p_1 under the null distribution of χ_1^2.
Estimate $\mathbf{A}(\boldsymbol{\theta}_0)$ based on $\hat{\sigma}_Y^2$ and $\hat{\boldsymbol{\Sigma}}_M$, and calculate its Q positive eigenvalues, which are then used to simulate the κ_Q distribution, and compute its p-value p_2.
Report $\max(p_1, p_2)$ as the final p-value.

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 those of two existing methods, namely, the PT-N and PT-NP tests proposed by (Huang and Pan (2016); Huang (2018, 2019a)), as well as with the high-dimensional multiple-testing (HDMT) method proposed by Dai, Stanford and LeBlanc (2020). The HDMT method was developed for the univariate screening of mediators with a controlled false discovery rate in genome studies, representing a typical kind of testing approach widely adopted in practice to avoid simultaneous inference. We present comparison results involving the HDMT method in the Supplementary Material S4 (Tables S3–S6).

The SEM is set up as follows. The exposure variable X is simulated from $N(0, 1)$, and the two confounding variables Z_1 and Z_2 are generated from $BVN(\mathbf{0}, I_2)$. Conditional on X and (Z_1, Z_2), in all simulation experiments in this section, the Q mediators M and outcome Y are generated according to the SEM (2.1), with $Q = 30$ or Q=60, $\gamma = -2$, $\boldsymbol{\eta} = (2, -3, 2)^\top$, and $\sigma_Y^2 = 1$. Here, 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 the matrix $\mathbf{\Sigma}_M$ of Q 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 $\rho = 0, 0.25, 0.75$; the results are summarized in the Supplementary Material S3 and S4.

5.2. Type I error

We consider the following four scenarios for the null hypotheses: (i) sparse pathways with no cancellation; (ii) sparse pathways with cancellation; (iii) non-sparse pathways with cancellation; and (iv) fully sparse pathways $\boldsymbol{\alpha} = \boldsymbol{\beta} = \mathbf{0}$. Here, sparsity refers to the number of zero parameters in $\boldsymbol{\alpha}$ and/or $\boldsymbol{\beta}$. For $Q = 30$, detailed specifications of $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ 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 and the PT-N and PT-NP tests show proper control of the type I error. In cases (i)–(iii), the three methods show empirical type I error rates close to the nominal level 0.05, as desired. In case (iv), they are all conservative, but our LR test appears to be the least conservative of the three. For 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 the power under the same basic model specifications above, in which $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ are specified in four sets of alternative scenarios different from the null hypothesis; see Table 1 for $Q = 30$. The designs for the four alternative hypotheses correspond to the following pathway scenarios: (v) both $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ are sparse; (vi) $\boldsymbol{\alpha}$ is sparse and $\boldsymbol{\beta}$ is not sparse; (vii) $\boldsymbol{\alpha}$ is not sparse and $\boldsymbol{\beta}$ is sparse; and (viii) $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ are both not sparse. Regardless of the specific settings, the overall absolute group-level effect is fixed at 0.16, that is, $|\boldsymbol{\alpha}^\top \boldsymbol{\beta}| = 0.16$. For $Q = 60$, we repeat the same patterns for $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ with a fixed size 0.16, in that the parameters are scaled by $\sqrt{2}$. Table 2 reports the estimated empirical power by the proportion of rejections to the null from 10,000 replicates.

We calculate the percentage power increase of the LR method over that of a competing method as (power of LR/power of competitor) $-$ 1. For all cases, our LR method demonstrates clearly higher power than that of the PT-N and

Table 1. Designed specifications for $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ for the null and alternative hypotheses.

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

PT-NP tests, especially when the sample sizes are small or moderate, say 500 or less. Furthermore, even though the mediation effect size is fixed constantly at 0.16 across the four cases, the power varies with the underlying parameter configurations and sparsity. The power also decreases as Q increases in each setting of the alternative hypothesis as the individual signal strengths decrease by a factor of $1/\sqrt{2}$. Case (vii) appears to be the most challenging scenario, where $\boldsymbol{\beta}$ is most sparse with a small magnitude of nonzero elements. To further examine the performance of these tests, in case (vii) with a sample size 200 and

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 is equal to 200, 500, or 1,000. The compound symmetry correlation of the mediators is set at 0.5. Power increase (%) =(power of LR test/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%
	1,000	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%
	1,000	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%

$Q = 30$, we set the single nonzero $\boldsymbol{\beta}$ coefficient to $0.2 + \delta$, with δ varying from 0 to 0.5 by an increment of 0.02, to illustrate the power increase pattern. Figure 2 shows that all three power curves increase to one when the size δ in the alternative hypothesis becomes further distant from the null hypothesis. Our LR test is more powerful than the competing tests. Empirically, the three tests are all consistent, because their power rises to one when the deviation from the null tends to infinity. In summary, these simulation results indicate that our LR test exhibits higher power than that of the PT-N and PT-NP tests, especially for 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, called the carbohydrate intake and fat intake, respectively, obtained from the food frequency questionnaires (Willett, Howe and Kushi (1997)). The outcome variable Y is a HOMA-CP score, defined by LaBarre et al. (2020), that measures insulin

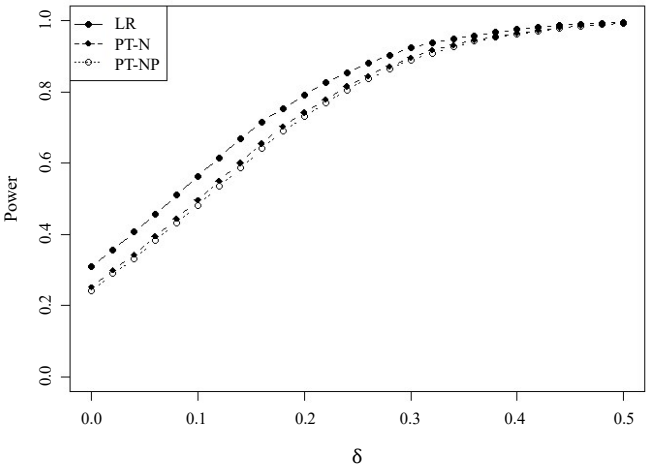

Figure 2. Power curves of three tests under the simulation case (vii).

resistance using the C-peptide biomarker produced by the pancreas. A higher HOMA-CP score means greater insulin resistance, leading to a potentially higher risk of developing diabetes in adulthood.

In this analysis, we study a cluster of seven metabolites of glucose metabolites and acetylamino acids that all passed data QC screening and were 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 has been found to be positively associated with dietary fiber intake (Lustgarten et al. (2014)) and negatively associated with the metabolic risk score (Perng et al. (2017)). The goal of interest is to test whether a cohesive cluster containing `N-acetylglycine` is involved as a group in a mediation pathway from dietary intake to the HOMA-CP score. This scientific question pertains to a hypothesis that food intake may change the metabolites and alter the function of the pancreas, thus elevating the risk of developing diabetes later in life.

In consultation with our collaborator, we chose a set of confounding variables, including age, gender, and puberty onset, and calculated the p-values for the null hypothesis $H_0 : \boldsymbol{\alpha}^T\boldsymbol{\beta} = \mathbf{0}$ with $Q = 7$ using the three methods LR, PT-N, and PT-NP. First, we test 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). With exposure of carbohydrate intake, we obtain the p-values 0.03 (LR), 0.04 (PT-N), and 0.04 (PT-N). For all three methods, with 95% confidence, this cluster of seven metabolites exhibits a significant group-level mediation effect on

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

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

the association between dietary intake and the HOMA-CP score. As expected, the LR test appears to have smaller p-values in both cases, consistent with the findings of the simulation studies.

Taking a closer look at 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 intake through the seven metabolites are -0.012 and 0.003, respectively. For fat intake, the negative mediation effect indicates that a greater fat intake helps to reduce insulin resistance through metabolites, where `N-acetyglicine` contributes most to the reduction of the insulin resistance score. In contrast, carbohydrate intake increases insulin resistance through metabolites, where again `N-acetyglicine` contributes most. In closing, we examine data quality issues, such as the truncation pattern due to the limit of detection and normality assumption, using QQ plots of the residuals from the respective regressions of mediator on exposure (i.e., fat and carbohydrate). There are no truncation patterns on the lower parts of the distributions, and all distributions look approximately normal. Refer to the Supplementary Materials S5.

7. Conclusion

We have provided an LR approach for testing a group-level mediation effect with multiple mediators. We were able to overcome a key technical challenge arising from the constrained MLE under irregular parameter spaces. In particular, we used the Lagrange multiplier method to carry out the constrained optimization using an efficient block coordinate decent algorithm required to implement our LR test statistic. The associated computational cost is negligible, on average 0.15 seconds for a data set of size 1,000. 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 proper control of the type I error. In both simulation studies and a data application, our LR method was less conservative and exhibited higher power than that of two existing methods, the PT-N test and the PT-NP test, especially when the sample size was moderate or small.

We have not attempted to develop a solution to differentiating different null parameter configurations arising from the composite null hypothesis. The LR test approach attempts to solve this conservatism problem using the kappa distribution, which was found to be the limiting distribution of Wilks' generalized LR statistic for the null case of $\boldsymbol{\alpha} = \boldsymbol{\beta} = \mathbf{0}$, being different from the chi-square distribution under the other null cases. This technical contribution serves as an important technical preparation, because once a new method enables us to differentiate null parameter configurations, we can apply the respective limiting distributions of the LR test statistic to achieve the optimal solution (i.e., a desirable size alpha LR test).

To apply our LR approach to test for a cluster of high-dimensional potential mediators, one needs to first divide them into subgroups based on existing scientific knowledge or clustering techniques, and then test for a group-level mediation effect, each for one subgroup of mediators. In future work, we would like to extend the current framework to the case of high-dimensional mediators with no need to divide them into subgroups. In addition, to deal with the dimensionality and complex patterns arising from the simultaneous testing setup (e.g., 243 possible null parameter configurations for $Q = 5$), alternative solutions are worth exploring, such as an extension of the approach of Dai, Stanford and LeBlanc (2020).

All test methods, including our LR test, appear to be conservative for the null case of $\boldsymbol{\alpha} = \boldsymbol{\beta} = \mathbf{0}$. This is an open problem in the theory of statistical inference for the mediation effect, even in the setting of one mediator. The technical difficulty pertains to the presence of multiple null parameter configurations, each giving rise to a specific distribution for the test statistic. However, the lack of knowledge about 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 of being selected. However, as shown in the simulation study, when $\boldsymbol{\alpha} = \boldsymbol{\beta} = \mathbf{0}$, this improvement is moderate, and the type I error rate is still below 0.05. Deriving better solutions that overcome this conservatism is left to future work.

Supplementary Material

The Supplementary Material includes additional simulation results in Sections S1-S4, and additional figures for data analysis in Section S5.

Acknowledgments

This research work was supported by grants NSF DMS1811734, DMS2113564, and NIH R24ES028502. The authors thank AE and the three anonymous reviewers for their constructive comments and suggestions. They also thank Drs. Jennifer LaBarre and Karen Peterson for their constructive discussions on the metabolomics data application.

A. Appendix

A.1. Information matrix

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

where

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

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

A.2. Proof of Lemma 1

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

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

where

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

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

Since $\mathbf{H}^2(\boldsymbol{\theta})$ is a diagonal matrix, and it has $2Q$ 1's and $(LQ + L + 1)$ 0's

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

Now we prove part (ii) of Lemma 1. From Theorem 1.4 in (Lu and Pearce (2000)), matrix $\mathbf{A}(\boldsymbol{\theta})=\mathbf{I}(\boldsymbol{\theta})^{-1/2}\mathbf{H}(\boldsymbol{\theta})\mathbf{I}(\boldsymbol{\theta})^{-1/2}$ has Q positive eigenvalues, Q negative eigenvalues and the rest eigenvalues are zero since the eigenvalues of $\mathbf{I}(\boldsymbol{\theta})^{-1/2}$ are all positive. Thus, the $2Q$ nonzero eigenvalues of $\mathbf{A}(\boldsymbol{\theta})$, $\upsilon_1\geq\upsilon_2\geq\cdots\geq\upsilon_Q>0>\upsilon_{Q+1}\geq\cdots\geq\upsilon_{2Q}$. Let $\mathbf{I}_{11}=(1/n)\boldsymbol{\Sigma}_M^{-1}\otimes\mathbf{B}^\top\mathbf{B}$ and $\mathbf{I}_{22}=(1/n\sigma_Y^2)\mathrm{E}(\mathbf{W}^\top\mathbf{W})$. Writing $\mathbf{I}(\boldsymbol{\theta})=\text{Block-diag}\left(\mathbf{I}_{11},\mathbf{I}_{22}\right)$, we have

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

Consequently, $tr(\mathbf{A}(\boldsymbol{\theta}))=0$, and $(\mathbf{I}_{11}^{-1/2}\tilde{\mathbf{H}}\mathbf{I}_{22}^{-1/2})^\top=\mathbf{I}_{22}^{-1/2}\tilde{\mathbf{H}}^\top\mathbf{I}_{11}^{-1/2}$. Let $\mathbf{I}_{11}^{-1/2}\tilde{\mathbf{H}}\mathbf{I}_{22}^{-1/2}=\mathbf{C}$. We have $\mathbf{A}^2(\boldsymbol{\theta})=\text{Block-diag}\left(\mathbf{C}\mathbf{C}^\top,\mathbf{C}^\top\mathbf{C}\right)$. The eigenvalues of $\mathbf{A}^2(\boldsymbol{\theta})$ are $\lambda(\mathbf{A}^2(\boldsymbol{\theta}))=(\lambda(\mathbf{C}\mathbf{C}^\top),\lambda(\mathbf{C}^\top\mathbf{C}))$, where the non-zero eigenvalues of $\mathbf{C}\mathbf{C}^\top$ and $\mathbf{C}^\top\mathbf{C}$ are the same. This indicates $\upsilon_1^2=\upsilon_{2Q}^2,\upsilon_2^2=\upsilon_{2Q-1}^2,\ldots,\upsilon_Q^2=\upsilon_{Q+1}^2$. In summary, $\mathbf{A}(\boldsymbol{\theta})$ has $2Q$ nonzero eigenvalues in a descending order $\upsilon_1\geq\upsilon_2\geq\cdots\geq\upsilon_Q>0>\upsilon_{Q+1}\geq\cdots\geq\upsilon_{2Q}$, satisfying $\sum_{i=1}^{2Q}\upsilon_i=tr(\mathbf{A}(\boldsymbol{\theta}))=0$. This implies that $\upsilon_1=-\upsilon_{2Q},\upsilon_2=-\upsilon_{2Q-1},\ldots,\upsilon_Q=-\upsilon_{Q+1}$.

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, $(1/\sqrt{n})\mathbf{u}(\boldsymbol{\theta}_0)\xrightarrow{d}\mathrm{N}\{0,\mathbf{I}(\boldsymbol{\theta}_0)\}$. Moreover, by the Law of Large Number, $(-1/n)\mathbf{U}(\boldsymbol{\theta}_0)\xrightarrow{p}\mathbf{I}(\boldsymbol{\theta}_0)$. Let $\{\tilde{\boldsymbol{\theta}},\tilde{\lambda}\}$ be the solution of the Lagrange multiplier equation (2.4). Then, they satisfy the following two equations:

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

It is easy to show that the k-th order ($k\geq3$) 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 $\boldsymbol{\theta}_0$, $h(\tilde{\boldsymbol{\theta}})=h(\boldsymbol{\theta}_0)+\dot{h}(\boldsymbol{\theta}_0)^\top(\tilde{\boldsymbol{\theta}}-\boldsymbol{\theta}_0)+(1/2)(\tilde{\boldsymbol{\theta}}-\boldsymbol{\theta}_0)^\top\mathbf{H}(\boldsymbol{\theta}_0)(\tilde{\boldsymbol{\theta}}-\boldsymbol{\theta}_0)$. Since $\boldsymbol{\alpha}=\boldsymbol{\beta}=\mathbf{0}$, $h(\boldsymbol{\theta}_0)=h(\tilde{\boldsymbol{\theta}})=0$ and $\dot{h}(\boldsymbol{\theta}_0)=\mathbf{0}_{2Q+p}$, then

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

A.4. Proof of Theorem 1

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

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

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

$$T_n = -2\{\ell(\tilde{\boldsymbol{\theta}}) - \ell(\hat{\boldsymbol{\theta}})\} \tag{A.4}$$

$$\approx \sqrt{n}(\tilde{\boldsymbol{\theta}} - \hat{\boldsymbol{\theta}})^\top \left\{ -\frac{\mathbf{U}(\boldsymbol{\theta}_0)}{n} \right\}^{-1} \sqrt{n}(\tilde{\boldsymbol{\theta}} - \hat{\boldsymbol{\theta}})$$

$$\approx \tilde{\lambda}^2 \frac{\mathbf{u}(\boldsymbol{\theta}_0)}{\sqrt{n}}^\top \mathbf{I}(\boldsymbol{\theta}_0)^{-1} \mathbf{H}(\boldsymbol{\theta}_0) \mathbf{I}(\boldsymbol{\theta}_0)^{-1} \mathbf{H}(\boldsymbol{\theta}_0) \mathbf{I}(\boldsymbol{\theta}_0)^{-1} \frac{\mathbf{u}(\boldsymbol{\theta}_0)}{\sqrt{n}}$$

$$= \tilde{\lambda}^2 \frac{\mathbf{u}(\boldsymbol{\theta}_0)}{\sqrt{n}}^\top \mathbf{I}(\boldsymbol{\theta}_0)^{-1/2} \mathbf{A}(\boldsymbol{\theta}_0)^2 \mathbf{I}(\boldsymbol{\theta}_0)^{-1/2} \frac{\mathbf{u}(\boldsymbol{\theta}_0)}{\sqrt{n}}.$$

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

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

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

References

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

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**, 1173–1182.

Bollen, K. A. and Stine, R. (1990). Direct and indirect effects: Classical and bootstrap estimates of variability. *Sociological Methodology* **20**, 115–140.

Dai, J. Y., Stanford, J. L. and LeBlanc, M. (2020). A multiple-testing procedure for high-dimensional mediation hypotheses. *Journal of the American Statistical Association* **117**, 198–213.

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

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

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

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.

LaBarre, J. L., Peterson, K. E., Kachman, M. T., Perng, W., Tang, L., Hao, W. et al. (2020). Mitochondrial nutrient utilization underlying the association between metabolites and insulin resistance in adolescents. *The Journal of Clinical Endocrinology & Metabolism* **105**, 2442–2455.

Liu, Z., Shen, J., Barfield, R., Schwartz, J., Baccarelli, A. A. and Lin, X. (2022). Large-scale hypothesis testing for causal mediation effects with applications in genome-wide epigenetic studies. *Journal of the American Statistical Association* **117**, 67–81.

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

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

MacKinnon, D. P., Lockwood, C. M., Hoffman, J. M., West, S. G. and Sheets, V. (2002). A comparison of methods to test mediation and other intervening variable effects. *Psychological Methods* **7**, 83–104.

Neyman, J. and Pearson, E. S. (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**, 289–337.

Pearl, J. (2001). Direct and indirect effects. In *Proceedings of the 17th Conference on Uncertainty in Artificial Intelligence*, 411–420.

Perng, W., Hector, E. C., Song, P. X. K., Tellez Rojo Solis, M. M., Raskind, S., Kachman, M. et al. (2017). Metabolomic determinants of metabolic risk in Mexican adolescents. *Obesity* **25**, 1594–1602.

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

Rubin, D. B. (1974). Estimating causal effects of treatments in randomized and nonrandomized studies. *Journal of educational Psychology* **66**, 688–701.

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

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

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

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

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

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

Wei Hao

Department of Biostatistics, School of Public Health, University of Michigan, Ann Arbor, MI 48109, USA.

E-mail: weihao@umich.edu

Peter X. K. Song

Department of Biostatistics, School of Public Health, University of Michigan, Ann Arbor, MI 48109, USA.

E-mail: pxsong@umich.edu

(Received January 2021; accepted January 2022)