Taylor & Francis
Taylor & Francis Group

Journal of the American Statistical Association

ISSN: (Print) (Online) Journal homepage: www.tandfonline.com/journals/uasa20

De-confounding Causal Inference Using Latent Multiple-Mediator Pathways

Yubai Yuan & Annie Qu

To cite this article: Yubai Yuan & Annie Qu (2024) De-confounding Causal Inference Using Latent Multiple-Mediator Pathways, Journal of the American Statistical Association, 119:547, 2051-2065, DOI: 10.1080/01621459.2023.2240461

To link to this article: https://doi.org/10.1080/01621459.2023.2240461

JOURNAL OF THE AMERICAN STATISTICAL ASSOCIATION
2024, VOL. 119, NO. 547, 2051–2065: Theory and Methods
https://doi.org/10.1080/01621459.2023.2240461

De-confounding Causal Inference Using Latent Multiple-Mediator Pathways

Yubai Yuan[a] and Annie Qu[b]

[a]Department of Statistics, The Pennsylvania State University, State College, PA; [b]Department of Statistics, University of California, Irvine, Irvine, CA

ABSTRACT

Causal effect estimation from observational data is one of the essential problems in causal inference. However, most estimation methods rely on the strong assumption that all confounders are observed, which is impractical and untestable in the real world. We develop a mediation analysis framework inferring the latent confounder for debiasing both direct and indirect causal effects. Specifically, we introduce generalized structural equation modeling that incorporates structured latent factors to improve the goodness-of-fit of the model to observed data, and deconfound the mediators and outcome simultaneously. One major advantage of the proposed framework is that it uses the causal pathway structure from cause to outcome via multiple mediators to debias the causal effect without requiring external information on latent confounders. In addition, the proposed framework is flexible in terms of integrating powerful nonparametric prediction algorithms while retaining interpretable mediation effects. In theory, we establish the identification of both causal and mediation effects based on the proposed deconfounding method. Numerical experiments on both simulation settings and a normative aging study indicate that the proposed approach reduces the estimation bias of both causal and mediation effects. Supplementary materials for this article are available online.

ARTICLE HISTORY
Received December 2021
Accepted July 2023

KEYWORDS
Causal identification;
Generalized additive model;
Latent factor modeling;
Mediation analysis;
Sequential ignorability

1. Introduction

Causal inference is one of the most essential tasks for much scientific research to infer whether certain predictors, that is, treatments have causal effects on outcomes. For complex studies, it is more critical to further identify the mechanism that explains *how* the treatment affects the outcome. One important direction is to investigate how the treatment and outcome relation is transmitted through intermediate variables. Specifically, causal mediation analysis identifies the causal mechanism by delineating pathways from treatment to outcome via mediators (Imai, Keele, and Tingley 2010). The basic paradigm of a classic mediation analysis is illustrated in Figure 1, which is widely used in psychological, sociological, epidemiological and biological studies (Hicks and Tingley 2011).

In various domain applications, the treatment-outcome mechanism is often complicated and might not be fully captured by a single-mediator model. For example, in study of educational of prevention strategies reducing students' drug addiction, the causal effect of education is explained by various mediators such as resistance skills, social norms, attitudes about drugs, and communication skills (MacKinnon 2012). Therefore, the multiple-mediator analysis is often more useful as the causal effects can be decomposed to a number of different mediation pathways, and provides a more accurate assessment and more meaningful interpretation of mediation effects (Shi and Li 2021; Cai, Song, and Lu 2020).

The main challenges in both causal inference and mediation analysis are the rise of confounders which could be intervene between treatments and outcomes. Specifically, confounders introduce noncausal associations between treatments and outcomes, therefore potentially inducing bias in causal-related inference. To solve the confounding issue, most of the existing causal inference methods assume that all confounders are observed. Under this assumption of unconfoundedness (Imbens and Rubin 2015), both the causal and mediation effects can be estimated unbiasedly via adjusting the observed data on confounders. For example, adjustment can be achieved via linear regression (Baron and Kenny 1986; Zhang et al. 2016a), random forest (Wager and Athey 2018), or other supervised learning methods (Künzel et al. 2019). However, the assumption of unconfoundedness might not be satisfied in practice, and could also be difficult to verify.

To relax the unconfoundedness assumption, extensive methods have been developed to allow causal identification given the existence of latent confounders. For example, Spirtes, Meek, and Richardson (2013), Wang, Franks, and Oh (2021), and Ogarrio, Spirtes, and Ramsey (2016) study conditions where the existence of causal paths from one variable to another can be identified without adjusting the latent confounding effects. However, these methods cannot remove estimation bias on causal effects (Spirtes, Meek, and Richardson 2013; Ogarrio, Spirtes, and Ramsey 2016), and further require the confounding effect is either significantly stronger or weaker than the causal effect

CONTACT Annie Qu aqu2@uci.edu Department of Statistics, University of California, Irvine, Irvine, CA
Supplementary materials for this article are available online. Please go to www.tandfonline.com/r/JASA.

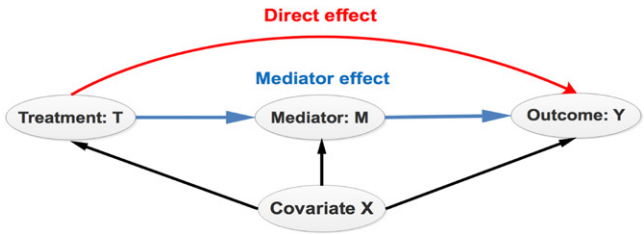

Figure 1. The causal mediation pathway with a single mediator and observed confounder X.

(Wang, Franks, and Oh 2021). Dukes, Shpitser, and Tchetgen (2021), Miao, Geng, and Tchetgen Tchetgen (2018), Louizos et al. (2017), and Tan (2006) focus on a series of proximal causal inference methods assuming that proxies of the underlying confounders are observed, and the proxy variables are used to reduce confounding bias in observational studies.

Recently, a new direction of deconfounding methodology has been developed to adjust the latent confounders via utilizing structural information from specific causal pathways and confounding effects. Specifically, Ranganath and Perotte (2018), Wang and Blei (2019) establish a two-stage deconfounding algorithm using the structure of multiple treatments sharing the same confounder. Similarly, Zhou, Kong, and Wang (2020) proposes to leverage the structure of multiple outcomes for identification and estimation of causal effects regardless of latent confounders. In addition, Witty et al. (2020) proposes to reduce the estimation bias of causal effects via incorporating the hierarchical structure of confounding effects where confounders are shared across subjects from the same subgroup. On the other hand, although mediation analysis is important and widely used in real applications, few papers discuss adjusting the latent confounding to reduce estimation bias.

To address this issue, we develop a mediation analysis framework which allows the identification of causal effects and mediation effects through the latent confounders. Compared with the treatment-outcome pathway considered in Ranganath and Perotte (2018), Wang and Blei (2019), and Zhou, Kong, and Wang (2020), adjusting latent confounders within causal mediation pathways could be more complex and challenging due to that the confounder-mediator relations introduce an additional layer of confounding between treatments and outcomes. On the other hand, the mediators themselves contain the latent confounders' information, and can be used in a principled way to reduce the biases in estimating causal effects and mediation effects. This motivates us to develop a new deconfounding framework. Specifically, we develop a confounder-sharing structure among multi-mediators, and use latent variables to aggregate the information of latent confounders from the shared variations in multiple mediators and outcomes. In addition, the confounder-sharing structure also allows the conditional independence among multiple mediators, which enables us to further estimate the confounder-relevant latent variable through a series of latent modelings. One key innovation is that the proposed method identifies causal and mediation effects without recovering true latent confounders. Instead, the deconfounding latent variables capture the confounder information, and also serve as surrogate confounders to adjust confounding effects.

One advantage of the proposed framework is that it does not require external proxies for latent confounders as in Louizos et al. (2017), Miao, Geng, and Tchetgen Tchetgen (2018), and Dukes, Shpitser, and Tchetgen (2021). In addition, our method is not restricted to any specific structure in latent confounders themselves, such as the hierarchical structure in Witty et al. (2020), and does not impose assumptions on the distribution of latent confounders. Furthermore, the proposed deconfounding algorithm can integrates various nonparametric estimators to infer underlying complex confounding effects on multiple mediators and outcomes, while still retaining the interpretability of the mediation effects. In theory, we show that the causal effects estimation is unbiased using the proposed surrogate confounders. We also establish the causality identification conditions for our method under different confounder-mediator structures of causal mediation pathways. Numerically, both simulations and the real data application indicate the effectiveness of the proposed method in reducing estimation biases of both causal effects and mediation effects.

This article is organized as follows: Section 2 introduces the background of the causal mediation analysis. Section 3 introduces the proposed deconfounding method. Section 4 provides an algorithm and implementation strategies. Section 5 establishes the theoretical properties of causal identification in the proposed method. Section 6 demonstrates simulation studies. Section 7 presents an application to a NIH normative aging study. The last section provides conclusions and some further discussion.

2. Background and Notation

Let $\boldsymbol{Y} = \{Y_i\}_{i=1}^N$ denote the set of observed outcomes from N subjects, where Y_i is a one-dimensional outcome from the ith subject. Denote $\boldsymbol{T} = \{T_i\}_{i=1}^N$ as the set of treatment assignments. We observe multiple mediators for each subject $\boldsymbol{M} = \{M_i^{(j)}\}$, $i = 1, \ldots, N, j = 1, \ldots, k$, and covariates $\boldsymbol{X} = \{X_i\}_{i=1}^N$ where $X_i \in R^p$, and k, p are the number of mediators and covariates, respectively. To formulate the causal mediation inference mathematically, we adopt the potential outcomes framework (Rubin 2005). Specifically, let $M_i^{(j)}(t)$ denote the potential value of the jth mediator for the ith subject when the subject takes treatment $t \in \{0, 1\}$. Similarly, we use $Y_i(t, m)$ to represent the potential outcome for the ith subject given that the subject takes treatment $T_i = t$ with mediators $M_i = (M_i^{(1)}, \ldots, M_i^{(k)})$ are $(m_1, \ldots, m_k)$. Notice that observed data can be denoted as $\{M_i(T_i), Y_i(T_i, M_i(T_i))\}$. Under the potential outcomes framework with binary treatment, the average total treatment effect can be defined as $\tau = \mathbb{E}\{Y_i(1, M_i(1)) - Y_i(0, M_i(0))\}$. Based on the total effect, we can specify the portion of the treatment effect through mediators, which is referred to as the average treatment mediation effect (ACME); that is, $\delta(t) = \mathbb{E}\{Y_i(t, M_i(1)) - Y_i(t, M_i(0))\}$. Accordingly, the direct treatment effects are defined as $\zeta(t) = \mathbb{E}\{Y_i(1, M_i(t)) - Y_i(0, M_i(t))\}$. It can be shown that the average total treatment effect $\tau = \delta(t) + \zeta(1 - t)$, for $t = 0, 1$.

One essential problem of the mediation analysis in Figure 1 is to directly estimate the total treatment effect and mediation effect from the observed data. One well-established

sufficient condition is to require that joint distribution of $\{Y(t,m), M^{(1)}(t), \ldots, M^{(k)}(t)\}$ be independent from the distribution of T conditioning on covariates X, and the distribution of $Y(t,m)$ be independent with the distribution of $\left\{M^{(1)}(t), \ldots, M^{(k)}(t)\right\}$ conditioning on (X, T) (Imai, Keele, and Yamamoto 2010). In other words, the above assumptions require that all the confounders of both direct and indirect associations between T and Y are observed, which can be stringent in real applications of mediation analysis. Therefore, the unmeasured confounder would introduce noncausal association among treatment, mediators, and outcome, leading to biased estimation for the causal treatment effect and mediation effect.

3. Methodology

In this section, we develop a de-confounder method for a broad class of mediation analysis featuring multiple mediators given the existence of latent confounders. Specifically, we consider the mediation causal pathway where the treatment can have both direct effect and indirect effects via the path of multiple mediators on the outcome, and latent confounders simultaneously affect treatment, outcome, and multiple mediators. The key structure is that the multiple mediators might not be causally dependent on each other. The causal mediation pathway is illustrated in Figure 2.

3.1. Joint Debiasing on Multiple-Mediator Pathway

In the following, we formulate the causal mediation pathway in Figure 2 via the potential outcome framework omitting the covariates X for ease of notation. The formulation can be generalized with covariates X. We assume that the pretreatment confounder $\boldsymbol{U}$ within the causal mediation pathway satisfies the following condition.

Sequential ignorability for multiple mediators:

$$\left\{Y\left(t', m\right), \boldsymbol{M}(t)\right\} \perp T \mid \boldsymbol{U} = u, \tag{1}$$

$$Y\left(t', m\right) \perp \boldsymbol{M}(t) \mid T = t, \boldsymbol{U} = u, \tag{2}$$

$$(\textbf{parallel mediators}): \; M^{(i)} \perp \boldsymbol{M}^{(-i)} | T = t, \boldsymbol{U} = u, \quad i \in \{1, \ldots, k\}, \; t \in \{0, 1\}, \tag{3}$$

where $\boldsymbol{M}^{(-i)}$ denotes the set of mediators excluding the ith mediator. The conditions (1) and (2), referred to as *sequential ignorability*, are standard assumptions in causal identification for mediation analysis (Imai, Keele, and Yamamoto 2010; Forastiere, Mattei, and Ding 2018). These two conditions guarantee the identification of mediation effects and direct effects (Imai, Keele, and Yamamoto 2010) via requiring that the joint distribution of mediators and outcome is conditional independent from the treatment assignment, and the distribution of outcomes is conditional independent from the distribution of mediators given the latent confounder $\boldsymbol{U}$. The conditions (1)–(2) together imply that $\boldsymbol{U}$ is the source of unobserved confounding between multiple mediators and outcomes.

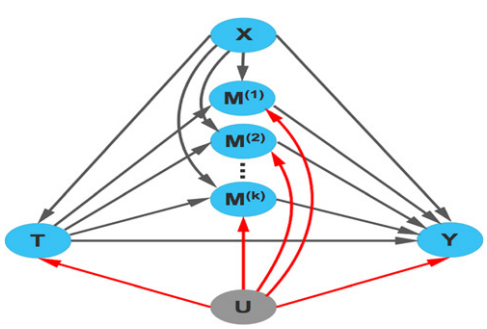

Figure 2. The causal mediation pathway with latent confounder U affecting both treatment T, outcome Y, and multiple mediators $\{M^{(j)}\}_{j=1}^{k}$. The multiple mediators are not causally dependent on each other. The observed data is colored in blue, and the latent confounder is colored in gray.

The condition (3) indicates a parallel-mediator structure in Figure 2 such that the mediators are independent to each other given the latent confounder $\boldsymbol{U}$. The corresponding causal mediation pathway generalizes a broad class of causal inference on observational data in many fields, especially biological and social sciences. For example, it is desirable to uncover how alcohol consumption affects blood pressure via mediators such as body mass index and various enzymes, where certain underlying genes might influence alcohol consumption, blood pressure, and mediators simultaneously (MacKinnon 2012). Additionally, it is also of scientific and social interest to investigate how primary prevention programs can reduce drug use via affecting socioeconomic mediators such as resistance skills and social norms. Confounders, such as students' personality, might also influence the causal mediation pathway.

Since $\boldsymbol{U}$ is not directly observed, the conditions (1)–(3) are not directly applicable for the identification of mediation effects and direct effects. However, the parallel mediator structure allows us to search a surrogate confounder $\hat{\boldsymbol{U}} = \{\hat{U}_i \in R^r\}_{i=1}^N$ where r is a dimension of the latent vector. With the surrogate confounder, the sequential ignorability of (1)–(2) can be approximately established conditioning on $\hat{\boldsymbol{U}}$ instead of the unobserved $\boldsymbol{U}$.

In the following, we introduce a latent factor model based on the surrogate confounder $\hat{\boldsymbol{U}}$ to fit the observed data. We denote $\boldsymbol{M} = (M^{(1)}, \ldots, M^{(k)})$ and the latent factor model as follows:

$$\hat{U}_i \in R^r \sim \mathcal{P}(\hat{\boldsymbol{U}}), \; T_i \sim \mathcal{P}(T|\hat{U}_i), \; M_i^{(j)} \sim \mathcal{P}(M^{(j)}|T_i, \hat{U}_i), \quad Y_i \sim \mathcal{P}(Y|T_i, \boldsymbol{M}_i, \hat{U}_i), \tag{4}$$

for $i = 1, \ldots, N$ and $j = 1, \ldots, k$. The individual latent factors $\{\hat{U}_i\}_{i=1}^N$ introduce the underlying confoundness across $(T, \boldsymbol{M}, Y)$, which cannot be adjusted from observed data. Assume the latent factor model captures the joint distribution of observed data $\mathcal{P}(T, \boldsymbol{M}, Y)$ in that

$$\mathcal{P}(T, \boldsymbol{M}, Y|\hat{\boldsymbol{U}}) = \mathcal{P}(T|\hat{\boldsymbol{U}}) \prod_{j=1}^{k} \mathcal{P}(M^{(j)}|T, \hat{\boldsymbol{U}}) \mathcal{P}(Y|T, \boldsymbol{M}, \hat{\boldsymbol{U}}). \tag{5}$$

Given that the true confounder $\boldsymbol{U}$ affects multiple mediators $M^{(1)}, \ldots, M^{(k)}$, the decomposition in (5) implies the sequential ignorability of (1) and (2) based on $\hat{\boldsymbol{U}}$:

$$\left\{Y\left(t', m\right), M^{(j)}(t)\right\} \perp T \mid \hat{\boldsymbol{U}} = u, \quad Y\left(t', m\right) \perp M^{(j)}(t) \mid T = t, \hat{\boldsymbol{U}} = u,$$

indicating that $\hat{\boldsymbol{U}}$ can serve as a surrogate confounder. The argument is based on a contradiction from the existence

of confoundness when (5) holds and sequential ignorability is conditioning on $\boldsymbol{U}$ in (1)–(3). Specifically, if there still exists a latent confounder $\boldsymbol{U}$ that simultaneously affects $(T, M^{(1)}(t), \ldots, M^{(k)}(t), Y(t,m))$ and is not captured by $\hat{\boldsymbol{U}}$, then the multiple mediators remain dependent even conditioning on $\hat{\boldsymbol{U}}$ and T due to the parallel mediator structure (3). Therefore, the decomposition fails in that $\mathcal{P}(\boldsymbol{M}|T,\hat{\boldsymbol{U}}) \neq \prod_{j=1}^{k} \mathcal{P}(M^{(j)}|T,\hat{\boldsymbol{U}})$, and leads to a contradiction in that (5) does not hold. In other words, if the latent factor model fits the observed data well and (5) holds, then we can use $\hat{\boldsymbol{U}}$ as the surrogate confounder that contains the information of true confounder $\boldsymbol{U}$.

Accordingly, we propose a deconfounding strategy through fitting the observed data of the causal mediation pathway (5). We first adopt the additive model to formulate the components $\mathcal{P}(T|\boldsymbol{U})$, $\mathcal{P}(M^{(j)}|T,\boldsymbol{U})$, and $\mathcal{P}(Y|T,\boldsymbol{M},\boldsymbol{U})$ incorporating covariate X as follows:

$$\begin{aligned} Y_i &= f_Y(T_i, M_i^{(1)}, \ldots, M_i^{(k)}, X_i) + g_Y(U_i) + \epsilon_{Y,i}, \\ M_i^{(j)} &= f_{M^{(j)}}(T_i, X_i) + g_{M^{(j)}}(U_i) + \epsilon_{M,i}^{(j)}, \\ T_i &= f_T(X_i) + g_T(U_i) + \epsilon_{T,i},\ i = 1, \ldots, N;\ j = 1, \ldots, k, \end{aligned} \tag{6}$$

where $f_T(\cdot) : R^p \to R$, $f_Y(\cdot) : R^{k+p+1} \to R$ and $f_{M^{(j)}}(\cdot) : R^{p+1} \to R$ are the output functions explaining the treatment assignment T, outcome Y and multiple mediators $\{M^{(j)}\}$, respectively. We concatenate the confounding functions $\boldsymbol{G}(\cdot) := \{g_T(\cdot), g_{M^{(1)}}(\cdot), \ldots, g_{M^{(k)}}(\cdot), g_Y(\cdot)\}$, which model the effects of individual confounder U_i on each element in $(T_i, M_i^{(1)}, \ldots, M_i^{(k)}, Y_i)$, and therefore $\boldsymbol{G}(\cdot) : R^r \to R^{k+2}$ encodes the confoundness patterns among the causal mediation pathway through the shared confounder $\boldsymbol{U}$. In addition, we assume that the permutations $(\epsilon_Y, \epsilon_M^{(1)}, \ldots, \epsilon_M^{(k)}, \epsilon_T)$ are random variables with zero-mean and unknown constant variances. Notice that the functional forms of $\{f_T(\cdot), f_Y(\cdot), f_{M^{(j)}}(\cdot), g_Y(\cdot), g_{M^{(j)}}(\cdot), g_T(\cdot)\}$ are not specified. In other words, we do not need to specify explicitly how the underlying true confounder U acts on treatment, mediators, and outcome via $\boldsymbol{G}(\cdot)$. Our method works when the model (6) provides a good approximation to the population distribution of observed data.

Based on (6), we can evaluate the fitness of the latent factor model (5) for the observed $(T, \boldsymbol{M}, Y)$ as follows:

$$\begin{aligned} &\sum_{i=1}^{N} \left\{T_i - f_T(X_i) - g_T(U_i)\right\}^2, \\ &\sum_{i=1}^{N} \left\{M_i^{(j)} - f_{M^{(j)}}(T_i, X_i) - g_{M^{(j)}}(U_i)\right\}^2, \\ &\sum_{i=1}^{N} \left\{Y_i - f_Y(T_i, \boldsymbol{M}_i, X_i) - g_Y(U_i)\right\}^2. \end{aligned} \tag{7}$$

Due to the additive modeling for $\mathcal{P}(T|\boldsymbol{U})$, $\{\mathcal{P}(M^{(j)}|T,\boldsymbol{U})\}_{j=1}^{k}$, and $\mathcal{P}(Y|T,\boldsymbol{M},\boldsymbol{U})$ in (6), the multiplicity in latent factor model (5) requires the independence as follows:

$$\epsilon_T \perp \epsilon_Y,\ \epsilon_T \perp \{\epsilon_M^{(j)}\}_{j=1}^{k},\ \epsilon_Y \perp \{\epsilon_M^{(j)}\}_{j=1}^{k},\ \epsilon_M^{(i)} \perp \epsilon_M^{(j)},\ i \neq j. \tag{8}$$

In the following, we denote $\boldsymbol{E}_{N\times(k+2)} = (\epsilon_T, \epsilon_M^{(1)}, \ldots, \epsilon_M^{(k)}, \epsilon_Y)$ and $\boldsymbol{F} = \{f_T, f_{M^{(1)}}, \ldots, f_{M^{(k)}}, f_Y\}$. Combining the observations (7) and (8), the proposed framework via the surrogate confounder can be formulated as follows:

$$\begin{aligned} (\hat{\boldsymbol{U}}, \hat{\boldsymbol{F}}, \hat{\boldsymbol{G}}) = \underset{\boldsymbol{U},\boldsymbol{F},\boldsymbol{G}}{\arg\min}\ & \|T - f_T(X) - \boldsymbol{G}_1(\boldsymbol{U})\|^2 \\ &+ \sum_{j=1}^{k} \|M^{(j)} - f_{M^{(j)}}(T, X) - \boldsymbol{G}_{j+1}(\boldsymbol{U})\|^2 \\ &+ \|Y - f_Y(T, X, \boldsymbol{M}) - \boldsymbol{G}_{k+2}(\boldsymbol{U})\|^2, \\ &\times \text{ s.t. } \operatorname{corr}(\boldsymbol{E})_{(k+2)\times(k+2)} = \boldsymbol{I}_{(k+2)\times(k+2)}, \end{aligned} \tag{9}$$

where $\operatorname{corr}(\boldsymbol{E}) = \left(\operatorname{corr}(\boldsymbol{E}_{\cdot i}, \boldsymbol{E}_{\cdot j})\right)$ is the correlation matrix of $\boldsymbol{E}$. Notice that the residuals $\boldsymbol{E}$ can be represented via $(\boldsymbol{U}, \boldsymbol{F}, \boldsymbol{G})$ as

$$\begin{aligned} \epsilon_T = T - f_T(X) - g_T(\boldsymbol{U}),\ \epsilon_Y = Y - f_Y(T, X, \boldsymbol{M}) - \boldsymbol{G}_{k+2}(\boldsymbol{U}), \\ \epsilon_M^{(j)} = M^{(j)} - f_{M^{(j)}}(T, X) - \boldsymbol{G}_j(\boldsymbol{U}),\ j = 1, \ldots, k. \end{aligned}$$

Therefore, the loss function in (9) is optimized jointly over $(\boldsymbol{U}, \boldsymbol{F}, \boldsymbol{G})$ for the surrogate confounder, output functions, and confounding effect, respectively. The proposed deconfounding strategy (9) enables us to construct a surrogate confounder $\hat{\boldsymbol{U}}$ and identify the confounding effect $\hat{\boldsymbol{G}}(\hat{\boldsymbol{U}})$ simultaneously. The former can correct the bias in estimating the causal mediation effect and direct treatment effect, and the latter captures the heterogeneity of mediators and outcomes across different subjects which might not be fully explained by the observed treatments and covariates.

One advantage of the proposed framework is the flexibility in customizing the functional relations of observed data and latent confounder within outputs to accommodate with different application scenarios. For example, for the parts involving observed data, we can restrict $(f_T, \boldsymbol{f}_M, f_Y)$ as parametric model or nonparametric models. After identifying and separating the confounding effects $\boldsymbol{G}(\boldsymbol{U})$ from mediators and outcome, the average treatment mediation effect and direct effect can be estimated as $\delta(t) = \frac{1}{N}\sum_{i=1}^{N}\left[f_Y\left\{t, M^{(1)}(1), \ldots, M^{(k)}(1), X_i\right\} - f_Y\left\{t, M^{(1)}(0), \ldots, M^{(k)}(0), X_i\right\}\right]$, and then $\zeta(t) = \frac{1}{N}\sum_{i=1}^{N}\left[f_Y\left\{1, M^{(1)}(t), \ldots, M^{(k)}(t), X_i\right\} - f_Y\left\{0, M^{(1)}(t), \ldots, M^{(k)}(t), X_i\right\}\right]$. Notice that the formation of $\delta(t)$ and $\zeta(t)$ can be further simplified based on specific functional forms of $\{f_M^{(j)}\}$ and f_Y. In the following, we provide detailed discussion on the modeling of latent confounding effects $\boldsymbol{G}(\boldsymbol{U})$.

3.2. Latent Confounding Effect Modeling

One key step in the proposed deconfounding strategy is to infer the surrogate confounder $\hat{\boldsymbol{U}}$ from the fitted confounding effects $\hat{\boldsymbol{G}}(\hat{\boldsymbol{U}})$, which does not require the underlying true confounding pattern $\boldsymbol{G}(\cdot) = \left(g_T(\cdot), g_{M^{(1)}}(\cdot), \ldots, g_{M^{(k)}}(\cdot), g_Y(\cdot)\right)$. In general, we can constrain $\hat{\boldsymbol{G}}(\cdot)$ within a class of multivariable functions such that the function class is sufficiently large to approximate $\boldsymbol{G}(\boldsymbol{U})$. In practice, utilizing the structure in $\boldsymbol{G}(\boldsymbol{U})$ allows us to determine an appropriate modeling of $\hat{\boldsymbol{G}}$.

Notice that $\boldsymbol{G}(\boldsymbol{U})$ can be formulated as a $N \times (k+2)$ matrix with $\boldsymbol{U} = \{U_i\}_{i=1}^{N}$. If the relation among components in $\boldsymbol{G}(\cdot) = \left(g_T(\cdot), g_{M^{(1)}}(\cdot), \ldots, g_{M^{(k)}}(\cdot), g_Y(\cdot)\right)$ is governed by linearity, then

the $\boldsymbol{G}(\boldsymbol{U})$ has a low-rank structure in that $\boldsymbol{G}_i(\boldsymbol{U})$ can be approximated by a linear combination of several $\{\boldsymbol{G}_j(\boldsymbol{U}),\ j \neq i\}$. Therefore, the $\boldsymbol{G}(\boldsymbol{U})$ can be simplified to a latent factor model:

$$\boldsymbol{G}(\boldsymbol{U}) \approx \hat{\boldsymbol{G}}(\hat{\boldsymbol{U}}) = \hat{\boldsymbol{U}}_{N\times r}\boldsymbol{A}_{r\times(k+2)}, \tag{10}$$

where $\boldsymbol{A}$ is the loading matrix to be estimated, and $r < k+2$ denotes the rank of the latent confounding matrix. Given that $\hat{\boldsymbol{U}}\boldsymbol{A}$ captures the subject-wise heterogeneity originating from the variation within $\boldsymbol{U}$, then $\hat{\boldsymbol{U}}$ would contain the information on the distribution of $\boldsymbol{U}$, therefore, serving as a confounder surrogate. Here the low-rank structure in $\boldsymbol{G}(\boldsymbol{U})$ can be verified from the observed data. We fit the mediator and outcome models $(\boldsymbol{f}_M, f_Y)$ on the observed data, and then perform the PCA on the residuals $\big(M^{(1)} - \hat{f}_M^{(1)}(T,X), \ldots, M^{(k)} - \hat{f}_M^{(k)}(T,X), Y - \hat{f}_Y(T,X,\boldsymbol{M})\big)_{N\times(k+1)}$. If only several leading principle components dominate the variation of residuals, we can choose the latent factor modeling (10) for $\boldsymbol{G}(\cdot)$. Accordingly, the rank of surrogate confounder r can be determined via the largest eigengap between two successive eigenvalues.

In many applications, the underlying confounding pattern $\boldsymbol{G}(\cdot)$ might be more complex than the linear relations, as the confounding effects $\{g_T(\boldsymbol{U}), g_M^{(1)}(\boldsymbol{U}), \ldots, g_M^{(k)}(\boldsymbol{U}), g_Y(\boldsymbol{U})\}$ on treatment, mediators and outcome could be nonlinear to each other. For example, in the causal effect of alcohol consumption on blood pressure, the various enzymes are treated as mediators, and the expression levels of specific genes might confound with the concentration of enzymes while the gene-enzyme relations could be significantly different across different enzymes. In this case, the latent factor model likely fails as the confounding matrix $\boldsymbol{G}(\boldsymbol{U})$ is full-rank, and the low-rank structure only preserves the latent space expanded by columns of the nonlinear-transformed $\boldsymbol{G}(\boldsymbol{U})$.

Alternatively, we adopt the autoencoder (Ballard 1987) which serves as a nonlinear generalization of PCA to infer the information of $\boldsymbol{U}$ from $\boldsymbol{G}(\boldsymbol{U})$. An autoencoder consists of an encoder $\Phi_{\text{encoder}}(\cdot)$ and a decoder $\Phi_{\text{decoder}}(\cdot)$ where the former performs transformation on the input to extract important features from the observed data and the latter reconstructs the input data based on the extracted features. We estimate the autoencoder via

$$(\hat{\Phi}_{\text{encoder}}, \hat{\Phi}_{\text{decoder}}) = \underset{\Phi_{\text{encoder}},\Phi_{\text{decoder}}}{\arg\min} \Big\| \boldsymbol{G}(\boldsymbol{U}) - \Phi_{\text{decoder}}\Big[\Phi_{\text{encoder}}\big\{\boldsymbol{G}(\boldsymbol{U})\big\}\Big]\Big\|^2. \tag{11}$$

Both Φ_{encoder} and Φ_{decoder} can be a class of composition function $\{\phi^{(L)} \circ \cdots \phi^{(l)} \circ \cdots \phi^{(1)} : \phi^{(l)}(\boldsymbol{x}) = \varphi(W^{(l)}\boldsymbol{x} + b^{(l)})\}$, where φ is a nonlinear activation function, $W^{(l)}$ is a weighting matrix, $b^{(l)}$ is a bias vector, and L is the number of layers. With the trained autoencoder $(\hat{\Phi}_{\text{encoder}}, \hat{\Phi}_{\text{decoder}})$, we are able to extract the confounding information within $\boldsymbol{U}$ from the encoder $\hat{\boldsymbol{U}} = \hat{\Phi}_{\text{encoder}}\big\{\boldsymbol{G}(\boldsymbol{U})\big\}$. Note that a good fitting of $\boldsymbol{G}(\boldsymbol{U})$ from (11) indicates $\boldsymbol{G}(\boldsymbol{U}) \approx \Phi_{\text{decoder}}(\hat{\boldsymbol{U}})$, which leads to $\hat{\boldsymbol{U}} \approx \Phi_{\text{decoder}}^{-1}\{\boldsymbol{G}(\boldsymbol{U})\}$. Here the cardinality of the function space of Φ_{decoder} increases rapidly as the number of layers L in composition increases. Therefore, $\Phi_{\text{decoder}}(\cdot)$ is capable of reconstructing $\boldsymbol{G}$ with a higher resolution, in that $\Phi_{\text{decoder}}(B) \subset \boldsymbol{G}(B)$ for any measurable set B in the input space. Consequently, the σ-field $\sigma(\hat{\boldsymbol{U}})$ is larger than the true confounder $\sigma(\boldsymbol{U})$, implying that $\hat{\boldsymbol{U}}$ can serve as a surrogate confounder to correct all the confounding among treatment assignments, mediators and outcomes to satisfy conditions (1) and (2).

4. Algorithm and Implementation

In this section, we develop a deconfounding algorithm for inferring the surrogate confounder via optimizing the proposed objective function (9). In general, we sequentially update the output functions $(\hat{f}_T, \hat{\boldsymbol{f}}_M, \hat{f}_Y)$, confounding function estimator $\hat{\boldsymbol{G}}$ and surrogate confounder $\hat{\boldsymbol{U}}$ at each iteration. Different models for the latent confounding effect $\boldsymbol{G}(\boldsymbol{U})$ lead to the different optimization strategies for estimating $\hat{\boldsymbol{G}}$ and $\hat{\boldsymbol{U}}$.

For the simplification of presentation, we introduce the $L_{\text{obs}}^{(s)}$ to denote the residuals of observed data at the sth step as

$$L_{\text{obs}}^{(s)} = \big(T - \hat{f}_T^{(s)}(X), M^{(1)} - \hat{f}_{M^{(1)}}^{(s)}(T,X), \ldots, M^{(k)} - \hat{f}_{M^{(k)}}^{(s)}(T,X), Y - \hat{f}_Y^{(s)}(T,X,\boldsymbol{M})\big)_{N\times(k+2)}.$$

To incorporate orthogonality across the residuals in (9), we use the method of Lagrange multipliers and transform the constraints into a penalty function as $\|\text{corr}(\boldsymbol{E}^{(s)}) - \boldsymbol{I}\|_F^2$, where the residuals at the sth step are formulated as $\boldsymbol{E}^{(s)} = L_{\text{obs}}^{(s)} - \hat{\boldsymbol{G}}^{(s)}(\hat{\boldsymbol{U}}^{(s)})$. Therefore, the loss function at the sth step is

$$\begin{aligned}\text{Loss}^{(s)} = {} & \|T - \hat{f}_T^{(s)}(X) - \hat{\boldsymbol{G}}_1^{(s)}(\hat{\boldsymbol{U}}^{(s)})\|_2^2 \\ & + \sum_{j=1}^{k} \|M^{(j)} - \hat{f}_{M^{(j)}}^{(s)}(T,X) - \hat{\boldsymbol{G}}_{j+1}^{(s)}(\hat{\boldsymbol{U}}^{(s)})\|_F^2 \\ & + \|Y - \hat{f}_Y^{(s)}(T,X,\boldsymbol{M}) - \hat{\boldsymbol{G}}_{k+2}^{(s)}(\hat{\boldsymbol{U}}^{(s)})\|_2^2 \\ & + \lambda\|\text{corr}(\boldsymbol{E}^{(s)}) - \boldsymbol{I}\|_F^2,\end{aligned}$$

where λ is the Lagrange multiplier. We first illustrate the deconfounding algorithm given that confounding effect $\boldsymbol{G}(\boldsymbol{U})$ is approximated by the latent factor model, that is, we replace $\boldsymbol{G}(\boldsymbol{U}) = \boldsymbol{U}_{N\times r}\boldsymbol{A}_{r\times(k+2)}$ in (9). Notice that with the latent factor modeling, both gradients of the loss function (9) in terms of $\boldsymbol{A}$ and $\boldsymbol{U}$ have explicit forms.

For the output functions $(f_T, f_{\boldsymbol{M}}, f_Y)$, we can choose various models such as linear regression, spline, and random forest according to a specific application. For example, when dealing with binary treatment, we adopt the logistic regression model for f_T, and change the square loss in (9) to a negative log-likelihood loss. In addition, we can initialize $(f_T, f_{\boldsymbol{M}}, f_Y)$ via fitting each of them on the observed data (X,T), $(T,X,\{M^{(j)}\}_{j=1}^k)$, and $(T,X,\boldsymbol{M},Y)$, respectively. Similarly, $\hat{\boldsymbol{G}}$ and $\hat{\boldsymbol{U}}$ can be initialized via performing the PCA on $(\boldsymbol{M},Y)$. Notice that Algorithm 1 can be generalized to the other confounding-effect models where the gradients of the latent confounding effect have explicit forms. In addition, Algorithm 1 can be modified to capture the nonlinear confounding patterns via autoencoder as Section 3.2. The detailed autoencoder-based algorithm is provided in the Section 5 in supplementary materials.

Algorithm 1 Blockwise backfitting for latent factor confounding modeling

1. *(Initialization)* Input the initialization of $(\hat{f}_T^{(0)}, \hat{f}_{\boldsymbol{M}}^{(0)}, \hat{f}_Y^{(0)})$, $\hat{\boldsymbol{G}}^{(0)}$, $\hat{\boldsymbol{U}}^{(0)}$, Lagrange multiplier λ, learning rate η, and stopping threshold γ.
2. *(Backfitting updates)* At the sth iteration ($s \geq 1$).

   (i) Given $(\hat{f}_T^{(s-1)}, \hat{f}_{\boldsymbol{M}}^{(s-1)}, \hat{f}_Y^{(s-1)})$, $\hat{\boldsymbol{A}}^{(s-1)}$, $\hat{\boldsymbol{U}}^{(s-1)}$ update the residuals of observed data $L_{\text{obs}}^{(s-1)}$. Then update the surrogate confounder via:

$$\hat{\boldsymbol{U}}^{(s)} = \hat{\boldsymbol{U}}^{(s-1)} - 2\eta\Big\{\big(L_{\text{obs}}^{(s-1)} - \hat{\boldsymbol{U}}^{(s-1)}\hat{\boldsymbol{A}}^{(s-1)}\big)\hat{\boldsymbol{A}}^{(s-1)T} + \lambda\frac{\partial\|\text{corr}(\boldsymbol{E}^{(s-1)}) - \boldsymbol{I}\|_F^2}{\partial \boldsymbol{U}}\Big\}.$$

   (ii) Given $\hat{\boldsymbol{U}}^{(s)}$, update the output models $(f_T, f_{\boldsymbol{M}}, f_Y)$:

$$\hat{f}_T^{(s)} \leftarrow \arg\min_{f_T} \boldsymbol{E}_{\hat{\boldsymbol{U}}^{(s)}}\big\{\|T - f_T(X)\|^2 | \hat{\boldsymbol{U}}^{(s)}\big\},$$
$$\hat{f}_{M^{(j)}}^{(s)} \leftarrow \arg\min_{f_{M^{(j)}}} \boldsymbol{E}_{\hat{\boldsymbol{U}}^{(s)}}\big\{\|M^{(j)} - f_{M^{(j)}}(T, X)\|^2 | \hat{\boldsymbol{U}}^{(s)}\big\},$$
$$j = 1, \ldots, k$$
$$\hat{f}_Y^{(s)} \leftarrow \arg\min_{f_Y} \boldsymbol{E}_{\hat{\boldsymbol{U}}^{(s)}}\big\{\|Y - f_Y(T, X, \boldsymbol{M})\|^2 | \hat{\boldsymbol{U}}^{(s)}\big\}.$$

   (iii) Given $(\hat{f}_T^{(s)}, \hat{f}_{\boldsymbol{M}}^{(s)}, \hat{f}_Y^{(s)})$, and $\hat{\boldsymbol{U}}^{(s)}$, update the loading matrix $\boldsymbol{A}$ via:

$$\hat{\boldsymbol{A}}_{\cdot 1}^{(s)} \leftarrow \arg\min_{v_{r\times 1}} \|T - f_T^{(s)}(X) - \hat{\boldsymbol{U}}^{(s)} v\|^2,$$
$$\hat{\boldsymbol{A}}_{\cdot (j+1)}^{(s)} \leftarrow \arg\min_{v_{r\times 1}} \|M^{(j)} - f_{M^{(j)}}^{(s)}(T, X) - \hat{\boldsymbol{U}}^{(s)} v\|^2,$$
$$j = 1, \ldots, k,$$
$$\hat{\boldsymbol{A}}_{\cdot k+2}^{(s)} \leftarrow \arg\min_{v_{r\times 1}} \|Y - f_Y^{(s)}(T, X, \boldsymbol{M}) - \hat{\boldsymbol{U}}^{(s)} v\|^2.$$

3. *(Stopping Criterion)* Stop backfitting updates if $\frac{|\text{Loss}^{(s)} - \text{Loss}^{(s-1)}|}{\text{Loss}^{(s-1)}} < \gamma$. Set $\hat{f}_T = \hat{f}_T^{(s)}$, $\hat{f}_{\boldsymbol{M}} = \hat{f}_{\boldsymbol{M}}^{(s)}$, $\hat{f}_Y = \hat{f}_Y^{(s)}$, $\hat{\boldsymbol{A}} = \hat{\boldsymbol{A}}^{(s)}$, and $\hat{\boldsymbol{U}} = \hat{\boldsymbol{U}}^{(s)}$. Otherwise set $s \leftarrow s + 1$ and iterate Step 2.

5. Theoretical Results

This section establishes the theoretical properties for the proposed deconfouding method. Specifically, we show that the surrogate confounder plays a deconfounding role on treatment assignment, mediators and outcome. In addition, we show that both the causal and mediation effect can be identified by incorporating the surrogate confounders. We also discuss identifiability under different causal pathway structures. The proposed deconfounding method relies on the structure of confounder-sharing among mediators as follows:

Assumption 1 (Mediator-sharing confounder). There exits a pre-treatment random variable $\boldsymbol{U}$ satisfying the following requirements:

1. Together with X and T, $\boldsymbol{U}$ generates the smallest σ-algebra such that individual distributions of mediators are independent from each other

$$M^{(i)} \perp \boldsymbol{M}^{(-i)} | T, \boldsymbol{U}, X. \tag{12}$$

2. Together with X and T, $\boldsymbol{U}$ is the σ-algebra to satisfy sequential ignorability

$$\big\{Y\left(t', m\right), \boldsymbol{M}(t)\big\} \perp T \mid (\boldsymbol{U}, X), \text{ and } Y\left(t', m\right) \perp \boldsymbol{M}(t) \mid (T, \boldsymbol{U}, X). \tag{13}$$

The mediator-sharing confounder assumption paraphrases the causal mediation pathway in Figure 2. Due to the conditional independence among mediators, $\boldsymbol{U}$ has to contain the information of all the multi-mediator confounders that only affect a subset of mediators $\{M^{(j)}\}_{j\in J}$, $J \subset \{1, \ldots, k\}$, $2 \leq |J| \leq k$. On the other hand, the concept of smallest σ-algebra guarantees that $\boldsymbol{U}$ only includes confounders affecting multiple mediators and excludes the latent confounders that only affect a single mediator. The structure of parallel mediators (12) is common in many scientific and social studies where the multiple mediators are conditionally independent given the treatment and confounders. For instance, in the study of causal relations between mindfulness and emotional distress, the negative cognitive bias and perceived stress are identified as independent mediators (Ford and Shook 2019). In addition, it is shown that interpersonal and intrapersonal factors are parallel independent mediators in the causal effect of HIV stigma on therapy adherence (Seghatol-Eslami et al. 2017). Sequential ignorability (13) is a standard condition in causal mediation inference (Imai, Keele, and Yamamoto 2010). The first part assumes that the treatment assignment is ignorable given the confounders, which can be satisfied when the treatments are randomly assigned. The second part assumes the ignorability of mediators conditioning on pre-treatment covariates, which is not directly testable from observed data in general (Manski 2009). However, a set of sensitivity analyses exists to quantify the robustness of causal effect estimation to the potential violation of the ignorability assumption (Imai, Keele, and Tingley 2010).

Assumption 2 (Overlap). $\mathcal{P}(T = t|\boldsymbol{U}, X) > 0$, and $\mathcal{P}(\boldsymbol{M} = \boldsymbol{m}|\boldsymbol{U}, X) > 0$ for all t and $\boldsymbol{m}$.

The overlap is a standard condition in the causal inference literature (Imai, Keele, and Tingley 2010; Imai, Keele, and Yamamoto 2010), which allows each treatment assignment and mediators' value to have a certain probability to be observed when controlling the confounders. This ensures that the potential outcomes $\{Y(t, m)\}$ can be identifiable. In the following, we establish the causal identification results based on the surrogate confounder $\hat{\boldsymbol{U}}$. Recall that estimating $\hat{\boldsymbol{U}}$ via the objective function (9) is equivalent to fitting the latent factor model (4) to the observed data. Specifically, the surrogate $\hat{\boldsymbol{U}}$ satisfies

$$\mathcal{P}(T, \boldsymbol{M}|X) = \int \prod_{j=1}^{k} \mathcal{P}(M^{(j)}|T, \hat{\boldsymbol{U}}, X)\mathcal{P}(T|\hat{\boldsymbol{U}}, X)\mathcal{P}(\hat{\boldsymbol{U}})\mathbf{d}\hat{\boldsymbol{U}}, \tag{14}$$

$$\mathcal{P}(T,\boldsymbol{M},Y|X) = \int \mathcal{P}(Y|T,\boldsymbol{M},\hat{\boldsymbol{U}},X)\prod_{j=1}^{k}\mathcal{P}(M^{(j)}|T,\hat{\boldsymbol{U}},X)\mathcal{P}(T|\hat{\boldsymbol{U}},X)\mathcal{P}(\hat{\boldsymbol{U}})\mathbf{d}\hat{\boldsymbol{U}}, \tag{15}$$

where $\mathcal{P}(T,\boldsymbol{M}|X)$ and $\mathcal{P}(T,\boldsymbol{M},Y|X)$ are the distributions of the observed data while the distributions involving $\hat{\boldsymbol{U}}$ under the integrations are inferred from the proposed mediation pathway modeling (6). Intuitively, with the surrogate $\hat{\boldsymbol{U}}$, we can decompose the joint distribution of mediators into multiple conditionally independent components to satisfy (12). Note that there exist different $\hat{\boldsymbol{U}}$ compatible with $\mathcal{P}(T,\boldsymbol{M}|X)$ and $\mathcal{P}(T,\boldsymbol{M},Y|X)$ on the observed data, while the causal identification requires the removal of uncertainty in estimating the latent factor model. Therefore, we allow that $\hat{\boldsymbol{U}}$ can be determined by the observed data.

Assumption 3 (Consistent surrogate confounder). All latent variables $\hat{U}$ satisfying (12), that is, $M^{(i)} \perp \boldsymbol{M}^{(-i)} \mid (T,\hat{U})$, can be consistently identified by the observed treatment and mediators $(T,\boldsymbol{M})$ in that for any $\varepsilon > 0$,

$$\lim_{n\to\infty}\mathcal{P}(|\hat{U} - f(T,\boldsymbol{M})| > \varepsilon) = 0, \text{ or}$$
$$\lim_{k\to\infty}\mathcal{P}(|\hat{U} - f(T,\boldsymbol{M})| > \varepsilon) = 0, \tag{16}$$

where $f(\cdot)$ is a deterministic function dependent on the factor models in (10), n and k are the sample size and number of mediators.

Notice that this assumption does not require $\hat{\boldsymbol{U}}$ to coincide with the true latent confounder $\boldsymbol{U}$. Instead, $\hat{\boldsymbol{U}}$ only needs to be consistently identified by the observed data. This assumption gives us flexibility in that many choices of low-rank latent models can be used for modeling $\hat{\boldsymbol{G}}(\hat{\boldsymbol{U}})$ in Section 3.2. For instance, PCA, matrix factorization have the identification property of latent factors when the number of mediators k and the number of samples N are large (Fu et al. 2019; Chen, Li, and Zhang 2020; Gresele et al. 2020). In addition, many practical applications involve a large number of mediators, especially epigenomic studies such as studying the relation between smoking and the risk of lung cancer via high-dimensional DNA methylation markers.

On the other hand, the Assumption 3 can also be satisfied when the number of mediators is finite or small. In this case, we can consistently estimate the surrogate confounder via using prior knowledge about the distribution of confounder U. For example, when there exist underlying K subgroups among subjects, we can assume $U_i = a_k \mathbb{1}_{\{i\in\text{group}\}}$, $k \in \{1,2,\ldots,K\}$. Or when mediators change smoothly across similar subjects, one can impose smoothness on the surrogate confounders. The distributional information of U can be incorporated via adding regularizations on the surrogate confounders to encourage a piece-wise constant structure (Tang, Xue, and Qu 2021), or control the variations among U from similar subjects (Hong and Lian 2013; Wang and Huang 2017). Under these constraints, the surrogate confounders can be consistently identified as the number of subjects increases with only limited available mediators or a few of mediators simultaneously affected by U. For nonlinear latent factor models, recent developments in identifiability for autoencoder and nonlinear ICA models implies that a surrogate confounder $\hat{U}$ can be consistently estimated by the autoencoder method via good data fitting. Specifically, when the underlying clustering structure exists in $\hat{U}$, then $\hat{U}$ can be consistently identifiable up to the affine transformation (Hyvarinen, Sasaki, and Turner 2019; Khemakhem et al. 2020; Willetts and Paige 2021). We perform a theoretical sensitivity analysis for the effects of estimation consistency of the surrogate confounder on causal estimation. See the detailed theorem and discussion of results' in Section 16 of the supplemental materials.

Lemma 5.1. Under Assumptions 1–3, the treatment T, mediators $\boldsymbol{M}$, and outcome Y are sequential deconfounded given covariate X and surrogate $\hat{\boldsymbol{U}}$ satisfying (14) and (15):

$$\left\{Y\left(t',m\right),\boldsymbol{M}(t)\right\} \perp T \mid \hat{\boldsymbol{U}},X, \quad Y\left(t',m\right) \perp \boldsymbol{M}(t) \mid T=t,\hat{\boldsymbol{U}},X,$$

if the distribution of $(T,\boldsymbol{M},Y)$ can be represented as a latent factor model (4).

The above Lemma 5.1 provides a theoretical justification of the proposed deconfounding scheme in that the surrogate confounder $\hat{\boldsymbol{U}}$ contains all the confounding information if the proposed latent modeling captures the distribution of $(T,\boldsymbol{M},Y)$. Based on the Lemma 5.1, we establish identification of the average causal effect and mediation effect.

Theorem 5.1. Given that Assumptions 1–3, and the following two conditions holds: (a) the response has additive forms as $\mathbb{E}(Y(t,\boldsymbol{m})|X=x,\boldsymbol{U}=u) = f_1(t,\boldsymbol{m},x) + f_2(u)$ and $\mathbb{E}(Y|T=t,\boldsymbol{M}=\boldsymbol{m},X=x,\hat{\boldsymbol{U}}=u) = f_3(t,\boldsymbol{m},x) + f_4(u)$, where f_1,f_3 are continuous functions and f_2,f_4 are piecewise functions; (b) the mediators follow the generalized additive model in (6). Then the average mediation effect $\delta(t)$ and average direct treatment effect $\zeta(t)$ can be identified through the surrogate confounder $\hat{U}$

$$\delta(t) = \iint\Big\{\int \mathbb{E}\left(Y \mid T=t,\boldsymbol{M}=\boldsymbol{m},X=x,\hat{\boldsymbol{U}}=u\right)\mathbf{d}F_{\boldsymbol{M}}^{(1)}(\boldsymbol{m}) - \int \mathbb{E}\left(Y \mid T=t,\boldsymbol{M}=\boldsymbol{m},X=x,\hat{\boldsymbol{U}}=u\right)\mathbf{d}F_{\boldsymbol{M}}^{(0)}(\boldsymbol{m})\Big\}\mathbf{d}F_X(x)\mathbf{d}F_{\hat{U}}(u), \tag{17}$$

$$\zeta(t) = \iint\Big\{\int \mathbb{E}\left(Y \mid T=1,\boldsymbol{M}=\boldsymbol{m},X=x,\hat{\boldsymbol{U}}=u\right)\mathbf{d}F_{\boldsymbol{M}}^{(t)}(\boldsymbol{m}) - \int \mathbb{E}\left(Y \mid T=0,\boldsymbol{M}=\boldsymbol{m},X=x,\hat{\boldsymbol{U}}=u\right)\mathbf{d}F_{\boldsymbol{M}}^{(t)}(\boldsymbol{m})\Big\}\mathbf{d}F_X(x)\mathbf{d}F_{\hat{U}}(u), \tag{18}$$

where $F_{\boldsymbol{M}}^{(t)} = F_{\boldsymbol{M}|T=t,X=x,\hat{U}=u}(\cdot)$, $F_W(\cdot)$ and $F_{V|W}(\cdot)$ represent the distribution function of a random variable W and the conditional distribution function of V given W. Then the average causal effect can be identified as $\tau = \delta(t) + \zeta(1-t),\ t=0,1$.

Theorem 5.1 shows that the surrogate confounder enables the average mediation effect and direct effect to be unbiasedly

estimated from the observed data. The result requires additional technical conditions to solve the violation of overlap in that $\mathcal{P}(\boldsymbol{M} = \boldsymbol{m}|T, X, \hat{\boldsymbol{U}})$ might be zero for some $\boldsymbol{m}$ due to the deterministic assumption $\hat{\boldsymbol{U}}$ in (16). However, we show that the identification of the average mediation effect for subsets of the mediators can be achieved under fewer constraints on the distribution of $(T, \boldsymbol{M}, Y)$. Notice that the proposed method also applies to the causal pathway where latent confounder only affects multiple mediators, but not affects treatment and outcome. In this case, the surrogate confounder $\hat{U}$ is consistently independent from T and $Y(t, \boldsymbol{m})$ since $\hat{U}$ can be consistently captured by the distribution of $\boldsymbol{M}(t)$, which is independent to $\{T, Y(t, \boldsymbol{m})\}$. Then there does not exist confounding among observed data $(T, \boldsymbol{M}, Y)$ conditioning on $\hat{U}$.

Corollary 5.1. Given Assumptions 1–3, the average mediation effect for subsets of mediators $\boldsymbol{M}^{(J)} = \{M^{(j)}\}_{j\in J}, J \subset \{1, \ldots, k\}, |J| < k$ denoted as $\delta^{(J)}(t) = \mathbb{E}\{Y(t, \boldsymbol{M}^{(J)}(1), \boldsymbol{M}^{(-J)})\} - \mathbb{E}\{Y(t, \boldsymbol{M}^{(J)}(0), \boldsymbol{M}^{(-J)})\}$ can be identifiable as

$$\begin{aligned}\delta^{(J)}(t) = \iint \Bigg\{ \int \mathbb{E}\Big(Y \mid T = t, \boldsymbol{M} = \boldsymbol{m}, X = x, \hat{\boldsymbol{U}} = u\Big) \\ \mathbf{d}F^{(1)}_{\boldsymbol{M}^{(J)}}(\boldsymbol{m}) \\ - \int \mathbb{E}\Big(Y \mid T = t, \boldsymbol{M} = \boldsymbol{m}, X = x, \hat{\boldsymbol{U}} = u\Big) \\ \mathbf{d}F^{(0)}_{\boldsymbol{M}^{(J)}}(\boldsymbol{m}) \Bigg\} \mathbf{d}F_X(x)\mathbf{d}F_{\hat{U}}(u),\end{aligned}$$

where $\boldsymbol{M}^{(-J)} = \{M^{(j)}\}_{j\in J^c}$ denotes the complement set of mediators and the conditional distribution $F^{(t)}_{\boldsymbol{M}^{(J)}} = F_{\boldsymbol{M}^{(J)}|T=t,X=x,\hat{U}=u}(\cdot)$. The identification holds when $\boldsymbol{M}^{(J)}$ satisfies the overlap condition in that $\mathcal{P}(\boldsymbol{M}^{(J)} = \boldsymbol{m}|T, X, \hat{\boldsymbol{U}}) > 0$ for any $\boldsymbol{m}$.

Corollary 5.1 shows that we can estimate the average mediation effect of a subset of mediators unbiasedly using surrogate confounder $\hat{\boldsymbol{U}}$ based on observed data. Specifically, it enables us to estimate the mediation effect for each individual mediator. Compared with Theorem 5.1, the identification can be established under weaker assumptions such that the regularity assumptions on the outcome model and mediators are replaced by the overlap condition on subsets of mediators. Notice that the overlap condition on the subset of mediators is compatible with Assumption 3 of the deterministic surrogate confounder in (16), since the consistency $\mathcal{P}(\hat{\boldsymbol{U}}|T, \boldsymbol{M}, Y) = \delta_{f(T,\boldsymbol{M}(T))}$ only imposes constraint on the nonzero support of $\mathcal{P}(\boldsymbol{M}|T, X, \hat{\boldsymbol{U}})$ while the nonzero support of a subset $\boldsymbol{M}^{(J)}$ can still be unconstrained.

The above identification results are established under Assumption 3 of the deterministic surrogate confounder, which might not be satisfied when the number of mediators is not large enough to exclude the uncertainty in inferring $\hat{\boldsymbol{U}}$ from observed data. In the following, we allow an alternative pathway structure for identifying causal mediation and direct effects without Assumption 3. We first introduce the concept of *null mediator* which indicates a mediator M that does not affect the outcome such that $Y \perp M|T, \boldsymbol{U}, X$. Null mediators are widespread in many scientific studies of delineating sparse causal pathways through high-dimensional mediators, such as genes and brain neuroimaging (Zhao and Luo 2016; Huang 2019; Zhao, Lindquist, and Caffo 2020). Given a large set of mediators, only a small subset has nonzero mediation effects between treatments and outcomes. In the following, we introduce the alternative identification condition:

Assumption 4 (Null mediator). There exists more than one null mediator which do not have causal effect on outcome Y, denoted as $\boldsymbol{M}^{\text{null}} = \{M^{(j)}_{\text{null}} : Y \perp M^{(j)}_{\text{null}}|T, \boldsymbol{U}, X\}$ and $|\boldsymbol{M}^{\text{null}}| \geq 2$. And the null mediators are conditional independent to other non-null mediators $\{M^{(i)}\}_{i=1}^{k}$ as

$$M^{(i)} \perp M^{(j)}_{\text{null}} \mid T, \boldsymbol{U}, X,\ i = 1, \ldots, k, j = 1, \ldots, |\boldsymbol{M}^{\text{null}}|.$$

The null mediators $M^{(k-1)}$ and $M^{(k)}$ are observable descendants of the latent confounder, and therefore can serve as the proxy variables for $\boldsymbol{U}$. Notice that the null mediator condition does not require the information as to which mediators are null. Given the existence of null mediators, we can establish the identification of average causal effect as follows. Notice that we do not require the proportion of null mediators for the whole mediator set. To validate the conditional independence among mediators in Assumptions 3 and 4, we can first obtain the surrogate confounder $\hat{U}$ and then adopt the conditional independent test as in (Su and White 2008; Huang 2010; Zhang et al. 2012; Cai, Li, and Zhang 2022).

Theorem 5.2. Given Assumptions 1, 2, and 4 and weak regularity conditions, the surrogate confounder $\hat{U}$ satisfying (14) and (15) identifies the average mediation effect $\delta(t)$ as (17) and average direct treatment effect $\zeta(t)$ as (18). And the causal effect can be identified as $\tau = \delta(t) + \zeta(1 - t),\ t = 0, 1$.

Compared with Theorem 5.1, Theorem 5.2 shows that with null mediators, identifying causal mediation and direct effect is still possible even when the surrogate confounder $\hat{\boldsymbol{U}}$ is a random variable of the observed data. In another words, it is unnecessary for $\hat{\boldsymbol{U}}$ to be completely identified by the observed data to play the deconfounding role. The identification in Theorem 5.2 leverages the proxy variable strategy for direct causal effect (Kuroki and Pearl 2014; Miao, Geng, and Tchetgen Tchetgen 2018) where the distribution of debiased causal effect is identifiable when two proxies of the latent confounder are observed. In the case of multiple mediators, the null mediators can serve as proxies for the shared latent confounder $\boldsymbol{U}$ to identify other causal mediation pathways. Therefore, unlike the case of no mediator or a single mediator, we do not need to observe external proxy variables for $\boldsymbol{U}$. Notice that Theorem 5.2 still holds when the null mediators are correlated to each other conditioning on unobserved confounder U and treatment T in Assumption 4.

6. Numerical Study

In this section, we conduct simulations to investigate the performance of the proposed deconfounding algorithm on debiasing treatment effect estimation, and perform numerical comparisons with existing causal inference and mediation analysis methods. Specifically, our method is compared with existing mediation analysis and causal inference methods:

linear structural equation modeling (LSEM), high-dimensional mediation analysis (HIMA), random-forests based causal effect inference (Causal Forest), and meta-learning method (XLearner). These four methods are popular and widely used in many applications. We investigate the debias performance under the settings of linear and nonlinear confounding effects, which are two representative situations of how the latent confounders affect mediators and outcomes.

6.1. Linear Confounding Effect

In this section, we investigate the performance of causal effect estimation based on different methods when the confounding effects on multiple mediators and outcomes are linear representations of each other. Specifically, we assume the subject-wise latent confounders $\boldsymbol{U} = \{U_i\}_{i=1}^N$ are randomly generated from a mixture of Guassian distributions as $U_i \sim \omega * N(-2, 1.5) + (1-\omega) * N(2, 1.5)$ where $\omega \sim \text{Bern}(1, 0.5)$. The covariates $X_{N\times p} = \{X_i\}_{i=1}^N$ are generated from $N(\mathbf{0}_{1\times p}, \boldsymbol{I}_{p\times p})$. With the latent confounders $\boldsymbol{U}$ and X, the observations of $(T, \boldsymbol{M}, Y)$ are generated via a series of additive models as follows:

$$\begin{aligned} Y &= \alpha^{(Y)}T + X\gamma^{(Y)} + \boldsymbol{M}\beta^{(Y)} + g_Y(\boldsymbol{U}) + \epsilon_Y, \\ M^{(j)} &= T\beta_j^{(\boldsymbol{M})} + X\Gamma_{\cdot j}^{(\boldsymbol{M})} + g_{M^{(j)}}(\boldsymbol{U}) + \epsilon_M^{(j)}, \\ T_i &\sim \text{Bern}\big(\sigma(0.4 * U_i)\big), j = 1, \ldots, k, \end{aligned} \tag{19}$$

where $\sigma(\cdot)$ is the logistic link function, $(g_Y(\boldsymbol{U}), g_{M^{(1)}}(\boldsymbol{U}), \ldots, g_{M^{(k)}}(\boldsymbol{U})) = (\eta_0\boldsymbol{U}, \ldots, \eta_k\boldsymbol{U})$ imposes linear confounding effects on mediators and outcome. In addition, $\{\alpha^{(Y)} \in R, \beta^{(\boldsymbol{M})} = (\beta_1^{(\boldsymbol{M})}, \ldots, \beta_k^{(\boldsymbol{M})}) \in R^k\}$, $\{\gamma^{(Y)} \in R^p, \Gamma^{(\boldsymbol{M})} \in R^{p\times k}\}$, $\beta^{(Y)} \in R^k$, and $\eta = (\eta_0, \ldots, \eta_k) \in R^{k+1}$ denotes the coefficients of treatment, covariates, mediators, and latent confounders, respectively. Accordingly, the direct causal effect, causal mediation effect and total causal effect can be formulated as $\alpha^{(Y)}$, $(\beta^{(M)})^T\beta^{(Y)}$, and $\alpha^{(Y)} + (\beta^{(M)})^T\beta^{(Y)}$. In the following simulations, we set the sample size $N = 200, 800, 2000$, and the number of mediators $k = 2, 5$. We set coefficients $\alpha^{(Y)} = 1$, $\beta^{(M)} = \mathbf{1}_{1\times k}$, and $\beta^{(Y)} = 0.5 \times \mathbf{1}_{1\times k}$.

Given that the proposed method, LSEM, and HIMA directly estimate the coefficients $\alpha^{(Y)}$, $\beta^{(M)}$, and $\beta^{(Y)}$, their estimations of causal effect and mediation effect are represented as $\hat{\alpha}^{(Y)} + (\hat{\beta}^{(M)})^T\hat{\beta}^{(Y)}$ and $(\hat{\beta}^{(M)})^T\hat{\beta}^{(Y)}$. For nonparametric methods such as Causal Forest and XLearner, we obtain the estimation of causal effect via averaging the estimated individual treatment effects. Given that the Causal Forest and XLearner are not designed to directly utilize information of mediators, we combine the mediators $\boldsymbol{M}$ with X such that $(X, \boldsymbol{M})$ serve as new covariates for model training. In this way, Causal Forest and XLearner include the same amount of information from input as other competing methods for the causal effect estimation to make a fair comparison. The performance is evaluated by the estimation bias of causal effect ($\text{Bias}_{\text{total}}$) and mediation effect ($\text{Bias}_{\text{med}}$). Besides the causal effect estimation, we also investigate the performance of predictions of outcomes based on different methods. See Section 7 in supplemental materials for detailed implementation for prediction procedure.

The performances of causal effect estimations from different methods are illustrated in Figure 3 and Table 1, where the proposed method utilizing factor modeling (10) and the autoencoder (11) are denoted as **Prop FM** and **Prop AE**, respectively. Figure 3 and Table 1 illustrate that the proposed method can consistently achieve a lower estimation bias of causal effect compared with competing methods. As the number of mediators and sample size increase, the proposed method achieves a more significant debias in estimation compared to other methods. In particular, in comparison with the best competing method LSEM, our algorithm reduces the causal effect estimation bias by 36% when $k = 2$ and by 50% when $k = 5$. In addition, the proposed method reduces the estimation bias by 50% to 89% compared to LSEM when the sample size increases from $N = 200$ to $N = 2000$. Although other methods incorporate information of multiple mediators into causal effect estimation procedure, they do not use the confounder-sharing structure among mediators and outcomes. As a consequence, the estimation bias accumulates as the number of mediators and sample size increase. However, the proposed method incorporates

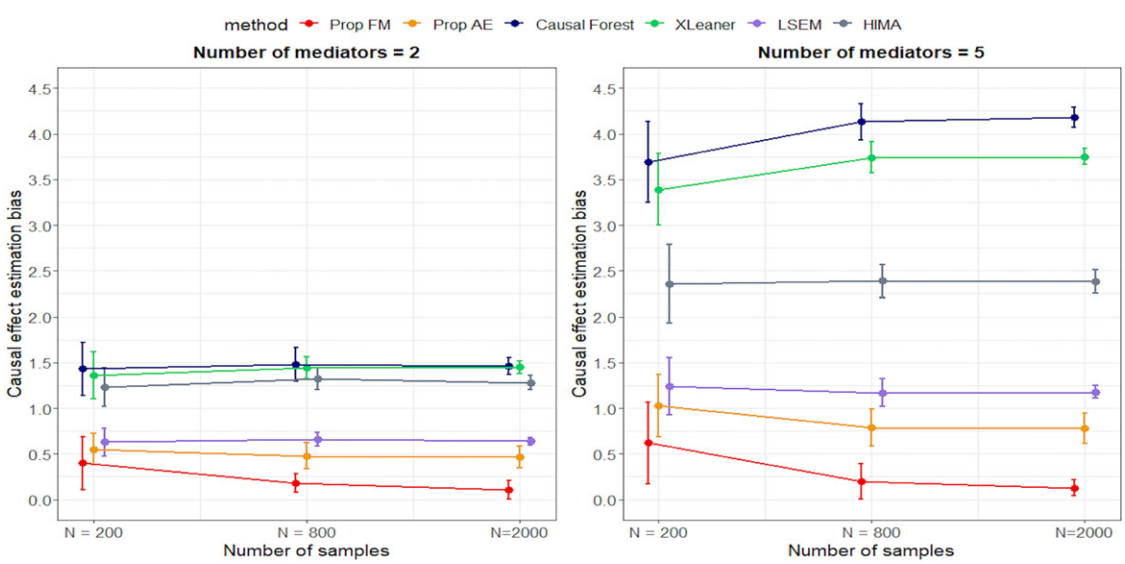

Figure 3. The bias of causal effect estimation from different methods under the setting of linear confounding effects on multiple mediators and outcomes.

Table 1. The performance of causal effect estimation and outcome prediction from different methods under the setting in Section 6.1.

Sample size	Method	Number of mediators k					
		$k=2$			$k=5$		
		$\text{Bias}_{\text{total}}$	Bias_{med}	MSE	$\text{Bias}_{\text{total}}$	Bias_{med}	MSE
$N=200$	Prop AE	0.53(0.18)	0.33(0.21)	0.89(0.11)	1.10(0.27)	0.80(0.62)	0.99(0.10)
	Prop FM	0.35(0.21)	0.26(0.18)	0.89(0.10)	0.54(0.48)	0.68(0.60)	1.00(0.10)
	Causal Forest	1.41(0.28)	–	1.13(0.15)	3.77(0.49)	–	1.46(0.12)
	XLearner	1.36(0.26)	–	1.11 (0.13)	3.33(0.37)	–	1.49(0.19)
	LSEM	0.64(0.13)	1.14(0.13)	0.96(0.10)	1.19(0.23)	2.76(0.27)	1.09(0.11)
	HIMA	1.29(0.24)	1.44(0.21)	1.06(0.13)	2.42(0.52)	3.59(0.47)	1.21(0.11)
$N=800$	Prop AE	0.51(0.12)	0.14(0.13)	0.86(0.04)	0.81(0.20)	0.32(0.19)	0.96(0.03)
	Prop FM	0.18(0.15)	0.12(0.08)	0.86(0.04)	0.21(0.17)	0.26(0.19)	0.95(0.03)
	Causal Forest	1.48(0.11)	–	1.00(0.05)	4.09(0.17)	–	1.22(0.06)
	XLearner	1.44(0.12)	–	0.99(0.04)	3.71(0.14)	–	1.24(0.06)
	LSEM	0.66(0.05)	1.15(0.06)	0.93(0.04)	1.18(0.14)	2.73(0.13)	1.06(0.03)
	HIMA	1.26(0.12)	1.38(0.09)	1.04(0.07)	2.31(0.19)	3.49(0.19)	1.15(0.05)
$N=2000$	Prop AE	0.43(0.10)	0.13(0.08)	0.87(0.02)	0.75(0.20)	0.21(0.16)	0.94(0.04)
	Prop FM	0.09(0.09)	0.12(0.10)	0.87(0.03)	0.12(0.09)	0.24(0.18)	0.96(0.04)
	Causal Forest	1.46(0.08)	–	0.96(0.03)	4.16(0.11)	–	1.15(0.04)
	XLearner	1.45(0.07)	–	0.96(0.03)	3.76(0.07)	–	1.17(0.05)
	LSEM	0.64(0.04)	1.13(0.04)	0.93(0.03)	1.17(0.06)	2.73(0.07)	1.04(0.03)
	HIMA	1.27(0.07)	1.42(0.06)	1.04(0.04)	2.34(0.12)	3.51(0.10)	1.15(0.04)

NOTE: The number of mediators are $k=2$ and $k=5$, and the sample size varies from $N=200$ to $N=2000$.

the inferred confounder to disentangle the confounding among mediators and outcome, and leads to a lower estimation bias. Notice that for the proposed approach, the debias performance based on factor model is better than autoencoder. This is because the factor model can better capture the linear structure of confounding effects on mediators with a lower model complexity than the autoencoder under the current setting. Given that Causal Forest and XLeaner do not provide the decomposition of causal effect, we compare the proposed method with HIMA and LSEM regarding the mediation effect estimation. The results are illustrated in Table 1. Similar to the performance in causal effect estimation, the estimation bias of mediation effect from HIMA and LSEM arises significantly as the number of mediators increases while the proposed method retains a low estimation bias. In addition, our method benefits from increasing sample size in terms of reducing bias. Our simulations show the advantage of incorporating a surrogate confounder, which can significantly reduce the confoundness among multiple mediation pathways. In Table 1, the estimation bias of the treatment effect increases as the number of mediators increases is due to the estimation bias of each individual mediator's effect accumulating when estimating the total treatment effect. However, the average estimation bias $\frac{1}{k+1}|(\hat{\alpha}-\alpha)+\sum_{i=1}^{k}(\hat{\beta}_i^M\hat{\beta}_i^Y-\beta_i^{(M)}\beta_i^{(Y)})|$ decreases as the number of mediators increases.

Furthermore, the performance comparison on outcome prediction is presented in Table 1. The proposed method consistently achieves a lower prediction error on the testing dataset, and its improvement over the best competing method LSEM increases as the sample size N increases. Table 1 shows that our deconfounding algorithm can utilize multiple mediators to recover partial unobserved effects of latent confounders from outcomes. We conduct simulation studies to investigate robustness of the proposed method when the underlying causal mediation pathway in our setting is misspecified. Specifically, we consider the case when multiple mediators are dependent among each other conditioning on underlying confounder U, and the case when U does not affect treatment or outcome. The following simulation results show that our method can achieve lower estimation bias compared with competing methods, and is relatively robust to the misspecification of confounding structure when the sample size is large. The detailed settings and results are provided in Sections 8 and 12 of supplemental materials.

6.2. Nonlinear Confounding Effect

In this section, we perform numerical comparisons when the observed data $(T, \boldsymbol{M}, Y)$ are generated following the sequential models (19) in Section 6.1, and the confounding effects on multiple mediators and outcome $\boldsymbol{G}(\boldsymbol{U}) = \{g_{M^{(1)}}(\boldsymbol{U}), \ldots, g_{M^{(k)}}(\boldsymbol{U}), g_Y(\boldsymbol{U})\}$ are generated as different nonlinear functions of $\boldsymbol{U}$. We first consider the low-rank confounding effect as follows:

$$\begin{aligned} g_{M^{(k)}}(\boldsymbol{U}) &= \text{Piecewise}(a^{(k)}, b^{(k)}), k=1,2,3, \\ g_{M^{(4)}}(\boldsymbol{U}) &= \sin(\boldsymbol{U}), \\ g_{M^{(5)}}(\boldsymbol{U}) &= g_{M^{(1)}}(\boldsymbol{U}) \times \cos(\boldsymbol{U}), \\ g_Y(\boldsymbol{U}) &= g_{M^{(1)}}(\boldsymbol{U}) \times \exp(-\boldsymbol{U}/6), \end{aligned} \tag{20}$$

where $\text{Piecewise}(a,b)$ is denoted as the piece-wise function on $\boldsymbol{U}$ that is, and $\text{Piecewise}(a,b) = \sum_{l=1}^{|a|} a_l \mathbb{1}_{\{b_l \le \boldsymbol{U} < b_{l+1}\}}$, with a, b are the piece-wise function values and cutoffs. We set $a^{(1)} = (1,2,-1,-2,-3)$, $a^{(2)} = (-2,0.5,1,2,3,4)$, $a^{(3)} = (-1,2,3)$; and $b^{(1)} = (-\infty,-3,-1,1,3,\infty)$, $b^{(2)} = (-\infty,-4,-2,0,2,4,\infty)$, $b^{(3)} = (-\infty,-3,3,\infty)$ according to the above data generating process. In addition to the above settings in (20), we also investigate other settings of nonlinear confounding effect as follows:

$$\begin{aligned} g_{M^{(1)}}(\boldsymbol{U}) &= \mathbb{1}_{\{T=0\}}\left\{\text{Piecewise}(a^{(4)}, b^{(1)})\right\} + \mathbb{1}_{\{T=0\}} \\ &\quad \left\{\text{Piecewise}(a^{(5)}, b^{(1)})\right\}, \\ g_{M^{(2)}}(\boldsymbol{U}) &= (0.5+T) \times \text{Piecewise}(a^{(6)}, b^{(2)}), \\ g_{M^{(3)}}(\boldsymbol{U}) &= 2 \times \mathbb{1}_{\{T=0\}} \\ &\quad \left\{\text{Piecewise}(a^{(7)}, b^{(3)})\right\} + 2 \times \mathbb{1}_{\{T=0\}}\left\{\text{Piecewise}(a^{(8)}, b^{(3)})\right\}, \end{aligned}$$

$$g_{M^{(4)}}(\boldsymbol{U}) = 2T \times \left\{ \sin(\boldsymbol{U}) + 0.2 \right\},\ g_{M^{(5)}}(\boldsymbol{U}) = g_{M^{(1)}}(\boldsymbol{U}) \times \left\{ \cos(\boldsymbol{U}) + 0.5 \right\}, g_Y(\boldsymbol{U}) = \exp(-\boldsymbol{U}/8 + 0.9T), \tag{21}$$

with $a^{(4)} = (1, 2, -2, -1, 1)$, $a^{(5)} = (2, 3, 0, -1, 2)$, $a^{(6)} = (-2, 0.5, 1, 2, 1, -1)$, $a^{(7)} = (-1, 0, -1)$, and $a^{(8)} = (-0.5, 1 - 0.5)$. Although both (20) and (21) are highly nonlinear in terms of $\boldsymbol{U}$, the former setting leads to a low-rank structure in $\boldsymbol{G}(\boldsymbol{U})$ in that the first principle component contributes about 55% total variation to $\{\boldsymbol{G}(U_i)\}_{i=1}^N$, while the proportion of the first component only explains about 38% of the total variation of $\{\boldsymbol{G}(U_i)\}_{i=1}^N$ for the latter setting. In other words, the confounder $\boldsymbol{U}$ can be captured via a linear combination of mediator-wise confounding effects for the former setting. In addition, setting (21) introduces the interaction between treatment and confounder in the confounding effects to mimic the post-treatment effect, where association between treatment and mediators, and between mediators and response are influenced by the treatment. The sample size varies from $N = 200, 800, 2000$ and $N = 1000, 2000, 3000$ for settings (20) and (21), respectively.

The performance of causal effect estimation and mediation effect estimation are illustrated in Figure 4 and Table 2. Under the low-rank setting, the proposed method with factor modeling achieves a lower estimation bias, and its improvement over the best competing method LSEM increases as the sample size increases. The improvement patterns are similar to the ones under linear confounding effects in Section 6.1, showing that the proposed factor modeling is able to capture the variation of subject-wise confounders regardless of the functionality of confounding effects under the low-rank structure of $\boldsymbol{G}(\boldsymbol{U})$. Due to the interactions between T and U, the data-generating process under setting (21) is in fact misspecified for the deconfounding framework here which is satisfied under a homogeneous treatment effect model (6). Correspondingly, the estimation biases from Prop FM and Prop AE both increase while these biases still remain lower than other competing methods. Our simulation results suggest that the proposed method is robust against model misspecificaiton, and still effective under the heterogeneous treatment effect setting to some extent. Different from the low-rank setting (20), the Prop AE produces smaller bias than Prop FM under the setting (21) when the relation among confounding effects of mediators and outcome is more complex, and beyond linearity. Through strong representation power from the autoencoder, Prop AE extracts more of the confounding information of $\boldsymbol{U}$ shared by $\{g_{M^{(1)}}(\boldsymbol{U}), \ldots, g_{M^{(k)}}(\boldsymbol{U}), g_Y(\boldsymbol{U})\}$ compared with factor modeling.

In terms of outcome prediction, Table 2 shows that the nonparametric methods Causal Forest and XLearner provide smaller prediction MSE than the two parametric methods HIMA and LSEM, since the tree-based methods such as Causal Forest and XLearner can better capture the nonlinear confounding effects on outcome. However, the proposed method still outperforms compared to the Causal Forest and XLearner, as we can incorporate the structural information of parallel mediators via the surrogate confounders.

7. Real Data Example

In this section, we apply the proposed deconfounding algorithm to the Normative Aging Study data obtained from the NIH dbGaP database (https://www.ncbi.nlm.nih.gov/gap/) under phs000853.v1.p1. The Normative Aging Study (NAS) is a longitudinal study conducted by the United States Department of Veterans Affairs starting from 1963, which collects phenotype data and genotype data from 657 male participants via regular physical examinations and laboratory tests. The phenotype data consists of status of coronary heart disease, diabetes, hypertension, Apolipoprotein E4 protein, and different types of white blood cells from individual blood tests. In addition, the phenotype data include the basic sociodemographic information of smoking status as a binary variable, age at death, and years of education. The genotype data contains the DNA methylation levels at 26,987 individual CpG sites for each participant, which are measured via the Infinium Human Methylation450 technique (Dedeurwaerder et al. 2014).

Studies have shown that smoking is hazardous to individual health regarding adverse effects on the quality of life and effects on death risk, and deterioration of the individuals' health status

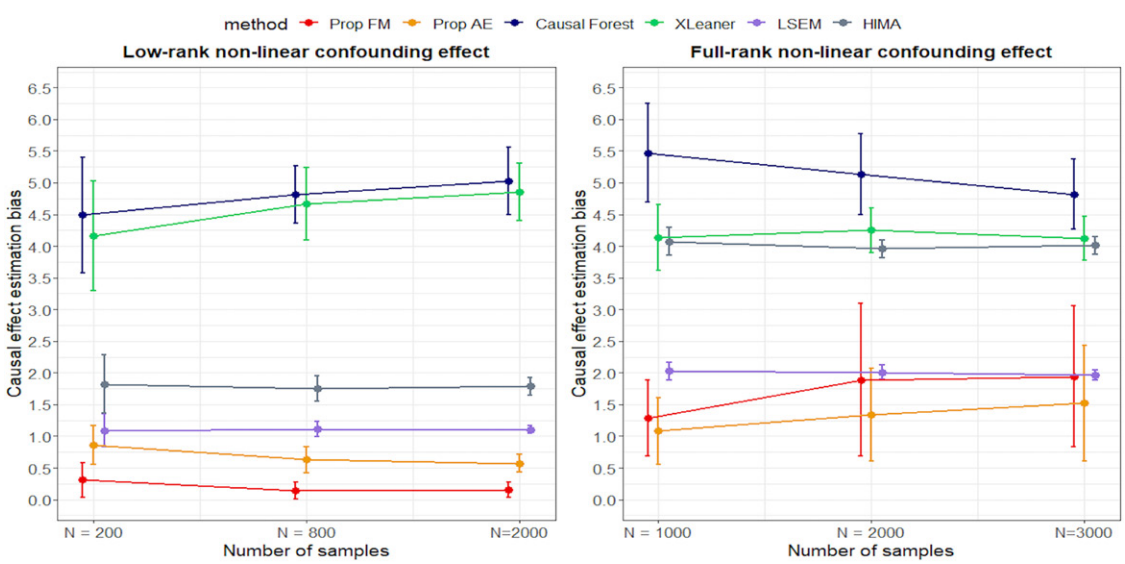

Figure 4. The bias of causal effect estimation from different methods under the setting of nonlinear confounding effects on multiple mediators and outcome.

Table 2. The performance of causal effect estimations and outcome predictions from different methods under the setting in Section 6.2.

Method	Low-rank nonlinear confounding effect				Full-rank nonlinear confounding effect			
	Sample size	$\text{Bias}_{\text{total}}$	Bias_{med}	MSE	Sample size	$\text{Bias}_{\text{total}}$	Bias_{med}	MSE
Prop AE	$N = 200$	$0.86_{0.31}$	$0.94_{0.93}$	$1.84_{0.56}$	$N = 1000$	$1.08_{0.52}$	$0.97_{0.72}$	$2.28_{0.11}$
Prop FM		$0.31_{0.27}$	$1.01_{0.78}$	$1.66_{0.57}$		$1.29_{0.60}$	$1.25_{0.43}$	$2.23_{0.11}$
Causal Forest		$4.49_{0.91}$	–	$2.23_{0.50}$		$5.47_{0.88}$	–	$2.42_{0.08}$
XLearner		$4.16_{0.86}$	–	$2.19_{0.07}$		$4.14_{0.52}$	–	$2.38_{0.09}$
LSEM		$1.09_{0.26}$	$3.55_{0.44}$	$2.53_{0.66}$		$2.03_{0.14}$	$3.50_{0.23}$	$2.59_{0.10}$
HIMA		$1.82_{0.46}$	$2.39_{0.86}$	$2.53_{0.66}$		$4.07_{0.22}$	$4.58_{0.72}$	$2.65_{0.18}$
Prop AE	$N = 800$	$0.63_{0.21}$	$0.79_{0.51}$	$1.64_{0.27}$	$N = 2000$	$1.34_{0.73}$	$1.37_{0.82}$	$2.21_{0.09}$
Prop FM		$0.14_{0.13}$	$0.79_{0.56}$	$1.64_{0.27}$		$1.89_{1.20}$	$1.93_{0.86}$	$2.19_{0.08}$
Causal Forest		$4.81_{0.45}$	–	$1.92_{0.24}$		$5.13_{0.64}$	–	$2.32_{0.09}$
XLearner		$4.67_{0.57}$	–	$1.88_{0.22}$		$4.25_{0.35}$	–	$2.30_{0.08}$
LSEM		$1.11_{0.12}$	$3.62_{0.18}$	$2.05_{0.23}$		$2.01_{0.11}$	$3.62_{0.13}$	$2.56_{0.05}$
HIMA		$1.75_{0.20}$	$2.82_{0.30}$	$2.19_{0.23}$		$3.96_{0.14}$	$4.48_{0.34}$	$2.62_{0.09}$
Prop AE	$N = 2000$	$0.57_{0.14}$	$0.57_{0.40}$	$1.44_{0.17}$	$N = 3000$	$1.52_{0.91}$	$1.59_{0.66}$	$2.21_{0.06}$
Prop FM		$0.16_{0.12}$	$0.47_{0.42}$	$1.42_{0.17}$		$1.94_{1.11}$	$1.83_{0.86}$	$2.18_{0.07}$
Causal Forest		$5.03_{0.53}$	–	$1.75_{0.15}$		$4.82_{0.55}$	–	$2.28_{0.06}$
XLearner		$4.85_{0.45}$	–	$1.72_{0.16}$		$4.12_{0.35}$	–	$2.27_{0.06}$
LSEM		$1.10_{0.06}$	$3.63_{0.13}$	$2.03_{0.13}$		$1.96_{0.08}$	$3.54_{0.17}$	$2.56_{0.06}$
HIMA		$1.79_{0.14}$	$2.77_{0.20}$	$2.27_{0.17}$		$4.01_{0.14}$	$4.56_{0.26}$	$2.63_{0.08}$

NOTE: The number of mediators is $k = 5$.

through various pathways in lifestyles and diseases (Doll and Hill 1956; Peto 1994; Carbone, Kverndokk, and Røgeberg 2005). In addition, recent studies discovered that smoking causes extensive genome-wide changes in DNA methylation (Zeilinger et al. 2013; Lee and Pausova 2013), which plays a critical role in the development and progression of cancers, and immune-system-related complex diseases (Suarez-Alvarez et al. 2012; Jin and Liu 2018). Motivated by these findings, our goal is to investigate whether the smoking habit further affects the lifespan based on the NAS data, and whether there are causal mediation pathways from smoking to lifespan through DNA methylation. Specifically, we focus on estimating the effect of smoking status on reducing the participants' lifespan, and identify mediation effects by DNA methylation levels at different CpG sites.

Given that the original DNA methylation data is high-dimensional and contains methylation levels from CpG sites which could be irrelevant to the mediation pathway, we first preprocess data to select significant mediators. Specifically, we select methylation levels at 22 CpG sites serving as mediators in the following analysis. The detailed preprocessing are provided in Section 9 in supplementary materials. We investigate the smoking effect on lifespan predictions based on the proposed method and four competing methods: LSEM, HIMA, Causal Forest, and XLearner, as detailed in Section 6. The prediction mean square error is evaluated via 5-fold cross-validation. We provide the implementation of our deconfounding algorithm and other methods in the supplementary materials. The results are provided in Table 3 where $\text{Treat}_{\text{total}}$, $\text{Treat}_{\text{dir}}$, and $\text{Treat}_{\text{med}}$ denote the estimation of total treatment effect, direct effect, and mediation effect. The prediction error indicates the medians of lifespan prediction mean square errors from multiple repeated measurements. Compared with existing methods, the proposed method detects a significantly stronger adverse effect of smoking on lifespan via incorporating the latent confounding effects. Specifically, the proposed method using either latent factor modeling or autoencoder estimates, shows that the smoking habit reduces the lifespan by about 8 years for participants in the NAS study, while estimations from other methods are less than 2 years. On the other hand, a national study based on the

Table 3. Estimations of total treatment effect, direct effect, mediation effect, and prediction error of lifespan estimations from different methods for the NAS data where Prop FM and Prop AE stand for the proposed methods modeling confounding effect with factor model and autoencoder, respectively.

	$\text{Treat}_{\text{total}}$	$\text{Treat}_{\text{dir}}$	$\text{Treat}_{\text{med}}$	Prediction error
Prop FM	$-8.40_{1.01}$	$-1.95_{0.87}$	$-6.45_{0.60}$	$2.94_{0.27}$
Prop AE	$-8.30_{1.06}$	$-1.58_{0.37}$	$-6.72_{1.09}$	$3.15_{0.33}$
Causal Forest	$-1.60_{0.34}$	–	–	$3.21_{0.30}$
XLearner	$-1.59_{0.22}$	–	–	$3.13_{0.27}$
LSEM	$-1.86_{0.19}$	$-1.69_{0.22}$	$-0.16_{0.22}$	$3.03_{0.29}$
HIMA	$-1.97_{0.26}$	$-1.51_{0.20}$	$-0.45_{0.14}$	$39.57_{15.58}$

2004 U.S. Census data concludes that the effect of smoking on increasing the mortality risk is similar to reducing 5–10 years in lifespan (Woloshin, Schwartz, and Welch 2008). In addition, other smoking-cancer association studies show that the average loss of life for smokers is 8 years in Europe and United States (Boyle 1997). Therefore, our method produces a treatment effect of smoking on lifespan, which is more consistent with existing data. In addition, clinical studies found that smoking can significantly increase the risk of premature death (Gavin 2004; Jha et al. 2008), which also supported a stronger negative effect of smoking on lifespan.

In addition to the direct treatment effect, both the proposed method, LSEM, and HIMA can identify the roles of DNA methylation levels as mediators to transmit the indirect effect of smoking on lifespan, supported by scientific evidence that smoking-induced DNA methylation also increases the risks of metabolic disorders, chronic diseases, diabetes, and cancers (Besingi and Johansson 2014; Tsai et al. 2018; Maas et al. 2020; Jamieson et al. 2020). However, in contrast to the existing LSEM and HIMA, the proposed method can capture stronger mediation effects of DNA methylation in that the proportion of total treatment effect of smoking is almost fully mediated by DNA methylation levels. The DNA methylation level is found to have a full mediation effect of smoking on epigenetic aging (Lei et al. 2020), 90% on bladder cancer risk (Jordahl et al. 2019), and 55% on lung functional degradation (De Vries et al. 2018). In addition, (Zhang et al. 2016b) found that the DNA methylation level is the

most informative biomarker for predicting risks from all causes and cardiovascular mortality associated with smoking (Zhang et al. 2016b). These studies indicate a significant and dominant role of DNA methylation levels in conveying the smoking effect reduction of lifespan, which is also consistent with the estimation of mediation effects from the proposed method. The distinct gaps between our deconfounding method and existing non-deconfounding methods on both total treatment effect and mediation effects suggest the usefulness of latent confounders for the NAS study. Furthermore, our deconfounding algorithm with latent factor modeling also produces a lower prediction error of lifespan estimation through incorporating partial confounding effects on lifespan. We also provide the interpretation for the latent confounder inferred from the 22 CpG mediators, and the sensitivity analysis of the causal estimation on the choice of factor models in Section 9 of the supplementary material.

8. Discussion

In this article, we propose a novel deconfounding method and algorithm to debias causal effect and mediation effect estimation in causal mediation analysis. Specifically, we consider the causal pathways with parallel mediators such that mediators are causally independent conditioning on the shared latent confounders. Our method generalizes the classic mediation analysis paradigm, and is applicable for a wide range of applications. The proposed method uses the confounder-mediator structure in multi-channel mediation pathways to infer the information of latent confounders. In addition, we provide flexible modeling on the confounding mechanism regarding the effects of latent confounders from treatment, mediators, and outcomes. The principle idea of our method is to construct surrogate confounders incorporating the confounding information instead of recovering the original confounders. In theory, we establish sequential ignorability via incorporating surrogate confounders. Accordingly, we show that both the causal effect and mediation effect can be identified based on the joint distribution of observed data and surrogate confounders. In particular, we provide identification conditions for causal effect estimation under different mediation pathway structures. Our numeric studies also confirm that the proposed method reduces the estimation bias of causal effect and mediation effect under various confounding mechanisms through the confounder-sharing structure of multiple mediators. In this article, the surrogate confounders are jointly estimated. In order to incorporate new observations, we are required to re-train the model on the entire dataset to update surrogate confounders for both historical and new observations. The proposed method can be modified using an online learning scheme in that deconfounding for new observations can be computationally independent from historical data.

Supplementary Materials

The supplementary materials provide proofs of the Lemma 5.1, Theorem 5.1, Corollary 5.1, Theorem 5.2, deconfounding algorithm using autoencoder, illustrations of numerical comparisons in Section 6, and data preprocessing, implementation and interpretation for real data application, sensitivity analysis, simulations under misspecified causal pathways, discussion on examples of real applications, and discussion on relation to deconfounder for multiple causes.

Acknowledgments

The authors thank the Associate Editor and anonymous reviewers for their suggestions and helpful feedback which improved the article significantly.

Disclosure Statement

The authors report there are no competing interests to declare.

Funding

This work was supported by NSF Grants DMS 1952406 and DMS 2210640.

References

Ballard, D. H. (1987), "Modular Learning in Neural Networks," in *AAAI* (Vol. 647), pp. 279–284. [2055]

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. [2051]

Besingi, W., and Johansson, Å. (2014), "Smoke-Related DNA Methylation Changes in the Etiology of Human Disease," *Human Molecular Genetics*, 23, 2290–2297. [2062]

Boyle, P. (1997), "Cancer, Cigarette Smoking and Premature Death in Europe: A Review Including the Recommendations of European Cancer Experts Consensus Meeting, Helsinki, October 1996," *Lung Cancer*, 17, 1–60. [2062]

Cai, H., Song, R., and Lu, W. (2020), "Anoce: Analysis of Causal Effects with Multiple Mediators via Constrained Structural Learning," in *International Conference on Learning Representations*. [2051]

Cai, Z., Li, R., and Zhang, Y. (2022), "A Distribution Free Conditional Independence Test with Applications to Causal Discovery," *Journal of Machine Learning Research*, 23, 1–41. [2058]

Carbone, J. C., Kverndokk, S., and Røgeberg, O. J. (2005), "Smoking, Health, Risk, and Perception," *Journal of Health Economics*, 24, 631–653. [2062]

Chen, Y., Li, X., and Zhang, S. (2020), "Structured Latent Factor Analysis for Large-Scale Data: Identifiability, Estimability, and their Implications," *Journal of the American Statistical Association*, 115, 1756–1770. [2057]

De Vries, M., van der Plaat, D. A., Nedeljkovic, I., Verkaik-Schakel, R. N., Kooistra, W., Amin, N., van Duijn, C. M., Brandsma, C.-A., van Diemen, C. C., Vonk, J. M., et al. (2018), "From Blood to Lung Tissue: Effect of Cigarette Smoke on DNA Methylation and Lung Function," *Respiratory Research*, 19, 1–9. [2062]

Dedeurwaerder, S., Defrance, M., Bizet, M., Calonne, E., Bontempi, G., and Fuks, F. (2014), "A Comprehensive Overview of Infinium Humanmethylation450 Data Processing," *Briefings in Bioinformatics*, 15, 929–941. [2061]

Doll, R., and Hill, A. B. (1956), "Lung Cancer and Other Causes of Death in Relation to Smoking," *British Medical Journal*, 2, 1071. [2062]

Dukes, O., Shpitser, I., and Tchetgen, E. J. T. (2021), "Proximal Mediation Analysis," arXiv preprint arXiv:2109.11904. [2052]

Forastiere, L., Mattei, A., and Ding, P. (2018), "Principal Ignorability in Mediation Analysis: Through and Beyond Sequential Ignorability," *Biometrika*, 105, 979–986. [2053]

Ford, C. G., and Shook, N. J. (2019), "Negative Cognitive Bias and Perceived Stress: Independent Mediators of the Relation between Mindfulness and Emotional Distress," *Mindfulness*, 10, 100–110. [2056]

Fu, X., Huang, K., Sidiropoulos, N. D., and Ma, W.-K. (2019), "Nonnegative Matrix Factorization for Signal and Data Analytics: Identifiability,

Algorithms, and Applications," *IEEE Signal Processing Magazine*, 36, 59–80. [2057]

Gavin, A. (2004), "Smoking is a Major Cause of Premature Death Worldwide," *Evidence-based Healthcare*, 8, 95–96. [2062]

Gresele, L., Rubenstein, P. K., Mehrjou, A., Locatello, F., and Schölkopf, B. (2020), "The Incomplete Rosetta Stone Problem: Identifiability Results for Multi-View Nonlinear ICA," in *Uncertainty in Artificial Intelligence*, pp. 217–227, PMLR. [2057]

Hicks, R., and Tingley, D. (2011), "Causal Mediation Analysis," *The Stata Journal*, 11, 605–619. [2051]

Hong, Z., and Lian, H. (2013), "Sparse-Smooth Regularized Singular Value Decomposition," *Journal of Multivariate Analysis*, 117, 163–174. [2057]

Huang, T.-M. (2010), "Testing Conditional Independence Using Maximal Nonlinear Conditional Correlation," *The Annals of Statistics*, 38, 2047–2091. [2058]

Huang, Y.-T. (2019), "Genome-Wide Analyses of Sparse Mediation Effects Under Composite Null Hypotheses," *The Annals of Applied Statistics*, 13, 60–84. [2058]

Hyvarinen, A., Sasaki, H., and Turner, R. (2019), "Nonlinear ICA Using Auxiliary Variables and Generalized Contrastive Learning," in *The 22nd International Conference on Artificial Intelligence and Statistics*, pp. 859–868, PMLR. [2057]

Imai, K., Keele, L., and Tingley, D. (2010), "A General Approach to Causal Mediation Analysis," *Psychological methods*, 15, 309–334. [2051,2056]

Imai, K., Keele, L., and Yamamoto, T. (2010), "Identification, Inference and Sensitivity Analysis for Causal Mediation Effects," *Statistical Science*, 25, 51–71. [2053,2056]

Imbens, G. W., and Rubin, D. B. (2015), *Causal Inference in Statistics, Social, and Biomedical Sciences*, Cambridge: Cambridge University Press. [2051]

Jamieson, E., Korologou-Linden, R., Wootton, R. E., Guyatt, A. L., Battram, T., Burrows, K., Gaunt, T. R., Tobin, M. D., Munafò, M., Smith, G. D., et al. (2020), "Smoking, DNA Methylation, and Lung Function: A Mendelian Randomization Analysis to Investigate Causal Pathways," *The American Journal of Human Genetics*, 106, 315–326. [2062]

Jha, P., Jacob, B., Gajalakshmi, V., Gupta, P. C., Dhingra, N., Kumar, R., Sinha, D. N., Dikshit, R. P., Parida, D. K., Kamadod, R., et al. (2008), "A Nationally Representative Case–Control Study of Smoking and Death in India," *New England Journal of Medicine*, 358, 1137–1147. [2062]

Jin, Z., and Liu, Y. (2018), "DNA Methylation in Human Diseases," *Genes and Diseases*, 5, 1–8. [2062]

Jordahl, K. M., Phipps, A. I., Randolph, T. W., Tindle, H. A., Liu, S., Tinker, L. F., Kelsey, K. T., White, E., and Bhatti, P. (2019), "Differential DNA Methylation in Blood as a Mediator of the Association between Cigarette Smoking and Bladder Cancer Risk among Postmenopausal Women," *Epigenetics*, 14, 1065–1073. [2062]

Khemakhem, I., Kingma, D., Monti, R., and Hyvarinen, A. (2020), "Variational Autoencoders and Nonlinear ICA: A Unifying Framework," in *International Conference on Artificial Intelligence and Statistics*, pp. 2207–2217, PMLR. [2057]

Künzel, S. R., Sekhon, J. S., Bickel, P. J., and Yu, B. (2019), "Metalearners for Estimating Heterogeneous Treatment Effects Using Machine Learning," *Proceedings of the National Academy of Sciences*, 116, 4156–4165. [2051]

Kuroki, M. and Pearl, J. (2014), "Measurement Bias and Effect Restoration in Causal Inference," *Biometrika*, 101, 423–437. [2058]

Lee, K. W., and Pausova, Z. (2013), "Cigarette Smoking and DNA Methylation," *Frontiers in Genetics*, 4, 132. [2062]

Lei, M.-K., Gibbons, F. X., Simons, R. L., Philibert, R. A., and Beach, S. R. (2020), "The Effect of Tobacco Smoking Differs across Indices of DNA Methylation-based Aging in an African American Sample: DNA Methylation-based Indices of Smoking Capture these Effects," *Genes*, 11, 311. [2062]

Louizos, C., Shalit, U., Mooij, J., Sontag, D., Zemel, R., and Welling, M. (2017), "Causal Effect Inference with Deep Latent-Variable Models," arXiv preprint arXiv:1705.08821. [2052]

Maas, S. C., Mens, M. M., Kühnel, B., van Meurs, J. B., Uitterlinden, A. G., Peters, A., Prokisch, H., Herder, C., Grallert, H., Kunze, S., et al. (2020), "Smoking-Related Changes in DNA Methylation and Gene Expression are Associated with Cardio-Metabolic Traits," *Clinical Epigenetics*, 12, 1–16. [2062]

MacKinnon, D. P. (2012), *Introduction to Statistical Mediation Analysis*, New York: Routledge. [2051,2053]

Manski, C. F. (2009), *Identification for Prediction and Decision*, Cambridge, MA: Harvard University Press. [2056]

Miao, W., Geng, Z., and Tchetgen Tchetgen, E. J. (2018), "Identifying Causal Effects with Proxy Variables of an Unmeasured Confounder," *Biometrika*, 105, 987–993. [2052,2058]

Ogarrio, J. M., Spirtes, P., and Ramsey, J. (2016), "A Hybrid Causal Search Algorithm for Latent Variable Models," in *Conference on Probabilistic Graphical Models*, pp. 368–379. PMLR. [2051]

Peto, R. (1994), "Smoking and Death: The Past 40 Years and the Next 40," *BMJ*, 309, 937–939. [2062]

Ranganath, R., and Perotte, A. (2018), "Multiple Causal Inference with Latent Confounding," arXiv preprint arXiv:1805.08273. [2052]

Rubin, D. B. (2005), "Causal Inference Using Potential Outcomes: Design, Modeling, Decisions," *Journal of the American Statistical Association*, 100, 322–331. [2052]

Seghatol-Eslami, V. C., Dark, H., Raper, J. L., Mugavero, M. J., Turan, J. M., and Turan, B. (2017), "Interpersonal and Intrapersonal Factors as Parallel Independent Mediators in the Association between Internalized HIV Stigma and Art Adherence," *Journal of Acquired Immune Deficiency Syndromes*, 74, e18. [2056]

Shi, C., and Li, L. (2021), "Testing Mediation Effects Using Logic of Boolean Matrices," *Journal of the American Statistical Association*, 117, 2014–2027. [2051]

Spirtes, P. L., Meek, C., and Richardson, T. S. (2013), "Causal Inference in the Presence of Latent Variables and Selection Bias," arXiv preprint arXiv:1302.4983. [2051]

Su, L. and White, H. (2008), "A Nonparametric Hellinger Metric Test for Conditional Independence," *Econometric Theory*, 24, 829–864. [2058]

Suarez-Alvarez, B., Rodriguez, R. M., Fraga, M. F., and López-Larrea, C. (2012), "DNA Methylation: A Promising Landscape for Immune System-Related Diseases," *Trends in Genetics*, 28, 506–514. [2062]

Tan, Z. (2006), "Regression and Weighting Methods for Causal Inference Using Instrumental Variables," *Journal of the American Statistical Association*, 101, 1607–1618. [2052]

Tang, X., Xue, F., and Qu, A. (2021), "Individualized Multidirectional Variable Selection," *Journal of the American Statistical Association*, 116, 1280–1296. [2057]

Tsai, P.-C., Glastonbury, C. A., Eliot, M. N., Bollepalli, S., Yet, I., Castillo-Fernandez, J. E., Carnero-Montoro, E., Hardiman, T., Martin, T. C., Vickers, A., et al. (2018), "Smoking Induces Coordinated DNA Methylation and Gene Expression Changes in Adipose Tissue with Consequences for Metabolic Health," *Clinical Epigenetics*, 10, 1–21. [2062]

Wager, S., and Athey, S. (2018), "Estimation and Inference of Heterogeneous Treatment Effects Using Random Forests," *Journal of the American Statistical Association*, 113, 1228–1242. [2051]

Wang, K., Franks, A., and Oh, S.-Y. (2021), "Learning Gaussian Graphical Models with Latent Confounders," arXiv preprint arXiv:2105.06600. [2051,2052]

Wang, W.-T., and Huang, H.-C. (2017), "Regularized Principal Component Analysis for Spatial Data," *Journal of Computational and Graphical Statistics*, 26, 14–25. [2057]

Wang, Y., and Blei, D. M. (2019), "The Blessings of Multiple Causes," *Journal of the American Statistical Association*, 114, 1574–1596. [2052]

Willetts, M., and Paige, B. (2021), "I don't Need U: Identifiable Nonlinear ICA Without Side Information," arXiv preprint arXiv:2106.05238. [2057]

Witty, S., Takatsu, K., Jensen, D., and Mansinghka, V. (2020), "Causal Inference Using Gaussian Processes with Structured Latent Confounders," in *International Conference on Machine Learning*, pp. 10313–10323, PMLR. [2052]

Woloshin, S., Schwartz, L. M., and Welch, H. G. (2008), "The Risk of Death by Age, Sex, and Smoking Status in the United States: Putting Health Risks in Context," *Journal of the National Cancer Institute*, 100, 845–853. [2062]

Zeilinger, S., Kühnel, B., Klopp, N., Baurecht, H., Kleinschmidt, A., Gieger, C., Weidinger, S., Lattka, E., Adamski, J., Peters, A., et al. (2013), "Tobacco Smoking Leads to Extensive Genome-Wide Changes in DNA Methylation," *PloS One*, 8, e63812. [2062]

Zhang, H., Zheng, Y., Zhang, Z., Gao, T., Joyce, B., Yoon, G., Zhang, W., Schwartz, J., Just, A., Colicino, E., et al. (2016a), "Estimating and Testing High-Dimensional Mediation Effects in Epigenetic Studies," *Bioinformatics*, 32, 3150–3154. [2051]

Zhang, K., Peters, J., Janzing, D., and Schölkopf, B. (2012), "Kernel-based Conditional Independence Test and Application in Causal Discovery," arXiv preprint arXiv:1202.3775. [2058]

Zhang, Y., Schöttker, B., Florath, I., Stock, C., Butterbach, K., Holleczek, B., Mons, U., and Brenner, H. (2016b), "Smoking-Associated DNA Methylation Biomarkers and their Predictive Value for All-Cause and Cardiovascular Mortality," *Environmental Health Perspectives*, 124, 67–74. [2062,2063]

Zhao, Y., Lindquist, M. A., and Caffo, B. S. (2020), "Sparse Principal Component based High-Dimensional Mediation Analysis," *Computational Statistics and Data Analysis*, 142, 106835. [2058]

Zhao, Y., and Luo, X. (2016), "Pathway Lasso: Estimate and Select Sparse Mediation Pathways with High Dimensional Mediators," arXiv preprint arXiv:1603.07749. [2058]

Zhou, Y., Kong, D., and Wang, L. (2020), "The Promises of Parallel Outcomes," arXiv preprint arXiv:2012.05849. [2052]