Submitted to the Annals of Statistics
arXiv: arXiv:0000.0000

EFFICIENT NONPARAMETRIC ESTIMATION OF CAUSAL MEDIATION EFFECTS

By K.C.G. Chan[†], K. Imai[‡], S.C.P. Yam[§], and Z. Zhang[¶]

University of Washington[†], Princeton University[‡],
The Chinese University of Hong Kong[§], The University of Hong Kong[¶]

An essential goal of program evaluation and scientific research is the investigation of causal mechanisms. Over the past several decades, causal mediation analysis has been used in medical and social sciences to decompose the treatment effect into the natural direct and indirect effects. However, all of the existing mediation analysis methods rely on parametric modeling assumptions in one way or another, typically requiring researchers to specify multiple regression models involving the treatment, mediator, outcome, and pre-treatment confounders. To overcome this limitation, we propose a novel nonparametric estimation method for causal mediation analysis that eliminates the need for applied researchers to model multiple conditional distributions. The proposed method balances a certain set of empirical moments between the treatment and control groups by weighting each observation; in particular, we establish that the proposed estimator is *globally* semiparametric efficient. We also show how to consistently estimate the asymptotic variance of the proposed estimator without additional efforts. Finally, we extend the proposed method to other relevant settings including the causal mediation analysis with multiple mediators.

1. Introduction. In program evaluation and scientific research, an essential goal is to understand why and how a treatment variable influences the outcomes of interest, going beyond the estimation of the average treatment effects. In this regard, causal mediation analysis plays an important role in the investigation of causal mechanisms by decomposing the treatment effect into the natural direct and indirect effects (Robins and Greenland, 1992;

[*]K.C.G. Chan thanks the United States National Institutes of Health for support (R01 HL122212, R01 AI121259) ; Kosuke Imai thanks the United States National Science Foundation for support (SES–0918968); Phillip Yam acknowledges the financial support from the Hong Kong RGC GRF 14301015, and Direct Grant for Research 2014/15 with project code: 4053141 offered by CUHK; Zheng Zhang acknowledges the financial support from the Chinese University of Hong Kong and the University of Hong Kong; the present work constitutes part of his research study leading to his Ph.D thesis in the Chinese University of Hong Kong.

MSC 2010 subject classifications: Primary 60K35, 60K35; secondary 60K35

Keywords and phrases: Exponential tilting, Natural direct effects, Natural indirect effects, Treatment effects, Semiparametric efficiency

Pearl, 2001; Robins, 2003). Such an approach has been widely used in a number of disciplines in medical and social sciences (see e.g., Baron and Kenny, 1986; Imai et al., 2011; MacKinnon, 2008; VanderWeele, 2015). The methodological literature on causal mediation analysis has also rapidly grown over the last decade and produced numerous approaches and extensions (see, for example, Albert, 2008; Geneletti, 2007; Imai, Keele and Yamamoto, 2010; Jo, 2008; Joffe et al., 2007; Sobel, 2008; Ten Have et al., 2007; Tchetgen Tchetgen and Shpitser, 2012; VanderWeele, 2009; VanderWeele and Vansteelandt, 2010).

In this article, we here contribute to this fast growing literature by developing a new efficient nonparametric estimation method for causal mediation analysis. All of the existing mediation analysis methods rely on parametric modeling assumptions in one way or another, typically requiring researchers to specify multiple regression models involving the treatment T, mediator M, outcome Y, and pre-treatment confounders X. For example, the standard approach based on the so called "mediation formula" require the specification of two or three conditional distributions, i.e. $f_{Y|M,T,X}$, $f_{M|T,X}$, and possibly $f_{T|X}$ (for example, Imai, Keele and Tingley, 2010; Pearl, 2012; VanderWeele, 2009). Inference under this standard approach is only valid when both the outcome model $f_{Y|M,T,X}$ and the mediator model $f_{M|T,X}$ are correctly specified. Our proposed method eliminates the need for applied researchers to model these multiple conditional distributions a priori.

We are inspired by the recent work of Tchetgen Tchetgen and Shpitser (2012) who develop a robust semiparametric estimation procedure to allow for possible model misspecification. The authors show that their proposed estimator is consistent when any two out of three chosen models are correctly specified and is locally semiparametric efficient whenever all three models are correct. While this estimator represents an important advance in the literature, its validity still relies upon the correct specification of multiple parametric or semiparametric models. We improve this estimator by proposing a *globally* semiparametric efficient estimator that attains the semiparametric efficiency bound, derived by Tchetgen Tchetgen and Shpitser (2012), without imposing the additional structural assumptions required for the existing semiparametric estimator. To the best of our knowledge, no globally semiparametric efficient estimator has been proposed in the causal mediation literature.

Our proposed estimator is based on a strategy of balancing covariates by weighting each observation, which has recently become popular when estimating the average treatment effects (for example, Chan, Yam and Zhang, 2015; Hainmueller, 2012; Graham, Pinto and Egel, 2012; Imai and Ratkovic, 2014). We combine this idea with the construction of globally semiparametric efficient estimation of the average treatment effects (see, Chen, Hong and Tarozzi, 2008; Hahn, 1998; Hirano, Imbens and Ridder, 2003; Imbens, Newey and Ridder, 2005). Unlike these plugin-type globally semiparametric efficient estimators that require semi-parametric estimation of the propensity score or the outcome regression function, we adopt the nonparametric calibration approach developed by Chan, Yam and Zhang (2015) that constructs observation-specific weights only from covariate balancing conditions. This is a significant advantage in causal mediation analysis because the plugin-type globally semiparametric estimators would require the semi-parametric estimation of three conditional distributions, which is a difficult task in practice yielding more doubt on the robustness of the estimators.

The rest of the paper is organized as follows. In Section 2, we describe the proposed estimation method, which matches the certain moment conditions of the mediator and pre-treatment covariates between the treatment and control groups. We then show how to consistently estimate the asymptotic variance of the proposed estimator without additional functional estimation. In Section 3, we extend our method to the case of multiple mediators studied in Imai and Yamamoto (2013). We then discuss two related estimation problems, namely the estimation of pure indirect effects and natural direct effect of the untreated. Finally, we apply the proposed methods to two data sets in Section 4 and offer concluding remarks in Section 5.

2. The Proposed Methodology.

In this section, we first consider the efficient nonparametric estimation of the average natural direct and indirect effects. In Theorem 1, we shall show that the proposed nonparametric estimator is consistent, asymptotically normal, and globally semiparametric efficient. We then demonstrate how to nonparametrically estimate the asymptotic variance of the proposed estimator.

2.1. *The framework.*

Suppose that we have a binary treatment variable $T \in \{0, 1\}$. Under the standard framework of causal inference, we let $M(t)$ denote a potential mediating variable, which represents the value of the mediator if the treatment variable is equal to $t \in \{0, 1\}$. Similarly, let $Y(t, m)$ represent the potential outcome variable under the scenario where the treatment and mediator variables take the value t and m, respectively. Then, the observed mediator M is given by $M = TM(1) + (1 - T)M(0)$ whereas the observed outcome is equal to $Y = TY(1, M(1)) + (1 - T)Y(0, M(0))$. We assume that we have a simple random sample of size N from a population and therefore observe the *i.i.d.* realizations of these random variables, $\{T_i, M_i, Y_i, X_i\}_{i=1}^N$ where X is a vector of pretreatment covariates.

A primary goal of causal mediation analysis is the following decomposition of the average treatment effect into the average natural indirect effect (or average causal mediation effect) and the average natural direct effect (Robins and Greenland, 1992; Pearl, 2001; Robins, 2003)

$$\begin{aligned} & \mathbb{E}[Y(1, M(1)) - Y(0, M(0))] \\ = \;& \mathbb{E}[Y(1, M(1)) - Y(1, M(0))] + \mathbb{E}[Y(1, M(0)) - Y(0, M(0))] \end{aligned} \tag{2.1}$$

The average natural indirect effect, which is the first term in this equation, is the average difference between the potential outcome under the treatment condition and the counterfactual outcome under the treatment condition where the mediator is equal to the value that would have realized under the control condition. This quantity represents the average difference that would result if the mediator value changes from $M(0)$ to $M(1)$ while holding the treatment variable constant at $T = 1$. In contrast, the average natural direct effect, which is the second term in equation (2.1), represents the average treatment effect when the mediator is held constant at $M(0)$. Therefore, this decomposition enables researchers to explore how much of the treatment effect is due to the change in the mediator.

Note that the following alternative decomposition for causal mediation analysis is also possible,

$$\begin{aligned} & \mathbb{E}[Y(1, M(1)) - Y(0, M(0))] \\ = \;& \mathbb{E}[Y(0, M(1)) - Y(0, M(0))] + \mathbb{E}[Y(1, M(1)) - Y(0, M(1))] \end{aligned} \tag{2.2}$$

where the treatment variable is held constant at $T = 0$ for the natural indirect effect and the mediator is fixed at $M(1)$ for the natural direct effect. Robins (2003) called this version of the natural indirect effect as the *pure indirect effect* while referring the natural indirect effect given in equation (2.1) as the *total indirect effect* since it is resulted from both treatment and mediator. Our proposed estimator is applicable to both cases as the difference between the two decompositions solely depends on the value at which the treatment is fixed.

To nonparametrically identify the average natural direct and indirect effects, we rely on the following set of assumptions as in Imai, Keele and Tingley (2010) and Tchetgen Tchetgen and Shpitser (2012).

Assumption 1.

1. *(Consistency) If $T = t$, then $M = M(t)$ with probability 1 for $t \in \{0, 1\}$. If $T = t$ and $M = m$, then $Y = Y(t, m)$ with probability 1 for $t \in \{0, 1\}$ and $m \in \mathcal{M}$, where $\mathcal{M}$ is the support of the distribution of M.*

2. *(Sequential Ignorability) Given* X, $\{Y(t',m), M(t)\}$ *is independent of* T *for* $t, t' \in \{0,1\}$. *Also, given* $T = t$ *and* X, $Y(t',m)$ *is independent of* $M(t)$ *for* $t, t' \in \{0,1\}$ *and* $m \in \mathcal{M}$.
3. *(Positivity) With probability 1 with respect to any* (t,x) *where* $t \in \{0,1\}$ *and* $x \in \mathcal{X}$, $f_{M|T,X}(m \mid t,x) > 0$ *for all* $m \in \mathcal{M}$ *where* $\mathcal{X}$ *is the support of* X. *With probability 1 with respect to any* $x \in \mathcal{X}$, $f_{T|X}(t \mid x) > 0$ *for all* $t \in \{0,1\}$.

The sequential ignorability assumption is a natural extension of the unconfoundedness assumption for the identification of the average treatment effect except that it requires the "cross-world" independence between $Y(t',m)$ and $M(t)$ (see, Richardson and Robins, 2013). Several researchers have proposed different sensitivity analysis techniques for estimating the bias that arises when this assumption is violated (see, Imai, Keele and Yamamoto, 2010; VanderWeele, 2010; Tchetgen Tchetgen and Shpitser, 2012). Under Assumption 1, Imai, Keele and Yamamoto (2010) showed that the average natural direct and indirect effects are nonparametrically identified. That is,

$$\begin{aligned}\theta_t &= \mathbb{E}(Y(1-t, M(t))) \\ &= \int\int \mathbb{E}(Y \mid T = 1-t, M = m, X = x) f_{M|T,X}(m \mid T = t, X = x) f_X(x) dx dm\end{aligned}$$

and

$$\delta_t = \mathbb{E}(Y(t, M(t))) = \int \mathbb{E}(Y \mid T = t, X = x) f_X(x) dx$$

for $t = 0, 1$.

Tchetgen Tchetgen and Shpitser (2012) made an important theoretical advance by showing that under Assumption 1 the efficient influence function of θ_t is given by,

$$\begin{aligned}S_{\theta_t} &= \frac{\mathbf{1}\{T = 1-t\} f_{M|T,X}(M \mid T = t, X)}{f_{T|X}(1-t \mid X) f_{M|T,X}(M \mid T = 1-t, X)}\{Y - \mathbb{E}(Y \mid X, M, T = 1-t)\} \\ &\quad + \frac{\mathbf{1}\{T = t\}}{f_{T|X}(t \mid X)}\{\mathbb{E}(Y \mid X, M, T = 1-t) - \eta(1-t, t, X)\} + \eta(1-t, t, X) - \theta_t \end{aligned} \tag{2.3}$$

where

$$\eta(t, t', X) = \int \mathbb{E}(Y \mid X, M = m, T = t) f_{M|T,X}(m \mid T = t', X) dm$$

for $t, t' \in \{0,1\}$. Hence, the definition of η implies that $\eta(1,1,X) = \mathbb{E}(Y \mid X, T=1)$ and $\eta(0,0,X) = \mathbb{E}(Y \mid X, T=0)$.

Furthermore, the efficient influence functions of the average natural direct effect when $t=0$, i.e., $\mathsf{NDE} = \theta_0 - \delta_0$, and the average natural indirect effect when $t=1$ (or the average total indirect effect), i.e., $\mathsf{NIE} = \delta_1 - \theta_0$, are $S_{\mathsf{NDE}} = S_{\theta_0} - S_{\delta_0}$ and $S_{\mathsf{NIE}} = S_{\delta_1} - S_{\theta_0}$, respectively, where S_{δ_1} and S_{δ_0} are the efficient influence functions for estimating δ_1 and δ_0. As shown in Robins, Rotnitzky and Zhao (1994) and Hahn (1998), these efficient influence functions are given by,

$$S_{\delta_t} = \frac{\mathbf{1}\{T=t\}}{f_{T|X}(t \mid X)}\{Y - \mathbb{E}(Y \mid X, T=t)\} + \mathbb{E}(Y \mid X, T=t) - \delta_t \ , \tag{2.4}$$

for $t = 0, 1$. Similarly, the average natural indirect effect when $t=0$ (or the pure indirect effect) is given by $\mathsf{PIE} = \theta_1 - \delta_0$ and its efficient function is equal to $S_{\theta_1} - S_{\delta_0}$. Therefore, the efficient estimations of the natural direct and indirect effects involve the efficient estimation of δ_t and θ_t for $t = 0, 1$.

2.2. *Efficient estimation of δ_1 and δ_0.* Before proposing a globally semiparametric efficient estimator of θ_0, which is one of the main contributions of our paper, we discuss the efficient estimation of δ_1 and δ_0, which is required for the efficient estimation of the natural direct and indirect effects. There exists an extensive literature on the globally efficient estimation of δ_0 and δ_1 in econometrics (see, for example, Chen, Hong and Tarozzi, 2008; Hahn, 1998; Hirano, Imbens and Ridder, 2003; Imbens, Newey and Ridder, 2005). However, many of these existing estimators require the semi-parametric estimation of propensity score or outcome regression model. In this paper, we focus on a globally efficient estimator recently proposed by Chan, Yam and Zhang (2015), which serves as a building block of our proposed estimator of θ_0 discussed below. Unlike the other estimators, this approach achieves the efficient nonparametric estimation by balancing covariates through weighting.

Let $p_0(x) \triangleq \frac{1}{N} f_{T|X}(1 \mid x)^{-1}$ and $q_0(x) \triangleq \frac{1}{N} f_{T|X}(0 \mid x)^{-1}$. Under Assumption 1, for any suitable integrable functions $u(x)$, the following important moment conditions hold,

$$\delta_1 = \mathbb{E}\left(\sum_{i=1}^N T_i p_0(X_i) Y_i\right) \tag{2.5}$$

$$\delta_0 = \mathbb{E}\left(\sum_{i=1}^N (1-T_i) q_0(X_i) Y_i\right) \tag{2.6}$$

$$\mathbb{E}(u(X)) = \mathbb{E}\left(\sum_{i=1}^{N} T_i p_0(X_i) u(X_i)\right) \tag{2.7}$$

$$\mathbb{E}(u(X)) = \mathbb{E}\left(\sum_{i=1}^{N} (1-T_i) q_0(X_i) u(X_i)\right) \tag{2.8}$$

The first two equalities represent the inverse-probability-weighting (IPW) estimators of the average potential outcomes. A number of scholars have exploited the covariate balance conditions in equations (2.7) and (2.8) in order to estimate the average treatment effects (e.g., Chan and Yam, 2014; Han and Wang, 2013; Imai and Ratkovic, 2014; Graham, Pinto and Egel, 2012; Qin and Zhang, 2007). These existing estimators are locally semiparametric efficient, yet all of these rely on parametric models in one way or another.

Our goal is, however, to develop a globally fully nonparametric efficient estimator. Thus, we utilize the nonparametric estimator proposed by Chan, Yam and Zhang (2015). Let $D(v, v')$ be a distance measure for $v, v' \in \mathbb{R}$. That is, we assume that $D(v, v')$ is continuously differentiable in $v \in \mathbb{R}$, non-negative, and strictly convex in v with $D(v, v) = 0$. Based on equations (2.7) and (2.8), Chan, Yam and Zhang (2015) constructs calibration weights by solving the following minimization problem subject to constraints that are empirical counterparts of equations (2.7) and (2.8):

$$\begin{aligned} &\text{Minimize} \quad \sum_{i=1}^{N} T_i D(N p_i, 1) \\ &\text{subject to} \quad \sum_{i=1}^{N} T_i p_i u_K(X_i) = \frac{1}{N}\sum_{i=1}^{N} u_K(X_i), \end{aligned} \tag{2.9}$$

and

$$\begin{aligned} &\text{Minimize} \quad \sum_{i=1}^{N} (1-T_i) D(N q_i, 1) \\ &\text{subject to} \quad \sum_{i=1}^{N} (1-T_i) q_i u_K(X_i) = \frac{1}{N}\sum_{i=1}^{N} u_K(X_i), \end{aligned} \tag{2.10}$$

where u_K is a $K(N)$-dimensional function of X, whose components form a set of orthonormal polynomials, here $K(N)$ increases to infinity when N goes to infinity yet with $K(N) = o(N)$. Furthermore, note that all these u_K's have to form a basis on L^∞ as K goes to infinity.

Furthermore, to gain computational efficiency for implementation, they consider the dual problems of equations (2.9) and (2.10). While the primal problems given in equations (2.9) and (2.10) are convex separable programming with linear constraints, Tseng and Bertsekas (1987) showed that the dual problems are unconstrained convex maximization problems, which can be solved by efficient and stable numerical algorithms. With slight abuse of notation, denote $D(v) = D(v,1)$. For observations with $T_i = 1$, the dual solution is given by,

$$\hat{p}_K(X_i) \triangleq \frac{1}{N}\rho'\left(\hat{\phi}_K^T u_K(X_i)\right), \tag{2.11}$$

where ρ' is the first derivative of the following strictly concave function,

$$\rho(v) = D\left((D')^{-1}(-v)\right) + v \cdot (D')^{-1}(-v), \tag{2.12}$$

and $\hat{\phi}_K \in \mathbb{R}^K$ maximizes the following objective function,

$$\hat{F}_K(\phi) \triangleq \frac{1}{N}\sum_{i=1}^{N}\left\{T_i\rho\left(\phi^\top u_K(X_i)\right) - \phi^\top u_K(X_i)\right\}. \tag{2.13}$$

Similarly, for observations with $T_i = 0$,

$$\hat{q}_K(X_i) \triangleq \frac{1}{N}\rho'\left(\hat{\lambda}_K^T u_K(X_i)\right), \tag{2.14}$$

where $\hat{\lambda}_K \in \mathbb{R}^K$ maximizes the following objective function,

$$\widehat{G}_K(\lambda) \triangleq \frac{1}{N}\sum_{i=1}^{N}\left\{(1-T_i)\rho\left(\lambda^T u_K(X_i)\right) - \lambda^T u_K(X_i)\right\} . \tag{2.15}$$

According to the first order conditions for the maximizations given in equations (2.13) and (2.15), one can easily verify that the linear constraints in equations (2.9) and (2.10) are satisfied. Finally, Chan, Yam and Zhang (2015) proposed the following empirical covariate balancing estimator for δ_1 and δ_0,

$$\hat{\delta}_{1K} \triangleq \sum_{i=1}^{N} T_i \hat{p}_K(X_i) Y_i \quad \text{and} \quad \hat{\delta}_{0K} \triangleq \sum_{i=1}^{N}(1-T_i)\hat{q}_K(X_i)Y_i \tag{2.16}$$

The authors showed that $\hat{\delta}_{1K}$ and $\hat{\delta}_{0K}$ attain the semiparametric efficiency bounds given in equation (2.4) under mild regularity conditions.

2.3. *Efficient estimation of θ_0 and θ_1.* We begin by considering the efficient estimation of θ_0. As explained below, the same approach can be applied to efficiently estimate θ_1. The efficient influence function of θ_0 given in equation (2.3), involves three sets of nonparametric functions: $f_{T|X}(1 \mid X)$, $f_{M|T,X}(M \mid T = t, X)$ for $t = 0, 1$, and $\mathbb{E}(Y \mid X, M, T = 1)$. While it is possible to construct a globally efficient estimator of θ_0 by plugging the corresponding nonparametric estimates into equation (2.3), the performance of the resulting estimator may be poor because it is difficult to estimate the conditional density of a possibly continuous mediator and $f_{M|T,X}(M \mid 1, X)$ appears in the denominator of the first term of S_{θ_0}. The direct nonparametric estimation of $f_{M|T,X}$ usually results in extreme weights and the corresponding weighting estimator can become unstable.

Our goal is to construct a weighting estimator for θ_0. Let us represent θ_0 as a weighted average of Y among the treated,

$$\theta_0 = \mathbb{E}\left[\frac{TY}{f_{T|X}(1 \mid X)} \cdot \frac{f_{M|T,X}(M \mid 0, X)}{f_{M|T,X}(M \mid 1, X)}\right] . \tag{2.17}$$

Furthermore, define,

$$r_0(m, x) \triangleq \frac{f_{M|T,X}(m \mid 0, x)}{N f_{T|X}(1 \mid x) f_{M|T,X}(m \mid 1, x)}. \tag{2.18}$$

If $r_0(m, x)$ is a known function, then a natural estimator for θ_0 is $\tilde{\theta}_0 = \sum_{i=1}^{N} T_i r_0(M_i, X_i) Y_i$, which converges to θ_0 by the Law of Large Number. Since $r_0(m, x)$ is mostly unknown, we shall replace it by an estimated weight. To construct moment conditions for estimating $r_0(m, x)$, we need to develop a covariate balancing property extending equations (2.7) and (2.8). This result is given in the following lemma.

LEMMA 1. *With $q_0(x) = (N f_{T|X}(0 \mid x))^{-1}$ and $r_0(x, m)$ defined in equation (2.18), we have*

$$\mathbb{E}[T r_0(X, M) v(X, M)] = \mathbb{E}[(1 - T) q_0(X) v(X, M)]$$

for any suitable integrable functions v.

PROOF.

$$\begin{aligned}
&\mathbb{E}[T r_0(X, M) v(X, M)] \\
&= \mathbb{E}[r_0(X, M) v(X, M) f_{T|X,M}(1 \mid X, M)] \\
&= \mathbb{E}\left[\frac{1}{N f_{T|X}(1 \mid X)} \frac{f_{M|T,X}(M \mid 0, X)}{f_{M|T,X}(M \mid 1, X)} v(X, M) f_{T|X,M}(1 \mid X, M)\right]
\end{aligned}$$

$$
\begin{aligned}
&= \int_{\mathcal{M}}\int_{\mathcal{X}} \frac{1}{N f_{T|X}(1\mid x)} \frac{f_{M|T,X}(m\mid 0,x)}{f_{M|T,X}(m\mid 1,x)} v(x,m) f_{T|X,M}(1\mid x,m) f_{X,M}(x,m)dxdm \\
&= \int_{\mathcal{M}}\int_{\mathcal{X}} \frac{1}{N f_{T|X}(1\mid x)} \frac{f_{M|T,X}(m\mid 0,x)}{f_{M|T,X}(m\mid 1,x)} v(x,m) f_{T,X,M}(1,x,m)dxdm \\
&= \int_{\mathcal{M}}\int_{\mathcal{X}} \frac{f_{M|T,X}(m\mid 0,x)}{N f_{M,T|X}(m,1\mid x)} v(x,m) f_{T,M|X}(1,m\mid x) f_X(x)dxdm \\
&= \frac{1}{N}\int_{\mathcal{M}}\int_{\mathcal{X}} f_{M|T,X}(m|0,x) v(x,m) f_X(x)dxdm \\
&= \int_{\mathcal{M}}\int_{\mathcal{X}} \frac{v(x,m)}{N f_{T|X}(0\mid x)} f_{T|X,M}(0\mid x,m) f_{X,M}(x,m)dxdm \\
&= \mathbb{E}[q_0(X)v(X,M)f_{T|X,M}(0\mid X,M)] \\
&= \mathbb{E}[(1-T)q_0(X)v(X,M)].
\end{aligned}
$$

□

Lemma 1 motivates us to consider the empirical covariate balancing weights $\hat{r}_K(x,m)$ which solves for the following constrained optimization problem:

$$
\text{(2.19)}\quad \text{Minimize} \quad \sum_{i=1}^{N} T_i D(Nr_i,1) \qquad \text{subject to}
$$

$$
\sum_{i=1}^{N} T_i r_i v_K(X_i,M_i) = \sum_{i=1}^{N}(1-T_i)\hat{q}_K(X_i)v_K(X_i,M_i)\ .
$$

Note that $\hat{q}_K(x)$ is constructed from equations (2.14) and (2.15), and $v_K(x,m)$ with $K\in\mathbb{N}$ is a L-dimensional vector-valued function, whose components form a set of orthonormal polynomials, where $L=\mathcal{O}(K)$, i.e. L is of the same order as K. Furthermore, note that these v_K's have to form a basis on L^∞ as K goes to infinity.

The weights $\hat{r}_K(x,m)'$s are obtained by minimizing the aggregate distance between the final weights to a vector of constant working design weights, subject to an empirical analogue of the moment conditions given in Lemma 1. Unlike to the case of Deville and Särndal (1992) who used the true and known design weights commonly available in sample surveys, the true design weights for our problem $r_0(m,x)$ is unknown and is a function of two unknowns $f_{T|X}(1\mid X)$ and $f_{M|T,X}(M\mid T,X)$. Therefore, calibration of true design weights is impossible in this case. While the true weights are unavailable, we choose the uniform working design weights because they make it less likely to yield extreme weights. Even with misspecified design weights, however, we can still show that the proposed weighting estimator is globally semiparametric efficient.

Similar to equations (2.11) and (2.14), we can derive the dual solution for equation (2.19). For observations in the treatment group, i.e., $T_i = 1$,

$$\hat{r}_K(X_i, M_i) \triangleq \frac{1}{N}\rho'\left(\hat{\beta}_K^\top v_K(X_i, M_i)\right), \tag{2.20}$$

where ρ' is the first derivative of the function given in equation (2.12), and $\hat{\beta}_K$ maximizes the following objective function:

$$\widehat{H}_K(\beta) \triangleq \frac{1}{N}\sum_{i=1}^{N}\left\{T_i\rho\left(\beta^\top v_K(X_i, M_i)\right) - N(1-T_i)\hat{q}_K(X_i)\beta^\top v_K(X_i, M_i)\right\}. \tag{2.21}$$

From the first order condition of the maximization of $\widehat{H}_K$, we can check that

$$\widehat{H}'_K(\hat{\beta}_K) = \frac{1}{N}\sum_{i=1}^{N}T_i\rho\left(\hat{\beta}_K^\top v_K(X_i, M_i)\right)v_K(X_i, M_i) - \sum_{i=1}^{N}(1-T_i)\hat{q}_K(X_i)v_K(X_i, M_i) = 0. \tag{2.22}$$

Now, we define the proposed estimator for θ_0 to be

$$\hat{\theta}_{0K} \triangleq \sum_{i=1}^{N}T_i\hat{r}_K(X_i, M_i)Y_i. \tag{2.23}$$

Asymptotic properties of $\hat{\theta}_{0K}$ will be derived in the next subsection. We shall show that $\hat{\theta}_{0K}$ is globally semiparametric efficient under some mild regularity conditions. Therefore, the proposed estimators $\hat{\delta}_{1K} - \hat{\theta}_{0K}$ and $\hat{\theta}_{0K} - \hat{\delta}_{0K}$ are globally semiparametric efficient estimators for the average natural indirect and direct effects respectively.

The relationship given in equation (2.12) between $\rho(v)$ and $D(v)$ is derived in the supplementary materials. In the supplementary materials, we also show that the strict convexity of D is equivalent to the strict concavity of ρ. Since the dual formulation is equivalent to the primal problem, we shall express the proposed estimator in terms of $\rho(v)$ in the rest of the present paper. When $\rho(v) = -\exp(-v)$, the weights are equivalent to the implied weights of exponential tilting (Kitamura and Stutzer, 1997). When $\rho(v) = \log(1+v)$, the weights correspond to empirical likelihood (Qin and Lawless, 1994). When $\rho(v) = -(1-v)^2/2$, the weights are the implied weights of the continuous updating estimator of the generalized method of moments (Hansen, Heaton and Yaron, 1996).

The proposed estimator $\hat{\theta}_{0K}$ is constructed in a similar manner as done for $\hat{\delta}_{1K}$ and $\hat{\delta}_{0K}$ (see Section 2.2). However, there are some important differences. First, $\hat{\delta}_{1K}$ and $\hat{\delta}_{0K}$ only require balancing pre-treatment variables X, but $\hat{\theta}_{0K}$ requires balancing both pre-treatment variables X and post-treatment mediators M. Moreover, in equation (2.21), ρ appears explicitly in the first term and implicitly in the second term through the dependency of $\hat{q}_K(x)$ on ρ. This creates new challenges to the establishment of theoretical results. Note that although we consider the weights estimated through a specific $\rho(v)$, the functional form of the true weights $r_0(x,m)$ is unspecified. It will be shown later that any function $r_0(x,m)$, satisfying a mild differentiability assumption, can be approximated arbitrarily well by $\hat{r}_K(x,m)$ uniformly as the sample size increases, so long as $\rho(v)$ also satisfies a mild regularity condition.

We can apply the same methodology to the estimation of θ_1. Note that

$$\theta_1 = \mathbb{E}\left[\frac{(1-T)Y}{f_{T|X}(1 \mid X)} \cdot \frac{f_{M|T,X}(M \mid 1, X)}{f_{M|T,X}(M \mid 0, X)}\right] \tag{2.24}$$

Define

$$w_0(x,m) \triangleq \frac{f_{M|T,X}(m \mid 1, x)}{N f_{T,M|X}(0, m \mid x)}$$

For any suitable integrable $v(x,m)$, we have

$$\begin{aligned}
&\mathbb{E}\left[(1-T)\frac{f_{M|T,X}(M|1,X)}{f_{T,M|X}(0,M \mid X)}v(X,M)\right] \\
&= \int_{\mathcal{M}}\int_{\mathcal{X}} \frac{f_{M|T,X}(m|1,x)}{f_{T,M|X}(0,m \mid x)} v(x,m) f_{T,M,X}(0,m,x)\,dx\,dm \\
&= \int_{\mathcal{M}}\int_{\mathcal{X}} \frac{f_{M,T,X}(m,1,x)}{f_{T,X}(1,x)} v(x,m) f_X(x)\,dx\,dm \\
&= \mathbb{E}\left[\frac{Tv(X,M)}{f_{T|X}(1 \mid X)}\right] .
\end{aligned}$$

Therefore, we can construct empirical covariate balancing weights $\hat{w}_K$ from the following constrained maximization problem:

$$\text{Minimize} \quad \sum_{i=1}^{N} (1-T_i) D\left(N w_i, 1\right) \quad \text{subject to}$$

$$\sum_{i=1}^{N} (1-T_i) w_i v_K(X_i, M_i) = \sum_{i=1}^{N} T_i \hat{p}_K(X_i) v_K(X_i, M_i) .$$

Its dual solution is given by,

$$\hat{w}_K(x,m) \triangleq \frac{1}{N}\rho'\left(\hat{\gamma}_K^\top v_k(x,m)\right) ,$$

where $\hat{\gamma}_K$ maximizes the following objective function,

$$\hat{J}_K(\beta) \triangleq \frac{1}{N}\sum_{i=1}^{N}\Big[(1-T_i)\rho(\gamma^\top v_K(X_i,M_i)) - NT_i\hat{p}_K(X_i)\gamma^\top v_K(X_i,M_i)\Big]$$

Then, we can now suggest the estimator of $\mathsf{PIE} = \theta_1 - \delta_0$ by:

(2.25)

$$\widehat{\mathsf{PIE}} \triangleq \hat{\theta}_{1K} - \hat{\delta}_{0K} = \sum_{i=1}^{N}(1-T_i)\hat{w}_K(X_i,M_i)Y_i - \sum_{i=1}^{N}(1-T_i)\hat{q}_K(X_i)Y_i \ .$$

2.4. *Asymptotic properties.* To derive the asymptotic properties of the proposed estimator, we list all additional technical assumptions that are required beyond Assumption 1.

Assumption 2. $\mathbb{E}(Y^2 \mid T=0) < \infty$ *and* $\mathbb{E}(Y^2 \mid T=1) < \infty$.

Assumption 3.

1. *The support $\mathcal{X}$ of r_1-dimensional covariate X is a Cartesian product of r_1 compact intervals.*
2. *The support $\mathcal{M}$ of r_2-dimensional mediating variable M is a Cartesian product of r_2 compact intervals.*

Denote $r \triangleq r_1 + r_2$.

Assumption 4. *There exist some constants $\eta_1, \eta_2, \eta_3, \eta_4, \eta_5, \eta_6$ such that the following inequalities hold:*

$$\begin{aligned}
0 < \frac{1}{\eta_1} \leq \quad & f_{T|X}(0 \mid x) & \leq \frac{1}{\eta_2} < 1,\\
0 < \frac{1}{\eta_3} \leq \quad & f_{M|T,X}(m \mid 0,x) & \leq \frac{1}{\eta_4} < 1,\\
0 < \frac{1}{\eta_5} \leq \quad & f_{M|T,X}(m \mid 1,x) & \leq \frac{1}{\eta_6} < 1.
\end{aligned}$$

Assumption 5. *The functions, $q(x)$ and $r(x,m)$, are s-times and s'-times continuously differentiable, respectively, where $s' > 19r > 0$ and $s > 16r_1 > 0$.*

ASSUMPTION 6. *The function* $\mathbb{E}(Y \mid T=1, M=m, X=x)$ *is* t*-times jointly continuously differentiable with respect to* (x,m)*, and* $\eta(1,0,x)$ *is* d'*-times continuously differentiable w.r.t.* x*, where* $d > 3r/2$ *and* $d' > 3r_1/2$*.*

ASSUMPTION 7. $K = O(N^\nu)$ *and* $\left(\frac{1}{s'/r-2} \vee \frac{1}{s/r_1+1}\right) < \nu < \frac{1}{17}$.

ASSUMPTION 8. $\rho \in C^3(\mathbb{R})$ *is a strictly concave function defined on* $\mathbb{R}$ *i.e.,* $\rho''(\gamma) < 0,\ \forall \gamma \in \mathbb{R}$*, and the range of* ρ' *contains the following subset of the positive real line:*

$$[\eta_2, \eta_1] \cup \left[\frac{\eta_1}{\eta_1 - 1}, \frac{\eta_2}{\eta_2 - 1}\right] \cup \left[\frac{\eta_1\eta_6}{(\eta_1 - 1)\,\eta_3}, \frac{\eta_2\eta_5}{(\eta_2 - 1)\,\eta_4}\right] .$$

Assumptions 1–7 or similar assumptions also appeared in the literature (e.g., Chan, Yam and Zhang, 2015; Hahn, 1998; Hirano, Imbens and Ridder, 2003; Imbens, Newey and Ridder, 2005). As explained earlier, Assumption 1 is used for the identification of the natural direct and indirect effects. Assumption 2 is required for the finiteness of asymptotic variance. Assumptions 3 and 4 are needed to establish the uniform boundedness of approximations. Assumptions 5 and 6 are required for controlling the remainder of approximations with a given set of basis functions. Assumption 8 is required for controlling the stochastic order of the reminder terms, which is satisfied by commonly used ρ functions as discussed above. This final assumption imposes a mild regularity condition on ρ. Chan, Yam and Zhang (2015) maintains the same assumption.

Two intermediate lemmas are needed to prove the main theorem. We define the following intermediate quantities that are probability limits of $\widehat{F}_K, \hat{\phi}_K, \hat{p}_K, \widehat{G}_K, \hat{\lambda}_K, \hat{q}_K, \widehat{H}_k, \hat{\beta}_K$ and $\hat{r}_K$ for each fixed K:

$$\begin{aligned}
F_K^*(\phi) &\triangleq \mathbb{E}\left[T\rho\left(\phi^\top u_K(X)\right) - \phi^\top u_K(X)\right] = \mathbb{E}\left(\widehat{F}_K(\phi)\right) , \\
\phi_K^* &\triangleq \arg\max_{\phi \in \mathbb{R}^K} F_K^*(\phi) , \\
p_K^*(x) &\triangleq \frac{1}{N}\rho'\left((\phi_K^*)^\top u_K(x)\right) , \\
G_K^*(\lambda) &\triangleq \mathbb{E}\left[(1-T)\rho\left(\lambda^\top u_K(X)\right) - \lambda^\top u_K(X)\right] = \mathbb{E}\left(\widehat{G}_K(\lambda)\right) , \\
\lambda_K^* &\triangleq \arg\max_{\lambda \in \mathbb{R}^K} G_K^*(\lambda) , \\
q_K^*(x) &\triangleq \frac{1}{N}\rho'\left((\lambda_K^*)^\top u_K(x)\right) ,
\end{aligned}$$

$$\begin{aligned} H_K^*(\beta) &\triangleq \mathbb{E}\left[T\rho\left(\beta^\top v_K(X,M)\right) - (1-T)(f_{T|X}(0 \mid X))^{-1}\beta^\top v_K(X,M)\right] , \\ \beta_K^* &\triangleq \arg\max_{\beta\in\mathbb{R}^K} H_K^*(\beta) , \\ r_K^*(x,m) &\triangleq \frac{1}{N}\rho'\left((\beta_K^*)^\top v_K(x,m)\right) . \end{aligned}$$

Also, let $\zeta(K) = \sup_{x\in\mathcal{X}} ||u_K(x)||$. The following lemma establishes the approximation of functions $p_0(x)$, $q_0(x)$, $r_0(x,m)$ by $p_K^*(x)$, $q_K^*(x)$ and $r_K^*(x,m)$.

LEMMA 2. *Under Assumptions 3, 4, 5, and 7, we have,*

$$\begin{aligned} \sup_{x\in\mathcal{X}} |Np_0(x) - Np_K^*(x)| &= O\left(K^{-\frac{s}{2r_1}}\zeta(K)\right) , \\ \sup_{x\in\mathcal{X}} |Nq_0(x) - Nq_K^*(x)| &= O\left(K^{-\frac{s}{2r_1}}\zeta(K)\right) , \\ \sup_{(x,m)\in\mathcal{X}\times\mathcal{M}} |Nr_0(x,m) - Nr_K^*(x,m)| &= O\left(K^{-\frac{s'}{2r}}\zeta(K)\right) . \end{aligned}$$

PROOF. The proof is given in the supplementary material. □

The other lemma is about the performance of the estimated auxiliary parameters, $\hat{\phi}_K$, $\hat{\lambda}_K$, and $\hat{\beta}_K$ that maximize equations (2.13), (2.15), and (2.21) respectively.

LEMMA 3. *Under Assumptions 3, 4, 5, and 7, we have*

$$\begin{aligned} \|\hat{\phi}_K - \phi_K^*\| &= O_p\left(\sqrt{\frac{K}{N}}\right) , \\ \|\hat{\lambda}_K - \lambda_K^*\| &= O_p\left(\sqrt{\frac{K}{N}}\right) , \\ \|\hat{\beta}_K - \beta_K^*\| &= O_p\left(\sqrt{\frac{K^5}{N}}\right) . \end{aligned}$$

PROOF. The proof is given in the supplementary material. □

The following theorem shows that $\hat{\theta}_{0K}$ is consistent, asymptotically normal, and globally semiparametric efficient.

THEOREM 1. *Under Assumptions 1–8, $\hat{\theta}_{0K}$ has the following properties:*

1. $\hat{\theta}_{0K} = \sum_{i=1}^{N} T_i \hat{r}_K(X_i, M_i) Y_i \xrightarrow{p} \theta_0$;
2. $\sqrt{N}\left(\sum_{i=1}^{N} T_i \hat{r}_K(X_i, M_i) Y_i - \theta_0\right) \xrightarrow{d} \mathcal{N}(0, V_{\theta_0})$, *where* $V_{\theta_0} = \mathbb{E}\left(S_{\theta_0}^2\right)$, *attains the semi-parametric efficiency bound (Tchetgen Tchetgen and Shpitser, 2012) with the definition of* S_{θ_0} *as given in equation* (2.3).

Proof. The proof is given in the supplementary material. □

The following corollary establishes the large sample properties of the estimated average natural indirect effect, $\widehat{\mathsf{NIE}}_K = \hat{\delta}_{1K} - \hat{\theta}_{0K}$, and of the estimated average natural direct effect, $\widehat{\mathsf{NDE}}_K = \hat{\theta}_{0K} - \hat{\delta}_{0K}$.

Corollary 2. *Under Assumptions 1–8,* $\widehat{\mathsf{NIE}}_K$ *and* $\widehat{\mathsf{NDE}}_K$ *have the following properties:*

1. *Consistency.*

$$\begin{aligned}
\widehat{\mathsf{NIE}}_K &= \sum_{i=1}^{N} T_i \hat{p}_K(X_i) Y_i - \sum_{i=1}^{N} T_i \hat{r}_K(X_i, M_i) Y_i \xrightarrow{p} \mathsf{NIE} = \delta_{1K} - \theta_{0K} \ , \\
\widehat{\mathsf{NDE}}_K &= \sum_{i=1}^{N} (1 - T_i) \hat{r}_K(X_i, M_i) Y_i - \sum_{i=1}^{N} (1 - T_i) \hat{q}_K(X_i) Y_i \xrightarrow{p} \mathsf{NDE} = \theta_{0K} - \delta_{0K} \ .
\end{aligned}$$

2. *Asymptotic normality and semiparametric efficiency.*

$$\begin{aligned}
\sqrt{N}\left(\widehat{\mathsf{NIE}}_K - \mathsf{NIE}\right) &\xrightarrow{d} \mathcal{N}(0, V_{\mathsf{NIE}}) \ , \\
\sqrt{N}\left(\widehat{\mathsf{NDE}}_K - \mathsf{NDE}\right) &\xrightarrow{d} \mathcal{N}(0, V_{\mathsf{NDE}}) \ ,
\end{aligned}$$

where $V_{\mathsf{NIE}} = \mathbb{E}\left[(S_{\delta_1} - S_{\theta_0})^2\right]$ *and* $V_{\mathsf{NDE}} = \mathbb{E}\left[(S_{\theta_0} - S_{\delta_0})^2\right]$, *with* S_{θ_0}, S_{δ_1}, *and* S_{δ_0} *as defined in equations* (2.3), *and* (2.4), *respectively.*

Similarly, we can derive the analogous asymptotic properties for the pure indirect effect.

Theorem 3. *Under Assumptions 1–8,* $\widehat{\mathsf{PIE}}$ *defined in equation* (2.25) *has the following properties:*

1. $\widehat{\mathsf{PIE}} \xrightarrow{p} \mathsf{PIE}$

2. $\sqrt{N}\left(\widehat{\mathsf{PIE}}-\mathsf{PIE}\right)\xrightarrow{d}\mathcal{N}(0,V_{\mathsf{PIE}})$*, where* $V_{\mathsf{PIE}}=\mathbb{E}\left[(S_{\mathsf{PIE}})^2\right]$ *attains the semi-parametric efficiency bound (Tchetgen Tchetgen and Shpitser, 2012). The definition of* S_{PIE} *is given by,*

$$\begin{aligned}S_{\mathsf{PIE}} &= \frac{\mathbf{1}\{T=0\}f_{M|T,X}(M\mid 1,X)}{f_{T|X}(0\mid X)f_{M|T,X}(M\mid 0,X)}\left\{Y-\mathbb{E}\left(Y\mid X,M,T=0\right)\right\}\\ &+\frac{\mathbf{1}\{T=1\}}{f_{T|X}(1\mid X)}\left\{\mathbb{E}\left(Y\mid X,M,T=0\right)-\eta(0,1,X)\right\}\\ &-\frac{\mathbf{1}\{T=0\}}{f_{T|X}(0\mid X)}\left\{Y-\eta(0,0,X)\right\}+\eta(0,1,X)-\eta(0,0,X)-\mathsf{PIE}.\end{aligned}$$

2.5. *Nonparametric estimation of the asymptotic variance.* We have shown that the proposed estimator attains the semiparametric efficiency bound whose efficient influence function depends on three sets of nonparametric functions. In this section, we propose a consistent variance estimator that does not require additional nonparametric function estimates and is easy to compute. Define the following quantities:

$$\begin{aligned}\tau &\triangleq (\phi^\top,\lambda^\top,\beta^\top,\delta,\delta',\theta)^\top\ ,\\ g_{K1}(T,X;\phi) &\triangleq T\rho'\left(\phi^\top u_K(X)\right)u_K(X)-u_K(X)\ ,\\ g_{K2}(T,X;\lambda) &\triangleq (1-T)\rho'\left(\lambda^\top u_K(X)\right)u_K(X)-u_K(X)\ ,\\ g_{K3}(T,X,M;\lambda,\beta) &\triangleq T\rho'\left(\beta^\top v_K(X,M)\right)v_K(X,M)-(1-T)\rho'\left(\lambda^\top u_K(X)\right)v_K(X,M)\ ,\\ g_{K4}(T,X,Y;\phi,\delta) &\triangleq T\rho'\left(\phi^\top u_K(X)\right)Y-\delta\ ,\\ g_{K5}(T,X,Y;\lambda,\delta) &\triangleq (1-T)\rho'\left(\lambda^\top u_K(X)\right)Y-\delta'\ ,\\ g_{K6}(T,X,M,Y;\beta,\theta) &\triangleq T\rho'\left(\beta^\top v_K(X,M)\right)Y-\theta\ ,\\ g_K(T,X,M,Y;\tau) &\triangleq \left(g_{K1}^\top,g_{K2}^\top,g_{K3}^\top,g_{K4},g_{K5},g_{K6}\right)^\top\ ,\\ \hat{\tau}_K &\triangleq \left(\hat{\phi}_K^\top,\hat{\lambda}_K^\top,\hat{\beta}_K^\top,\hat{\delta}_{1K},\hat{\delta}_{0K},\hat{\theta}_{0K}\right)^\top\ ,\\ \tau_K^* &\triangleq \left((\phi_K^*)^\top,(\lambda_K^*)^\top,(\beta_K^*)^\top,\delta_{1K}^*,\delta_{0K}^*,\theta_{0K}^*\right)^\top\ ,\end{aligned}$$

where $\delta,\delta',\theta\in\mathbb{R}$, $\delta_{1K}^*\triangleq\mathbb{E}\left[TNp_K^*(X)Y\right]$, $\delta_{0K}^*\triangleq\mathbb{E}\left[(1-T)Nq_K^*(X)Y\right]$ and $\theta_K^*\triangleq\mathbb{E}\left[TNr_K^*(X,M)Y\right]$. Note that by definition $\hat{\tau}_K$ satisfies:

$$\frac{1}{N}\sum_{i=1}^{N}g_K(T_i,X_i,M_i,Y_i;\hat{\tau}_K)\ =\ 0\ . \tag{2.26}$$

Applying the Taylor's theorem for (2.26) at τ_K^* yields:
(2.27)

$$0 = \frac{1}{N}\sum_{i=1}^N g_K(T_i,X_i,M_i,Y_i;\tau_K^*)+\frac{1}{N}\sum_{i=1}^N \frac{\partial g_K(T_i,X_i,M_i,Y_i;\tilde{\tau}_K)}{\partial \tau}(\hat{\tau}_K-\tau_K^*)\ ,$$

where $\tilde{\tau}_K$ lies on the line joining $\hat{\tau}_K$ with τ_K^*. In the supplementary material, we show that:

(2.28)

$$\frac{1}{N}\sum_{i=1}^N \frac{\partial g_K(T_i,X_i,M_i,Y_i;\tilde{\tau}_K)}{\partial \tau} = \mathbb{E}\left[\frac{\partial g_K(T,X,M,Y;\tau_K^*)}{\partial \tau}\right] + o_p(1)\ ,$$

where

$$\mathbb{E}\left[\frac{\partial g_K(T,X,M,Y;\tau_K^*)}{\partial \tau}\right] = \begin{pmatrix} A_{3K\times 3K}, & B_{3K\times 3} \\ C_{3\times 3K}, & D_{3\times 3}\end{pmatrix}$$

and

$$\begin{aligned}
A_{3K\times 3K} &\triangleq \begin{pmatrix} A_{11} & 0_{K\times K} & 0_{K\times K} \\ 0_{K\times K} & A_{22} & 0_{K\times K} \\ 0_{K\times K} & A_{32} & A_{33}\end{pmatrix}\ , \\
B_{3K\times 3} &\triangleq 0_{3K\times 3}\ , \\
C_{3\times 3K} &\triangleq \begin{pmatrix} C_{11} & 0_{1\times K} & 0_{1\times K} \\ 0_{1\times K} & C_{22} & 0_{1\times K} \\ 0_{1\times K} & 0_{1\times K} & C_{33}\end{pmatrix}\ , \\
D_{3\times 3} &\triangleq -I_{3\times 3}\ , \\
A_{11} &= \mathbb{E}\left[T\rho''((\phi_K^*)^\top u_K(X))u_K(X)u_K^\top(X)\right]\ , \\
A_{22} &= \mathbb{E}[(1-T)\rho''((\lambda_K^*)^\top u_K(X))u_K(X)u_K^\top(X)]\ , \\
A_{32} &= -\mathbb{E}[(1-T)\rho''((\lambda_K^*)^\top u_K(X))v_K(X,M)u_K^\top(X)]\ , \\
A_{33} &= \mathbb{E}[T\rho''((\beta_K^*)^\top v_K(X,M))v_K(X,M)v_K^\top(X,M)]\ , \\
C_{11} &= \mathbb{E}[T\rho''((\phi_K^*)^\top u_K(X))Yu_K^\top(X)]\ , \\
C_{22} &= \mathbb{E}[(1-T)\rho''(-(\lambda_K^*)^\top u_K(X))Yu_K^\top(X)]\ , \\
C_{33} &= \mathbb{E}[T\rho''((\beta_K^*)^\top v_K(X,M))Yv_K^\top(X,M)]\ .
\end{aligned}$$

The parameters of interest are

$$\begin{pmatrix}\mathsf{NIE}\\ \mathsf{NDE}\\ \theta_0\end{pmatrix} = \begin{pmatrix}\delta_1-\theta_0\\ \theta_0-\delta_0\\ \theta_0\end{pmatrix} = \begin{pmatrix}1 & 0 & -1\\ 0 & -1 & 1\\ 0 & 0 & 1\end{pmatrix}\pi$$

where $\pi \triangleq (\delta_1, \delta_0, \theta_0)^\top$. Note that asymptotic variance of estimators $\hat{\pi}_K \triangleq (\hat{\delta}_{1K}, \hat{\delta}_{0K}, \hat{\theta}_{0K})^\top$, which is the lower right corner element of $\lim_{N\to\infty} \mathbb{V}(\sqrt{N}(\hat{\tau}_K - \tau_K^*))$. Consider a sub-matrix formed by the last three rows of $\mathbb{E}\left[\frac{\partial g_K(T,X,M,Y;\tau_K^*)}{\partial \tau}\right]^{-1}$, which is

$$L_K \triangleq \left(C_{3\times 3K} A_{3K\times 3K}^{-1}, -I_{3\times 3}\right).$$

Algebraic manipulation yields:

$$L_K = \begin{pmatrix} L_{11K} & 0_{1\times K} & 0_{1\times K} & \\ 0_{1\times K} & L_{22K} & 0_{1\times K} & -I_{3\times 3} \\ 0_{1\times K} & L_{32K} & L_{33K} & \end{pmatrix},$$

where

$$\begin{aligned}
L_{11K} &= \mathbb{E}[T\rho''((\phi_K^*)^\top u_K(X))Y u_K(X)^\top] \cdot \mathbb{E}[T\rho''((\phi_K^*)^\top u_K(X)) u_K(X) u_K(X)^\top]^{-1}, \\
L_{22K} &= \mathbb{E}[(1-T)\rho''((\lambda_K^*)^\top u_K(X))Y u_K(X)^\top] \cdot \mathbb{E}[(1-T)\rho''((\lambda_K^*)^\top u_K(X)) u_K(X) u_K(X)^\top]^{-1}, \\
L_{32K} &= \mathbb{E}[T\rho''((\beta_K^*)^\top v_K(X,M))Y v_K(X,M)^\top] \\
&\quad \cdot \mathbb{E}\left[T\rho''((\beta_K^*)^\top v_K(X,M)) v_K(X,M) v_K(X,M)^\top\right]^{-1} \\
&\quad \cdot \mathbb{E}[(1-T)\rho''((\lambda_K^*)^\top u_K(X)) v_K(X,M) u_K(X)^\top] \\
&\quad \cdot \mathbb{E}[(1-T)\rho''((\lambda_K^*)^\top u_K(X)) u_K(X) u_K(X)^\top]^{-1}, \\
L_{33K} &= \mathbb{E}[T\rho''((\beta_K^*)^\top v_K(X,M))Y v_K(X,M)^\top] \\
&\quad \cdot \mathbb{E}[T\rho''((\beta_K^*)^\top v_K(X,M)) v_K(X,M) v_K(X,M)^\top]^{-1}.
\end{aligned}$$

Applying Lemmas 2 and 3, we show in the supplementary material that

$$\|\mathbb{E}[g_K(T,X,M,Y;\tau_K^*)]\|^2 = O\left(K^{-\frac{s}{r_1}+1}\right) + O\left(K^{-\frac{s'}{r}+1}\right), \tag{2.29}$$

$$\|L_K\|^2 = O(K^4). \tag{2.30}$$

By (2.27), (2.28) we can have

(2.31)

$$\hat{\tau}_K - \tau_K^* = \mathbb{E}\left[\frac{\partial g_K(T,X,M,Y;\tau_K^*)}{\partial \tau}\right]^{-1} \left(\frac{1}{N}\sum_{i=1}^N g_K(T_i, X_i, M_i, Y_i; \hat{\tau}_K)\right).$$

Also, by equations (2.29) and (2.30), we can show:

(2.32)

$$\lim_{K\to\infty} Var(\sqrt{N}(\hat{\pi}_K - \pi_K^*)) = \lim_{K\to\infty} L_K \mathbb{E}[g_K(T,X,M,Y;\tau_K^*) g_K(T,X,M,Y;\tau_K^*)^\top] L_K^\top$$

$$= \lim_{K\to\infty} L_K P_K L_K^\top ,$$

where $\pi_K^* \triangleq (\delta_{1K}^*, \delta_{0K}^*, \theta_K^*)^\top$ and $P_K \triangleq \mathbb{E}[g_K(T,X,M,Y;\tau_K^*)g_K(T,X,M,Y;\tau_K^*)^\top]$. This leads to the following estimator of the asymptotic variance:

$$\widehat{V}_K \triangleq \widehat{L}_K \widehat{P}_K \widehat{L}_K^\top, \tag{2.33}$$

where

$$\begin{aligned}
\hat{L}_k &= \begin{pmatrix} \widehat{L}_{11K} & 0_{1\times K} & 0_{1\times K} & \\ 0_{1\times K} & \widehat{L}_{22K} & 0_{1\times K} & -I_{3\times 3} \\ 0_{1\times K} & \widehat{L}_{32K} & \widehat{L}_{33K} & \end{pmatrix} , \\
\widehat{L}_{11K} &\triangleq \left[\frac{1}{N}\sum_{i=1}^N T_i \rho''(\hat{\phi}_K^\top u_K(X_i))u_K(X_i)^\top Y_i\right] \cdot \left[\frac{1}{N}\sum_{i=1}^N T_i \rho''(\hat{\phi}_K^\top u_K(X_i))u_K(X_i)^\top u_K(X_i)\right]^{-1} , \\
\widehat{L}_{22K} &\triangleq \left[\frac{1}{N}\sum_{i=1}^N (1-T_i) \rho''(\hat{\lambda}_K^\top u_K(X_i))u_K(X_i)^\top Y_i\right] \\
&\quad \cdot \left[\frac{1}{N}\sum_{i=1}^N (1-T_i) \rho''(\hat{\lambda}_K^\top u_K(X_i))u_K(X_i)^\top u_K(X_i)\right]^{-1} , \\
\widehat{L}_{32K} &\triangleq \left[\frac{1}{N}\sum_{i=1}^N T_i \rho''(\hat{\beta}_K^\top v_K(X_i,M_i))v_K(X_i,M_i)^\top Y_i\right] \\
&\quad \cdot \left[\frac{1}{N}\sum_{i=1}^N T_i \rho''(\hat{\beta}_K^\top v_K(X_i,M_i))v_K(X_i,M_i)v_K(X_i,M_i)^\top\right]^{-1} \\
&\quad \cdot \left[\frac{1}{N}\sum_{i=1}^N (1-T_i) \rho''(\hat{\lambda}_K^\top u_K(X_i))v_K(X_i,M_i)u_K(X_i)^\top\right] \\
&\quad \cdot \left[\frac{1}{N}\sum_{i=1}^N (1-T_i) \rho''(\hat{\lambda}_K^\top u_K(X_i))u_K(X_i)u_K(X_i)^\top\right]^{-1} , \\
\widehat{L}_{33K} &\triangleq \left[\frac{1}{N}\sum_{i=1}^N T_i \rho''(\hat{\beta}_K^\top v_K(X_i,M_i))v_K(X_i,M_i)^\top Y_i\right] \\
&\quad \cdot \left[\frac{1}{N}\sum_{i=1}^N T_i \rho''(\hat{\beta}_K^T v_K(X_i,M_i))v_K(X_i,M_i)^\top v_K(X_i,M_i)\right]^{-1} , \\
\widehat{P}_K &\triangleq \frac{1}{N}\sum_{i=1}^N g_K(T_i,X_i,M_i,Y_i;\hat{\tau}_K)g_K(T_i,X_i,M_i,Y_i;\hat{\tau}_K)^\top .
\end{aligned}$$

Finally, the following theorem states that this estimator of the asymptotic variance $\widehat{V}_K$ is consistent, indeed.

THEOREM 4. *Let* $k_1 = (0, 0, 1), k_2 = (1, 0, -1)$ *and* $k_3 = (0, -1, 1)$. *Under Assumptions 1–8,* $k_1\widehat{V}_K k_1^\top$ *is a consistent estimator for* $\mathbb{E}\left[(S_{\theta_0})^2\right]$, $k_2\widehat{V}_K k_2^\top$ *is a consistent estimator for* $\mathbb{E}\left[(S_{\mathsf{NIE}})^2\right]$ *and* $k_3\widehat{V}_K k_3^\top$ *is a consistent estimator for* $\mathbb{E}\left[(S_{\mathsf{NDE}})^2\right]$, *where* $S_{\mathsf{NIE}} = S_{\delta_1} - S_{\theta_0}$, $S_{\mathsf{NDE}} = S_{\theta_0} - S_{\delta_0}$ *and* $S_{\theta_0}, S_{\delta_1}, S_{\delta_0}$ *are defined in equations* (2.3) *and* (2.4), *respectively.*

PROOF. The proof is given in the supplementary material. ☐

3. Extensions of our Proposed Methodology. In this section, we consider several extensions of the proposed methodology presented in the previous section. We first study the case of multiple mediators, and then discuss the estimation of pure indirect effects and natural indirect effects for the untreated.

3.1. *Multiple mediators.* We show that the proposed methodology introduced in the previous section does not require the mediator to be univariate. Instead, we consider the situation where multiple mediators exist and are not causally independent. Specifically, we study the setting considered by Imai and Yamamoto (2013) who proposed a semiparametric estimation method within a linear structural equation modeling framework. We improve this existing method by considering the efficient nonparametric estimation.

Consider the setting with a binary treatment variable $T \in \{0, 1\}$ and two mediators W and M with W being causally prior to M. Let $W(t)$ be a potential mediator variable, which represents the value of the mediator W when the treatment variable is set to $t \in \{0, 1\}$. Similarly, denote the potential mediator variable of M by $M(t, w)$ which represents the value of the mediator M when T and W are set to (t, w). Finally, the potential outcome variable can be defined as $Y(t, m, w)$, which represents the value of the outcome variable Y when the treatment and the two mediators are set to (t, m, w).

We consider an extension of the sequential ignorability assumption given in Assumption 1.

ASSUMPTION 9. *The following three conditional independence statements hold:*

1. $\{Y(t, m, w), M(t, w), W(t)\} \perp\!\!\!\perp T \mid X = x$,
2. $\{Y(t', m, w), M(t', w)\} \perp\!\!\!\perp W(t) \mid T = t, X = x$,
3. $Y(t', m, w) \perp\!\!\!\perp M(t, w') \mid T = t, X = x, W(t) = w'$;

for any $t, t' \in \{0,1\}$ *and* $(m, w, w', x) \in \mathcal{M} \times \mathcal{W} \times \mathcal{W} \times \mathcal{X}$*, where* $\mathcal{W}$ *is the support of* W*.*

This assumption is stronger than the one considered by Imai and Yamamoto (2013), which avoids the "cross-world" independence but results in partial identification. Finally, we also make the relevant consistency and positivity assumptions as done in Assumption 1. The following lemma establishes an important connection between two sequential ignorability assumptions given in Assumptions 1 and 9.

LEMMA 4. *Under the first two conditions of Assumption 9, the third condition of Assumption 9 holds if and only if:*

$$Y(t', w, m) \perp\!\!\!\perp \{W(t), M(t, W(t))\} \mid T = t, X = x \tag{3.1}$$

for any $t, t' \in \{0,1\}$ *and* $(m, w, x) \in \mathcal{M} \times \mathcal{W} \times \mathcal{X}$*.*

PROOF. Under the first two conditions of Assumption 9, we first show that the equation (3.1) implies the third condition of Assumption 9. For any test functions $\phi_1(y)$, $\phi_2(m)$, and $\phi_3(w)$, by using (3.1) and Assumption 9 (2), we have

$$\begin{aligned}
&\mathbb{E}\left[\phi_1(Y(t',m,w))\phi_2(M(t,W(t)))\phi_3(W(t))|T=t,X=x\right]\\
=&\mathbb{E}\left[\phi_1(Y(t',m,w))|T=t,X=x\right]\cdot\mathbb{E}\left[\phi_2(M(t,W(t)))\phi_3(W(t))|T=t,X=x\right]\\
=&\mathbb{E}\Bigg[\mathbb{E}\left[\phi_1(Y(t',m,w))|T=t,X=x\right]\\
&\qquad\cdot\mathbb{E}\left[\phi_2(M(t,W(t)))\phi_3(W(t))|T=t,X=x,W(t)\right]\Bigg|T=t,X=x\Bigg]\\
=&\mathbb{E}\Bigg[\mathbb{E}\left[\phi_1(Y(t',m,w))|T=t,X=x,W(t)\right]\\
&\quad\cdot\mathbb{E}\left[\phi_2(M(t,W(t)))|T=t,X=x,W(t)\right]\cdot\phi_3(W(t))\Bigg|T=t,X=x\Bigg]\ ,
\end{aligned}$$

where the second equality follows from applying the Tower property to the second term and then plugging the first term inside the second one; the last equality follows from Assumption 9 (2). Therefore, we can get that

$$Y(t', m, w) \perp\!\!\!\perp M(t, W(t)) \mid T = t, X = x, W(t)\ .$$

Conversely, we shall show that under the first two conditions of Assumption 9, Assumption 9 (3) implies (3.1). For any test functions $\phi_1(y)$ and $\phi_4(w, m)$, we have

$$\mathbb{E}\left[\phi_1(Y(t',w,m))\phi_4(W(t),M(t,W(t)))|T=t,X=x\right]$$

$$
\begin{aligned}
=&\mathbb{E}\left[\mathbb{E}\left[\phi_1(Y(t',w,m))\phi_4(W(t),M(t,W(t)))|T=t,X=x,W(t)\right]|T=t,X=x\right]\\
=&\mathbb{E}\left[\mathbb{E}\left[\phi_1(Y(t',w,m))|T=t,X=x,W(t)\right]\right.\\
&\quad\left.\cdot\,\mathbb{E}\left[\phi_4(W(t),M(t,W(t)))|T=t,X=x,W(t)\right]|T=t,X=x\right]\\
=&\mathbb{E}\left[\phi_1(Y(t',w,m))|T=t,X=x\right]\\
&\quad\cdot\,\mathbb{E}\left[\mathbb{E}\left[\phi_4(W(t),M(t,W(t)))|T=t,X=x,W(t)\right]|T=t,X=x\right]\\
=&\mathbb{E}\left[\phi_1(Y(t',w,m))|T=t,X=x\right]\cdot\mathbb{E}\left[\phi_4(W(t),M(t,W(t)))|T=t,X=x\right]\ ,
\end{aligned}
$$

where the second equality follows from Assumption 9 (3) and the third equality follows from Assumption 9 (2), which finally yields (3.1). □

Lemma 4, together with the discussion in the previous section, implies that we can estimate $\mathbb{E}[Y(1,W(0),M(0,W(0)))]$ using the proposed methodology that calibrates the functions of X, W and M. To study path-specific effects, we consider the following decomposition of the average treatment effects:

$$
\begin{aligned}
&\mathbb{E}[Y(1,W(1),M(1,W(1)))-Y(0,W(0),M(0,W(0)))]\\
=&\ \mathbb{E}[Y(1,W(1),M(1,W(1)))-Y(1,W(0),M(1,W(0)))] &\quad(3.2)\\
&+\mathbb{E}[Y(1,W(0),M(1,W(0)))-Y(1,W(0),M(0,W(0)))] &\quad(3.3)\\
&+\mathbb{E}[Y(1,W(0),M(0,W(0)))-Y(0,W(0),M(0,W(0)))]\ . &\quad(3.4)
\end{aligned}
$$

This decomposition is also studied by Avin, Shpitser and Pearl (2005) and VanderWeele, Vansteelandt and Robins (2014). The first term represents a partial mediation effect through W, the second term is the partial mediation effect through M, and the third term is the direct effect of the treatment that does not go through W or M. Although the above decomposition is not the only way to define partial mediation effects, it has several advantages. First, the sum of the two partial mediation effects equals the joint natural indirect effects through both mediators. Moreover, the first term and the sum of the second and third terms are identified even if M is not observed.

To estimate the two partial mediation effects (3.2) and (3.3), we show below that $\mathbb{E}[Y(1,W(0),M(1,W(0)))]$ is estimable by adapting the proposed methodology under Assumption 9. The following lemma presents the results in the current setting that are analogous to those given in Lemma 1 in the univariate mediator case.

LEMMA 5. *Let* $q_0(X)=(Nf_{T|X}(0\mid X))^{-1}$ *and* $r_0(X,W)=f_{T|X,W}(0\mid X,W)\cdot(Nf_{T|X,W}(1\mid X,W)f_{T|X}(0\mid X))^{-1}$*. Under Assumption 9, for any suitable function* $v(X,M)$*, we have the following properties:*

$$
\mathbb{E}[Y(1,W(0),M(1,W(0)))]\ =\ \mathbb{E}[TNr_0(X,W)Y]\ , \qquad (3.5)
$$

$$\mathbb{E}[TNr_0(X,W)v(X,W)] \ = \ \mathbb{E}[(1-T)Nq_0(X)v(X,W)] \ . \tag{3.6}$$

PROOF. We begin by noting that

$$\begin{aligned}
&\mathbb{E}\left[Y(1,W(0),M(1,W(0))) \mid X=x\right] \\
=&\int_{\mathcal{W}\times\mathcal{M}} \mathbb{E}\left[Y(1,w,m) \mid X=x, W(0)=w, M(1,W(0))=m\right] \\
&\quad \cdot f_{W(0),M(1,W(0))|X}(w,m \mid x)dwdm
\end{aligned} \tag{3.7}$$

We can express each term in this equation using the observed data,

$$\begin{aligned}
&\mathbb{E}\left[Y(1,w,m) \mid X=x, W(0)=w, M(1,W(0))=m\right] \\
=&\mathbb{E}\left[Y(1,w,m) \mid X=x, W(0)=w, M(1,w)=m\right] \\
=&\mathbb{E}\left[Y(1,w,m) \mid X=x, W(0)=w, M(1,w)=m, T=0\right] \quad \text{(by Assumption 9 (1))} \\
=&\mathbb{E}\left[Y(1,w,m) \mid X=x, M(1,w)=m, T=0\right] \quad \text{(by Assumption 9 (2))} \\
=&\mathbb{E}\left[Y(1,w,m) \mid X=x, M(1,w)=m, T=1\right] \quad \text{(by Assumption 9 (1))} \\
=&\mathbb{E}\left[Y(1,w,m) \mid X=x, W(1)=w, M(1,w)=m, T=1\right] \quad \text{(by Assumption 9 (2))} \\
=&\mathbb{E}\left[Y(1,w,m) \mid X=x, W(1)=w, M(1,W(1))=m, T=1\right] \\
=&\mathbb{E}\left[Y \mid X=x, W=w, M=m, T=1\right] \ .
\end{aligned} \tag{3.8}$$

Similarly, we have

$$\begin{aligned}
&f_{W(0),M(1,W(0))|X}(w,m \mid x) \\
=&f_{M(1,W(0))|X,W(0)}(m \mid x,w) f_{W(0)|X}(w \mid x) \\
=&f_{M(1,w)|X,W(0)}(m \mid x,w) f_{W(0)|X}(w \mid x) \\
=&f_{M(1,w)|X,W(0),T}(m \mid x,w,0) f_{W(0)|X}(w \mid x) \quad \text{(by Assumption 9 (1))} \\
=&f_{M(1,w)|X,W(0),T}(m \mid x,w,0) f_{W(0)|X,T}(w \mid x,0) \quad \text{(by Assumption 9 (1))} \\
=&f_{M(1,w)|X,T}(m \mid x,0) f_{W(0)|X,T}(w \mid x,0) \quad \text{(by Assumption 9 (2))} \\
=&f_{M(1,w)|X,T}(m \mid x,1) f_{W(0)|X,T}(w \mid x,0) \quad \text{(by Assumption 9 (1))} \\
=&f_{M(1,w)|X,W(1),T}(m \mid x,w,1) f_{W(0)|X,T}(w \mid x,0) \quad \text{(by Assumption 9 (2))} \\
=&f_{M(1,W(1))|X,W(1),T}(m \mid x,w,1) f_{W(0)|X,T}(w \mid x,0) \\
=&f_{M|X,W,T}(m \mid x,w,1) f_{W|X,T}(w \mid x,0) \ .
\end{aligned} \tag{3.9}$$

Then, equations (3.7), (3.8) and (3.9) imply that

$$\mathbb{E}\left[Y(1,W(0),M(1,W(0))) \mid X=x\right]$$

$$
\begin{aligned}
&= \int_{\mathcal{W}\times\mathcal{M}} \mathbb{E}(Y \mid X=x, W=w, M=m, T=1) f_{M|X,W,T}(m \mid x,w,1) f_{W|X,T}(w \mid x,0) dwdm \\
&= \sum_{t=0}^{1} \int_{\mathcal{Y}\times\mathcal{W}\times\mathcal{M}} ty \cdot f_{Y|X,W,M,T}(y \mid x,w,m,t) f_{M|X,W,T}(m \mid x,w,1) f_{W|X,T}(w \mid x,0) dydwdm \\
&= \sum_{t=0}^{1} \int_{\mathcal{Y}\times\mathcal{W}\times\mathcal{M}} ty \cdot \frac{f_{Y,X,W,M,T}(y,x,w,m,t)}{f_{X,W,M,T}(x,w,m,1)} \cdot \frac{f_{M,X,W,T}(m,x,w,1)}{f_{X,W,T}(x,w,1)} \cdot \frac{f_{W,X,T}(w,x,0)}{f_{X,T}(x,0)} dydwdm \\
&= \sum_{t=0}^{1} \int_{\mathcal{Y}\times\mathcal{W}\times\mathcal{M}} ty \cdot \frac{f_{Y,X,W,M,T}(y,x,w,m,t)}{f_{X,W,T}(x,w,1)} \cdot \frac{f_{W,X,T}(w,x,0)}{f_{X,T}(x,0)} dydwdm \ .
\end{aligned}
$$

Therefore, we can have

$$
\begin{aligned}
&\mathbb{E}\left[Y(1, W(0), M(1, W(0)))\right] \\
&= \sum_{t=0}^{1} \int_{\mathcal{X}\times\mathcal{W}\times\mathcal{M}\times\mathcal{Y}} ty \cdot \frac{f_{Y,X,W,M,T}(y,x,w,m,t)}{f_{X,W,T}(x,w,1)} \cdot \frac{f_{W,X,T}(w,x,0)}{f_{X,T}(x,0)} f(x) dxdwdmdy \\
&= \mathbb{E}\left[TY \frac{f_{X,W,T}(X,W,0)}{f_{X,W,T}(X,W,1) f_{T|X}(0 \mid X)} \right] \\
&= \mathbb{E}\left[TYNr_0(X,W) \right] \ .
\end{aligned}
$$

This proves equation (3.5). For equation (3.6), we can show that the term on the left-hand side is equal to,

$$
\begin{aligned}
&\mathbb{E}[TNr_0(X,W)v(X,W)] \\
&= \int_{\mathcal{X}\times\mathcal{W}} \frac{f_{T|X,W}(0 \mid x,w)}{f_{T|X,W}(1 \mid x,w) f_{T|X}(0 \mid x)} v(x,w) f_{T,X,W}(1,x,w) dxdw \\
&= \int_{\mathcal{X}\times\mathcal{W}} \frac{f_{T|X,W}(0 \mid x,w)}{f_{T|X}(0 \mid x)} v(x,w) f_{X,W}(x,w) dxdw \\
&= \int_{\mathcal{X}\times\mathcal{W}} \frac{1}{f_{T|X}(0 \mid x)} v(x,w) f_{T,X,W}(0,x,w) dxdw \\
&= \sum_{t=0}^{1} \int_{\mathcal{X}\times\mathcal{W}} \frac{1-t}{f_{T|X}(0 \mid x)} v(x,w) f_{T,X,W}(t,x,w) dxdw \\
&= \mathbb{E}\left[(1-T) Nq_0(X) v(X,W)\right] \ .
\end{aligned}
$$

This proves equation (3.6). □

In Lemma 5, equation (3.5) expresses the average potential outcome, $\mathbb{E}[Y(1, W(0), M(1, W(0)))]$, in terms of observable variables; and equation (3.6) gives an identification condition of the unknown weight function $r_0(x,w)$. Therefore, $\mathbb{E}[Y(1,W(0),M(1,W(0)))]$ can be estimated by the proposed methodology described in the previous section that calibrates the functions of X and W but not M.

3.2. *Natural direct effect for the untreated.* Lendle, Subbaraman and van der Laan (2013) considered the estimation of the natural direct effect for the untreated (NDEU), which is defined as $\mathsf{NDEU} = \mathbb{E}[Y(1,M(0)) - Y(0,M(0)) \mid T = 0] = \theta_1' - \delta_0'$. Lendle, Subbaraman and van der Laan (2013) showed that the identification of NDEU requires the following assumption (along with Assumption 1).

Assumption 10. *For any $t \in \{0,1\}$, $m, m' \in \mathcal{M}$ and $x \in \mathcal{X}$, the following equality holds,*

$$\mathbb{E}\left[Y(t,m) - Y(0,m) \mid M(0) = m', X = x\right] \;=\; \mathbb{E}\left[Y(t,m) - Y(0,m) \mid X = x\right].$$

Under Assumption 10, we have

$$\begin{aligned}\mathsf{NDEU} \;&\triangleq\; \mathbb{E}\left[Y(1,M(0)) - Y(0,M(0)) \mid T = 0\right] \\ &=\; \mathbb{E}\left[\frac{TYf_{M,X|T}(M,X \mid 0)}{f_{T,M,X}(1,M,X)}\right] - \frac{\mathbb{E}\left[(1-T)Y\right]}{\mathbb{E}(1-T)} \;.\end{aligned}$$

Define

$$\tilde{r}_0(x,m) \;=\; \frac{f_{M,X|T}(m,x \mid 0)}{Nf_{T,M,X}(1,m,x)} \;,$$

for any integrable function $v(x,m)$, we have,

$$\begin{aligned}\mathbb{E}\left[Tv(X,M)N\tilde{r}_0(X,M)\right] \;&=\; \int v(x,m)\frac{f_{M,X|T}(m,x \mid 0)}{f_{T,M,X}(1,m,x)} \cdot f_{T,M,X}(1,m,x)dmdx \\ &=\; \int v(x,m)f_{M,X|T}(m,x \mid 0)dmdx \\ &=\; \frac{\mathbb{E}[(1-T)v(X,M)]}{\mathbb{E}(1-T)}.\end{aligned}$$

Therefore, we define

$$\hat{\tilde{r}}_K(x,m) \;=\; \frac{1}{N}\rho'\left(\hat{\beta}_{1K}^{\top}v_K(x,m)\right) \;,$$

where $\hat{\beta}_{1K}$ maximizes the following general empirical likelihood function $\widehat{H}_{1K}(\beta)$,

$$\widehat{H}_{1K}(\beta) \;\triangleq\; \frac{1}{N}\sum_{i=1}^{N}\left\{T_i\rho\left(\beta^{\top}v_K(X_i,M_i)\right) - \frac{(1-T_i)\beta^{\top}v_K(X_i,M_i)}{1-\overline{T}}\right\} \;,$$

where $\overline{T} \triangleq \frac{1}{N}\sum_{i=1}^{N} T_i$. Therefore, we can define the estimators of θ_1' and δ_0' to be,

$$\hat{\theta}_{1K}' \triangleq \sum_{i=1}^{N} T_i \hat{\hat{r}}_K(X_i, M_i) Y_i \ ,$$

$$\hat{\delta}_0' \triangleq \frac{\sum_{i=1}^{N}(1-T_i)Y_i}{\sum_{i=1}^{N}(1-T_i)} \ .$$

Given these estimators, the proposed estimator for the NDEU is given by,

$$\widehat{\mathsf{NDEU}} \triangleq \hat{\theta}_{1K}' - \hat{\delta}_0'$$

The proposed estimator $\widehat{\mathsf{NDEU}}$ extends the estimators of average treatment effects on the treated studied by Hainmueller (2012) and Chan, Yam and Zhang (2015) to causal mediation analysis. The following theorem summarizes the asymptotic properties of this estimator.

THEOREM 5. *Under Assumptions 1-8 and 10, $\widehat{\mathsf{NDEU}}$ has the following properties:*

1. $\widehat{\mathsf{NDEU}} \xrightarrow{p} \mathsf{NDEU}$
2. $\sqrt{N}\left(\widehat{\mathsf{NDEU}} - \mathsf{NDEU}\right) \xrightarrow{d} \mathcal{N}(0, V_{\mathsf{NDEU}})$, *where* $V_{\mathsf{NDEU}} = \mathbb{E}\left(S_{\mathsf{NDEU}}^2\right)$ *attained the semi-parametric efficiency bound (Lendle, Subbaraman and van der Laan, 2013), where*

$$S_{\mathsf{NDEU}} = \left\{\frac{\mathbf{1}\{T=1\}}{f_T(0)}\frac{f_{T|X,M}(0 \mid X, M)}{f_{T|X,M}(1 \mid X, M)} - \frac{\mathbf{1}\{T=0\}}{f_T(0)}\right\}\{Y - \mathbb{E}(Y \mid T, M, X)\} + \frac{\mathbf{1}\{T=0\}}{f_T(0)}\{\mathbb{E}(Y \mid T=1, M, X) - \mathbb{E}(Y \mid T=0, M, X) - \mathsf{NDEU}\} \ .$$

4. Empirical Applications. In this section, we briefly describe the application of the proposed methodology to two data sets, one regarding the evaluation of a job training program, while another concerning a psychological experiment from political science.

4.1. *Evaluation of a job training program.* We reanalyze the data from the JOBSII intervention study for unemployed job seekers (Vinokur and Schul, 1997). JOBSII is a randomized intervention study that investigates the efficacy of a job training program. In the study, 1801 unemployed workers were randomly assigned into two groups, with 1249 workers in the intervention arm and 552 workers in the control arm. The intervention group received five job training workshops that had a specific focus on improving a general sense of mastery of job seekers, which is a composite measure including confidence, internal locus of control, and self-esteem. It was hypothesized that the sense of mastery was a key mediator of a successful reemployment. The control group received a booklet but not the workshop sessions. Vinokur and Schul (1997) used a linear structural equation model and found that the enhanced sense of mastery had a significant mediating effect on reemployment.

To relax the linearity assumption, we reanalyze the JOBSII data using the present proposed nonparametric method. We include the following baseline covariates in the analysis: financial strain, depressive symptoms, age, sex, race, education, marital status, previous income, and previous occupation. The function $u(X)$ is set to be linear in covariates, and $v(X, M)$ is linear in covariates and the mediator. We find that the overall average treatment effect is 6.4% with a 95% confidence interval of [0.9%, 11.8%]. The estimated average natural indirect effect is 0.8% (resp. [0.0%, 1.6%]), while the estimated average direct effect is 5.6% (resp. [0.1%, 11.1%]). Our result suggests that the improvement in the sense of mastery contributed approximately 12.5% to the total average treatment effect. This finding is stronger than the original study, which reported the estimated proportion of this mediating effect to be 7.2%.

4.2. *A political framing study.* Next, we reanalyze the data from an experiment conducted by Brader, Valentino and Suhay (2008), who studied the role of emotions for the effect of news stories on the preferences of immigration policies. The authors randomly assigned 351 White non-Latino adults into two groups. In one group, respondents were shown a newspaper article with a picture of a Hispanic immigrant. In the other group, the same article was shown except with a picture of an European immigrant. The original authors studied the mediation of the framing effect through two psychological mechanisms: the perceived harm mechanism and the anxiety mechanism. They assumed that the two mediators are independent but the assumption may not be plausible because an increased level of perceived harm of immigration can cause more anxiety about immigration policies. Therefore, we apply the proposed methodology described in Section 3.1, which allows for multiple mediators. Specifically, we use the level of perceived harm as W while the level of anxiety is M. We include the same set of covariates as in the original paper, which includes education, age, income, and gender.

We find that the news article with a picture of a Latino immigrant leads to a 0.430 percentage point increase (relative to the same story with a picture of an European immigrant) in the opposition of increased immigration on the five-point scale, with a 95% confidence interval of [0.199, 0.661]. The combined mediation effect from the perceived harm and anxiety mechanisms is estimated to be 0.164 (resp. [0.001, 0.327]). This implies that the two mediators together account for 38% of the total average treatment effect. We further decompose this combined mediation effect into two partial mediation effects. The partial indirect effect concerning the perceived harm mechanism is estimated to be 0.115 (resp. [−0.03, 0.263]) whereas the estimated partial mediation effect for the anxiety mechanism is 0.049 (resp. [−0.159, 0.257]). This finding suggests that the anxiety mechanism may play only a secondary role in the public opinion about immigration policy. Our results contradict with the original study, which concluded that the anxiety plays an essential role by assuming that the two mediators are causally independent.

5. Concluding remarks. In this paper, we proposed a novel methodology for causal mediation analysis. We establish that the proposed estimator is fully nonparametric and globally semiparametric efficient. This improves the existing estimators, which rely upon parametric assumptions and is only locally efficient. Furthermore, we show how to consistently estimate the asymptotic variance of this proposed estimator without requiring additional nonparametric estimation. Another advantage is the availability of efficient and stable numerical algorithms that can be used to compute the proposed estimator and its estimated variance. We show how to extend our methodology to a setting with multiple mediators and the estimation of related causal quantities of interest. While causal mediation analysis has gained popularity in a variety of disciplines, applied researchers have largely relied upon parametric methods. We believe that our nonparametric estimator has a potential to significantly improve the credibility of causal mediation analysis in scientific research.

References.

Albert, J. M. (2008). Mediation analysis via potential outcomes models. *Statistics in Medicine* **27** 1282–1304.

Avin, C., Shpitser, I. and Pearl, J. (2005). Identifiability of Path-Specific Effects. In *Proceedings of the Nineteenth International Joint Conference on Artificial Intelligence* 357–363. Morgan Kaufmann, Edinburgh, Scotland.

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.

Brader, T., Valentino, N. and Suhay, E. (2008). What Triggers Public Opposition to Immigration? Anxiety, Group Cues, and Immigration Threat. *American Journal of Political Science* **52** 959–978.

Chan, K. C. G. and Yam, S. C. P. (2014). Oracle, Multiple Robust and Multipurpose Calibration in a Missing Response Problem. *Statistical Science* **29** 380–396.

Chan, K., Yam, S. and Zhang, Z. (2015). Globally Efficient Nonparametric Inference of Average Treatment Effects by Empirical Balancing Calibration Weighting. *To appear in Journal of Royal Statistical Society: Series B.*

Chen, X., Hong, H. and Tarozzi (2008). Semiparametric efficiency in GMM models with auxiliary data. *The Annals of Statistics* **36** 808-843.

Deville, J.-C. and Särndal, C.-E. (1992). Calibration estimators in survey sampling. *Journal of the American statistical Association* **87** 376–382.

Geneletti, S. (2007). Identifying direct and indirect effects in a non-counterfactual framework. *Journal of the Royal Statistical Society, Series B (Statistical Methodology)* **69** 199–215.

Graham, B. S., Pinto, C. and Egel, D. (2012). Inverse probability tilting for moment condition models with missing data. *Review of Economic Studies* **79** 1053–1079.

Hahn, J. (1998). On the role of the propensity score in efficient semiparametric estimation of average treatment effects. *Econometrica* 315–331.

Hainmueller, J. (2012). Entropy Balancing for Causal Effects: Multivariate Reweighting Method to Produce Balanced Samples in Observational Studies. *Political Analysis* **20** 25–46.

Han, P. and Wang, L. (2013). Estimation with missing data: beyond double robustness. *Biometrika* **100** 417–430.

Hansen, L. P., Heaton, J. and Yaron, A. (1996). Finite-sample properties of some alternative GMM estimators. *Journal of Business & Economic Statistics* **14** 262–280.

Hirano, K., Imbens, G. and Ridder, G. (2003). Efficient estimation of average treatment effects using the estimated propensity score. *Econometrica* **71** 1161–1189.

Imai, K., Keele, L. and Tingley, D. (2010). A General Approach to Causal Mediation Analysis. *Psychological Methods* **15** 309–334.

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

Imai, K. and Ratkovic, M. (2014). Covariate Balancing Propensity Score. *Journal of the Royal Statistical Society, Series B (Statistical Methodology)* **76** 243–263.

Imai, K. and Yamamoto, T. (2013). Identification and Sensitivity Analysis for Multiple Causal Mechanisms: Revisiting Evidence from Framing Experiments. *Political Analysis* **21** 141–171.

Imai, K., Keele, L., Tingley, D. and Yamamoto, T. (2011). Unpacking the Black Box of Causality: Learning about Causal Mechanisms from Experimental and Observational Studies. *American Political Science Review* **105** 765–789.

Imbens, G. W., Newey, W. K. and Ridder, G. (2005). Mean-square-error calculations for average treatment effects. *IEPR working paper.*

Jo, B. (2008). Causal Inference in Randomized Experiments with Mediational Processes. *Psychological Methods* **13** 314–336.

Joffe, M. M., Small, D., Hsu, C.-Y. et al. (2007). Defining and estimating intervention effects for groups that will develop an auxiliary outcome. *Statistical Science* **22** 74–97.

Kitamura, Y. and Stutzer, M. (1997). An information-theoretic alternative to generalized method of moments estimation. *Econometrica* **65** 861–874.

Lendle, S. D., Subbaraman, M. S. and van der Laan, M. J. (2013). Identification and efficient estimation of the natural direct effect among the untreated. *Biometrics* **69** 310–317.

MacKinnon, D. P. (2008). *Introduction to statistical mediation analysis*. Routledge.

Pearl, J. (2001). Direct and Indirect Effects. *Proceedings of the seventeenth conference* *on uncertainty in articial intelligence* 411–420.

PEARL, J. (2012). The Causal Mediation Formula: A Guide to the Assessment of Pathways and Mechanisms. *Prevention Science* **13** 426–436.

QIN, J. and ZHANG, B. (2007). Empirical-likelihood-based inference in missing response problems and its application in observational studies. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **69** 101–122.

RICHARDSON, T. S. and ROBINS, J. M. (2013). Single World Intervention Graphs (SWIGs): A Unification of the Counterfactual and Graphical Approaches to Causality Technical Report, University of Washington.

ROBINS, J. M. (2003). Semantics of causal DAG models and the identification of direct and indirect effects. In *Highly Structured Stochastic Systems (eds., P.J. Green, N.L. Hjort, and S. Richardson)* 70–81. Oxford University Press, Oxford.

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

ROBINS, J. M., ROTNITZKY, A. and ZHAO, L. P. (1994). Estimation of Regression Coefficients When Some Regressors are Not Always Observed. *Journal of the American Statistical Association* **89** 846–866.

SOBEL, M. E. (2008). Identification of Causal Parameters in Randomized Studies with Mediating Variables. *Journal of Educational and Behavioral Statistics* **33** 230–251.

TCHETGEN TCHETGEN, E. J. and SHPITSER, I. (2012). Semiparametric theory for causal mediation analysis: efficiency bounds, multiple robustness and sensitivity analysis. *The Annals of Statistics* **40** 1816–1845.

TEN HAVE, T. R., JOFFE, M. M., LYNCH, K. G., BROWN, G. K., MAISTO, S. A. and BECK, A. T. (2007). Causal mediation analyses with rank preserving models. *Biometrics* **63** 926–934.

TSENG, P. and BERTSEKAS, D. P. (1987). Relaxation methods for problems with strictly convex separable costs and linear constraints. *Mathematical Programming* **38** 303–321.

VANDERWEELE, T. J. (2009). Marginal structural models for the estimation of direct and indirect effects. *Epidemiology* **20** 18–26.

VANDERWEELE, T. J. (2010). Bias formulas for sensitivity analysis for direct and indirect effects. *Epidemiology* **21** 540–551.

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

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

VANDERWEELE, T. J., VANSTEELANDT, S. and ROBINS, J. M. (2014). Effect decomposition in the presence of an exposure-induced mediator-outcome confounder. *Epidemiology* **25** 300–306.

VINOKUR, A. D. and SCHUL, Y. (1997). Mastery and inoculation against setbacks as active ingredients in the JOBS intervention for the unemployed. *Journal of consulting and clinical psychology* **65** 867–877.

K.C.G. CHAN
DEPARTMENT OF BIOSTATISTICS
UNIVERSITY OF WASHINGTON
E-MAIL: kcgchan@uw.edu

K. IMAI
DEPARTMENT OF POLITICS
CENTER FOR STATISTICS AND MACHINE LEARNING
PRINCETON UNIVERSITY
E-MAIL: kimai@princeton.edu

S.C.P. YAM
DEPARTMENT OF STATISTICS
THE CHINESE UNIVERSITY OF HONG KONG
E-MAIL: scpyam@sta.cuhk.edu.hk

Z. ZHANG
DEPARTMENT OF STATISTICS AND ACTUARIAL SCIENCE
THE UNIVERSITY OF HONG KONG
E-MAIL: zzhang1989@gmail.com