Harvard University
Harvard University Biostatistics Working Paper Series

Year 2011 *Paper* 130

Semiparametric Theory for Causal Mediation Analysis: efficiency bounds, multiple robustness, and sensitivity analysis

Eric J. Tchetgen Tchetgen* Ilya Shpitser†

*Harvard University, etchetge@hsph.harvard.edu
†

Semiparametric Theory for Causal Mediation Analysis: efficiency bounds, multiple robustness, and sensitivity analysis

Eric J. Tchetgen Tchetgen and Ilya Shpitser

Abstract

Whilst estimation of the marginal (total) causal effect of a point exposure on an outcome is arguably the most common objective of experimental and observational studies in the health and social sciences, in recent years, investigators have also become increasingly interested in mediation analysis. Specifically, upon establishing a non-null total effect of the exposure, investigators routinely wish to make inferences about the direct (indirect) pathway of the effect of the exposure not through (through) a mediator variable that occurs subsequently to the exposure and prior to the outcome. Although powerful semiparametric methodologies have been developed to analyze observational studies, that produce double robust and highly efficient estimates of the marginal total causal effect, similar methods for mediation analysis are currently lacking. Thus, this paper develops a general semiparametric framework for obtaining inferences about so-called marginal natural direct and indirect causal effects, while appropriately accounting for a large number of pre-exposure confounding factors for the exposure and the mediator variables. Our analytic framework is particularly appealing, because it gives new insights on issues of efficiency and robustness in the context of mediation analysis. In particular, we propose new multiply robust locally efficient estimators of the marginal natural indirect and direct causal effects, and develop a novel double robust sensitivity analysis framework for the assumption of ignorability of the mediator variable.

Semiparametric Theory for Causal Mediation Analysis: efficiency bounds, multiple robustness, and sensitivity analysis

by

Eric J. Tchetgen Tchetgen[#†*] and Ilya Shpitser[†]

Departments of Epidemiology[†] and Biostatistics[#], Harvard University

Abstract

Whilst estimation of the marginal (total) causal effect of a point exposure on an outcome is arguably the most common objective of experimental and observational studies in the health and social sciences, in recent years, investigators have also become increasingly interested in mediation analysis. Specifically, upon establishing

Key Words and Phrases: Natural direct effects, Natural indirect effects; double robust;mediation analysis, local efficiency

AMS 1991 Subject Classifications. Primary: 62G05 .

[*]Supported by NIH grant #R21ES019712

a non-null total effect of the exposure, investigators routinely wish to make inferences about the direct (indirect) pathway of the effect of the exposure not through (through) a mediator variable that occurs subsequently to the exposure and prior to the outcome. Although powerful semiparametric methodologies have been developed to analyze observational studies, that produce double robust and highly efficient estimates of the marginal total causal effect, similar methods for mediation analysis are currently lacking. Thus, this paper develops a general semiparametric framework for obtaining inferences about so-called marginal natural direct and indirect causal effects, while appropriately accounting for a large number of pre-exposure confounding factors for the exposure and the mediator variables. Our analytic framework is particularly appealing, because it gives new insights on issues of efficiency and robustness in the context of mediation analysis. In particular, we propose new multiply robust locally efficient estimators of the marginal natural indirect and direct causal effects, and develop a novel double robust sensitivity analysis framework for the assumption of ignorability of the mediator variable.

1 Introduction

Estimation of the total causal effect of a given point exposure, treatment or intervention on an outcome of interest is arguably the most common objective of experimental and observational studies in the fields of epidemiology, biostatistics and in the social

sciences. However, in recent years, investigators in these various fields have become increasingly interested in making inferences about the direct (indirect) pathway of the exposure effect not through (through) a mediator variable that occurs subsequently to the exposure and prior to the outcome. Recently, the counterfactual language of causal inference has proven particularly useful for formalizing mediation analysis. Causal inference indeed offers a formal mathematical framework for defining varieties of direct and indirect effects, and for establishing necessary and sufficient identifying conditions of these effects. A notable contribution of causal inference to the literature on mediation analysis is the key distinction drawn between so-called controlled direct and indirect effects versus natural direct and indirect effects. In words, the controlled direct effect refers to the exposure effect that arises upon intervening to set the mediator to a fixed level that may differ from its actual observed value (Robins and Greenland, 1992, Pearl, 2001, Robins, 2003). In contrast, the natural (also known as pure) direct effect captures the effect of the exposure when one intervenes to set the mediator to the (random) level it would have been in the absence of exposure(Robins and Greenland, 1992, Pearl 2001). The controlled direct effect combines with the controlled indirect effect to produce the joint effect of the exposure and the mediator, whereas, the natural direct and indirect effects combine to produce the exposure total effect. As noted by Pearl (2001), controlled direct and indirect effects are particularly relevant for policy making whereas natural direct and indirect effects are more useful

for understanding the underlying mechanism by which the exposure operates.

In an effort to account for confounding bias when estimating causal effects, say for instance when estimating the marginal total effect of the exposure from non-experimental data, investigators routinely collect and adjust for in data analysis, a large number of confounding factors. Because of the curse of dimensionality, non-parametric methods of estimation are typically not practical in such settings, and one usually resorts to one of two dimension-reduction strategies. The first strategy uses a parametric or semiparametric model relating the confounders to the outcome controlling for the exposure, whereas the second strategy uses a parametric or semi-parametric model for the conditional density of the exposure given the confounders, i.e. the propensity score, to recover an adjusted estimate of the total causal effect. An important drawback of either of these two approaches is a strong reliance on the corresponding modeling assumption, which when incorrect, can produce severely biased effect estimates and ultimately lead to the incorrect inferences about the causal effect of interest. As a remedy, powerful semiparametric methods have recently been developed to analyze observational studies, that produce so-called double robust and highly efficient estimates of the exposure total causal effect (Robins, 1999, Scharfstein, Rotnitzky and Robins, 1999, Bang and Robins, 2005, Tsiatis, 2006) and similar methods have also been developed to estimate controlled direct and indirect effects (Goetgeluk, Vansteelandt and Goetghebeur, 2008). An important advantage of a dou-

ble robust method is that it carefully combines both of the aforementioned dimension reduction strategies for confounding adjustment, to produce an estimator of the causal effect that remains consistent and asymptotically normal provided at least one of the two strategies is correct, without necessarily knowing which strategy is indeed correct (van der Laan and Robins, 2003). Unfortunately, similar methods for making semiparametric inferences about marginal natural direct and indirect effects are currently lacking. Thus, this paper develops a general semiparametric framework for obtaining inferences about marginal natural direct and indirect effects on the mean of an outcome, while appropriately accounting for a large number of confounding factors for the exposure and the mediator variables.

Our semiparametric framework is particularly appealing, as it gives new insight on issues of efficiency and robustness in the context of mediation analysis. Specifically, in Section 2, we adopt the sequential ignorability assumption of Imai et al (2010) under which, in conjunction with the standard consistency and positivity assumptions, we derive the efficient influence function and thus obtain the semiparametric efficiency bound for the natural direct and natural indirect marginal mean causal effects, in the nonparametric model $\mathcal{M}_{\text{nonpar}}$ in which the observed data likelihood is left unrestricted. We further show that in order to conduct mediation inferences in $\mathcal{M}_{\text{nonpar}}$, one must estimate at least a subset of the following quantities:

(i) the conditional expectation of the outcome given the mediator, exposure and

confounding factors;

(ii) the density of the mediator given the exposure and the confounders;

(iii) the density of the exposure given the confounders.

Ideally, to minimize the possibility of modeling bias, one may wish to estimate each of these quantities nonparametrically; however, as discussed in the previous paragraph, when as we assume throughout we observe a high dimensional vector of confounders of the exposure and the mediator, such nonparametric estimates will likely perform poorly in finite samples. Thus, in Section 2.3 we develop an alternative multiply robust strategy. To do so, we propose to model (i), (ii) and (iii) parametrically (or semiparametrically), but rather than obtaining mediation inferences that rely on the correct specification of a specific subset of these models, instead we carefully combine these three models to produce estimators of the marginal mean direct and indirect effects that remain consistent and asymptotically normal (CAN) in a union model where at least one but not necessarily all of the following conditions hold:

(a) the parametric models for the conditional expectation of the outcome (i) and for the conditional density of the mediator (ii) are correctly specified;

(b) the parametric models for the conditional expectation of the outcome (i) and for the conditional density of the exposure (iii) are correctly specified

(c) the parametric models for the conditional densities of the exposure and the mediator (ii) and (iii) are correctly specified.

Accordingly, we define submodels $\mathcal{M}_a, \mathcal{M}_b$ and $\mathcal{M}_c$ of $\mathcal{M}_{\text{nonpar}}$ corresponding to models (a), (b) and (c) respectively. Thus, the proposed approach is triply robust as it produces valid inferences about natural direct and indirect effects in the union model $\mathcal{M}_{\text{union}} = \mathcal{M}_a \cup \mathcal{M}_b \cup \mathcal{M}_c$. Furthermore, as we later show in Section 2.3, the proposed estimators are also locally semiparametric efficient in the sense that they achieve the respective efficiency bounds for estimating the natural direct and indirect effects in $\mathcal{M}_{\text{union}}$, at the intersection submodel $\mathcal{M}_a \cap \mathcal{M}_b \cap \mathcal{M}_c = \mathcal{M}_a \cap \mathcal{M}_c = \mathcal{M}_a \cap \mathcal{M}_b = \mathcal{M}_b \cap \mathcal{M}_c \subset \mathcal{M}_{\text{union}} \subset \mathcal{M}_{\text{nonpar}}$.

In Section 2.4, we compare the proposed methodology to the prevailing estimators in the literature. Based on this comparison, we conclude that the new approach should generally be preferred because an inference under the proposed method is guaranteed to remain valid under many more data generating laws than an inference based on each of the other existing approaches. In particular, as we argue below the approach of van der Laan and Petersen (2005) is not entirely satisfactory because, despite producing a CAN estimator of the marginal direct effect under the union model $\mathcal{M}_a \cup \mathcal{M}_c$ (and therefore an estimator that is double robust), their estimator requires a correct model for the density of the mediator. Thus unlike the direct effect estimator developed in this paper, the van der Laan estimator fails to be consistent

under the submodel $\mathcal{M}_b \subset \mathcal{M}_{\text{union}}$. Nonetheless, the estimator of van der Laan is in fact locally efficient in model $\mathcal{M}_a \cup \mathcal{M}_c$, provided the model for the mediator's conditional density is either known, or can be efficiently estimated. This property is confirmed in Section 3, where we also provide a general map that relates the efficient influence function for model $\mathcal{M}_{\text{union}}$ to the corresponding influence function for model $\mathcal{M}_a \cup \mathcal{M}_c$ assuming an arbitrary parametric or semiparametric model for the mediator conditional density is correctly specified. In Section 4, we describe a novel double robust sensitivity analysis framework to assess the impact on inferences about the natural direct effect, of a departure from the ignorability assumption of the mediator variable. We conclude with a brief discussion.

2 The nonparametric mediation functional

2.1 Identification

Suppose i.i.d data on $O = (Y, E, M, X)$ is collected for n subjects. Here, Y is an outcome of interest, E is a binary exposure variable, M is a mediator variable with support $\mathcal{S}$, known to occur subsequently to E and prior to Y, and X is a vector of pre-exposure variables with support $\mathcal{X}$ that confound the association between (E, M) and Y. The overarching goal of this paper is to provide some theory of inference about the fundamental functional of mediation analysis which Judea Pearl calls "the

mediation causal formula" (Pearl, 2010) and which expressed on the mean scale, is :

$$\theta_0 = \iint\limits_{\mathcal{S}\times\mathcal{X}} \mathbb{E}\left(Y|E=1,M=m,X=x\right) f_{M|E,X}\left(m|E=0,X=x\right) f_X(x) d\mu(m,x) \quad (1)$$

where $\mathbb{E}$ stands for expectation, $f_{M|E,X}$ and f_X are respectively the conditional density of the mediator M given (E,X) and the density of X, and μ is a dominating measure for the distribution of (M,X). Hereafter, to keep with standard statistical parlance, we shall simply refer to θ_0 as the "mediation functional" or "M-functional" since it is formally a functional on the nonparametric statistical model $\mathcal{M}_{nonpar} = \{F_O(\cdot) : F_O$ unrestricted$\}$ of all regular laws F_O of the observed data O that satisfy the positivity assumption given below; i.e. $\theta_0 = \theta_0(F_O) : \mathcal{M}_{nonpar} \rightarrow \mathcal{R}$, with $\mathcal{R}$ the real line. The functional θ_0 is of keen interest here because it arises in the estimation of natural direct and indirect effects which we describe next. To do so, we assume for each level $E = e$, $M = m$, there exist a counterfactual variable $Y_{e,m}$ corresponding to the outcome had possibly contrary to fact the exposure and mediator variables taken the value (e,m) and for $E = e$, there exist a counterfactual variable M_e corresponding to the mediator variable had possibly contrary to fact the exposure variable taken the value e.

Furthermore, we make the consistency assumption:

Consistency

$$\text{if } E = e \,, \text{then } M_e = M \text{ w.p.1}$$

$$\text{and if } E = e \quad \text{and } M = m \quad \text{then } Y_{e,m} = Y \text{ w.p.1}$$

In addition, we adopt the sequential ignorability of Imai et al (2010) which states that for $e, e' \in \{0, 1\}$:

Sequential ignorability

$$\{Y_{e',m}, M_e\} \amalg E|X$$

$$Y_{e'm} \amalg M|E = e, X$$

paired with the following

positivity :

$$f_{M|E,X}\left(m|E, X\right) > 0 \text{ w.p.1 for each } m \in \mathcal{S}$$

$$\text{and } f_{E|X}\left(e|X\right) > 0 \text{ w.p.1 for each } e \in \{0, 1\}$$

Then, under the consistency, sequential ignorability and positivity assumptions, Imai et al (2010) showed that:

$$\theta_0 = \mathbb{E}\left(Y_{1,M_0}\right), \text{ and}$$

$$\begin{aligned}\delta_e &\equiv \int_{\mathcal{X}} \mathbb{E}\left(Y|E=e, X=x\right) f_X(x) d\mu(x) \\ &= \iint_{\mathcal{S}\times\mathcal{X}} \mathbb{E}\left(Y|E=e, M=m, X=x\right) f_{M|E,X}\left(m|E=e, X=x\right) f_X(x) d\mu(m,x) \quad (2) \\ &= \mathbb{E}\left(Y_e\right) = \mathbb{E}\left(Y_{e,M_e}\right); \; e=0,1\end{aligned}$$

so that $\mathbb{E}\left(Y_{1,M_0}\right)$ and $\mathbb{E}\left(Y_j\right)$ are identified from the observed data, and so is the mean natural direct effect (on the mean difference scale)

$$\mathbb{E}\left(Y_{1,M_0}\right) - \mathbb{E}\left(Y_0\right) = \theta_0 - \delta_0$$

and the mean natural indirect effect (on the mean difference scale) :

$$\mathbb{E}\left(Y_1\right) - \mathbb{E}\left(Y_{1,M_0}\right) = \delta_1 - \theta_0$$

For binary Y, one might alternatively consider the natural direct effect on the risk ratio scale

$$\mathbb{E}\left(Y_{1,M_0}\right)/\mathbb{E}\left(Y_0\right) = \theta_0/\delta_0$$

or on the odds ratio scale

$$\frac{\mathbb{E}\left(Y_{1,M_0}\right)\mathbb{E}\left(1-Y_0\right)}{\mathbb{E}\left(1-Y_{1,M_0}\right)\mathbb{E}\left(Y_0\right)} = \frac{\left\{\theta_0\left(1-\delta_0\right)\right\}}{\left\{\delta_0\left(1-\theta_0\right)\right\}}$$

and similarly defined natural indirect effects on the risk ratio and odds ratio scales. It is instructive to contrast the expression (1) for $\mathbb{E}\left(Y_{1,M_0}\right)$ with the expression (2) for $e = 1$ corresponding to $\mathbb{E}\left(Y_1\right)$, and to note that the two expressions bare a striking resemblance except the density of the mediator in the first expression conditions on the unexposed (with $E = 0$) whereas in the second expression, the mediator density is conditional on the exposed (with $E = 1$). As we demonstrate below, this subtle difference has remarkable implications for inference.

Pearl (2001) was the first to derive the M-functional $\theta_0 = \mathbb{E}\left(Y_{1,M_0}\right)$ under a different set of assumptions. Others have since contributed alternative sets of identifying assumptions. In this paper, we have chosen to work under the sequential ignorability assumption of Imai et al but note that alternative related assumptions exist in the literature (Robins and Greenland,1992, Pearl, 2001, Petersen and van der Laan, 2005, Hafeman and Vanderweele, 2010). Although we note that Robins and Richardson (2010) disagree with the label "sequential ignorability" because its terminology has previously carried a different interpretation in the literature. Nonetheless, the assumption entails two ignorability-like assumptions that are made sequentially. First, given the observed preexposure confounders, the exposure assignment is assumed to be ignorable, that is, statistically independent of potential outcomes and potential mediators. The second part of the assumption states that the mediator is ignorable given the observed exposure and preexposure confounders. Specifically, the second part of

the sequential ignorability assumption is made conditional on the observed value of the ignorable treatment and the observed pretreatment confounders. We note that the second part of the sequential ignorability assumption is particularly strong and must be made with care. This is partly because, it is always possible that there might be unobserved variables that confound the relationship between the outcome and the mediator variables even upon conditioning on the observed exposure and covariates. Furthermore, the confounders X must all be pre-exposure variables, i.e. they must precede E. In fact, Avin et al (2005) proved that without additional assumptions, one cannot identify natural direct and indirect effects if there are confounding variables that are affected by the exposure even if such variables are observed by the investigator. This implies that similar to the ignorability of the exposure in observational studies, ignorability of the mediator cannot be established with certainty even after collecting as many pre-exposure confounders as possible. Furthermore, as Robins and Richardson (2010) point out, whereas the first part of the sequential ignorability assumption could in principle be enforced in a randomized study, by randomizing E within levels of X; the second part of the sequential ignorability assumption cannot similarly be enforced experimentally, even by randomization. And thus for this latter assumption to hold, one must entirely rely on expert knowledge about the mechanism under study. For this reason, it will be crucial in practice to supplement mediation analyses with a sensitivity analysis that accurately quantifies the degree to which

results are robust to a potential violation of the sequential ignorability assumption. For this reason, later in the paper, we develop a set of sensitivity analyses that will allow the analyst to quantify the degree to which his or her mediation analysis results are robust to a potential violation of the sequential ignorability assumption.

2.2 Semiparametric efficiency bounds for $\mathcal{M}_{\text{nonpar}}$

In this section, we derive the efficient influence function for the M-functional θ_0 in $\mathcal{M}_{nonpar}$, this result is then combined with the efficient influence function for the functional δ_e (Robins, Rotnitzky and Zhao, 1994, Hahn, 1998) to obtain the efficient influence function for the natural direct and indirect effects, on the mean difference scale. Thus, in the following, we shall use the efficient influence function $S_{\delta e}^{eff,nonpar}(\delta_e)$ of δ_e which is well known to be:

$$\frac{I(E=e)}{f_{E|X}(e|X)}\{Y-\eta(e,e,X)\}+\eta(e,e,X)+\delta_e$$

where for $e, e^* \in \{0,1\}$, we define

$$\eta(e,e^*,X)=\int_{\mathcal{S}}\mathbb{E}(Y|X,M=m,E=e)\,f_{M|E,X}(m|E=e^*,X)\,d\mu(m)$$

so that $\eta(e,e,X)=\mathbb{E}(Y|X,E=e)$, $e=0,1$

The following theorem is proved in the appendix

Theorem 1: Under the consistency, sequential ignorability and positivity assumptions, the efficient influence function of the M-functional θ_0 in model $\mathcal{M}_{nonpar}$ is given by:

$$
\begin{aligned}
S_{\theta_0}^{eff,nonpar}(\theta_0) &= s_{\theta_0}^{eff,nonpar}(O;\theta_0) \\
&= \frac{I\{E=1\} f_{M|E,X}(M|E=0,X)}{f_{E|X}(1|X) f_{M|E,X}(M|E=1,X)} \{Y - \mathbb{E}(Y|X,M,E=1)]\} \\
&+ \frac{I(E=0)}{f_{E|X}(0|X)} \{\mathbb{E}(Y|X,M,E=1) - \eta(1,0,X)\} \\
&+ \eta(1,0,X) - \theta_0
\end{aligned}
$$

and the efficient influence function of the natural direct and indirect effects on the mean difference scale in model $\mathcal{M}_{nonpar}$ *are respectively given by:*

$$
\begin{aligned}
S_{NDE}^{eff,nonpar}(\theta_0,\delta_0) &= s_{NDE}^{eff,nonpar}(O;\theta_0,\delta_0) = S_{\theta_0}^{eff,nonpar}(\theta_0) - S_{\delta_0}^{eff,nonpar}(\delta_0) \\
&= \frac{I\{E=1\} f_{M|E,X}(M|E=0,X)}{f_{E|X}(1|X) f_{M|E,X}(M|E=1,X)} \{Y - \mathbb{E}(Y|X,M,E=1)]\} \\
&+ \frac{I(E=0)}{f_{E|X}(0|X)} \{\mathbb{E}(Y|X,M,E=1) - Y - \eta(1,0,X) + \eta(0,0,X)\} \\
&+ \eta(1,0,X) - \eta(0,0,X) - \theta_0 + \delta_0
\end{aligned}
$$

$$
\begin{aligned}
S_{NIE}^{eff,nonpar}(\delta_1,\theta_0) &= s_{NIE}^{eff,nonpar}(O;\delta_1,\theta_0) = S_{\theta_0}^{eff,nonpar}(\theta_0) - S_{\delta_1}^{eff,nonpar}(\delta_1) \\
&= \frac{I(E=1)}{f_{E|X}(1|X)} \left\{ \begin{array}{c} Y - \eta(1,1,X) \\ -\frac{f_{M|E,X}(M|E=0,X)}{f_{M|E,X}(M|E=1,X)} \{Y - \mathbb{E}(Y|X,M,E=1)]\} \end{array} \right\} \\
&- \frac{I(E=0)}{f_{E|X}(0|X)} \{\mathbb{E}(Y|X,M,E=1) - \eta(1,0,X)\} \\
&+ \eta(1,1,X) - \eta(1,0,X) + \theta_0 - \delta_1
\end{aligned}
$$

Thus the semiparametric efficiency bound for estimating the natural direct and the natural indirect effects in M_{nonpar} *are respectively given by* $\mathbb{E}\left\{S_{NDE}^{eff,nonpar}\left(\theta_0,\delta_0\right)^2\right\}$ *and* $\mathbb{E}\left\{S_{NIE}^{eff,nonpar}\left(\delta_1,\theta_0\right)^2\right\}$

Although not presented here, Theorem 1 is easily extended to obtain the efficient influence functions and the respective semiparametric efficiency bounds for the direct and indirect effects on the risk ratio and the odds ratio scales by a straightforward application of the delta method. An important implication of the theorem is that all regular and asymptotically linear (RAL) estimators of $\theta_0, \delta_1-\theta_0$ and $\theta_0-\delta_0$ in model $\mathcal{M}_{nonpar}$ share the common influence functions $S_{\theta_0}^{eff,nonpar}\left(\theta_0\right), S_{NDE}^{eff,nonpar}\left(\theta_0,\delta_0\right)$ and $S_{NIE}^{eff,nonpar}\left(\delta_1,\theta_0\right)$ respectively. Specifically, any RAL estimator $\widehat{\theta}_0$ of the M-functional θ_0 in model $\mathcal{M}_{nonpar}$, shares a common asymptotic expansion

$$n^{1/2}\left(\widehat{\theta}_0-\theta_0\right)=n^{1/2}\mathbb{P}_n S_{\theta_0}^{eff,nonpar}\left(\theta_0\right)+o_P\left(1\right)$$

where $\mathbb{P}_n\left[\cdot\right]=n^{-1}\sum_i\left[\cdot\right]_i$. To illustrate this property of nonparametric RAL estimators and as a motivation for multiply robust estimation when nonparametric methods are not appropriate, we provide a detailed study of three nonparametric strategies for estimating the M-functional in a simple yet instructive setting in which X and M are both discrete with finite support.

Strategy 1: The first strategy entails obtaining the maximum likelihood estimator

upon evaluating the M-functional under the empirical law of the observed data:

$$\widehat{\theta}_0^{ym} = \mathbb{P}_n \sum_{m \in \mathcal{S}} \widehat{\mathbb{E}}\left(Y|E=1, M=m, X\right) \widehat{f}_{M|E,X}\left(m|E=0, X\right)$$

where $\widehat{f}_{Y|E,M,X}$ and $\widehat{f}_{M|E,X}$ are the empirical probability mass functions, and $\widehat{\mathbb{E}}(Y|E = e,\ M = m, X = x)$ is the expectation of Y under $\widehat{f}_{Y|E,M,X}$.

Strategy 2: The second strategy is based on the following alternative representation of the M-functional

$$\begin{aligned}
&\iint_{\mathcal{S}\times\mathcal{X}} \mathbb{E}\left(Y|E=1, M=m, X=x\right) dF_{M|E}\left(m|E=0, X=x\right) dF_X\left(x\right) \\
&= \sum_{e=0}^{1} \iint_{\mathcal{S}\times\mathcal{X}} \mathbb{E}\left(Y|E=1, M=m, X=x\right) \frac{I(e=0)}{f_{E|X}(e|X=x)} dF_{M,E,X}\left(m, e, x\right) \\
&= \mathbb{E}\left\{ \frac{I(E=0)}{f_{E|X}(0|X)} \mathbb{E}\left(Y|E=1, M, X\right) \right\}
\end{aligned}$$

Thus, our second estimator takes the form:

$$\widehat{\theta}_0^{ye} = \mathbb{P}_n \left\{ \frac{I(E=0)}{\widehat{f}_{E|X}(0|X)} \widehat{\mathbb{E}}\left(Y|E=1, M, X\right) \right\}$$

with $\widehat{f}_{E|X}$ the empirical estimate of the probability mass function $f_{E|X}$.

Strategy 3: The last strategy is based on a third representation of the M-functional

$$\begin{aligned}
&\iint_{\mathcal{S}\times\mathcal{X}} \mathbb{E}\left(Y|E=1,M=m,X=x\right) dF_{M|E}\left(m|E=0,X=x\right) dF_X\left(x\right) \\
&= \sum_{e=0}^{1} \iiint_{\mathcal{Y}\times\mathcal{S}\times\mathcal{X}} y \frac{I(e=1)}{f_{E|X}(e|X=x)} \frac{f_{M|E,X}\left(M|E=0,X\right)}{f_{M|E,X}\left(M|E,X\right)} dF_{Y,M,E,X}\left(y,m,e,x\right) \\
&= \mathbb{E}\left\{ Y \frac{I(E=1)}{f_{E|X}(E|X)} \frac{f_{M|E,X}\left(M|E=0,X\right)}{f_{M|E,X}\left(M|E,X\right)} \right\}
\end{aligned}$$

Thus, our third estimator takes the form:

$$\widehat{\theta}_0^{em} = \mathbb{P}_n \left\{ Y \frac{I(E=1)}{\widehat{f}_{E|X}(E|X)} \frac{\widehat{f}_{M|E,X}\left(M|E=0,X\right)}{\widehat{f}_{M|E,X}\left(M|E,X\right)} \right\}$$

At first glance the three estimators $\widehat{\theta}_0^{em}, \widehat{\theta}_0^{ye}$ and $\widehat{\theta}_0^{ym}$ might appear to be distinct, however, we observe that provided the empirical distribution function $\widehat{F}_O = \widehat{F}_{Y|E,M,X} \times \widehat{F}_{M|E,X} \times \widehat{F}_{E|X} \times \widehat{F}_X$ satisfies the positivity assumption, and thus $\widehat{F}_O \in \mathcal{M}_{nonpar}$, then actually $\widehat{\theta}_0^{em} = \widehat{\theta}_0^{ye} = \widehat{\theta}_0^{ym} = \theta_0\left(\widehat{F}_O\right)$ since the three representations agree on the nonparametric model $\mathcal{M}_{nonpar}$. Therefore we may conclude that these three estimators are in fact asymptotically efficient in $\mathcal{M}_{nonpar}$ with common influence function $S_{\theta_0}^{eff,nonpar}\left(\theta_0\right)$. Furthermore, from this observation, one further concludes that (asymptotic) inferences obtained using one of the three representations are identical to inferences using either of the other two representations.

At this juncture, we note that the above equivalence no longer applies when as we have previously argued will likely occur in practice, (M, X) contains 3 or

more continuous variables and/or X is too high dimensional for models to be saturated or nonparametric, and thus parametric (or semiparametric) models are specified for dimension reduction. Specifically, for such settings, we observe that three distinct modeling strategies are available. Under the first strategy, the estimator $\widehat{\theta}_0^{ym,par}$ is obtained as $\widehat{\theta}_0^{ym,par}$ using parametric model estimates $\widehat{\mathbb{E}}^{par}(Y|E,M,X)$ and $\widehat{f}^{par}_{M|E,X}(m|E,X)$ instead of their nonparametric counterparts; similarly under the second strategy, the estimator $\widehat{\theta}_0^{ye,par}$ is obtained as $\widehat{\theta}_0^{ye}$ using estimates of parametric models $\widehat{\mathbb{E}}^{par}(Y|E=1,M=m,X)$ and $\widehat{f}^{par}_{E|X}(e|X)$ and finally, under the third strategy, $\widehat{\theta}_0^{em,par}$ is obtained as $\widehat{\theta}_0^{em}$ using $\widehat{f}^{par}_{E|X}(e|X)$ and $\widehat{f}^{par}_{M|E,X}(m|E,X)$. Then, it follows that $\widehat{\theta}_0^{ym,par}$ is CAN under the submodel $\mathcal{M}_a$, but is generally inconsistent if either $\widehat{\mathbb{E}}^{par}(Y|E,M,X)$ or $\widehat{f}^{par}_{M|E,X}(m|E,X)$ fails to be consistent. Similarly, $\widehat{\theta}_0^{ye,par}$ and $\widehat{\theta}_0^{em,par}$ are respectively CAN under the submodels $\mathcal{M}_b$ and $\mathcal{M}_c$, but each estimator generally fails to be consistent outside of the corresponding submodel. In the next section, we propose an approach that produces a triply robust estimator by combining the above three strategies so that only one of models $\mathcal{M}_a, \mathcal{M}_b$ and $\mathcal{M}_c$ needs to be valid for consistency of the estimator.

2.3 Triply robust estimation

The proposed triply robust estimator $\widehat{\theta}_0^{triply}$ solves

$$\mathbb{P}_n \widehat{S}^{eff,nonpar}_{\theta_0}\left(\widehat{\theta}^{triply}_0\right) = 0$$

where $\widehat{S}^{eff,nonpar}_{\theta_0}(\theta)$ is equal to $S^{eff,nonpar}_{\theta_0}(\theta)$ evaluated at $\{\widehat{\mathbb{E}}^{par}(Y|E,M,X), \widehat{f}^{par}_{M|E,X}(m|E,X), \widehat{f}^{par}_{E|X}(e|X)\}$ instead of $\{\mathbb{E}(Y|E,M,X), f_{M|E,X}(m|E,X), f_{E|X}(e|X)\}$; that is

$$\widehat{\theta}^{triply}_0 = \mathbb{P}_n \left[\begin{array}{c} \frac{I\{E=1\}\widehat{f}^{par}_{M|E,X}(M|E=0,X)}{\widehat{f}^{par}_{E|X}(1|X)\widehat{f}^{par}_{M|E,X}(M|E=1,X)} \left\{ Y - \widehat{\mathbb{E}}^{par}(Y|X,M,E=1)] \right\} \\ + \frac{I(E=0)}{\widehat{f}^{par}_{E|X}(0|X)} \left\{ \widehat{\mathbb{E}}^{par}(Y|X,M,E=1) - \widehat{\eta}^{par}(1,0,X) \right\} + \widehat{\eta}^{par}(1,0,X) \end{array} \right]$$

is CAN in model $\mathcal{M}_{\text{union}} = \mathcal{M}_a \cup \mathcal{M}_b \cup \mathcal{M}_c$, where $\widehat{\eta}^{par}(e,e^*,X) = \int_{\mathcal{S}} \widehat{\mathbb{E}}^{par}(Y|X,M=m,E=e)$ $\widehat{f}^{par}_{M|E,X}(m|E=e^*,X)\, d\mu(m)$. In the next theorem, the estimator in the above display is combined with a doubly robust estimator $\widehat{\delta}^{doubly}_e$ of δ_e (see van der Laan and Robins, 2003 or Tsiatis, 2006), to obtain multiply-robust estimators of natural direct and indirect effects, where

$$\widehat{\delta}^{doubly}_e = \mathbb{P}_n \left[\frac{I(E=e)}{\widehat{f}^{par}_{E|X}(e|X)} \{Y - \widehat{\eta}^{par}(e,e,X)\} + \widehat{\eta}^{par}(e,e,X) \right]$$

To state the result, we set $\widehat{\mathbb{E}}^{par}(Y|X,M,E) = \mathbb{E}^{par}\left(Y|X,M,E;\widehat{\beta}_y\right) = g^{-1}\left(\widehat{\beta}_y^T h(X,M,E)\right)$ where g is a known link function h is a user specified function of (X,M,E) so that $\mathbb{E}^{par}\left(Y|X,M,E;\beta_y\right) = g^{-1}\left(\beta_y^T h(X,M,E)\right)$ entails a working regression model for $\mathbb{E}(Y|X,M,E)$ and $\widehat{\beta}_y$ solves the estimating equation

$$0 = \mathbb{P}_n\left[S_y\left(\widehat{\beta}_y\right)\right] = \mathbb{P}_n\left[h(X,M,E)\left(Y - g^{-1}\left(\widehat{\beta}_y^T h(X,M,E)\right)\right)\right]$$

Similarly, we set $\widehat{f}^{par}_{M|E,X}(m|E,X) = f^{par}_{M|E,X}\left(m|E,X;\widehat{\beta}_m\right)$ for $f^{par}_{M|E,X}(m|E,X;\beta_m)$ a parametric model for the density of $[M|E,X]$ with $\widehat{\beta}_m$ solving

$$0 = \mathbb{P}_n\left[S_m\left(\widehat{\beta}_m\right)\right] = \mathbb{P}_n\left[\frac{\partial}{\partial\beta_m}\log f^{par}_{M|E,X}\left(M|E,X;\widehat{\beta}_m\right)\right]$$

and we set $\widehat{f}^{par}_{E|X}(e|X) = f^{par}_{E|X}(e|X;\widehat{\beta}_e)$ for $f^{par}_{E|X}(e|X;\beta_e)$ a parametric model for the density of $[E|X]$ with $\widehat{\beta}_e$ solving

$$0 = \mathbb{P}_n\left[S_e\left(\widehat{\beta}_e\right)\right] = \mathbb{P}_n\left[\frac{\partial}{\partial\beta_e}\log f^{par}_{E|X}\left(E|X;\widehat{\beta}_e\right)\right]$$

Theorem 2: Suppose that the assumptions of Theorem 1 hold, and that the regularity conditions stated in the appendix hold and that β_m, β_e and β_y *are variation independent.*

(i)Mediation functional: Then, $\sqrt{n}(\widehat{\theta}_0^{triply} - \theta_0)$ *is RAL under model* $\mathcal{M}_{union}$ *with influence function*

$$\begin{aligned} & S^{union}_{\theta_0}(\theta_0,\beta^*) \\ = \; & S^{eff,nonpar}_{\theta_0}(\theta_0,\beta^*) - \frac{\partial\mathbb{E}\left\{S^{eff,nonpar}_{\theta_0}(\theta_0,\beta)\right\}}{\partial\beta^T}|_{\beta^*}\mathbb{E}\left\{\frac{\partial S_\beta(\beta)}{\partial\beta^T}|_{\beta^*}\right\}^{-1} S_\beta(\beta^*) \end{aligned}$$

and thus converges in distribution to a $N(0,\Sigma_{\theta_0})$*, where*

$$\Sigma_{\theta_0}(\theta_0,\beta^*) = \mathbb{E}\left(S^{union}_{\theta_0}(\theta_0,\beta^*)^{\otimes 2}\right)$$

with $\beta^T = \left(\beta_m^T,\beta_e^T,\beta_y^T\right)$ *and* $S_\beta(\beta) = \left(S_m^T(\beta_m), S_e^T(\beta_e), S_y^T\left(\beta_y\right)\right)^T$*, and with* β^* *denoting the probability limit of the estimator* $\widehat{\beta} = \left(\widehat{\beta}_m^T,\widehat{\beta}_e^T,\widehat{\beta}_y^T\right)^T$

(ii)Natural direct effect:Then, $\sqrt{n}(\widehat{\theta}_0^{triply} - \widehat{\delta}_0^{doubly} - \theta_0 + \delta_0)$ *is RAL under model* $\mathcal{M}_{union}$ *with influence function*

$$\begin{aligned} & S_{NDE}^{union}(\theta_0, \delta_0, \beta^*) \\ = \; & S_{NDE}^{eff,nonpar}(\theta_0, \delta_0, \beta^*) - \frac{\partial \mathbb{E}\left\{ S_{NDE}^{eff,nonpar}(\theta_0, \delta_0, \beta) \right\}}{\partial \beta^T}|_{\beta^*} \mathbb{E}\left\{ \frac{\partial S_\beta(\beta)}{\partial \beta^T}|_{\beta^*} \right\}^{-1} S_\beta(\beta^*) \end{aligned}$$

and thus converges in distribution to a $N(0, \Sigma_{\theta_0 - \delta_0})$, where

$$\Sigma_{\theta_0 - \delta_0}(\delta_1, \theta_0, \beta^*) = \mathbb{E}\left(S_{NDE}^{union}(\theta_0, \delta_0, \beta^*)^{\otimes 2} \right)$$

(iii)Natural indirect effect:Then, $\sqrt{n}(\widehat{\delta}_1^{doubly} - \widehat{\theta}_0^{triply} - (\delta_1 - \theta_0))$ *is RAL under model* $\mathcal{M}_{union}$ *with influence function*

$$\begin{aligned} & S_{NIE}^{union}(\delta_1, \theta_0, \beta^*) \\ = \; & S_{NIE}^{eff,nonpar}(\delta_1, \theta_0, \beta^*) - \frac{\partial \mathbb{E}\left\{ S_{NIE}^{eff,nonpar}(\delta_1, \theta_0, \beta) \right\}}{\partial \beta^T}|_{\beta^*} \mathbb{E}\left\{ \frac{\partial S_\beta(\beta)}{\partial \beta^T}|_{\beta^*} \right\}^{-1} S_\beta(\beta^*) \end{aligned}$$

and thus converges in distribution to a $N(0, \Sigma_{\delta_1 - \theta_0})$, *where*

$$\Sigma_{\delta_1 - \theta_0}(\delta_1, \theta_0, \beta^*) = \mathbb{E}\left(S_{NIE}^{union}(\delta_1, \theta_0, \beta^*)^{\otimes 2} \right)$$

iv)$\widehat{\theta}_0^{triply}$, $\widehat{\theta}_0^{triply} - \widehat{\delta}_0^{doubly}$ *and* $\widehat{\delta}_1^{doubly} - \widehat{\theta}_0^{triply}$ *are semiparametric locally efficient in the sense that they are RAL under model* $\mathcal{M}_{union}$ *and respectively achieve the semiparametric efficiency bound for* $\theta_0, \theta_0 - \delta_0$, and $\delta_1 - \theta_0$ *under model* $\mathcal{M}_{union}$ *at the*

intersection submodel $\mathcal{M}_a \cap \mathcal{M}_b \cap \mathcal{M}_c$*, with respective efficient influence functions:* $S_{\theta_0}^{eff,nonpar}(\theta_0, \beta^*), S_{NDE}^{eff,nonpar}(\theta_0, \delta_0, \beta^*)$ *, and* $S_{NIE}^{eff,nonpar}(\delta_1, \theta_0, \beta^*)$.

Empirical versions of $\Sigma_{\theta_0 - \delta_0}(\delta_1, \theta_0, \beta^*)$ and $\Sigma_{\delta_1 - \theta_0}(\delta_1, \theta_0, \beta^*)$ are easily obtained, and the corresponding Wald type confidence intervals can be used to make formal inferences about natural direct and indirect effects. It is also straightforward to extend the approach to the risk ratio and odds ratio scales for binary Y. By a theorem due to Robins and Rotnitzky (2001), part iv) of the theorem implies that when all models are correct. $\widehat{\theta}_0^{triply}$, $\widehat{\theta}_0^{triply} - \widehat{\delta}_0^{doubly}$ and $\widehat{\delta}_1^{doubly} - \widehat{\theta}_0^{triply}$ are semiparametric efficient in model $\mathcal{M}_{\text{nonpar}}$ at the intersection submodel $\mathcal{M}_a \cap \mathcal{M}_b \cap \mathcal{M}_c$.

2.4 A comparison to some existing estimators

In this section, we briefly compare the proposed approach to some existing estimators in the literature. Perhaps the most common approach for estimating direct and indirect effects when Y is continuous uses a system of linear structural equations; whereby, a linear structural equation for the outcome given the exposure, the mediator and the confounders is combined with a linear structural equation for the mediator given the exposure and confounders to produce an estimator of natural direct and indirect effects. The classical approach of Baron and Kenny (1986) is a particular instance of this approach. In recent work mainly motivated by Pearl's mediation functional, several authors (Imai et al, 2010, Pearl, 2010, VanderWeele, 2009, VanderWeele and

Vansteedlandt, 2010) have demonstrated how the simple linear structural equation approach generalizes to accommodate both, the presence of an interaction between exposure and mediator variables, and a nonlinear link function either in the regression model for the outcome or in the regression model for the mediator, or both. In fact, when the effect of confounders is also modeled in such structural equations, inferences based on the latter can be viewed as special instances of inferences obtained under a particular specification of model $\mathcal{M}_a$ for the outcome and the mediator densities. And thus, an estimator obtained under a system of structural equations will generally fail to produce a consistent estimator of natural direct and indirect effects when model $\mathcal{M}_a$ is incorrect whereas, by using the proposed multiply robust estimator valid inferences can be recovered under the union model $\mathcal{M}_b \cup \mathcal{M}_c$, even if $\mathcal{M}_a$ fails.

A notable improvement on the system of structural equations approach is the double robust estimator of a natural direct effect due to van der Laan and Petersen (2005). They show their estimator remains CAN in the larger submodel $\mathcal{M}_a \cup \mathcal{M}_c$ and therefore, they can recover valid inferences even when the outcome model is incorrect, provided both the exposure and mediator models are correct. Unfortunately, the van der Laan estimator is still not entirely satisfactory because unlike the proposed multiply robust estimator, it requires that the model for the mediator density is correct. Nonetheless, if the mediator model is correct, the authors establish that their estimator achieves the efficiency bound for model $\mathcal{M}_a \cup \mathcal{M}_c$ at the intersection

submodel $\mathcal{M}_a \cap \mathcal{M}_c$ where all models are correct; and thus it is locally semiparametric efficient in $\mathcal{M}_a \cup \mathcal{M}_c$. Interestingly, as we report below, the semiparametric efficiency bounds for models $\mathcal{M}_a \cup \mathcal{M}_c$ and $\mathcal{M}_a \cup \mathcal{M}_b \cup \mathcal{M}_c$ are distinct, because the density of the mediator variable is not ancillary for inferences about the M-functional. Thus, any restriction placed on the mediator's conditional density can, when correct, produce improvements in efficiency. This is in stark contrast with the role played by the density of the exposure variable, which as in the estimation of the marginal causal effect, remains ancillary for inferences about the M-functional and thus the efficiency bound for the latter is unaltered by any additional information on the former (Robins et al 1994). In the next section, we provide a general functional map that relates the efficient influence function for the larger model $\mathcal{M}_a \cup \mathcal{M}_b \cup \mathcal{M}_c$ to the efficient influence for the smaller model $\mathcal{M}_a \cup \mathcal{M}_c$ where the model for the mediator may be parametric or semiparametric. Our map is instructive because it makes explicit using simple geometric arguments, the information that is gained from increasing restrictions on the law of the mediator. We illustrate the map by recovering the efficient influence function of van der Laan in the case of singleton model (i.e. a known conditional density) for the mediator and in the case of a parametric model for the mediator.

3 Estimation with a known model for the mediator density

Suppose we know a correct (possibly semiparametric) model for the law of the mediator variable given the exposure and confounding variables. Suppose Λ_M^{model} denotes the tangent space for this model, and let $\Lambda_M^{\text{nonpar}}$ denote the tangent space for the nonparametric model of the mediator law. Recall that the tangent space of a parametric, semiparametric or nonparametric model is defined as the $L_2(F_O)$ closure of the scores of regular parametric submodels in the model (Bickel et al 2003), and $L_2(F_O)$ is the Hilbert space of all functions of O with finite variance under F_O. In addition, given a Hilbert subspace H, we let $\Pi(\cdot|H)$ denote the L_2 projection into H.

Theorem 3: Under the consistency, sequential ignorability and positivity assumptions, the efficient influence function of the mediation functional, the natural direct and the natural indirect effects in model $\mathcal{M}_a \cup \mathcal{M}_c$ are respectively:

$$
\begin{aligned}
& S_{\theta_0}^{eff,\mathcal{M}_a \cup \mathcal{M}_c}(\theta_0) \\
= \; & S_{\theta_0}^{eff,\text{nonpar}}(\theta_0) - \Pi\left(S_{\theta_0}^{eff,\text{nonpar}}(\theta_0) \,|\Lambda_M^{\text{model},\perp} \cap \Lambda_M^{\text{nonpar}}\right)
\end{aligned}
$$

$$
\begin{aligned}
& S_{NDE}^{eff,\mathcal{M}_a \cup \mathcal{M}_c}(\theta_0, \delta_0) \\
= \; & S_{NDE}^{eff,\text{nonpar}}(\theta_0, \delta_0) - \Pi\left(S_{NDE}^{eff,\text{nonpar}}(\theta_0, \delta_0) \,|\Lambda_M^{\text{model},\perp} \cap \Lambda_M^{\text{nonpar}}\right)
\end{aligned}
$$

$$
\begin{aligned}
& S_{NIE}^{eff,\mathcal{M}_a \cup \mathcal{M}_c}(\delta_1, \theta_0) \\
= \; & S_{NIE}^{eff,\text{nonpar}}(\delta_1, \theta_0) - \Pi\left(S_{NIE}^{eff,\text{nonpar}}(\delta_1, \theta_0) \,|\Lambda_M^{\text{model},\perp} \cap \Lambda_M^{\text{nonpar}}\right)
\end{aligned}
$$

where $\Lambda_M^{\text{model},\perp}$ *is the orthogonal complement of* Λ_M^{model}*. Thus the semiparametric efficiency bound for estimating the natural direct and the natural indirect effects in* $\mathcal{M}_a \cup \mathcal{M}_c$ *are respectively given by*

$$
\begin{aligned}
& \mathbb{E}\left\{S_{NDE}^{eff,\mathcal{M}_a \cup \mathcal{M}_c}(\theta_0, \delta_0)^2\right\} \\
\leq \; & \mathbb{E}\left\{S_{NDE}^{eff,\text{nonpar}}(\theta_0, \delta_0)^2\right\}
\end{aligned}
$$

and

$$
\begin{aligned}
& \mathbb{E}\left\{S_{NIE}^{eff,\mathcal{M}_a \cup \mathcal{M}_c}(\delta_1, \theta_0)^2\right\} \\
\leq \; & \mathbb{E}\left\{S_{NIE}^{eff,\text{nonpar}}(\delta_1, \theta_0)^2\right\}
\end{aligned}
$$

The above theorem makes explicit the information that is gained by restricting the law of the mediator. The theorem shows that for any two models $\mathcal{M}_1$ and $\mathcal{M}_2$ that only differ in an assumption made about the law of the mediator variable, and that are otherwise nonparametric, with $\mathcal{M}_1 \subset \mathcal{M}_2$: $\mathbb{E}\left\{S_{NDE}^{eff,\mathcal{M}_1}(\theta_0, \delta_0)^2\right\} \leq \mathbb{E}\left\{S_{NDE}^{eff,\mathcal{M}_2}(\theta_0, \delta_0)^2\right\}$. This is because $\Lambda_M^{\mathcal{M}_1,\perp} \cap \Lambda_M^{\text{nonpar}} \supseteq \Lambda_M^{\mathcal{M}_2,\perp} \cap \Lambda_M^{\text{nonpar}}$ and thus $S_{NDE}^{eff,\mathcal{M}_1}(\theta_0, \delta_0)$ is the residual of the orthogonal projection of $S_{\theta_0}^{eff,\text{nonpar}}(\theta_0)$ onto a larger Hilbert subspace, resulting in a smaller $L_2(F_O)-$norm of the latter.

According to the theorem, evaluating the efficient influence function of a natural direct or indirect effect under the union model $\mathcal{M}_a \cup \mathcal{M}_c$ requires the evaluation of the Hilbert space projection $\Pi\left(U|\Lambda_M^{\text{model},\perp} \cap \Lambda_M^{\text{nonpar}}\right)$ for U in $\left\{S_{NDE}^{eff,\text{nonpar}}\left(\theta_0, \delta_0\right), S_{NIE}^{eff,\text{nonpar}}\left(\delta_1, \theta_0\right)\right\}$. Obviously, this projection will depend on the specific form of the model for the mediator, thus here we consider two instructive settings; in the first setting the model for the mediator is a singleton (and thus the law is known) whereas in the second case, the law $f_{M|E,X}(\cdot|E,X) = f_{M|E,X}^{par}(\cdot|E,X;\beta_m)$ is known up to a finite dimensional parameter, with score function S_{β_E}. We observe that in general, $\Pi\left(U|\Lambda_M^{\text{model},\perp} \cap \Lambda_M^{\text{nonpar}}\right) = \Pi\left(U|\Lambda_M^{\text{nonpar}}\right) - \Pi\left(U|\Lambda_M^{\text{model}}\right)$, with $\Lambda_E^{\text{nonpar}} = \{a(E,X,M) : \mathbb{E}\left[a(E,X,M)|E,X\right] = 0\} \cap L_2$. Thus in the first setting $\Lambda_M^{\text{model}} = \varnothing$ so that $\Pi\left(U|\Lambda_M^{\text{model},\perp} \cap \Lambda_M^{\text{nonpar}}\right) = \mathbb{E}(U|E,X,M) - \mathbb{E}(U|E,X)$. In contrast, in the second setting, $\Lambda_M^{\text{model}} = \{c^T S_{\beta_m} : c \in \mathcal{R}^{\dim(\beta_M)}\}$ so that $\Pi\left(U|\Lambda_M^{\text{nonpar}}\right) - \Pi\left(U|\Lambda_M^{\text{model}}\right) = \mathbb{E}(U|E,X,M) - \mathbb{E}(U|E,X) - \mathbb{E}\left(US_{\beta_m}\right)\mathbb{E}\left(S_{\beta_m}S_{\beta_m}^T\right)^- S_{\beta m}$.

Applying these results yields the following formulae for the efficient influence functions of the natural direct effect. In the case where the model for the mediator is a singleton (and thus the conditional density for the mediator is assumed known), one

obtains the following influence function

$$S^{eff,\mathcal{M}_a\cup\mathcal{M}_c}_{NDE,\text{singleton}}(\theta_0,\delta_0) = \frac{I\{E=1\}f_{M|E,X}(M|E=0,X)}{f_{E|X}(1|X)f_{M|E,X}(M|E=1,X)}\{Y-\mathbb{E}(Y|X,M,E=1)]\}$$
$$-\frac{I(E=0)}{f_{E|X}(0|X)}\{Y-\mathbb{E}(Y|X,M,E=0)\}$$
$$+\eta(1,0,X)-\eta(0,0,X)-\theta_0+\delta_0$$

which matches the influence function obtained by van der Laan and Petersen (2005) for this setting. In the second case in which the mediator is modeled parametrically, one obtains the following influence function

$$S^{eff,\mathcal{M}_a\cup\mathcal{M}_c}_{NDE,\text{param}}(\theta_0,\delta_0) = \frac{I\{E=1\}f_{M|E,X}(M|E=0,X)}{f_{E|X}(1|X)f_{M|E,X}(M|E=1,X)}\{Y-\mathbb{E}(Y|X,M,E=1)]\}$$
$$-\frac{I(E=0)}{f_{E|X}(0|X)}\{Y-\mathbb{E}(Y|X,M,E=0)\}$$
$$+\eta(1,0,X)-\eta(0,0,X)-\theta_0+\delta_0$$
$$+\mathbb{E}\left[S_{\beta_e}\frac{I(E=0)}{f_{E|X}(0|X)}\left\{\begin{array}{c}\mathbb{E}(Y|X,M,E=1)-\mathbb{E}(Y|X,M,E=0)\\ -\eta(1,0,X)+\eta(0,0,X)\end{array}\right\}\right]$$
$$\times\mathbb{E}\left(S_{\beta_e}S^T_{\beta_e}\right)^{-}S_{\beta_e}$$

For comparison, consider the influence function corresponding to the van der Laan estimator at the intersection submodel $\mathcal{M}_a\cap\mathcal{M}_c$ and assuming the unknown parameter β_m is estimated by maximum likelihood:

$$S^{eff,\mathcal{M}_a\cup\mathcal{M}_c}_{NDE}(\theta_0,\delta_0,\beta_m)+\mathbb{E}\left[\frac{\partial S^{eff,\mathcal{M}_a\cup\mathcal{M}_c}_{NDE}(\theta_0,\delta_0,\beta^*_m)}{\partial\beta^{*T}_m}|_{\beta_m}\right]\mathbb{E}\left(S_{\beta_m}S^T_{\beta_m}\right)^{-}S_{\beta_m}$$

where $S_{NDE}^{eff,\mathcal{M}_a \cup \mathcal{M}_c}(\theta_0, \delta_0, \beta_m)$ is equal to $S_{NDE,\text{singleton}}^{eff,\mathcal{M}_a \cup \mathcal{M}_c}(\theta_0, \delta_0)$ evaluated at the law $f_{M|E,X}(\cdot|X) = f_{M|E,X}^{par}(\cdot|E, X; \beta_m)$. Thus, we may conclude that the van der Laan estimator achieves the efficiency bound of $\mathcal{M}_a \cup \mathcal{M}_c$ at the intersection submodel only if

$$\begin{aligned} & \mathbb{E}\left[\frac{\partial S_{NDE}^{eff,\mathcal{M}_a \cup \mathcal{M}_c}(\theta_0, \delta_0, \beta_m)}{\partial \beta_m^T}\right] \\ = \quad & \mathbb{E}\left[S_{\beta_m} \frac{I(E=0)}{f_{E|X}(0|X)} \left\{ \begin{array}{c} \mathbb{E}(Y|X, M, E=1) - \mathbb{E}(Y|X, M, E=0) \\ -\eta(1, 0, X) + \eta(0, 0, X) \end{array} \right\}\right] \end{aligned}$$

We show that the equality in the above display holds by making the following observation:

$$\mathbb{E}_{\beta_m^*}\left\{S_{NDE}^{eff,\mathcal{M}_a \cup \mathcal{M}_c}(\theta_0(\beta_m^*), \delta_0, \beta_m^*)\right\} = 0 \text{ for all } \beta_m^*$$

where $\mathbb{E}_{\beta_m^*}(\cdot)$ is the expectation and $\theta_0(\beta_m^*)$ is the M-functional both evaluated at the mediator density $f_{M|E,X}^{par}(\cdot|E, X; \beta_m^*)$, with $\mathbb{E}_{\beta_m}(\cdot) = \mathbb{E}(\cdot)$ and $\theta_0(\beta_m) = \theta_0$. This in turn implies that

$$\frac{\partial \mathbb{E}_{\beta_m^*}\left\{S_{NDE}^{eff,\mathcal{M}_a \cup \mathcal{M}_c}(\theta_0(\beta_m^*), \delta_0, \beta_m^*)\right\}}{\partial \beta_m^*}|_{\beta_m} = 0$$

which implies

$$\begin{aligned} & -\mathbb{E}\left[\frac{\partial S_{NDE}^{eff,\mathcal{M}_a \cup \mathcal{M}_c}(\theta_0, \delta_0, \beta_m^*)}{\partial \beta_m^{*T}}|_{\beta_m}\right] \\ = \quad & \mathbb{E}\left[\left\{S_{NDE}^{eff,\mathcal{M}_a \cup \mathcal{M}_c}(\theta_0, \delta_0, \beta_m) + S_{NDE,\text{param}}^{eff,\mathcal{M}_a \cup \mathcal{M}_c}(\theta_0, \delta_0)\right\} S_{\beta_m}\right] \\ = \quad & \mathbb{E}\left[S_{\beta_m} \frac{I(E=0)}{f_{E|X}(0|X)} \{\mathbb{E}(Y|X, M, E=1) - \mathbb{E}(Y|X, M, E=0) - \eta(1, 0, X) + \eta(0, 0, X)\}\right] \end{aligned}$$

where the last equality is obtained upon noting that $\mathbb{E}\left[S_{NDE}^{eff,\mathcal{M}_a\cup\mathcal{M}_c}\left(\theta_0,\delta_0,\beta_m^*\right)S_{\beta_m}\right]=0$.

4 A semiparametric sensitivity analysis

We describe a semiparametric sensitivity analysis framework to assess the extent to which a violation of the ignorability assumption for the mediator might alter inferences about a natural direct effect. Although only results for the natural direct effect are given here, the extension for the indirect effect is easily deduced from the presentation. Let

$$t\left(e,m,x\right)=\mathbb{E}\left[Y_{1,m}|E=e,M=m,X=x\right]-\mathbb{E}\left[Y_{1,m}|E=e,M\neq m,X=x\right]$$

then

$$Y_{e',m}\not\perp\!\!\!\perp M|E=e,X$$

i.e. a violation of the ignorability assumption for the mediator variable, generally implies that

$$t\left(e,m,x\right)\neq 0 \text{ for some } \left(e,m,x\right)$$

Thus, we proceed as in Robins, Rotnitzky and Scharfstein (1999), and propose to recover inferences by assuming the selection bias function $t\left(e,m,x\right)$ is known, which encodes the magnitude and direction of the unmeasured confounding for the mediator. In the following, $\mathcal{S}$ is assumed to be finite. To motivate the proposed approach,

suppose for the moment that $f_{M|E,X}(M|E,X)$ is known, then under the assumption that the exposure is ignorable given X, we show in the appendix that:

$$\begin{aligned}
& \mathbb{E}[Y_{1,m}|M_0 = m, X = x] \\
= \; & \mathbb{E}[Y_{1,m}|E = 0, M = m, X = x] \\
= \; & \mathbb{E}[Y|E = 1, M = m, X = x] - t(1,m,x)\left(1 - f_{M|E,X}(m|E = 1, X = x)\right) \\
& + t(0,m,x)\left(1 - f_{M|E,X}(m|E = 0, X = x)\right)
\end{aligned}$$

and therefore the M-functional is identified by:

$$\mathbb{E}\left[\sum_{m\in\mathcal{S}}\left\{\begin{array}{c}\mathbb{E}[Y|E=1,M=m,X] \\ -\,t(1,m,X)\left(1 - f_{M|E,X}(m|E=1,X)\right) \\ +t(0,m,X)\left(1 - f_{M|E,X}(m|E=0,X)\right)\end{array}\right\} f_{M|E,X}(m|E=0,X)\right] \quad (3)$$

which is equivalently represented as:

$$\mathbb{E}\left[\frac{I\{E=1\} f_{M|E,X}(M|E=0,X)}{f_{E|X}(1|X) f_{M|E,X}(M|E=1,X)}\left\{\begin{array}{c} Y - t(1,M,X)\left(1 - f_{M|E,X}(m|E=1,X)\right) \\ +t(0,M,X)\left(1 - f_{M|E,X}(M|E=0,X\ )\right)\end{array}\right\}\right] \quad (4)$$

Below, these two equivalent representations (3) and (4) are carefully combined to obtain a double robust estimator of the M-functional assuming $t(\cdot,\cdot,\cdot)$ is known. A sensitivity analysis is then obtained by repeating this process and reporting inferences for each choice of $t(\cdot,\cdot,\cdot)$ in a finite set of user–specified functions $\mathcal{T} = \{\, t_\lambda(\cdot,\cdot,\cdot) : \lambda\}$ indexed by a finite dimensional parameter λ with $t_0(\cdot,\cdot,\cdot) \in \mathcal{T}$ corresponding to the

no unmeasured confounding assumption,i.e. $t_0(\cdot,\cdot,\cdot) \equiv 0$. Throughout, the model $f^{par}_{M|E,X}(\cdot|E,X;\beta_m)$ for the probability mass function of M is assumed to be correct. Thus, to implement the sensitivity analysis, we develop a semiparametric estimator of the natural direct effect in the union model $\mathcal{M}_a \cup \mathcal{M}_c$, assuming $t(\cdot,\cdot,\cdot) = t_{\lambda^*}(\cdot,\cdot,\cdot)$ for a fixed λ^*. The proposed doubly robust estimator of the natural direct effect is then given by $\widehat{\theta}_0^{doubly}(\lambda^*) - \widehat{\delta}_0^{doubly}$ where $\widehat{\delta}_0^{doubly}$ is as previously described, and

$$\widehat{\theta}_0^{doubly}(\lambda^*) = \mathbb{P}_n \left[\begin{array}{c} \frac{I\{E=1\}\widehat{f}^{par}_{M|E,X}(M|E=0,X)}{\widehat{f}^{par}_{E|X}(1|X)\widehat{f}^{par}_{M|E,X}(M|E=1,X)} \left\{ Y - \widehat{\mathbb{E}}^{par}(Y|X,M,E=1) \right\} \\ + \widetilde{\eta}^{par}(1,0,X;\lambda^*) \end{array} \right]$$

with

$$\begin{aligned} & \widetilde{\eta}^{par}(1,0,X;\lambda^*) \\ = & \sum_{m \in \mathcal{S}} \left\{ \begin{array}{c} \widehat{\mathbb{E}}^{par}(Y|X,M=m,E=1) \\ + t_{\lambda^*}(0,m,X)\left(1 - \widehat{f}^{par}_{M|E,X}(m|E=0,X)\right) \\ - t_{\lambda^*}(1,m,X)\left(1 - \widehat{f}^{par}_{M|E,X}(m|E=1,X)\right) \end{array} \right\} \widehat{f}^{par}_{M|E,X}(m|E=0,X) \end{aligned}$$

Our sensitivity analysis then entails reporting the set $\left\{ \widehat{\theta}_0^{doubly}(\lambda) - \widehat{\delta}_0^{doubly} : \lambda \right\}$ (and the associated confidence intervals) which summarizes how sensitive inferences are to a deviation from the ignorability assumption $\lambda = 0$. A theoretical justification for the approach is given by the following formal result which is proved in the appendix

Theorem 4:Suppose $t(\cdot,\cdot,\cdot) = t_{\lambda^*}(\cdot,\cdot,\cdot)$, *then under the consistency, positivity assumptions, and the ignorability assumption for the exposure,* $\widehat{\theta}_0^{doubly}(\lambda^*) - \widehat{\delta}_0^{doubly}$ *is a CAN estimator of the natural direct effect in* $\mathcal{M}_a \cup \mathcal{M}_c$.

The influence function of $\widehat{\theta}_0^{doubly}(\lambda^*)$ is provided in the appendix, and can be used to construct a corresponding confidence interval.

It is important to note that the sensitivity analysis technique presented here differs in crucial ways from previous techniques developed by Hafeman (2008), VanderWeele (2010) and Imai et al (2010a). First, the methodology of Vanderweele (2010) postulates the existence of an unmeasured confounder U (possibly vector valued) which when included in X recovers the sequential ignorability assumption. The sensitivity analysis then requires specification of a sensitivity parameter encoding the effect of the unmeasured confounder on the outcome within levels of (E, X, M), and another parameter for the effect of the exposure on the density of the unmeasured confounder given (X, M). This is a daunting task which renders the approach generally impractical, except perhaps in the simple setting where it is reasonable to postulate a single binary counfounder is unobserved, and one is willing to make further simplifying assumptions about the required sensitivity parameters (VanderWeele, 2010). In comparison, the proposed approach circumvents this difficulty by concisely encoding a violation of the ignorability assumption for the mediator through the selection bias function $t_\lambda(e, m, x)$. Thus the approach makes no reference and thus is agnostic about the existence, dimension, and nature of unmeasured confounders U. Furthermore, in our proposal, the ignorability violation can arise due to an unmeasured confounder of the mediator-outcome relationship that is also an effect of the exposure variable,

a setting not handled by the technique of VanderWeele (2010). The method of Hafeman(2008) which is restricted to binary data, shares some of the limitations given above. Finally, in contrast with our proposed double robust approach, a coherent implementation of the sensitivity analysis techniques of Imai et al (2010a, 2010b) and VanderWeele (2010) both rely on correct specification of all posited models. We refer the reader to VanderWeele (2010) for further discussion of Hafeman (2008) and Imai et al (2010a).

5 Discussion

The main contribution of the current paper is a theoretically rigorous yet practically relevant semiparametric framework for making inferences about natural direct and indirect causal effects in the presence of a large number of confounding factors. Semiparametric efficiency bounds are given for the nonparametric model, and multiply robust locally efficient estimators are developed that can be used when nonparametric estimation is not possible. For good finite sample performance, the proposed estimators which involve inverse probability weights for the exposure and mediator variables, appear to depend heavily on the positivity assumption. In fact, it was recently shown by Kang and Shafer (2007) that a practical violation of this assumption in data analysis can severely compromise inferences based on such methodology; although their analysis only considered the functional δ_0 and not the M-functional θ_0. In

future work, it will be crucial to critically examine by simulation the extent to which our proposed estimators are susceptible to a practical violation of the assumption , and we also plan to develop modifications of the methods along the lines of Robins et al (2007), Cao et al (2009) and Tan (2010), to improve their performance under such stressful conditions. In the meantime, the estimator presented herein could immediately be made more stable by substituting one of these improved estimators for $\widehat{\delta}_j^{doubly}$ while similar estimators of the M-functional are being developed.

Although the paper focuses on a binary exposure, we note that the extension to a polytomous exposure is trivial. In future work, we shall also consider other generalizations of the results given in the current paper. For instance, it is of interest to develop similar semiparametric methods for estimating models for conditional natural direct and indirect effects given a subset of pre-exposure variables. These models are particular important in making inferences about so-called moderated mediation effects, a topic of growing interest particularly in the field of psychology(Preacher, Rucker and Hayes, 2007).

Acknowledgement 1 *The authors would like to acknowledge Andrea Rotnitzky who provided invaluable comments that improved the presentation of the results given in Section 3 of the manuscript. The authors also thank James Robins and Tyler VanderWeele for useful comments that significantly improved the presentation of this article.*

References

Avin, C., I. Shpitser, and J. Pearl (2005). Identifiability of path-specific effects. In IJCAI-05, Proceedings of the Nineteenth International Joint Conference on Artificial Intelligence, Edinburgh, Scotland, UK, July 30-August 5, 2005, pp. 357–363.

Bang H, Robins J. (2005). Doubly robust estimation in Missing data and causal inference models. Biometrics, 61:692-972.

Baron, R. M., & 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.

Bickel, P., Klassen, C., Ritov, Y. and Wellner, J. (1993). Efficient and Adaptive Estimation for Semi-parametric Models. Springer, New York.

Cao, W., Tsiatis, A.A., and Davidian, M. (2009) Improving efficiency and robustness of the doubly robust estimator for a population mean with incomplete data. Biometrika 96, 732-734

Goetgeluk, S., Vansteelandt, S. and Goetghebeur, E. (2008). Estimation of controlled direct effects. Journal of the Royal Statistical Society – Series B, 70, 1049-1066.

Hafeman, D. Opening the Black Box: A Reassessment of Mediation from a Counterfactual Perspective[dissertation[. New York. Columbeia University, 2008.

Hafeman, D. and T. VanderWeele (2009). Alternative assumptions for the identification of direct and indirect effects. Epidemiology. In press.

Hahn, J. (1998): "On the Role of the Propensity Score in Efficient Semiparametric Estimation of Average Treatment Effects, "Econometrica, 66,pp. 315–331

van der Laan MJ, Robins JM. (2003). Unified Methods for Censored Longitudinal Data and Causality. Springer Verlag: New York.

van der Laan, M, Petersen,M.(2005) Direct Effect Models. U.C. Berkeley Division of Biostatistics Working Paper Series. Working Paper 187. http://www.bepress.com/ucbbiostat/paper187

Imai, K., Keele, L., and Yamamoto, T. (2010a). Identification, inference and sensitivity analysis for causal mediation effects. Statistical Science 25, 51–71.

Imai, K, Keele L and Tingley D. (2010b). "A General Approach to Causal Mediation Analysis." Psychological Methods, Vol. 15, No. 4 (December), pp. 309-334. (lead article)

Kang, J. D. Y. and Schafer, J. L. (2007). Demystifying double robustness: a

comparison of alternative strategies for estimating a population mean from incomplete data (with discussion). Statist. Sci. 22, 523–39.

Pearl, J. (2001). Direct and indirect effects. In Proceedings of the 17th Annual Conference on Uncertainty in Artificial Intelligence (UAI-01), San Francisco, CA, pp. 411–42. Morgan Kaufmann.

Pearl J (2011) The Mediation Formula: A guide to the assessment of causal pathways in nonlinear models.Technical report <http://ftp.cs.ucla.edu/pub/stat_ser/r379.pdf>

Preacher, K. J., Rucker, D. D. and Hayes, A. F. (2007). Assessing moderated mediation hypotheses: Strategies, methods, and prescriptions. Multivariate Behavioral Research, 42, 185–227.

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

Robins, J. M., Mark, S. D. and Newey, W. K. (1992), "Estimating exposure effects by modeling the expectation of exposure conditional on confounders," Biometrics, 48, 479-495.

Robins JM, Rotnitzky A, Scharfstein D. (1999). Sensitivity Analysis for Selection Bias and Unmeasured Confounding in Missing Data and Causal Inference Models. In: Statistical Models in Epidemiology: The Environment and Clinical

Trials. Halloran, M.E. and Berry, D., eds. IMA Volume 116, NY: Springer-Verlag, pp. 1-92.

Robins, J. (2003). Semantics of causal DAG models and the identification of direct and indirect effects. In P. Green, N. Hjort, and S. Richardson (Eds.), Highly Structured Stochastic Systems, pp. 70–81. Oxford, UK: Oxford University Press.

Robins JM, Rotnitzky A. (2001). Comment on the Bickel and Kwon article, "Inference for semiparametric models: Some questions and an answer" Statistica Sinica, 11(4):920-936.

Robins JM. (2000). Robust estimation in sequentially ignorable missing data and causal inference models. Proceedings of the American Statistical Association Section on Bayesian Statistical Science 1999, pp. 6-10.

Robins JM, Rotnitzky A, Zhao LP. (1994). Estimation of regression coefficients when some regressors are not always observed. Journal of the American Statistical Association, 89:846-866.

Robins JM, Sued M, Lei-Gomez Q, Rotnitsky A. (2007). Comment: Performance of double-robust estimators when "Inverse Probability" weights are highly variable. Statistical Science 22(4):544-559.

Robins JM, Richardson, TS. (2010). Alternative graphical causal models and the identification of direct effects. To appear in Causality and Psychopathology: Finding the Determinants of Disorders and Their Cures. P. Shrout, Editor. Oxford University Press

Scharfstein DO, Rotnitzky A, Robins JM. (1999). Rejoinder to comments on "Adjusting for non-ignorable drop-out using semiparametric non-response models". Journal of the American Statistical Association, 94:1096-1120. Journal of the American Statistical Association, 94:1121-1146.

Tsiatis AA (2006) Semiparametric Theory and Missing Data. Springer.Verlag: New York.

VanderWeele, T.J. (2009). Marginal structural models for the estimation of direct and indirect effects. Epidemiology, 20:18-26.

VanderWeele, T.J. and Vansteelandt, S. (2010). Odds ratios for mediation analysis for a dichotomous outcome - with discussion. American Journal of Epidemiology, 172, 1339-1348.

VanderWeele TJ. Bias formulas for sensitivity analysis for direct and indirect effects. Epidemiology. 2010;21:540–551.

APPENDIX

PROOF OF THEOREM 1:

Let $F_{O;t} = F_{Y|M,X,E;t} F_{M|E,X;t} F_{E|X;t} F_{X;t}$ denote a one dimensional regular parametric submodel of $\mathcal{M}_{\text{nonpar}}$, with $F_{O,0} = F_O$, and let

$$\theta_t = \theta_0 \left(F_{O;t}\right) = \iint_{\mathcal{S}\times\mathcal{X}} \mathbb{E}_t \left(Y|E=1, M=m, X=x\right) f_{M|E,X;t} \left(m|E=0, X=x\right) f_{X;t}(x) d\mu(m,x)$$

The efficient influence function $S_{\theta_0}^{eff,nonpar} \left(\theta_0\right)$ is the unique random variable to satisfy the following equation

$$\nabla_{t=0} \theta_t = \mathbb{E}\left\{ S_{\theta_0}^{eff,nonpar} \left(\theta_0\right) U \right\}$$

for U the score of $F_{O;t}$ at $t=0$, and $\nabla_{t=0}$ denoting differentiation wrt t at $t=0$. We observe that

$$\begin{aligned}
\frac{\partial \theta_t}{\partial t}|_{t=0} &= \iint_{\mathcal{S}\times\mathcal{X}} \nabla_{t=0} \mathbb{E}_t \left(Y|E=1, M=m, X=x\right) f_{M|E,X} \left(m|E=0, X=x\right) f_X(x) d\mu(m,x) \\
&+ \iint_{\mathcal{S}\times\mathcal{X}} \mathbb{E}\left(Y|E=1, M=m, X=x\right) \nabla_{t=0} f_{M|E,X;t} \left(m|E=0, X=x\right) f_X(x) d\mu(m,x) \\
&+ \iint_{\mathcal{S}\times\mathcal{X}} \mathbb{E}\left(Y|E=1, M=m, X=x\right) f_{M|E,X} \left(m|E=0, X=x\right) \nabla_{t=0} f_X(x) d\mu(m,x)
\end{aligned}$$

Consider the first term, it is straightforward to verify that:

$$\begin{aligned}
&\iint_{\mathcal{S}\times\mathcal{X}} \nabla_{t=0} \mathbb{E}_t \left(Y|E=1, M=m, X=x\right) f_{M|E,X} \left(m|E=0, X=x\right) f_X(x) d\mu(m,x) \\
&= \mathbb{E}\left[U \frac{I(E=1)}{f_{E|X}\left(E|X\right)} \left\{ Y - \mathbb{E}\left(Y|E, M=m, X=x\right) \right\} \frac{f_{M|E,X}\left(M|E=0, X\right)}{f_{M|E,X}\left(M|E=1, X\right)} \right]
\end{aligned}$$

Similarly, one can easily verify that

$$\begin{aligned} & \iint_{\mathcal{S}\times\mathcal{X}} \mathbb{E}\left(Y|E=1,M=m,X=x\right)\nabla_{t=0}f_{M|E,X;t}\left(m|E=0,X=x\right)f_X(x)d\mu(m,x) \\ = \;& \mathbb{E}\left[U\frac{I(E=0)}{f_{E|X}\left(E|X\right)}\left\{\mathbb{E}\left(Y|E=1,M=m,X=x\right)-\eta\left(1,0,X\right)\right\}\right] \end{aligned}$$

and finally, one can also verify that

$$\begin{aligned} & \iint_{\mathcal{S}\times\mathcal{X}} \mathbb{E}\left(Y|E=1,M=m,X=x\right)f_{M|E,X}\left(m|E=0,X=x\right)\nabla_{t=0}f_X(x)d\mu(m,x) \\ = \;& \mathbb{E}\left[U\left\{\eta\left(1,0,X\right)-\theta_0\right\}\right] \end{aligned}$$

Thus, we obtain

$$\nabla_{t=0}\theta_t = \mathbb{E}\left\{S_{\theta_0}^{eff,nonpar}\left(\theta_0\right)U\right\}$$

Given $S_{\delta e}^{eff,nonpar}\left(\delta_e\right)$, the results for the direct and indirect effect follow from the fact that the influence function of a difference of two functionals equals the difference of the respective influence functions. Because the model is nonparametric, there is a unique influence function for each functional, and it is efficient in the model leading to the efficiency bound results.

PROOF OF THEOREM 2:

We begin by showing that

$$\mathbb{E}\{S_{\theta_0}^{eff,nonpar}\left(\theta_0;\beta_m^*,\beta_e^*,\beta_y^*\right)\} = 0 \tag{5}$$

under model $\mathcal{M}_{\text{union}}$. First note that $\left(\beta_y^*, \beta_m^*\right) = \left(\beta_y, \beta_m\right)$ under model $\mathcal{M}_a$. Equality (5) now follows because $\mathbb{E}^{par}\left(Y|X, M, E=1; \beta_y\right) = \mathbb{E}\left(Y|X, M, E=1\right)$ and $\eta\left(1, 0, X; \beta_y, \beta_m\right) = \mathbb{E}\left[\left\{\mathbb{E}^{par}\left(Y|X, M, E=1; \beta_y\right)\right\}|E=0, X\right] = \eta\left(1, 0, X\right)$

$$
\begin{aligned}
&\mathbb{E}\{S_{\theta_0}^{eff,nonpar}\left(\theta_0; \beta_m, \beta_e^*, \beta_y\right)\} \\
&= \mathbb{E}\left[\frac{I\left\{E=1\right\} f_{M|E,X}^{par}\left(M|E=0, X; \beta_m\right)}{f_{E|X}^{par}(1|X; \beta_e^*) f_{M|E,X}^{par}\left(M|E=1, X; \beta_m\right)} \overbrace{\mathbb{E}\left\{Y - \mathbb{E}^{par}\left(Y|X, M, E=1; \beta_y\right)\right]|E=1, M, X\right\}}^{=0}\right] \\
&+ \mathbb{E}\left[\frac{I(E=0)}{f_{E|X}^{par}(1|X; \beta_e^*)} \overbrace{\mathbb{E}\left[\left\{\mathbb{E}^{par}\left(Y|X, M, E=1; \beta_y\right) - \eta\left(1, 0, X; \beta_y, \beta_m\right)\right\}|E=0, X\right]}^{=0}\right] \\
&+ \mathbb{E}\left[\eta\left(1, 0, X; \beta_y, \beta_m\right)\right] - \theta_0 \\
&= 0
\end{aligned}
$$

Second, $\left(\beta_y^*, \beta_e^*\right) = \left(\beta_y, \beta_e\right)$ under model $\mathcal{M}_b$. Equality (5) now follows because $\mathbb{E}^{par}\left(Y|X, M, E=1; \beta_y\right) = \mathbb{E}\left(Y|X, M, E=1\right)$ and $f_{E|X}^{par}(1|X; \beta_e) = f_{E|X}(1|X)$:

$$
\begin{aligned}
&\mathbb{E}\{S_{\theta_0}^{eff,nonpar}\left(\theta_0; \beta_m^*, \beta_e, \beta_y\right)\} \\
= \;& \mathbb{E}\left[\frac{I\left\{E=1\right\} f_{M|E,X}^{par}\left(M|E=0, X; \beta_m^*\right)}{f_{E|X}^{par}(1|X; \beta_e) f_{M|E,X}^{par}\left(M|E=1, X; \beta_m^*\right)} \overbrace{\mathbb{E}\left\{Y - \mathbb{E}^{par}\left(Y|X, M, E=1; \beta_y\right)\right]|E=1, M, X\right\}}^{=0}\right] \\
&+ \mathbb{E}\left[\frac{I(E=0)}{f_{E|X}^{par}(1|X; \beta_e)} \mathbb{E}\left[\left\{\mathbb{E}^{par}\left(Y|X, M, E=1; \beta_y\right) - \eta\left(1, 0, X; \beta_y, \beta_m^*\right)\right\}|E=0, X\right]\right] \\
&+ \mathbb{E}\left[\eta\left(1, 0, X; \beta_y, \beta_m^*\right)\right] - \theta_0 \\
= \;& \mathbb{E}\left[\mathbb{E}\left[\left\{\mathbb{E}^{par}\left(Y|X, M, E=1; \beta_y\right)\right\}|E=0, X\right]\right] - \theta_0 = 0
\end{aligned}
$$

Third, equality (5) holds under model $\mathcal{M}_c$ because

$$\begin{aligned}
& \mathbb{E}\{S_{\theta_0}^{eff,nonpar}\left(\theta_0;\beta_m,\beta_e,\beta_y^*\right)\} \\
= \; & \mathbb{E}\left[\frac{I\{E=1\}f_{M|E,X}^{par}(M|E=0,X;\beta_m)}{f_{E|X}^{par}(1|X;\beta_e)f_{M|E,X}^{par}(M|E=1,X;\beta_m)}\mathbb{E}\left\{Y-\mathbb{E}^{par}\left(Y|X,M,E=1;\beta_y^*\right)\right]\}\right] \\
& +\mathbb{E}\left[\frac{I(E=0)}{f_{E|X}^{par}(1|X;\beta_e)}\mathbb{E}\left[\left\{\mathbb{E}^{par}\left(Y|X,M,E=1;\beta_y^*\right)-\eta\left(1,0,X;\beta_y^*,\beta_m\right)\right\}|E=0,X\right]\right] \\
& +\mathbb{E}\left[\eta\left(1,0,X;\beta_y^*,\beta_m\right)\right]-\theta_0 \\
= \; & \mathbb{E}\left[\mathbb{E}\left[\left\{\mathbb{E}\left(Y|X,M,E=1\right)\right\}|E=0,X\right]\right]-\mathbb{E}\left[\mathbb{E}\left[\mathbb{E}^{par}\left(Y|X,M,E=1;\beta_y^*\right)|E=0,X\right]\right] \\
& +\mathbb{E}\left[\mathbb{E}\left[\mathbb{E}^{par}\left(Y|X,M,E=1;\beta_y^*\right)|E=0,X\right]\right]-\mathbb{E}\left[\eta\left(1,0,X;\beta_y^*,\beta_m\right)\right] \\
& +\mathbb{E}\left[\eta\left(1,0,X;\beta_y^*,\beta_m\right)\right]-\theta_0 \\
= \; & \mathbb{E}\left[\mathbb{E}\left[\left\{\mathbb{E}\left(Y|X,M,E=1\right)\right\}|E=0,X\right]\right]-\theta_0
\end{aligned}$$

Assuming that the regularity conditions of Theorem 1A in Robins, Mark and Newey (1992) hold for $S_{\theta_0}^{eff,nonpar}\left(\theta_0;\beta_m,\beta_e,\beta_y\right)$,$S_\beta\left(\beta\right)$; the expression for $S_{\theta_0}^{union}\left(\theta_0,\beta^*\right)$ follows by standard Taylor expansion arguments and it now follows that

$$\sqrt{n}(\widehat{\theta}_0^{triply}-\theta_0)=\frac{1}{n^{1/2}}\sum_{i=1}^{n}S_{\theta_0,i}^{union}\left(\theta_0,\beta^*\right)+o_p(1) \tag{6}$$

The asymptotic distribution of $\sqrt{n}(\widehat{\theta}_0^{triply}-\theta_0)$ under model $\mathcal{M}_{\text{union}}$ follows from the previous equation by Slutsky's Theorem and the Central Limit Theorem.

We note that $\widehat{\delta}_e^{doubly}$ is CAN in the union model $\mathcal{M}_{\text{union}}$ since it is CAN in the larger model where either the density for the exposure is correct, or the density of the mediator and the outcome regression are both correct and thus $\eta\left(e,e,X;\beta_y^*,\beta_m^*\right)=$

$\mathbb{E}(Y|X, E=e)$. This gives the multiply robust result for direct and indirect effects. The asymptotic distribution of direct and indirect effect estimates then follow from similar arguments as above.

At the intersection submodel

$$\frac{\partial \mathbb{E}\left\{S_{\theta_0}^{eff,nonpar}(\theta_0,\beta)\right\}}{\partial \beta^T} = 0$$

hence

$$S_{\theta_0}^{union}(\theta_0,\beta) = S_{\theta_0}^{eff,nonpar}(\theta_0,\beta).$$

The semiparametric efficiency claim then follows for $\widehat{\theta}_0^{triply}$ and a similar argument gives the result for direct and indirect effects.

PROOF OF THEOREM 3:

The set of influence functions in the restricted model is given by $S_{\theta_0}^{eff,\text{nonpar}}(\theta_0) \oplus \left[\Lambda_M^{\text{model},\perp} \cap \Lambda_M^{\text{nonpar}}\right]$, since $S_{\theta_0}^{eff,\text{nonpar}}(\theta_0)$ is certainly an influence function in the restricted model, and $\left[\Lambda_M^{\text{model},\perp} \cap \Lambda_M^{\text{nonpar}}\right]$ constitutes the set of scores for the law of the mediator that are now orthogonal to the tangent space for the restricted model (Bickel et al 1993). Therefore, the efficient influence function is the element of the above set with smallest norm, which we obtain by noting that for $R \in \left[\Lambda_M^{\text{model},\perp} \cap \Lambda_M^{\text{nonpar}}\right]$

$$\begin{aligned}
&\mathbb{E}\left[S_{\theta_0}^{eff,\text{nonpar}}(\theta_0) + R\right]^2 \\
= \;&\mathbb{E}\left[S_{\theta_0}^{eff,\text{nonpar}}(\theta_0) - \Pi\left(S_{\theta_0}^{eff,\text{nonpar}}(\theta_0)\,|\Lambda_M^{\text{model},\perp} \cap \Lambda_M^{\text{nonpar}}\right)\right]^2 \\
&+\mathbb{E}\left[R + \Pi\left(S_{\theta_0}^{eff,\text{nonpar}}(\theta_0)\,|\Lambda_M^{\text{model},\perp} \cap \Lambda_M^{\text{nonpar}}\right)\right]^2
\end{aligned}$$

which is minimized by the choice $R = -\Pi\left(S_{\theta_0}^{eff,\text{nonpar}}(\theta_0)\,|\Lambda_M^{\text{model},\perp} \cap \Lambda_M^{\text{nonpar}}\right)$ proving the result. The same approach gives the result for the direct and indirect effects.

PROOF OF THEOREM 4:

Note that

$$\begin{aligned}
& \mathbb{E}\left[Y_{1,m}|E=e,X=x\right] \\
= \;& \mathbb{E}\left[Y_{1,m}|E=e,M=m,X=x\right] f_{M|E,X}\left(m|E=e,X=x\right) \\
& +\mathbb{E}\left[Y_{1,m}|E=e,M\neq m,X=x\right]\left(1-f_{M|E,X}\left(m|E=e,X=x\right)\right) \\
= \;& \mathbb{E}\left[Y_{1,m}|E=e,M=m,X=x\right] \\
& -t\left(e,m,x\right)\left(1-f_{M|E,X}\left(m|E=e,X=x\right)\right)
\end{aligned}$$

then

$$\begin{aligned}
& \mathbb{E}\left[Y_{1,m}|E=0,M=m,X=x\right]-\mathbb{E}\left[Y_{1,m}|E=1,M=m,X=x\right] \\
= \;& \overbrace{\mathbb{E}\left[Y_{1,m}|E=0,X=x\right]-\mathbb{E}\left[Y_{1,m}|E=1,X=x\right]}^{=0 \text{ by ignorability of } E} \\
& +t\left(0,m,x\right)\left(1-f_{M|E,X}\left(m|E=0,X=x\right)\right) \\
& -t\left(1,m,x\right)\left(1-f_{M|E,X}\left(m|E=1,X=x\right)\right) \\
= \;& -t\left(1,m,x\right)\left(1-f_{M|E,X}\left(m|E=1,X=x\right)\right) \\
& +t\left(0,m,x\right)\left(1-f_{M|E,X}\left(m|E=0,X=x\right)\right)
\end{aligned}$$

First note that $\left(\beta_y^*, \beta_m^*\right) = \left(\beta_y, \beta_m\right)$ under model $\mathcal{M}_a$, so

$$
\begin{aligned}
& \mathbb{E}\left[\begin{array}{c} \frac{I\{E=1\} f^{par}_{M|E,X}(M|E=0,X;\beta_m)}{f^{par}_{E|X}(1|X;\beta_e^*) f^{par}_{M|E,X}(M|E=1,X;\beta_m)} \left\{Y - \mathbb{E}^{par}\left(Y|X,M,E=1;\beta_y\right)\right]\right\} \\ +\eta^{par}\left(1,0,X;\lambda^*,\beta_y,\beta_m\right) \end{array}\right] \\
= \; & \mathbb{E}\left[\eta^{par}\left(1,0,X;\lambda^*,\beta_y,\beta_m\right)\right] \\
= \; & \mathbb{E}\left[\sum_{m\in\mathcal{S}} \left\{\begin{array}{c} \mathbb{E}^{par}\left(Y|X,M=m,E=1;\beta_y\right) \\ +t_{\lambda^*}(0,m,X)\left(1 - f^{par}_{M|E,X}\left(m|E=0,X\right);\beta_m\right) \\ -t_{\lambda^*}(1,m,X)\left(1 - f^{par}_{M|E,X}\left(m|E=1,X;\beta_m\right)\right) \end{array}\right\} f^{par}_{M|E,X}\left(m|E=0,X;\beta_m\right)\right] \\
= \; & \mathbb{E}\left[\mathbb{E}\left[Y_{1,M_0}|M_0,X\right]\right]
\end{aligned}
$$

Second, note that $\left(\beta_y^*, \beta_e^*\right) = \left(\beta_y, \beta_e\right)$ under model $\mathcal{M}_c$, so

$$\begin{aligned}
& \mathbb{E}\left[\begin{array}{c} \frac{I\{E=1\} f_{M|E,X}^{par}(M|E=0,X;\beta_m)}{f_{E|X}^{par}(1|X;\beta_e) f_{M|E,X}^{par}(M|E=1,X;\beta_m)} \left\{Y - \mathbb{E}^{par}\left(Y|X,M,E=1;\beta_y^*\right)\right]\right\} \\ +\eta^{par}\left(1,0,X;\lambda^*,\beta_y^*,\beta_m\right) \end{array}\right] \\
= & \mathbb{E}\left[\begin{array}{c} \frac{I\{E=1\} f_{M|E,X}^{par}(M|E=0,X;\beta_m)}{f_{E|X}^{par}(1|X;\beta_e) f_{M|E,X}^{par}(M|E=1,X;\beta_m)} \left\{\begin{array}{c} \mathbb{E}\left(Y|X,M,E=1\right) \\ +t_{\lambda^*}\left(0,m,X\right)\left(1 - f_{M|E,X}^{par}\left(m|E=0,X\right);\beta_m\right) \\ -t_{\lambda^*}\left(1,m,X\right)\left(1 - f_{M|E,X}^{par}\left(m|E=1,X;\beta_m\right)\right) \\ -\mathbb{E}^{par}\left(Y|X,M,E=1;\beta_y^*\right)] \\ -t_{\lambda^*}\left(0,m,X\right)\left(1 - f_{M|E,X}^{par}\left(m|E=0,X\right);\beta_m\right) \\ +t_{\lambda^*}\left(1,m,X\right)\left(1 - f_{M|E,X}^{par}\left(m|E=1,X;\beta_m\right)\right) \end{array}\right\} \\ +\eta^{par}\left(1,0,X;\lambda^*,\beta_y^*,\beta_m\right) \end{array}\right] \\
= & \mathbb{E}\left[\mathbb{E}\left[Y_{1,M_0}|M_0,X\right]\right] \\
& -\mathbb{E}\left[\sum_{m\in\mathcal{S}}\left\{\begin{array}{c} \mathbb{E}^{par}\left(Y|X,M=m,E=1;\beta_y^*\right) \\ +t_{\lambda^*}\left(0,m,X\right)\left(1 - f_{M|E,X}^{par}\left(m|E=0,X\right);\beta_m\right) \\ -t_{\lambda^*}\left(1,m,X\right)\left(1 - f_{M|E,X}^{par}\left(m|E=1,X;\beta_m\right)\right) \end{array}\right\} f_{M|E,X}^{par}\left(m|E=0,X;\beta_m\right)\right] \\
& +\eta^{par}\left(1,0,X;\lambda^*,\beta_y^*,\beta_m\right) \\
= & \mathbb{E}\left[\mathbb{E}\left[Y_{1,M_0}|M_0,X\right]\right]
\end{aligned}$$

which establishes double robustness. Let

$$\begin{aligned}
Q\left(\theta_0;\beta_m,\beta_e,\beta_y,\lambda^*\right) = & \frac{I\left\{E=1\right\} f_{M|E,X}^{par}\left(M|E=0,X;\beta_m\right)}{f_{E|X}^{par}(1|X;\beta_e) f_{M|E,X}^{par}\left(M|E=1,X;\beta_m\right)}\left\{Y - \mathbb{E}^{par}\left(Y|X,M,E=1;\beta_y\right)]\right\} \\
& +\eta^{par}\left(1,0,X;\lambda^*,\beta_y,\beta_m\right) - \theta_0
\end{aligned}$$

Then, the asymptotic distribution of $\widehat{\theta}_0^{doubly}\left(\lambda^*\right)$ for fixed λ^* is obtained as in Theorem

2 upon replacing $S_{\theta_0}^{eff,nonpar}\left(\theta_0;\beta_m,\beta_e,\beta_y\right)$ with $Q\left(\theta_0;\beta_m,\beta_e,\beta_y,\lambda^*\right)$.