Harvard University
Harvard University Biostatistics Working Paper Series

Year 2011 *Paper* 129

Semiparametric Estimation of Models for Natural Direct and Indirect Effects

Eric J. Tchetgen Tchetgen* Ilya Shpitser†

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

Semiparametric Estimation of Models for Natural Direct and Indirect Effects

Eric J. Tchetgen Tchetgen and Ilya Shpitser

Abstract

In recent years, researchers in the health and social sciences have become increasingly interested in mediation analysis. Specifically, upon establishing a non-null total effect of an 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. Natural direct and indirect effects are of particular interest as they generally combine to produce the total effect of the exposure and therefore provide insight on the mechanism by which it operates to produce the outcome. A semiparametric theory has recently been proposed to make inferences about marginal mean natural direct and indirect effects in observational studies (Tchetgen Tchetgen and Shpitser, 2011), which delivers multiply robust locally efficient estimators of the marginal direct and indirect effects, and thus generalizes previous results for total effects to the mediation setting. In this paper we extend the new theory to handle a setting in which a parametric model for the natural direct (indirect) effect within levels of pre-exposure variables is specified and the model for the observed data likelihood is otherwise unrestricted. We show that estimation is generally not feasible in this model because of the curse of dimensionality associated with the required estimation of auxiliary conditional densities or expectations, given high-dimensional covariates. We thus consider multiply robust estimation and propose a more general model which assumes a subset but not all of several working models holds.

Semiparametric Estimation of Models for Natural Direct and Indirect Effects

by

Eric J. Tchetgen Tchetgen and Ilya Shpitser

Departments of Epidemiology and Biostatistics, Harvard University

Abstract

In recent years, researchers in the health and social sciences have become increasingly interested in mediation analysis. Specifically, upon establishing a non-null total effect of an 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. Natural direct and indirect effects are of particular interest as they generally combine to produce the total effect of the exposure and therefore provide insight on the mechanism by which it operates to produce the outcome. A semiparametric theory has recently been proposed to make inferences about marginal mean natural direct and indirect effects in observational studies (Tchetgen Tchetgen and Shpitser, 2011), which de-

livers multiply robust locally efficient estimators of the marginal direct and indirect effects, and thus generalizes previous results for total effects to the mediation setting. In this paper we extend the new theory to handle a setting in which a parametric model for the natural direct (indirect) effect within levels of pre-exposure variables is specified and the model for the observed data likelihood is otherwise unrestricted. We show that estimation is generally not feasible in this model because of the curse of dimensionality associated with the required estimation of auxiliary conditional densities or expectations, given high-dimensional covariates. We thus consider multiply robust estimation and propose a more general model which assumes a subset but not all of several working models holds.

1 Introduction

In recent years, researchers in the health and social sciences have become increasingly interested in mediation analysis. Specifically, upon establishing a non-null total effect of an 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. 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). Such an effect generally differs from the controlled direct effect which 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). 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.

A semiparametric theory has recently been proposed to make inferences about marginal mean natural direct and indirect effects in observational studies (Tchetgen Tchetgen and Shpitser, 2011), which delivers multiply robust locally efficient estimators of the marginal direct and indirect effects, and thus generalizes previous similar results for the marginal total effect to the mediation setting. In this paper we further extend the new theory to handle a setting in which a parametric model for the natural direct (indirect) causal effect conditional on a subset of pre-exposure covariates is specified and the model for the observed data likelihood is otherwise unrestricted. Conditional models for direct and indirect effects are particularly of interest in making inferences about so-called moderated mediation effects, a topic of growing interest particularly in the field of psychology (Muller, Judd and Yzerbyt, 2005, and Preacher, Rucker and Hayes, 2007, MacKinnon, 2008). That is these models are useful for assessing the extent to which a pre-exposure variable modifies the natural direct (indirect) causal effect of exposure.

We show that estimation of the parameter indexing the causal model is generally not feasible in this model because of the curse of dimensionality associated with the required estimation of auxiliary conditional densities or expectations, given high-dimensional covariates. We thus consider a multiply robust approach and propose a more general model which assumes a subset but not necessarily all of several working models holds. Interestingly, we recover the results of Tchetgen Tchetgen and Shpitser (2010) as a special case in which the causal model conditions on no covariates and thus the marginal direct (indirect) effect is obtained. Here, we characterize the efficiency bound for the finite dimensional parameter of a model for a conditional natural direct (indirect) effect and develop a corresponding multiply robust locally efficient estimator; that is, an estimator that is consistent and asymptotically normal in the more general semiparametric model and which achieves the efficiency bound for the model, at the intersection submodel where all models are correct. Specifically, below we adopt the sequential ignorability assumption of Imai et al (2010) under which, together with the standard consistency and positivity assumptions, we derive the set of all influence functions including the semiparametric efficient influence function for the parameter of a model for the natural direct (indirect) causal effects given a subset of baseline covariates, in the semiparametric model $\mathcal{M}_{np}$ in which the observed data likelihood is otherwise unrestricted. We further show that in order to make inferences about conditional mediation effects in $\mathcal{M}_{np}$, 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 mentioned 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 this paper, 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 given subset of these models, instead we carefully combine these three models to produce estimators of the conditional mean direct and indirect effects that remain consistent and asymptotically normal 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 $\mathcal{M}_{np}$ corresponding to models (a), (b) and (c) respectively. We also define the submodel $\mathcal{M}_y$ which assumes that the conditional expectation of the outcome (i) is correctly modeled. The classical approach of Baron and Kenny essentially assumes model $\mathcal{M}_a$. as do the parametric approaches considered in Imai et al. (2010a, 2010b) and VanderWeele and Vansteelandt (2010), whereas van der Laan and Petersen (2005) consider the union of $\mathcal{M}_a$ and $\mathcal{M}_c$.

We show that when the causal models for natural direct and indirect effects condition on a strict subset of the confounders, the approach proposed in this paper is triply robust as it produces valid inferences about natural direct and indirect effects in the union model $\mathcal{M}_{\text{union}}^{abc} = \mathcal{M}_a \cup \mathcal{M}_b \cup \mathcal{M}_c$. Whereas, when the causal models condition on all confounders, the proposed method for direct effect models is doubly robust as it delivers valid inferences in the larger union model $\mathcal{M}_{\text{union}}^{yc} = \mathcal{M}_y \cup \mathcal{M}_c \supset \mathcal{M}_{\text{union}}^{abc}$.

Furthermore, we develop locally semiparametric efficient estimators, that achieve the efficiency bound for estimating the natural direct and indirect effects in $\mathcal{M}_{\text{union}}^{abc}$ when the causal models condition on a subset of the confounders, and in $\mathcal{M}_{\text{union}}^{yc}$ when the causal models condition on all of the confounders, at the intersection submodel where all models are correct.

We later 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 previously argued in Tchetgen Tchetgen and Shpitser (2011) for marginal direct effects, we again argue below that the approach of van der Laan and Petersen (2005) which only applies for $V \subset X$, is not entirely satisfactory for estimating conditional direct effects, because, despite producing a consistent and asymptotically normal estimator of the conditional 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 $\mathcal{M}_b \subset \mathcal{M}_{\text{union}}^{abc}$. Finally, in this paper we develop a novel double robust sensitivity analysis framework to assess the impact on inferences about the conditional natural direct (indirect) effect, of a departure from the ignorability assumption of the mediator variable. Until otherwise stated, we shall assume exposure is binary and later relax this assumption to allow for polytomous exposure.

2 Semiparametric theory

2.1 Conditional natural direct effects

Suppose independent and identically distributed data on a vector $O = (Y, E, M, X)$ is collected for n subjects. Here, Y is an outcome of interest, E is the binary exposure variable, M is a mediator variable with support $\mathcal{S}$, known to occur subsequently to E and prior to Y, and $X = (V, L)$ is a vector of pre-exposure variables with support $\mathcal{X} = \mathcal{V} \times \mathcal{L}$ that confound the association between (E, M) and Y. 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. The main objective in this section is to provide some theory of inference about the unknown $p-$dimensional parameter ψ indexing a parametric model $\gamma_{DIR}(E, V; \psi)$ for the conditional mean natural direct effect:

$$\gamma_{DIR}(e, V) = g\{\mathbb{E}(Y_{e,M_0}|V)\} - g\{\mathbb{E}(Y_{0,M_0}|V)\} \quad (1)$$

where $\mathbb{E}$ stands for expectation and g is the identity or log link function. $\gamma_{DIR}(E, V; \cdot)$ is assumed to be a smooth function that satisfies $\gamma_{DIR}(E, V; 0) = \gamma_{DIR}(0, V; \cdot) = 0$ and thus $\psi = 0$ encodes the null hypothesis of no natural direct effect. A simple example of the contrast $\gamma_{DIR}(E, V; \psi)$ takes the familiar linear form

$$\psi E$$

which assumes the natural direct effect of E is constant across levels of V. An alternative model might posit $\log \gamma_{DIR}(E, V; \psi)$ takes the linear form

$$(E, E \times V_1)\, \psi$$

which encodes effect modification on the log scale of the natural direct effect of the exposure by V_1 a component of V.

The conditional causal effect $\gamma_{DIR}(e, V\ )$ is generally not identified without additional assumptions.

To proceed, we make the consistency assumption:

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

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

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

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

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

paired with the following positivity assumption:

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

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

Then, under the consistency, sequential ignorability and positivity assumptions, one can show as in Imai et al (2010), that:

$$\mathbb{E}(Y_{e,M_0}|V = v)$$
$$= \iint_{\mathcal{S}\times\mathcal{L}} \mathbb{E}(Y|E = e, M = m, L = l, V = v)\, f_{M|E,X}(m|E = 0, L = l, V = v)\, f_{L|V}(l|V = v)d\mu(m,l) \tag{2}$$

where $f_{M|E,X}$ and $f_{L|V}$ are respectively the conditional density of the mediator M given (E,X) and the density of L given V, and μ is a dominating measure for the distribution of $[M, L|V]$. Thus $\gamma_{DIR}(e,v)$ is identified from the observed data (See Pearl, 2011 and van der Laan and Petersen (2005) for related identification results). Tchetgen Tchetgen and Shiptser (2011) considered the special case where $V = \varnothing$ in which case $\gamma_{DIR}(e,V\ ) = \gamma_{DIR}(e)$ is a nonparametric functional. 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 adapt and extend the sensitivity analysis technique of Tchetgen Tchetgen and Shpitser (2011), 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.

We give our first result, which serves as motivation for our multiply robust approach. First, 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, If $\widehat{\psi}$ is a regular asymptotically linear estimator of ψ in model $\mathcal{M}_{np}$, then there exists a $p \times 1$ function $h(V)$ of V such that $\widehat{\psi}$ has influence function $S_{\psi}^{np}(h; \psi)$, where for g the identity link

$$S_{\psi}^{;np}(h; \psi) = h(V)\, U_1(\psi)$$

$$\text{where } U_1(\psi) = \frac{I(E = 1) f_{M|E,X}(M|E = 0, X)}{f_{E|X}(1|X) f_{M|E,X}(M|E = 1, X)} \{Y - \mathbb{E}(Y|X, M, E = 1)]\}$$
$$+ \frac{I(E = 0)}{f_{E|X}(0|X)} \{\mathbb{E}(Y|X, M, E = 1) - Y - \eta(1, 0, X) + \eta(0, 0, X)\}$$
$$+ \{\eta(1, 0, X) - \eta(0, 0, X) \;\; - \gamma_{DIR}(1, V \;; \psi)\}$$

and for g the log-link

$$S_{\psi}^{np}(h;\psi) = h(V)\,U_2(\psi)$$

where

$$\begin{aligned} U_2(\psi) = & \frac{I(E=1)\, f_{M|E,X}(M|E=0,X)}{f_{E|X}(1|X) f_{M|E,X}(M|E=1,X)} \{Y - \mathbb{E}(Y|X,M,E=1)]\} \exp\{-\gamma_{DIR}(1,V;\psi\ )\} \\ & + \ \frac{I(E=0)}{f_{E|X}(0|X)} \left\{ \begin{array}{c} \mathbb{E}(Y|X,M,E=1)\exp\{-\gamma_{DIR}(1,V;\ \psi)\} - Y \\ -\eta(1,0,X)\exp\{-\gamma_{DIR}(1,V;\ \psi)\} + \eta(0,0,X) \end{array} \right\} \\ & + \ \{\eta(1,0,X)\exp\{-\gamma_{DIR}(1,V;\psi\ )\} - \eta(0,0,X)\ \ \} \end{aligned}$$

That is, $n^{1/2}\left(\widehat{\psi} - \psi\right) = n^{-1/2}\sum_{i=1}^{n} S_{\psi,i}^{np}(h;\psi) + o_p(1)$. *In the special case where* $V = X$

$$\begin{aligned} U_1(\psi) = & \ \frac{I(E=1)\, f_{M|E,X}(M|E=0,X)}{f_{E|X}(1|X) f_{M|E,X}(M|E=1,X)} \{Y - \mathbb{E}(Y|X,M,E=1)]\} \\ & + \ \frac{I(E=0)}{f_{E|X}(0|X)} \{\mathbb{E}(Y|X,M,E=1) - Y - \gamma_{DIR}(1,X\ ;\psi)\} \end{aligned}$$

and

$$\begin{aligned} U_2(\psi) = & \frac{I(E=1)\, f_{M|E,X}(M|E=0,X)}{f_{E|X}(1|X) f_{M|E,X}(M|E=1,X)} \{Y - \mathbb{E}(Y|X,M,E=1)]\} \exp\{-\gamma_{DIR}(1,X;\psi\ )\} \\ & + \ \frac{I(E=0)}{f_{E|X}(0|X)} \{\mathbb{E}(Y|X,M,E=1)\exp\{-\gamma_{DIR}(1,X;\psi)\} - Y\} \end{aligned}$$

The efficient score of ψ *in model* $\mathcal{M}_{np}$ *is given by* $S_{\psi}^{eff,np}(\psi) = S_{\psi}^{np}(h_{opt};\psi)$ *where* $h_{opt}(V) = \mathbb{E}\left\{\frac{\partial U(\psi)}{\partial \psi}|V\right\}\mathbb{E}\left\{U(\psi)^2|V\right\}^{-1}$ *with* $U(\psi) = U_1(\psi)$ *in the case of the identity link and* $U(\psi) = U_2(\psi)$ *for the log-link.*

By standard results from semiparametric theory in Bickel et al. (1993), Theorem 1 implies that all regular and asymptotically linear estimators of ψ in model $\mathcal{M}_{np}$ can be obtained (up to asymptotic equivalence) as the solution $\widetilde{\psi}(h)$ to the equation

$$\mathbb{P}_n S_{\psi}^{eff,np}(h;\psi) = 0, \tag{3}$$

for some p-dimensional function h,where $\mathbb{P}_n(\cdot) = n^{-1}\sum_i (\cdot)_i$. The solution $\widetilde{\psi}(h)$ to this equation is an infeasible estimator as the set of functions $\left\{S_{\psi}^{eff,np}(h;\psi) : h\right\}$ with finite variance, depends on the unknown conditional expectation $\mathbb{E}(Y|X,M,E)$, and on the unknown density functions $f_{E|X}(\cdot|X)$ and $f_{M|E,X}(\cdot|E,X)$. A feasible regular and asymptotically linear estimator is not possible unless at least a subset of these unknown functions can be consistently estimated. While smoothing methods could in principle be used, as argued above, with the sample sizes found in practice, the data available to estimate either of the required functions will be sparse when X is a vector with more than two continuous components. As a consequence any feasible estimator of ψ in model $\mathcal{M}_{np}$ will exhibit poor finite sample performance when the predictor space is large. It follows that in general, inference about ψ in model $\mathcal{M}_{np}$ is infeasible due to the curse of dimensionality and that dimension-reducing (e.g. parametric) working models must be used to estimate the unknown auxiliary functions $\mathbb{E}(Y|X,M,E)$, $f_{M|E,X}(\cdot|E,X)$ and $f_{E|X}(\cdot|X)$. For this reason, in the following section, we consider inferences that employ parametric working models for each of these functions. Specifically, we assume a working regression

$\mathbb{E}^{par}\left(Y|X,M,E;\beta_y\right) = g^{-1}\left(\beta_y^T r(X,M,E)\right)$ for $\mathbb{E}\left(Y|X,M,E\right)$ where r is a user specified function of (X,M,E) and β_y is an unknown parameter estimated by $\widehat{\beta}_y$ that solves the estimating equation:

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

where g^* is a link function. Similarly, we set $\widehat{f}^{par}_{M|E,X}\left(m|E,X\right) = f^{par}_{M|E,X}\left(m|E,X;\widehat{\beta}_m\right)$ for $f^{par}_{M|E,X}\left(m|E,X;\beta_m\right)$ 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

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

We could in principle obtain inferences about ψ by only using two of these three working models, say for instance under the submodel $\mathcal{M}_a$, by obtaining $\widehat{\psi}_{\mathcal{M}_a}$ that solves the equation:

$$\mathbb{P}_n\left[h\left(V\right)\left\{\widehat{\eta}\left(1,0,X\right) - \widehat{\eta}\left(0,0,X\right) - \gamma_{DIR}\left(1,V;\widehat{\psi}_{\mathcal{M}_a}\right)\right\}\right] = 0$$

for g the identity link, and a user specified function h of dimension p; where $\widehat{\eta}^{par}\left(e,e^*,X\right) = \int_{\mathcal{S}}\widehat{\mathbb{E}}^{par}\left(Y|X,M=m,E=e\right)\ \widehat{f}^{par}_{M|E,X}\left(m|E=e^*,X\right)d\mu(m)$. Unfortunately, $\widehat{\psi}_{\mathcal{M}_a}$ will generally fail to be consistent if either $\mathbb{E}^{par}\left(Y|X,M,E;\beta_y\right)$ or $f^{par}_{M|E,X}\left(m|E,X;\beta_m\right)$

is incorrect, even if one of the two models is correct and $f^{par}_{E|X}(e|X;\beta_e)$ is correct. One of two alternative approaches might be considered, where an estimator is obtained, assuming either, both $\mathbb{E}^{par}\left(Y|X,M,E;\beta_y\right)$ and $f^{par}_{E|X}(E|X;\beta_e)$ are correct,i.e. $\mathcal{M}_b$ is correct, or one that assumes both $f^{par}_{M|E,X}(M|E,X;\beta_m)$ and $f^{par}_{E|X}(E|X;\beta_e)$ and thus $\mathcal{M}_c$ is correct. Both of these alternative approaches suffer from the same potential for yielding biased results under mis-specification of the required models, and therefore will not be pursued any further.

In the following section, to handle the setting where $V\subset X$, we develop a multiply robust approach that combines these three parametric models, and gives the correct answer under the union model $\mathcal{M}^{abc}_{\text{union}}=\mathcal{M}_a\cup\mathcal{M}_b\cup\mathcal{M}_c$ in which any one of the three working models (i),(ii) and (iii), can be incorrect provided the other two are correct; and remarkably, the analyst does not need to know which of the three models is incorrect to get a correct answer. A doubly robust estimators for direct effect models, that are consistent and asymptotically normal in $\mathcal{M}^{yc}_{\text{union}}$ are obtained for the setting where $V=X$.

2.2 Multiply robust estimation

The proposed estimator $\widehat{\psi}=\widehat{\psi}\,(h)$ solves

$$\mathbb{P}_n\widehat{S}^{eff,np}_{\psi}\left(h;\widehat{\psi}\right)=0$$

where h is user-specified, and $\widehat{S}_{\psi}^{eff,np}\left(h;\widehat{\psi}\right) = S_{\psi}^{eff,np}\left(h;\widehat{\beta}_m,\widehat{\beta}_e,\widehat{\beta}_y,\widehat{\psi}\right)$ is equal to $S_{\psi}^{eff,np}\left(h;\widehat{\psi}\ \right)$ evaluated at $\{\widehat{\mathbb{E}}^{par}\left(Y|E,M,X\right), \widehat{f}_{M|E,X}^{par}\left(m|E,X\right), \widehat{f}_{E|X}^{par}(e|X)\}$ instead of $\{\mathbb{E}\left(Y|E,M,X\right), f_{M|E,X}\left(m|E,X\right), f_{E|X}(e|X)\}$. That is $\widehat{\psi}$ is consistent and asymptotically normal in model $\mathcal{M}_{\text{union}}^{abc}$ when $V \subset X$ and in model $\mathcal{M}_{\text{union}}^{yc}$ when $V = X$. The following theorem states the result more formally.

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. Then,* $\sqrt{n}(\widehat{\psi}-\psi)$ *is regular and asymptotically linear respectively under model* $\mathcal{M}_{\text{union}}^{abc}\left(\mathcal{M}_{\text{union}}^{yc}\right)$ *when* $V \subset X$ $(V = X)$, *with influence function* $\mathbb{E}\left\{\frac{\partial S_{\psi}^{eff,np}\left(h;\beta,\overline{\psi}\right)}{\partial\overline{\psi}^T}|_{\psi}\right\}^{-1} S_{\psi}^{eff,union}\left(h;\beta^*,\psi\right)$ *where*

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

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

$$\Sigma_{\psi}\left(h;\psi,\beta^*\right) = \mathbb{E}\left\{S_{\psi}^{eff,union}\left(h;\beta^*,\psi\right)^{\otimes 2}\right\}$$

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

Suppose that $\widehat{h}_{opt}$ *denotes a consistent estimator of* h_{opt}, *then* $\widehat{\psi}_{eff} = \widehat{\psi}\left(\widehat{h}_{opt}\right)$ *is semiparametric locally efficient in the sense that it is regular and asymptotically linear in model* $\mathcal{M}_{\text{union}}^{abc}\left(\mathcal{M}_{\text{union}}^{yc}\right)$ *and it achieves the semiparametric efficiency bound* *for model* $\mathcal{M}_{\text{union}}^{abc}$ $(\mathcal{M}_{\text{union}}^{yc})$ *at the intersection submodel* $\mathcal{M}_a \cap \mathcal{M}_b \cap \mathcal{M}_c$*, with efficient influence function:* $\mathbb{E}\left\{\frac{\partial S_{\psi}^{eff,np}\left(h_{opt};\beta,\overline{\psi}\right)}{\partial \overline{\psi}^T}|_{\psi}\right\}^{-1} S_{\psi}^{eff,np}\left(h_{opt};\beta^*,\psi\right)$.

An empirical version of $\Sigma_{\psi}\left(h;\psi,\beta^*\right)$ is easily obtained, and the corresponding Wald type confidence interval can be used to make formal inferences about natural direct and indirect effects. By a theorem due to Robins and Rotnitzky (2001), the theorem implies that when all models are correct $\widehat{\psi}_{eff}$ is semiparametric efficient in model $\mathcal{M}_{\text{np}}$ at the intersection submodel $\mathcal{M}_a \cap \mathcal{M}_b \cap \mathcal{M}_c$, provided that $\widehat{h}_{opt}$ converges to h_{opt} in probability.

When $V = X$, we note from the previous section that a working model for the outcome regression $\mathbb{E}\left(Y|X,M,E\right)$ is only needed when $E = 1$, and therefore, $\mathbb{E}^{par}\left(Y|X,M,E;\beta_y\right)$ can be replaced by the more parsimonious model $\mathbb{E}^{par}\left(Y|X,M;\beta_y\right) = g^{-1}\left(\beta_y^T r(X,M)\right)$ for $\mathbb{E}\left(Y|X,M,E=1\right)$ where r is a user specified function of (X,M) and β_y is an unknown parameter estimated by $\widehat{\beta}_y$ that solves the estimating equation:

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

We note that local efficiency will generally involve additional modeling to obtain $\widehat{h}_{opt}$ than is required for multiple robustness. To clarify this point, consider the log-link. Then, some algebra reveals that

$$h_{opt}(V) = \frac{\partial \gamma_{DIR}\left(1,V;\psi\right)}{\partial \psi}\mathbb{E}\left\{\eta\left(0,0,X\right)|V\right\}\mathbb{E}\left\{U_2\left(\psi\right)^2|V\right\}^{-1}$$

and therefore,

$$\widehat{h}_{opt} = \frac{\partial \gamma_{DIR}\left(1, V; \widehat{\psi}_{\text{prelim}}\right)}{\partial \psi} \widehat{\mathbb{E}}\left\{\widehat{\eta}\left(0,0,X\right)|V\right\} \widehat{\mathbb{E}}\left\{U_2\left(\widehat{\psi}_{\text{prelim}}\right)^2 |V\right\}^{-1}$$

where $\widehat{\psi}_{\text{prelim}}$ is a preliminary (possibly multiply robust) estimator of ψ, $\widehat{\mathbb{E}}\left\{\widehat{\eta}\left(0,0,X\right)|V\right\}$ is an estimate of a parametric regression of $\eta\left(0,0,X\right)$ on V, and $\widehat{\mathbb{E}}\left\{U_2\left(\widehat{\psi}_{\text{prelim}}\right)^2 |V\right\}$ is an estimate of a parametric model for the variance of $U_2\left(\psi\right)$ given V. Thus, we may conclude that local efficiency is contingent on a correct specification of the latter two regression models. Similarly, additional modeling is required for local efficiency in the case of an identity link, details are omitted but are easily inferred from the above remark.

2.3 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

(Pearl 2011), several authors (Imai et al, 2010, Pearl, 2011 and VanderWeele, 2009) have demonstrated how the simple linear structural equation approach generalizes to accommodate both, the presence of an interaction between exposure and mediator variables, or a non-linear link 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 must be modeled, inferences based on parametric structural equations (Pearl 2011, Imai et al, 2010, Pearl, 2011, VanderWeele, 2009, VanderWeele and Vansteedlandt, 2010) 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 such a system of structural equations, whether linear or nonlinear, 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 estimator of a natural direct effect due to van der Laan and Petersen (2005) who only consider $V \subset X$. They show their estimator remains consistent and asymptotically normal 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. Furthermore, in contrast to van der Laan and Petersen who do not consider the estimation of natural indirect effect models, in the next section, we develop an analogous multiply robust approach to estimate the parameter indexing a model for a conditional natural indirect effect. We refer the reader to Tchetgen Tchetgen and Shpitster (2011) for an additional discussion of the van der Laan approach and of some important implications for efficiency of the assumption that a parametric or semiparametric model for the mediator conditional density is known.

2.4 Estimation of conditional natural indirect effects

In this section we develop a theory of estimation of the unknown $q-$dimensional parameter θ indexing a parametric model $\gamma_{IND}(E,V;\theta)$ for the conditional mean natural indirect effect:

$$\gamma_{IND}(e,V) = g\left\{\mathbb{E}\left(Y_{1,M_e}|V\right)\right\} - g\left\{\mathbb{E}\left(Y_{1,M_0}|V\right)\right\} \quad (4)$$

where g is again either the identity or log link function. $\gamma_{IND}(E,V;\cdot)$ is assumed to be a smooth function that satisfies $\gamma_{IND}(E,V;0) = \gamma_{IND}(0,V;\cdot) = 0$ and thus $\theta = 0$ encodes the null hypothesis of no natural indirect effect. A simple example of the contrast $\gamma_{IND}(E,V;\theta)$ takes the familiar form

$E\ \theta$

which assumes the natural indirect effect of E is constant across levels of V. An alternative model might posit $\log \gamma_{IND}(E, V; \theta)$ takes the form

$$(E, E \times V_1)\ \theta$$

which encodes effect modification on the log scale of the indirect effect of the exposure by V_1 a component of V.

The contrast $\gamma_{IND}(e, V)$ is identified under the consistency, positivity and sequential ignorability assumptions, since $\mathbb{E}(Y_{1,M_1}|V) = \mathbb{E}(Y_1|V)$ and $\mathbb{E}(Y_{1,M_0}|V)$ are both identified under the assumptions.

Theorem 3 below is the indirect effect analogue of Theorem 1.

Theorem 3: Under the consistency, sequential ignorability and positivity assumptions, If $\widehat{\theta}$ is a regular asymptotically linear estimator of θ in model $\mathcal{M}_{np}$, then there exists a $q \times 1$ function $h(V)$ of V such that $\widehat{\theta}$ has influence function $S_{\theta}^{np}(h; \theta)$, where for the identity link

$$S_{\theta}^{np}(h; \theta) = h(V)\, W_1(\theta)$$

$$\text{where } W_1(\theta) = \frac{I(E=1)}{f_{E|X}(1|X)} \left[\begin{array}{c} Y - \eta(1,1,X) \\ -\frac{f_{M|E,X}(M|E=0,X)}{f_{M|E,X}(M|E=1,X)} \{Y - \mathbb{E}(Y|X,M,E=1)]\} \end{array} \right]$$

$$- \frac{I(E=0)}{f_{E|X}(0|X)} \{\mathbb{E}(Y|X,M,E=1) - \eta(1,0,X)\}$$

$$+ \eta(1,1,X) - \eta(1,0,X) - \gamma_{IND}(1, V; \theta)$$

and for g the log-link

$$S_\theta^{np}(h;\theta) = h(V)\, W_2(\theta)$$

where

$$W_2(\theta) = \frac{I(E=1)}{f_{E|X}(1|X)} \left\{ \begin{array}{c} [Y - \eta(1,1,X)] \exp\{-\gamma_{IND}(1,V;\theta)\} \\ -\frac{f_{M|E,X}(M|E=0,X)}{f_{M|E,X}(M|E=1,X)} \{Y - \mathbb{E}(Y|X,M,E=1)]\} \end{array} \right\}$$
$$- \frac{I(E=0)}{f_{E|X}(0|X)} \{\mathbb{E}(Y|X,M,E=1) - \eta(1,0,X)\}$$
$$+ \eta(1,1,X)\exp\{-\gamma_{IND}(1,V;\theta)\} - \eta(1,0,X)$$

That is, $n^{1/2}\left(\widehat{\theta} - \theta\right) = n^{-1/2}\sum_{i=1}^{n} S_{\theta,i}^{np}(h;\theta) + o_p(1)$. *In the special case where* $V =$ X

$$W_1(\theta) = \frac{I(E=1)}{f_{E|X}(1|X)} \left[\begin{array}{c} Y - \eta(1,1,X) \\ -\frac{f_{M|E,X}(M|E=0,X)}{f_{M|E,X}(M|E=1,X)} \{Y - \mathbb{E}(Y|X,M,E=1)]\} \end{array} \right]$$
$$- \frac{I(E=0)}{f_{E|X}(0|X)} \{\mathbb{E}(Y|X,M,E=1) - \eta(1,1,X) + \gamma_{IND}(1,X;\theta)\}$$

and for g the log-link

$$W_2(\theta) = \frac{I(E=1)}{f_{E|X}(1|X)} \left[\begin{array}{c} [\{Y - \eta(1,1,X)\}] \exp\{-\gamma_{IND}(1,X;\theta)\} \\ -\frac{f_{M|E,X}(M|E=0,X)}{f_{M|E,X}(M|E=1,X)} \{Y - \mathbb{E}(Y|X,M,E=1)]\} \end{array} \right]$$
$$- \frac{I(E=0)}{f_{E|X}(0|X)} \{\mathbb{E}(Y|X,M,E=1) - \eta(1,1,X)\exp\{-\gamma_{IND}(1,X;\theta)\}\}$$

The efficient score of θ in model $\mathcal{M}_{np}$ is given by $S_\theta^{eff,np}(\theta) = S_\theta^{np}(h_{opt};\theta)$ *where* $h_{opt}(V) = \mathbb{E}\left\{\frac{\partial W(\theta)}{\partial \theta}|V\right\}\mathbb{E}\left\{W(\theta)^2|V\right\}^{-1}$ *where* $W(\theta) = W_1(\theta)$ *in the case of the identity link and* $W(\theta) = W_2(\theta)$ *for the log-link.*

As in the previous section, we propose to base inferences about θ on the triply robust estimator $\widehat{\theta} = \widehat{\theta}(h)$ that solves

$$\mathbb{P}_n \widehat{S}_{\theta}^{eff,np}\left(h;\widehat{\theta}\right) = 0$$

where h is user-specified of dimension q, and $\widehat{S}_{\theta}^{eff,np}\left(h;\widehat{\theta}\right) = S_{\psi}^{eff,np}\left(h;\widehat{\beta}_m,\widehat{\beta}_e,\widehat{\beta}_y,\widehat{\theta}\right)$ is equal to $S_{\theta}^{eff,np}\left(h;\widehat{\theta}\right)$ evaluated at $\{\widehat{\mathbb{E}}^{par}\left(Y|E,M,X\right), \widehat{f}_{M|E,X}^{par}\left(m|E,X\right), \widehat{f}_{E|X}^{par}(e|X)\}$ instead of $\{\mathbb{E}\left(Y|E,M,X\right),\ f_{M|E,X}\left(m|E,X\right),\ f_{E|X}(e|X)\}$. In fact, an analogue to Theorem 2 that states that $\widehat{\theta}$ is consistent and asymptotically normal in model $\mathcal{M}_{\text{union}}^{abc}$, can be established using a similar technique as in the proof of Theorem 2, and a locally efficient estimator is similarly obtained. As for direct effects, an essential condition for the result for indirect effects entails showing that the estimating function $S_{\psi}^{eff,np}\left(h;\beta_m,\beta_e,\beta_y,\theta\right)$ for $V \subset X$ is triply robust (doubly robust), and thus has mean zero in model $\mathcal{M}_{\text{union}}^{abc}$, a property that can be confirmed in a manner similar to the proof of Theorem 2. When $V = X$, we note that the remark made in Section 2.2 which states that one really only needs to model $\mathbb{E}\left(Y|E=1,M,X\right)$ instead of $\mathbb{E}\left(Y|E,M,X\right)$ equally applies here; but we also observe that unlike $\widehat{\psi}$, the triply robust estimator $\widehat{\theta}$ is unfortunately not doubly robust in this case.

We finally note that by definition

$$\begin{aligned} &\overbrace{g\left\{\mathbb{E}\left(Y_1|V\right)\right\} - g\left\{\mathbb{E}\left(Y_0|V\right)\right\}}^{=total\ effect} \\ = \ &\overbrace{g\left\{\mathbb{E}\left(Y_{1,M_1}|V\right)\right\} - g\left\{\mathbb{E}\left(Y_{1,M_0}|V\right)\right\}}^{=natural\ indirect\ effect} + \overbrace{g\left\{\mathbb{E}\left(Y_{1,M_0}|V\right)\right\} - g\left\{\mathbb{E}\left(Y_{0,M_0}|V\right)\right\}}^{=natural\ direct\ effect} \end{aligned}$$

and therefore $\gamma_{DIR}(E,V;\psi)$ and $\gamma_{IND}(E,V;\theta)$ combine to produce a model of the total exposure effect in terms of its direct and indirect components.

2.5 Polytomous exposure

In this section we show how the results from previous sections generalize for polytomous exposure. Suppose E has finite support $\mathcal{E}$, and as before we wish to make inferences about $\gamma_{DIR}(E,V;\psi)$ or $\gamma_{IND}(E,V;\theta)$. We outline how a generalization of the proposed method is obtained for estimating $\gamma_{DIR}(E,V;\psi)$ when $V\subset X$, and omit details for the other settings, although the corresponding extension can easily be inferred from the exposition. To obtain an estimator of ψ with the multiple robustness properties described in Section 2.2, we propose to solve $\mathbb{P}_n\widehat{S}_{\psi}^{eff,np}\left(h;\widehat{\psi}\right)=0$ with $S_{\psi}^{eff,np}(h;\psi)=h(V)U_1(\psi)$ upon redefining $U_1(\psi)$ to equal:

$$
\begin{aligned}
U_1(\psi) = & \sum_{e\in\mathcal{E}\backslash\{0\}} \frac{I(E=e)\, f_{M|E,X}(M|E=0,X)}{f_{E|X}(e|X) f_{M|E,X}(M|E=e,X)}\{Y-\mathbb{E}(Y|X,M,E=e)]\} \\
& + \sum_{e\in\mathcal{E}\backslash\{0\}} \frac{I(E=0)}{f_{E|X}(0|X)}\{\mathbb{E}(Y|X,M,E=e)-Y-\eta(e,0,X)+\eta(0,0,X)\} \\
& + \sum_{e\in\mathcal{E}\backslash\{0\}} \{\eta(e,0,X)-\eta(0,0,X) \quad -\gamma_{DIR}(e,V\ ;\psi)\}
\end{aligned}
$$

3 A semiparametric sensitivity analysis

We extend the semiparametric sensitivity analysis technique proposed by Tchetgen Tchetgen and Shpitser (2011), to assess the extent to which a violation of the ignorability assumption for the mediator might alter inferences about a conditional natural direct or indirect effect. Although only results for natural direct effects are presented, the extension for indirect effects is given in the appendix. Let

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

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 $t(e, m, x) \neq 0$ for some (e, m, x). Suppose M is binary and higher values of Y are beneficial for health, then if $t(e, 1, x) > 0$ but $t(e, 0, x) < 0$, then on average, individuals with $\{E = e, X = x\}$ and mediator value $M = 1$ have higher potential outcomes $\{Y_{11,} Y_{10}\}$ than individuals with $\{E = e, X = x\}$ but $M = 0$; i.e. healthier individuals are more likely to receive the mediator. On the other hand, if $t(e, 1, x) < 0$ but $t(e, 0, x) > 0$ suggests confounding by indication for the mediator variable; i.e. unhealthier individuals are more more likely to receive the mediating factor.

We proceed as in Robins et al (1999) who originally proposed using a selection bias function for the purposes of conducting a sensitivity analysis for total effects, and Tchetgen Tchetgen and Shpitser (2011) who adapted the approach for assessing the impact of unmeasured confounding on the estimation of a marginal natural direct effect. Here we propose to recover inferences about conditional natural direct effects by assuming the selection bias function $t(e,m,x)$ 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 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, Tchetgen Tchetgen and Shpitser (2011) establish:

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

and therefore $\mathbb{E}\left(Y_{1,M_0}|V\right)$ is identified by:

$$
\mathbb{E}\left(Y_{1,M_0}|V\right) = \mathbb{E}\left(\sum_{m\in\mathcal{S}}\left[\begin{array}{c}\mathbb{E}\left(Y|E=1,M=m,X\right)\\ -\,t(1,m,X)\left\{1-f_{M|E,X}\left(m|E=1,X\right)\right\}\\ +t(0,m,X)\left\{1-f_{M|E,X}\left(m|E=0,X\right)\right\}\end{array}\right] f_{M|E,X}\left(m|E=0,X\right)\middle| V\right) \quad (5)
$$

which is equivalently represented as:

$$
\mathbb{E}\left[\frac{I(E=1)\,f_{M|E,X}\left(M|E=0,X\right)}{f_{E|X}(1|X)f_{M|E,X}\left(M|E=1,X\right)}\left\{\begin{array}{c} Y-\; t(1,M,X)\left\{1-f_{M|E,X}\left(m|E=1,X\right)\right\}\\ +t(0,M,X)\left\{1-f_{M|E,X}\left(M|E=0,X\;\right)\right\}\end{array}\right\}\middle| V\right] \quad (6)
$$

Below, these two equivalent representations (5) and (6) are carefully combined to obtain a double robust estimator of ψ assuming $t(\cdot,\cdot,\cdot)$ is known. A sensitivity analysis is then obtained as in Tchetgen Tchetgen and Shpitser (2011) by repeating this process and by 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 ignorability assumption of M, 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 semi-parametric estimator of ψ 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^*. Suppose V is a proper subset of X. The proposed doubly robust estimator of the natural direct effect is then given by $\widehat{\psi}^{doubly}(\lambda^*) = \widehat{\psi}^{doubly}(h;\lambda^*)$ that solves, for g the identity link

$$\mathbb{P}_n\left\{\widehat{S}^{,doubly}_{\psi}(h;\psi,\lambda^*)\right\} = \mathbb{P}_n\left\{h(V)\,\widehat{U}_1(\psi,\lambda^*)\right\}$$

$$\begin{aligned}
\text{where } \widehat{U}_1(\psi,\lambda^*) = \; & \frac{I(E=1)\,\widehat{f}^{par}_{M|E,X}(M|E=0,X)}{\widehat{f}^{par}_{E|X}(1|X)\widehat{f}^{par}_{M|E,X}(M|E=1,X)}\left\{Y - \widehat{\mathbb{E}}^{par}(Y|X,M,E=1)\right\} \\
& - \frac{I(E=0)}{f_{E|X}(0|X)}\left\{Y - \widehat{\mathbb{E}}^{par}(Y|X,M,E=0)\right\} \\
& + \widetilde{\eta}^{par}(1,0,X;\lambda^*) - \widehat{\eta}(0,0,X) - \gamma_{DIR}(1,V\;;\psi)
\end{aligned}$$

where

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

A sensitivity analysis then entails reporting the set $\left\{\widehat{\psi}^{doubly}(\lambda):\lambda\right\}$ (and the associated confidence intervals) which summarizes how sensitive inferences are to a deviation from the ignorability assumption $\lambda=0$. The theoretical justification for the approach is given by the following result that generalizes Theorem 4 of Tchetgen Tchetgen and Shpitser (2011). Its proof is given 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,* $\widehat{\psi}^{doubly}(\lambda^*)$ *is a consistent and asymptotically normal estimator of* ψ *in* $\mathcal{M}_a\cup\mathcal{M}_c$.

The influence function of $\widehat{\psi}^{doubly}(\lambda^*)$ is given in the appendix which can in turn be used to construct corresponding confidence intervals. The appendix also gives an analogous double robust sensitivity analysis technique for direct effects when $V=X$ or g is the log-link, as well as corresponding methodology for indirect effects. Interestingly, as we note in the appendix, under the current assumption that we have correctly specified a model for the mediator density $f_{M|E,X}$, the proposed sensitivity analysis technique for indirect effects does not require additional working models for $\{f_{Y|M,E,X}, f_{E|X}\}$ when $V = X$; and thus, in this particular setting, unlike the methodology of Theorem 4, the approach is completely robust to mis-specification of these latter working models.

It is helpful for practice, to briefly consider some simple functional forms for $t_\lambda(\cdot,\cdot,\cdot)$ and their associated bias. In the simple case where M is binary, we consider

$$\begin{aligned} t_{\lambda,1}(e,m,x) &= \lambda(2m-1) \qquad t_{\lambda,2}(e,m,x) = \lambda m \\ t_{\lambda,3}(e,m,x) &= \lambda(2m-1)e \qquad t_{\lambda,4}(e,m,x) = \lambda me \\ t_{\lambda,5}(e,m,x) &= \lambda(2m-1)ex_1 \qquad t_{\lambda,6}(e,m,x) = \lambda mex_1 \end{aligned}$$

where for each of the above functional forms, the parameter λ encodes the magnitude and direction of unmeasured confounding for the mediator.

For instance, the choice $t_{\lambda,1}$ implies a bias for $\mathbb{E}(Y_{1,M_0}|X)$ due to unmeasured confounding of the form

$$\lambda \times \{1 - OR_{EM|X}(X)\} \times f_{M|E,X}(0|E=1,X) \times f_{M|E,X}(1|E=0,X)$$

where $OR_{EM|X}(X)$ is the $E-M$ odds ratio association within levels of X.Thus, assuming a positive $E-M$ association across levels of X, we may conclude that for positive λ, unmeasured confounding leads to a downward bias in the naive estimate of the conditional natural direct effect by an amount given in the above display. Therefore, we can further conclude that the conditional natural indirect effect is biased upward by an amount of the same magnitude.

For unmeasured confounding as in $t_{\lambda,2}$, the induced bias is of the form

$$\lambda \times \left\{ RR_{ME|X}(X) - 1 \right\} \times f^2_{M|E,X}(1|E=0,X)$$

where $RR_{ME|X}(X)$ is the $E-M$ risk ratio association within levels of X.Thus, assuming a positive $E-M$ association across levels of X, we may conclude that for positive λ, unmeasured confounding leads to an upward bias in the naive estimate of the conditional natural direct effect by an amount given in the above display. Therefore, we can further conclude that the conditional natural indirect effect is biased downward by the same amount. The functions $t_{\lambda,3}, t_{\lambda,4}, t_{\lambda,5}$ and $t_{\lambda,6}$ model interactions with the exposure variable and a component X_1 of X, thus allowing for heterogeneity in the selection bias function. Since the functional form of t_λ is not identified from the observed data, we generally recommend reporting results for a variety of functional forms.

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 and recovers a general expression for the bias due to U. 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 partially circumvents this difficulty by concisely encoding a violation of the ignorability assumption for the mediator through the selection bias function $t_{\lambda}(e, m, x)$, although in practice, a finite dimensional model must still be used for this quantity as illustrated in the previous section. Nonetheless, it is important to note that the approach makes no reference and thus is agnostic about the existence, dimension, and nature of unmeasured confounders U. Furthermore, in the current 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. In addition, in contrast with the 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. Furthermore, their methodologies have not been developed to handle a setting in which, as we have considered, natural direct and indirect effects are sought conditional on a subset V of confounders.

Finally, we note that while in the foregoing, the support of M is assumed finite, the proposed sensitivity analysis methodology can be extended to accomodate a continuous mediator by further adapting the approach of Robins et al (1999).

4 Discussion

The main contribution of the present paper is to present a formal and yet practically relevant semiparametric framework for making inferences about conditional natural direct and indirect causal effects in the presence of a large number of confounding factors. A large class of multiply robust estimators for the parameters indexing models for the natural direct and indirect effects is derived, that can be used when as will usually be the case in practice, nonparametric estimation is not feasible. 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 future work, it will be crucial to critically examine by extensive simulation study the extent to which the proposed estimators are susceptible to a practical violation of the assumption, and we plan to explore 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 unfavorable conditions.

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

MacKinnon, D.(2008) Statistical Mediation Analysis. Psychology Press. Taylor and Francis Group.

Muller, D., Judd C and Yzerbyt, V. (2005) When Moderation is Mediated and Mediation is Moderated. Journal of Personality and Psychology. Vol 89 No 6, pp. 852-863.

Newey, W.K. and McFadden, D. (1993) Estimation in large samples. Handbook of Econometrics, Vol. 4. Eds. McFadden, D., Engler, R. Amsterdam: North Holland.

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.

Tchetgen Tchetgen EJ, Shpitser I. (2011) Semiparametric Theory for Causal

Mediation Analysis: efficiency bounds, multiple robustness, and sensitivity analysis. Submitted

Tan, Z. (2010) Bounded, efficient, and doubly robust estimation with inverse weighting, Biometrika, 97, 661-68

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.

Vinokur, A. D., Price, R. H., & Schul, Y. (1995). Impact of the JOBS intervention on unemployed workers varying in risk for depression. American Journal of Community Psychology, 23(1), 39–74.

Vinokur, A. D., & 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.

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 $\mathcal{M}_{\text{np}}$, with $F_{O,0} = F_O$, suppose $V \subset X$ then define $\psi_t = \psi\left(F_{O;t}\right)$ such that under the model,

$$
\begin{aligned}
&\gamma_{DIR}\left(1, v; \psi_t\right) \\
=\ &\iint_{\mathcal{S}\times\mathcal{L}} \left\{ \begin{array}{c} \mathbb{E}_t\left(Y|E=1, M=m, X=x\right) \\ -\mathbb{E}_t\left(Y|E=0, M=m, X=x\right) \end{array} \right\} f_{M|E,X;t}\left(m|E=0, X=x\right) \\
&\times f_{L|V;t}(l|v) d\mu(m,l)
\end{aligned}
$$

is our sole restriction. Let $\nabla_{t=0}$ denote differentiation wrt t at $t = 0$.Then, assuming the order of differentiation and integration is interchangeable, we obtain from the above display:

$$
\begin{aligned}
&\nabla_{t=0}\gamma_{DIR}\left(1, v; \psi_t\right) \\
=\ &\iint_{\mathcal{S}\times\mathcal{L}} \nabla_{t=0} \left[\begin{array}{c} \left\{\mathbb{E}_t\left(Y|E=1, M=m, X=x\right) - \mathbb{E}_t\left(Y|E=0, M=m, X=x\right)\right\} \\ \times f_{M|E,X;t}\left(m|E=0, X=x\right) f_{L|V;t}(l|v) \end{array} \right] d\mu(m,l)
\end{aligned}
$$

which by Theorem 1 of Tchetgen Tchetgen and Shpitser (2011) yields:

$$\nabla_{t=0}\gamma_{DIR}\left(1,V;\psi_t\right)=\mathbb{E}\left\{U\left(\psi\right)Q|V\right\}$$

where Q is the score of $F_{O;t}$ at $t=0$. This in turn implies that the nuisance tangent Λ_{nuis} space of the model $\mathcal{M}_{\text{np}}$ is given by:

$$\Lambda_{nuis}=\left\{\begin{array}{c} Q+P:Q=q\left(O\right)\ \text{ with } \mathbb{E}\left\{U\left(\psi\right)Q|V\right\}=0 \\ \text{and } P=p(V) \text{ with } \mathbb{E}\left(P\right)=0\end{array}\right\}\cap L_2(F_O)$$

Recall that the nuisance tangent space of a parametric, semiparametric or nonparametric model is defined as the $L_2\left(F_O\right)$ closure of the nuisance scores of 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. Thus we may conclude that the orthocomplement to the nuisance tangent space is given by

$$\Lambda_{nuis}^{\perp}=\left\{h\left(V\right)U\left(\psi\right):h\right\}\cap L_2(F_O)$$

proving the first part of the theorem. The result for the case $V=X$ is immediate.

To handle the log-link, we observe that under the submodel:

$$\begin{aligned}0 &= \iint\limits_{\mathcal{S}\times\mathcal{L}}\left\{\mathbb{E}_t\left(Y|E=1,M=m,X=x\right)\exp\left\{-\gamma_{DIR}\left(1,v;\psi_t\right)\right\}-\mathbb{E}_t\left(Y|E=0,M=m,X=x\right)\right\} \\ &\quad\times f_{M|E,X;t}\left(m|E=0,X=x\right)f_{L|V;t}(l|v)d\mu(m,l)\end{aligned}$$

then by the chain rule, we obtain

$$
\begin{aligned}
&\nabla_{t=0}\gamma_{DIR}(1,v;\psi_t)\,\mathbb{E}\left(Y_{0,M_0}|V=v\right)\\
=&\iint_{\mathcal{S}\times\mathcal{L}}\nabla_{t=0}\left[\begin{array}{c}\left\{\begin{array}{c}\mathbb{E}_t\left(Y|E=1,M=m,X=x\right)\exp\left\{-\gamma_{DIR}(1,v;\psi)\right\}\\ -\mathbb{E}_t\left(Y|E=0,M=m,X=x\right)\end{array}\right\}\\ \times f_{M|E,X;t}\left(m|E=0,X=x\right)f_{L|V;t}(l|v)\end{array}\right]d\mu(m,l)
\end{aligned}
$$

which by an argument analogous to one used for the identity link, implies that

$$\nabla_{t=0}\gamma_{DIR}(1,v;\psi_t)=\mathbb{E}\left\{U_2(\psi)\,Q|V=v\right\}$$

and therefore again as before:

$$\Lambda_{nuis}^{\perp}=\left\{h(V)\,U_2(\psi):h\right\}\cap L_2(F_O)$$

proving the result. The result for the case $V=X$ is immediate

It is straightforward to verify that the efficient score $S_{\psi}^{eff,np}(\psi)=h_{opt}(V)\,U_2(\psi)$ satisfies the equation

$$\mathbb{E}\left\{h(V)\frac{\partial U_2(\psi)}{\partial\psi}\right\}=\mathbb{E}\left\{h_{opt}(V)\,U_2(\psi)\,h(V)\,U_2(\psi)\right\}\text{ for all }h$$

which establishes the semiparametric efficiency result by Theorem 5.3 of Newey and McFadden (1993).

PROOF OF THEOREM 2:

We begin by showing that

$$\mathbb{E}\{S_{\theta_0}^{eff,np}\left(h;\psi,\beta_m^*,\beta_e^*,\beta_y^*\right)\} \tag{7}$$

$$=0$$

under model $\mathcal{M}_{\text{union}}$ with $V\subset X$. First note that $\left(\beta_y^*,\beta_m^*\right)=\left(\beta_y,\beta_m\right)$ under model $\mathcal{M}_a$. Equality (7) 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 $\eta\left(1,0,X;\beta_y,\beta_m\right)=\mathbb{E}\left[\left\{\mathbb{E}^{par}\left(Y|X,M,E=1;\beta_y\right)\right\}|E=0,X\right]=\eta\left(1,0,X\right)$

$$\begin{aligned}
&\mathbb{E}\{U_1\left(\psi;\beta_m,\beta_e^*,\beta_y\right)|V\}\\
&=\mathbb{E}\left[\left.\begin{array}{c}\frac{I(E=1)f^{par}_{M|E,X}(M|E=0,X;\beta_m)}{f^{par}_{E|X}(1|X;\beta_e^*)f^{par}_{M|E,X}(M|E=1,X;\beta_m)}\\ \times\left[\overbrace{\mathbb{E}\left\{Y-\mathbb{E}^{par}\left(Y|X,M,E=1;\beta_y\right)\right]|E=1,M,X\right\}}^{=0}\right]\end{array}\right|V\right]\\
&+\mathbb{E}\left[\left.\frac{I(E=0)}{f^{par}_{E|X}(1|X;\beta_e^*)}\left(\overbrace{\mathbb{E}\left[\left\{\mathbb{E}^{par}\left(Y|X,M,E=1;\beta_y\right)-\eta\left(1,0,X;\beta_y,\beta_m\right)\right\}|E=0,X\right]}^{=0}\right)\right|V\right]\\
&-\mathbb{E}\left[\left.\frac{I(E=0)}{f^{par}_{E|X}(1|X;\beta_e^*)}\left(\overbrace{\mathbb{E}\left[\left\{\mathbb{E}\left(Y|X,M,E=0\right)-\eta\left(0,0,X;\beta_y,\beta_m\right)\right\}|E=0,X\right]}^{=0}\right)\right|V\right]\\
&+\mathbb{E}\left\{\eta\left(1,0,X;\beta_y,\beta_m\right)-\eta\left(0,0,X;\beta_y,\beta_m\right)-\gamma_{DIR}\left(1,V;\psi\right)|V\right\}\\
&=0
\end{aligned}$$

Second, $\left(\beta_y^*,\beta_e^*\right)=\left(\beta_y,\beta_e\right)$ under model $\mathcal{M}_b$. Equality (7) 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 $f^{par}_{E|X}(1|X;\beta_e)=f_{E|X}(1|X)$:

$$
\begin{aligned}
&\mathbb{E}\{U_1\left(\psi;\beta_m^*,\beta_e,\beta_y\right)\,\big|\,V\}\\
=\;&\mathbb{E}\left[\begin{array}{c}\frac{I(E=1)f^{par}_{M|E,X}(M|E=0,X;\beta_m^*)}{f^{par}_{E|X}(1|X;\beta_e)f^{par}_{M|E,X}(M|E=1,X;\beta_m^*)}\\ \times\overbrace{\mathbb{E}\left\{Y-\mathbb{E}^{par}\left(Y|X,M,E=1;\beta_y\right)]|E=1,M,X\right\}}^{=0}\end{array}\right|V\Bigg]\\
&+\mathbb{E}\left[\frac{I(E=0)}{f^{par}_{E|X}(1|X;\beta_e)}\mathbb{E}\left[\left\{\mathbb{E}^{par}\left(Y|X,M,E=1;\beta_y\right)-\eta\left(1,0,X;\beta_y,\beta_m^*\right)\right\}|E=0,X\right]\right|V\Bigg]\\
&-\mathbb{E}\left[\frac{I(E=0)}{f^{par}_{E|X}(1|X;\beta_e)}\left(\mathbb{E}\left[\left\{\mathbb{E}\left(Y|X,M,E=0\right)-\eta\left(0,0,X;\beta_y,\beta_m^*\right)\right\}|E=0,X\right]\right)\right|V\Bigg]\\
&+\mathbb{E}\left\{\eta\left(1,0,X;\beta_y,\beta_m^*\right)-\eta\left(0,0,X;\beta_y,\beta_m^*\right)-\gamma_{DIR}\left(1,V\;;\psi\right)|V\right\}\\
=\;&\mathbb{E}\left[\begin{array}{c}\mathbb{E}\left[\left\{\mathbb{E}^{par}\left(Y|X,M,E=1;\beta_y\right)-\mathbb{E}\left(Y|X,M,E=0\right)\right\}|E=0,X\right]\\ -\gamma_{DIR}\left(1,V\;;\psi\right)\end{array}\right|V\Bigg]=0
\end{aligned}
$$

Third, equality (7) holds under model $\mathcal{M}_c$ because

$$
\begin{aligned}
& \mathbb{E}\{U_1\left(\psi;\beta_m,\beta_e,\beta_y^*\right)\big| V\ \} \\
= \ & \mathbb{E}\left[\frac{I\left\{E=1\right\} f^{par}_{M|E,X}\left(M|E=0,X;\beta_m\right)}{f^{par}_{E|X}(1|X;\beta_e) f^{par}_{M|E,X}\left(M|E=1,X;\beta_m\right)}\mathbb{E}\left\{Y-\mathbb{E}^{par}\left(Y|X,M,E=1;\beta_y^*\right)\right]\right\}\Bigg| V\Bigg] \\
& +\mathbb{E}\left[\frac{I(E=0)}{f^{par}_{E|X}(1|X;\beta_e)}\mathbb{E}\left[\left\{\mathbb{E}^{par}\left(Y|X,M,E=1;\beta_y^*\right)-\eta\left(1,0,X;\beta_y^*,\beta_m\right)\right\}|E=0,X\right]\Bigg| V\right] \\
& -\mathbb{E}\left[\frac{I(E=0)}{f^{par}_{E|X}(1|X;\beta_e)}\left(\mathbb{E}\left[\left\{\mathbb{E}\left(Y|X,M,E=0\right)-\eta\left(0,0,X;\beta_y^*,\beta_m\right)\right\}|E=0,X\right]\right)\Bigg| V\right] \\
& +\mathbb{E}\left[\eta\left(1,0,X;\beta_y^*,\beta_m\right)-\eta\left(0,0,X;\beta_y^*,\beta_m\right)-\gamma_{DIR}\left(1,V\ ;\psi\right)\big| V\right] \\
= \ & \mathbb{E}\left[\mathbb{E}\left[\left\{\mathbb{E}\left(Y|X,M,E=1\right)\right\}|E=0,X\right]|V\right]-\mathbb{E}\left[\mathbb{E}\left[\mathbb{E}^{par}\left(Y|X,M,E=1;\beta_y^*\right)|E=0,X\right]|V\right] \\
& +\mathbb{E}\left[\mathbb{E}\left[\mathbb{E}^{par}\left(Y|X,M,E=1;\beta_y^*\right)|E=0,X\right]|V\right]-\mathbb{E}\left[\eta\left(1,0,X;\beta_y^*,\beta_m\right)|V\right] \\
& -\mathbb{E}\left[\ \left(\mathbb{E}\left[\left\{\mathbb{E}\left(Y|X,M,E=0\right)-\eta\left(0,0,X;\beta_y^*,\beta_m\right)\right\}|E=0,X\right]\right)|V\right] \\
& +\mathbb{E}\left[\eta\left(1,0,X;\beta_y^*,\beta_m\right)-\eta\left(0,0,X;\beta_y^*,\beta_m\right)-\gamma_{DIR}\left(1,V\ ;\psi\right)|V\right] \\
= \ & \mathbb{E}\left[\mathbb{E}\left[\left\{\mathbb{E}\left(Y|X,M,E=1\right)-\mathbb{E}\left(Y|X,M,E=0\right)\right\}|E=0,X\right]-\gamma_{DIR}\left(1,V\ ;\psi\right)|V\right] \\
= \ & 0
\end{aligned}
$$

Now suppose that $V = X$, then assuming $\mathbb{E}^{par}\left(Y|X, M, E=1; \beta_y\right)$ is correct,

$$
\begin{aligned}
&\mathbb{E}\{U_1\left(\psi; \beta_m^*, \beta_e^*, \beta_y\right)|X\} \\
&= \mathbb{E}\left[\left.\begin{array}{c} \frac{I\{E=1\} f^{par}_{M|E,X}(M|E=0,X;\beta_m)}{f^{par}_{E|X}(1|X;\beta_e^*) f^{par}_{M|E,X}(M|E=1,X;\beta_m^*)} \\ \times\left[\overbrace{\mathbb{E}\left\{Y - \mathbb{E}^{par}\left(Y|X, M, E=1; \beta_y\right)\right]|E=1, M, X\right\}}^{=0}\right] \end{array}\right| X\right] \\
&+ \mathbb{E}\left[\left.\frac{I(E=0)}{f^{par}_{E|X}(1|X;\beta_e^*)} \overbrace{\left(\begin{array}{c} \mathbb{E}\left[\mathbb{E}^{par}\left(Y|X, M, E=1; \beta_y\right) - \mathbb{E}\left(Y|X, M, E=0\right)|E=0, X\right] \\ -\gamma_{DIR}\left(1, X\ ; \psi\right) \end{array}\right)}^{=0}\right| X\right] \\
&= 0
\end{aligned}
$$

Next, equality (7) holds under model $\mathcal{M}_c$ because

$$
\begin{aligned}
&\mathbb{E}\{U_1\left(\psi; \beta_m, \beta_e, \beta_y^*\right)|X\} \\
&= \mathbb{E}\left[\left.\begin{array}{c} \frac{I\{E=1\} f^{par}_{M|E,X}(M|E=0,X;\beta_m)}{f^{par}_{E|X}(1|X;\beta_e) f^{par}_{M|E,X}(M|E=1,X;\beta_m)} \\ \times\left[\mathbb{E}\left\{Y - \mathbb{E}^{par}\left(Y|X, M, E=1; \beta_y^*\right)\right]|E=1, M, X\right\}\right] \end{array}\right| X\right] \\
&+ \mathbb{E}\left[\left.\frac{I(E=0)}{f^{par}_{E|X}(1|X;\beta_e)} \left(\begin{array}{c} \mathbb{E}\left[\mathbb{E}^{par}\left(Y|X, M, E=1; \beta_y^*\right) - \mathbb{E}\left(Y|X, M, E=0\right)|E=0, X\right] \\ -\gamma_{DIR}\left(1, X\ ; \psi\right) \end{array}\right)\right| X\right] \\
&= \mathbb{E}\left[\left. \mathbb{E}\left[\mathbb{E}\left(Y|X, M, E=1\right) - \mathbb{E}\left(Y|X, M, E=0\right)|E=0, X\right]\right| X\right] - \gamma_{DIR}\left(1, X\ ; \psi\right) = 0
\end{aligned}
$$

It is straightforward to verify the unbiasedness property holds for the log-link upon noting that

$$T_2\left(\psi;\beta_m^*,\beta_y^*,\beta_e^*\right) = \frac{I\left\{E=1\right\} f_{M|E,X}^{par}\left(M|E=0,X;\beta_m^*\right)}{f_{E|X}^{par}(1|X;\beta_e^*) f_{M|E,X}^{par}\left(M|E=1,X;\beta_m^*\right)}\left\{Y-\mathbb{E}^{par}\left(Y|X,M,E=1;\beta_y^*\right)]\right\}$$
$$+\ \frac{I(E=0)}{f_{E|X}^{par}(1|X;\beta_e^*)}\left\{\mathbb{E}^{par}\left(Y|X,M,E=1;\beta_y^*\right)-\eta\left(1,0,X;\beta_m^*,\beta_y^*\right)\right\}$$
$$+\ \eta\left(1,0,X;\beta_m^*,\beta_y^*\right)$$

is triply robust for $E(Y_{1M_0}|V)$ and therefore $\mathbb{E}\left[T_2\left(\psi;\beta_m^*,\beta_y^*,\beta_e^*\right)|V\right]\exp\left\{-\gamma_{DIR}\left(1,V;\ \psi\right)\right\} = E(Y_{0M_0}|V)$ so that

$$\begin{aligned} U_2\left(\psi;\beta_m^*,\beta_y^*,\beta_e^*\right) &= T_2\left(\psi;\beta_m^*,\beta_y^*,\beta_e^*\right)\exp\left\{-\gamma_{DIR}\left(1,V;\ \psi\right)\right\} \\ &\quad - \frac{I(E=0)}{f_{E|X}(0|X,\beta_e^*)}\left\{Y-\eta\left(0,0,X;\beta_m^*,\beta_y^*\right)\right\}-\eta\left(0,0,X;\beta_m^*,\beta_y^*\right) \end{aligned}$$

has mean zero under model $\mathcal{M}_{union}^{abc}$ when $V\subset X$. The same approach gives the result under model $\mathcal{M}_{union}^{yc}$ when $V=X$, details are omitted.

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

$$\sqrt{n}(\widehat{\psi}-\psi) = \frac{1}{n^{1/2}}\sum_{i=1}^{n}\mathbb{E}\left\{\frac{\partial S_{\psi,i}^{union}\left(h;\psi,\beta^*\right)}{\partial\psi}\right\}^{-1} S_{\psi,i}^{union}\left(h;\psi,\beta^*\right)+o_p(1) \qquad (8)$$

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

At the intersection submodel

$$\frac{\partial \mathbb{E}\left\{S_{\psi}^{eff,np}\left(h_{opt};\psi,\beta\ \right)\right\}}{\partial \beta^{T}}=0$$

hence

$$S_{\psi}^{union}\left(h;\psi,\beta\right)=S_{\psi}^{np}\left(h;\psi,\beta^{*}\right).$$

The semiparametric efficiency claim then follows for $\widehat{\psi}\left(\widehat{h}_{opt}\right)$.

PROOF OF THEOREM 3:

For the identity link, note that

$$\begin{aligned}
\gamma_{IND;t}\left(1,V\ \right) &= \mathbb{E}_{t}\left(Y_{1,M_{e}}|V\right)-\mathbb{E}_{t}\left(Y_{1,M_{0}}|V\right)\\
&= \mathbb{E}_{t}\left(Y_{1}|V\right)-\mathbb{E}_{t}\left(Y_{0}|V\right)\\
&\quad +\mathbb{E}_{t}\left(Y_{0,M_{0}}|V\right)-\mathbb{E}_{t}\left(Y_{1,M_{0}}|V\right)
\end{aligned}$$

Robins et al (1994) established

$$\begin{aligned}
&\nabla_{t=0}\left\{\mathbb{E}_{t}\left(Y_{1}|V\right)-\mathbb{E}_{t}\left(Y_{0}|V\right)\right\}\\
=\ &\mathbb{E}\left\{\left(R_{0}-R_{1}\right)Q|V\right\}
\end{aligned}$$

where recall Q is the score of $F_{O;t}$ at $t=0$, and

$$R_{e}=\frac{I(E=e)}{f_{E|X}(e|X)}\left\{Y-\mathbb{E}\left(Y|E=e,X\ \right)\right\}+\mathbb{E}\left(Y|E=e,X\ \right)-\mathbb{E}\left(Y_{e}|V\right)$$

thus, by Theorem 1, the nuisance tangent space of the model $\mathcal{M}_{\mathrm{np}}$ is given by:

$$\Lambda_{nuis}=\left\{\begin{array}{c} Q+P:Q=q\left(O\right)\ \text{ with } \mathbb{E}\left\{\left[R_{0}-R_{1}-U\right]Q|V\right\}=0\\ \text{and } P=p(V) \text{ with } \mathbb{E}\left(P\right)=0\end{array}\right\}\cap L_{2}(F_{O})$$

where U is defined as $U(\psi)$ but with $\gamma_{DIR}(1, V\ )$ replacing $\gamma_{DIR}(1, V\ ;\psi)$. Thus we may conclude that the orthocomplement to the nuisance tangent space is given by

$$\Lambda_{nuis}^{\perp} = \{h(V)[R_0 - R_1 - U] : h\} \cap L_2(F_O)$$

which gives the result.

For the log link, $\mathbb{E}_t(Y_{1,M_1}|V)\exp\{-\gamma_{IND}(1, V\ ;\theta_t)\} = \mathbb{E}_t(Y_{1,M_0}|V)$, thus

$$\nabla_{t=0}[\mathbb{E}_t(Y_{1,M_1}|V)\exp\{-\gamma_{IND}(1, V\ ;\theta_t)\} - \mathbb{E}_t(Y_{1,M_0}|V)] = 0$$

therefore by Theorem 1 of Tchetgen Tchetgen and Shpitser (2011), one can show that:

$$\Lambda_{nuis} = \left\{ \begin{array}{c} Q + P : Q = q(O) \\ \text{with } \mathbb{E}\{[R_1\exp\{-\gamma_{IND}(1, V\ ;\theta_t)\} - U]\,Q|V\} = 0 \\ \text{and } P = p(V) \text{ with } \mathbb{E}(P) = 0 \end{array} \right\} \cap L_2(F_O)$$

Thus we may conclude that the orthocomplement to the nuisance tangent space is given by

$$\Lambda_{nuis}^{\perp} = \{h(V)[R_1\exp\{-\gamma_{IND}(1, V\ ;\theta_t)\} - U] : h\} \cap L_2(F_O)$$

which gives the result.

PROOF OF THEOREM 4:

Note that by Theorem 4 of Tchetgen Tchetgen and Shpitser (2011)

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

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

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

which together with the ignorability assumption of the exposure and the fact that

$$\begin{aligned}
& \mathbb{E}\left[\left.\frac{I\left\{E=0\right\}}{f^{par}_{E|X}(0|X;\beta_e^*)}\left\{Y-\eta^{par}\left(0,0,X;\beta_y,\beta_m\right)\right]\right\}+\eta^{par}\left(0,0,X;\beta_y,\beta_m\right)\right|V\right] \\
= & \mathbb{E}_t\left(Y_{0,M_0}|V\right)
\end{aligned}$$

implies unbiasedness of the estimating function for ψ.

Second, note that $(\beta_m^*, \beta_e^*) = (\beta_m, \beta_e)$ under model $\mathcal{M}_c$, and thus

$$\mathbb{E}\left[\left.\begin{array}{c} \frac{I\{E=1\} f^{par}_{M|E,X}(M|E=0,X;\beta_m)}{f^{par}_{E|X}(1|X;\beta_e) f^{par}_{M|E,X}(M|E=1,X;\beta_m)} \left\{Y - \mathbb{E}^{par}\left(Y|X,M,E=1;\beta_y^*\right)]\right\} \\ +\eta^{par}\left(1,0,X;\lambda^*,\beta_y^*,\beta_m\right) \end{array}\right| V\right]$$

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

which together with the fact that

$$\begin{aligned}
& \mathbb{E}\left[\frac{I\left(E=0\ \right)}{f_{E|X}^{par}(0|X;\beta_e)}\left\{Y-\eta^{par}\left(0,0,X;\beta_y^*,\beta_m\right)\right]\right\}+\eta^{par}\left(0,0,X;\beta_y^*,\beta_m\right)\Bigg|V\Bigg] \\
= & \ \mathbb{E}\left(Y_{0,M_0}|V\right)
\end{aligned}$$

establishes double robustness.

Define $Q\left(\psi;\beta_m,\beta_e,\beta_y,\lambda^*\right)$ as $\widehat{S}_{\psi}^{doubly}\left(h;\psi,\lambda^*\right)$ evaluated at $\left(\beta_m,\beta_e,\beta_y\right)$ instead of $\left(\widehat{\beta}_m,\widehat{\beta}_e,\widehat{\beta}_y\right)$.The asymptotic distribution of $\widehat{\psi}^{doubly}\left(\lambda^*\right)$ for fixed λ^* is obtained as in Theorem 2 upon replacing $S_{\psi}^{np}\left(\psi;\beta_m,\beta_e,\beta_y\right)$ with $Q\left(\psi;\beta_m,\beta_e,\beta_y,\lambda^*\right)$.

ADDITIONAL SENSITIVITY ANALYSIS METHODOLOGY

Direct effect, $V=X$, **identity link:**

Although details are omitted, Theorem 4 can be shown to hold when $V=X$, provided $\widehat{\psi}^{doubly}\left(\lambda^*\right)$ solves

$$\mathbb{P}_n\left\{h\left(V\right)\widehat{U}_1\left(\widehat{\psi}^{doubly}\left(\lambda^*\right),\lambda^*\right)\right\}=0$$

where

$$\begin{aligned}
\widehat{U}_1\left(\psi,\lambda^*\right)= & \ \frac{I\left(E=1\right)\widehat{f}_{M|E,X}^{par}\left(M|E=0,X\right)}{\widehat{f}_{E|X}^{par}(1|X)\widehat{f}_{M|E,X}^{par}\left(M|E=1,X\right)}\left\{\widehat{Y}\left(\lambda^*\right)-\gamma_{DIR}\left(1,X\ ;\psi\right)-\widehat{\eta}^{par}\left(0,0,X\right)\right\} \\
& -\frac{I(E=0)}{\widehat{f}_{E|X}^{par}(0|X)}\left\{Y-\widehat{\eta}^{par}\left(0,0,X\right)\right\}
\end{aligned}$$

and $\widehat{Y}\left(\lambda^*\right)=Y+t_{\lambda^*}\left(0,M,X\right)\left\{1-\widehat{f}_{M|E,X}^{par}\left(M|E=0,X\right)\right\}$

$$-t_{\lambda^*}\left(1,m,X\right)\left\{1-\widehat{f}_{M|E,X}^{par}\left(m|E=1,X\right)\right\}$$

Direct effect, $V \subset X$,log link:

For g the log link, one can similarly show the theorem holds provided $\widehat{\psi}^{doubly}(\lambda^*)$ solves

$$\mathbb{P}_n\left\{h(V)\,h(V)\,\widehat{U}_2\left(\widehat{\psi}^{doubly}(\lambda^*),\lambda^*\right)\right\}$$

with: when $V \subset X$

$$\begin{aligned}\widehat{U}_2(\psi,\lambda^*) = &\frac{I(E=1)\,\widehat{f}^{par}_{M|E,X}(M|E=0,X)}{\widehat{f}^{par}_{E|X}(1|X)\widehat{f}^{par}_{M|E,X}(M|E=1,X)}\left\{Y-\widehat{\mathbb{E}}^{par}(Y|X,M,E=1)]\right\}\exp\{-\gamma_{DIR}(1,V;\psi\ )\}\\ &-\ \frac{I(E=0)}{f_{E|X}(0|X)}\{Y-\eta(0,0,X)\}\\ &+\ \widetilde{\eta}^{par}(1,0,X;\lambda^*)\exp\{-\gamma_{DIR}(1,V;\psi\ )\}-\eta(0,0,X)\end{aligned}$$

Direct effect, $V = X$, log link:

$$\begin{aligned}\widehat{U}_2(\psi,\lambda^*) = &\frac{I(E=1)\,\widehat{f}^{par}_{M|E,X}(M|E=0,X)}{\widehat{f}^{par}_{E|X}(1|X)\widehat{f}^{par}_{M|E,X}(M|E=1,X)}\left[\widehat{Y}(\lambda^*)\exp\{-\gamma_{DIR}(1,V;\psi\ )\}-\widehat{\eta}^{par}(0,0,X)\right]\\ &-\ \frac{I(E=0)}{\widehat{f}^{par}_{E|X}(0|X)}\{Y-\widehat{\eta}^{par}(0,0,X)\}\end{aligned}$$

Indirect effect, $V \subset X$, identity link:

Next, we briefly describe an analogous double-robust sensitivity analysis technique for indirect effects. Specifically, for fixed λ^*,under the assumptions of Theorem 4, we propose to construct a doubly robust estimator $\widehat{\theta}^{doubly}(\lambda^*)$ of θ for $V \subset X$ and the identity link, by using an empirical version of the estimating function $h(V)\,W_1^{doubly}(\theta)$:

$$W_1^{doubly}(\theta;\lambda^*) = \frac{I(E=1)}{f_{E|X}(1|X)} \left\{ \begin{array}{c} Y - \eta(1,1,X) \\ -\frac{f_{M|E,X}(M|E=0,X)}{f_{M|E,X}(M|E=1,X)} \{Y - \mathbb{E}(Y|X,M,E=1)]\} \end{array} \right\}$$

$$+ \eta(1,1,X) - \eta(1,0,X;\lambda^*) - \gamma_{IND}(1,V;\theta)$$

with

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

Indirect effect, $V = X$, identity link: we propose to use the estimating function $h(V)\, W_1^{doubly}(\theta)$, where:

$$W_1^{robust}(\theta;\lambda^*) = I(E=1)\left\{Y \; - \frac{f_{M|E,X}(M|E=0,X)}{f_{M|E,X}(M|E=1,X)} \{Y(\lambda^*) + \gamma_{IND}(1,X;\theta)\}\right\}$$

and

$$\begin{aligned} Y(\lambda^*) &= Y + t_{\lambda^*}(0,M,X)\left(1 - f_{M|E,X}(M|E=0,X)\right) \\ &\quad - t_{\lambda^*}(1,m,X)\left(1 - f_{M|E,X}(m|E=1,X)\right) \end{aligned}$$

We note here that under the current assumption that we have correctly specified a model for the mediator density $f_{M|E,X}$, unlike previous settings, $W_1^{robust}(\theta;\lambda^*)$ evaluated under $\widehat{f}_{M|E,X}$ is guaranteed to yield a consistent estimator of θ without requiring an additional working model for $\left\{f_{Y|M,E,X}, f_{E|X}\right\}$;therefore, it is completely robust to mis-specification of the latter.

Indirect effect, $V \subset X$**, log link:**

Similarly, for g the log-link, we propose to construct a doubly robust estimator $\widehat{\theta}^{doubly}(\lambda^*)$ of θ when $V \subset X$, based on the estimating function $h(V)\, W_2^{doubly}(\theta)$:

$$W_2^{doubly}(\theta;\lambda^*) = \frac{I(E=1)}{f_{E|X}(1|X)} \left\{ \begin{array}{c} Y - \eta(1,1,X) \\ -\frac{f_{M|E,X}(M|E=0,X)}{f_{M|E,X}(M|E=1,X)} \left\{ Y - \mathbb{E}(Y|X,M,E=1)] \right\} \end{array} \right\}$$
$$+ \eta(1,1,X) \exp\left\{-\gamma_{IND}(1,V;\theta)\right\} - \eta(1,0,X;\lambda^*)$$

Indirect effect, $V = X$**, log link:**

we propose the alternative estimating function $h(V)\, W_2^{robust}(\theta)$, where:

$$W_2^{robust}(\theta;\lambda^*) = I(E=1) \left\{ Y - \frac{f_{M|E,X}(M|E=0,X)}{f_{M|E,X}(M|E=1,X)} Y(\lambda^*) \exp\left\{\gamma_{IND}(1,X;\theta)\right\} \right\}$$

Interestingly, as in the case of an identity link, the estimator corresponding to the estimating function $h(X) W_1^{robust}(\theta;\lambda^*)$ evaluated under $\widehat{f}_{M|E,X}$ is guaranteed to yield a consistent estimator of θ without requiring an additional working model for $\left\{f_{Y|M,E,X}, f_{E|X}\right\}$, and therefore, it is completely robust to mis-specification of the latter.