Harvard University
Harvard University Biostatistics Working Paper Series

Year 2013 *Paper* 166

Alternative Identification and Inference for the Effect of Treatment on the Treated with an Instrumental Variable

Eric J. Tchetgen Tchetgen* Stijn Vansteelandt†

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

Alternative Identification and Inference for the Effect of Treatment on the Treated with an Instrumental Variable

Eric J Tchetgen Tchetgen[1] and Stijn Vansteelandt[2]
Departments of Biostatistics and Epidemiology, Harvard University[1]
Department of Applied Mathematics, Ghent University[2]

June 6, 2013

Abstract

This paper presents a new theoretical framework for identification and inference of the average causal effect of treatment on the treated in the presence of unobserved confounding, using an instrumental variable (IV) approach. The causal effect is shown to be nonparametrically identified under a homogeneity assumption about the nature of unobserved confounding. Specifically, the assumption, states that unobserved confounding is on average balanced across levels of the instrumental variable on the additive scale. This assumption is made more plausible by conditioning on enough baseline characteristics to render the degree of unmeasured confounding additively constant for different values of the IV. For inference, a straightforward likelihood approach is described, which generalizes a certain control function method popularized by economists. Double robust methodology is also described which presents appealing robustness properties compared to parametric maximum likelihood. Semiparametric efficiency theory is carefully studied in the special case of binary IV and exposure, and the proposed framework is shown to generalize in several directions of interest. The approach is developed in a general IV model in which a more abstract restriction can be placed on the nature of unobserved confounding to obtain identification. A sensitivity analysis is also described to relax certain assumptions needed for identification, and the approach is shown to extend to the context of complex longitudinal studies with time varying exposures, both observed and unobserved time-varying confounding, and time-varying IVs.

1 Introduction

Studies in the social and health sciences commonly aim to determine the causal effect of a point exposure. Although the double-blind randomized study design remains the gold standard for unbiased evaluation of the effects of an exposure, observational studies are often conducted for practical or ethical reasons. The main challenge with drawing causal inferences from an observational study stems from its inability as a study design, to categorically rule out the possibility that differences in outcome measures between exposed and unexposed persons, may be due to systematic background differences in selecting exposure status, that also predict the outcome. Such confounding can compromise conclusions drawn from an observational study, but may likewise operate in a randomized trial with non-compliance, where individuals may choose for reasons unknown to the investigator, not to adhere to their treatment assignment. Confounding bias is a major concern for evaluating effects of exposures in observational studies and randomized experiments with non-compliance, and the development of methodology to adequately address this issue remains a priority for several disciplines, including biostatistics, epidemiology, econometrics and sociology.

The instrumental variable (IV) approach refers to a particular set of methods that allow one to recover, under certain assumptions, a causal effect of an exposure in the presence of unmeasured confounding. The IV approach has a longstanding tradition in econometrics going back to the original work of Wright (1928) and Golderberger (1972) who initially developed the approach in the context of linear structural equation modeling, and these ideas have more recently been formalized using potential outcomes or counterfactual variables, by Angrist (1994), Robins (1994), Angrist, Imbens and Rubin (1996) and Heckman (1997). A key contribution of the counterfactual language to the IV approach is that it allows one to formally define the causal effect of interest and to clearly articulate assumptions needed to identify this effect. Fundamental to all IV methods is the key assumption that one has observed a pre-exposure unconfounded instrumental variable, known to satisfy the exclusion restriction, that the IV affects the outcome only through its effects on the exposure. A variable satisfying these assumptions can be hard to find, but if a valid IV is used, the approach is potentially unaffected by unmeasured confounding of the exposure-outcome relation, and thus may be used to reduce the evidentiary gap between observational and experimental study designs. A variety of causal estimands have been shown to be identified in the IV approach, but typically require making an additional assumption, even with a valid IV. Imbens and Angrist (1994) and Angrist, Imbens and Rubin (1996) formulate the IV approach using counterfactuals to define the effect of treatment on individuals whose treatment status can be manipulated by the IV, also known as the complier average treatment effect. They show that a monotonicity assumption about the effects of the IV on the exposure is sufficient for nonparametric identification of this particular causal effect. This strand of work has been extended in recent years by Abadie (2003), Abadie, Angrist and Imbens (2002), Carneiro, Heckman and Vytlacil (2003), Tan (2006) and further development tied to specific applications abound (Hirano, Imbens, Rubin and Zhou, 2000, Little and Yau, 1998, Barnard, Frangakis, Hill and Rubin, 2003, Frangakis et al, 2004).

In a separate strand of work, Robins (1994) formulates the IV approach using counterfactuals to define the effect of treatment on treated individuals conditional on the instrumental variable. He shows that this effect is identified assuming no heterogeneity with respect to the IV in a structural mean model, which he calls the "no current-treatment value interaction" assumption. Building on these initial results, Joffe et al (2003), Vansteelandt and Goetghebeur (2003), Robins and Rotnitzky (2004) and Tan (2010) further develop under similar identifying assumptions, the IV approach using additive, multiplicative and logistic structural nested models, also see Hernan and Robins (2006). Robins (2000), Okui et al (2012) and Tan (2010) describe doubly robust estimation in the context of structural mean models. Richardson and Robins (2010) give a detailed study of the binary IV model.

As stated before, the complier average treatment effect is nonparametrically identified with an IV by the monotonicity assumption, and therefore no additional assumption restricting the functional form of this local causal effect is necessary for inference. In contrast, the no-current treatment value interaction of Robins necessarily restricts the functional form of the treatment effect on the treated, by ruling out heterogeneity of the effect measure (either on the additive, multiplicative, or logistic scale) by the IV. In Section 3 of this paper, we extend the approach of Robins by describing an alternative assumption that delivers nonparametric identification of the average additive effects of treatment on the treated, thus allowing the causal effects to vary with the value of the IV. The key idea of our approach is to replace Robins' identifying assumption with an assumption of "homogeneous selection bias" due to unobserved confounding with respect to the observed value of the IV. The new assumption essentially restricts the nature of unobserved confounding while allowing causal effects to remain unrestricted. Specifically, homogeneous selection bias on the additive scale entails an assumption of a constant average additive association between the observed exposure and the counterfactual outcome were exposure set to its baseline value, conditional on the IV.

For inference under our identifying assumptions, we describe in Section 4.1, a semiparametric approach for a model $\mathcal{M}_{\text{np}}$ of the effect of treatment on the treated that assumes the likelihood for the observed data is otherwise unrestricted. We show that inference in this semiparametric model requires estimates of certain parts of the likelihood not directly of scientific interest. Ideally, one may wish to minimize the possibility of bias due to model mis-specification by making inferences under $\mathcal{M}_{\text{np}}$ using saturated or nonparametric models for estimating these auxiliary models. However, in order to make plausibe the assumption of homogeneous selection bias, it may be necessary to condition on a moderate to large number of baseline covariates, which in turn may render a nonparametric estimation strategy impractical for the small to moderate sample size often encountered in practice due to the curse of dimensionality. Thus, we develop a simple parametric likelihood approach in Section 4.2 easily implemented using off-the-shelf statistical software. We show that this parametric approach is closely related to a particular control function method sometimes used in econometrics, whereby the residual of the exposure regression on the IV and covariates is entered as a covariate into the regression of the outcome on the exposure and covariates, using linear models for both stages (Wooldrige, 2002). Thus, the proposed IV framework gives new justification for the control function method, which equally applies whether the exposure is binary or continuous, based on straightforward assumptions about the nature of selection bias. The proposed likelihood approach also facilitates an evaluation of whether the IV approach is strictly necessary, i.e. whether unmeasured confounding is present, and the methodology is extended in a sensitivity analysis to assess the extent to which a violation of various identifying assumptions might impact inference.

In addition, we develop a framework for inference which offers a compromise between a fully parametric likelihood approach and a fully nonparametric approach. In Section 4.4, we develop using semiparametric theory, a large class of doubly robust estimating equations for the effect of treatment on the treated. The corresponding estimators remain regular and asymptotically linear (RAL), assuming the model for the treatment effect on the treated is correctly specified, a model for the treatment propensity score conditional on the IV and covariates is correctly specified and at least one of two working models involving different parts of the likelihood for the observed data is correctly specified. We show that a generalization of Robins g-estimation is recovered as a special case. We also characterize the semiparametric efficiency bound for model $\mathcal{M}_{\text{np}}$ and describe for the case of binary exposure and IV, a doubly robust estimator that achieves the efficiency bound of $\mathcal{M}_{\text{np}}$ at the intersection of the union model where all working models are correct. We also present a semiparametric theory of inference for a more general IV model, in which the exposure and IV may be vector valued with continuous and discrete components, and more general restrictions are imposed on the structure of unobserved confounding, for identification.

An important contribution of Robins' seminal work on structural nested models is that the approach can be used to analyze longitudinal studies with time-varying exposure and both observed and unobserved time-varying confounding, using time-varying IVs. In Section 7 of this paper, we extend our results to this more general longitudinal context.

2 Framework and Review

Suppose we observe independent and identically distributed data $\mathbf{O} = (\mathbf{C}, Z, X, Y)$, consisting of a dichotomous exposure X , a set of pre-exposure variables $(Z, \mathbf{C})$, and an outcome Y. To proceed, we require a definition of counterfactual outcomes for different hypothetical interventions. In this vein, let Y_x denote the counterfactual outcome one would observe were exposure set to x. For binary X, we define the effect of treatment on the treated within levels of z and $\mathbf{c}$ as followed:

$$\gamma(z, \mathbf{c}) = \mathbb{E}(Y_1 - Y_0 | X = 1, Z = z, \mathbf{C} = \mathbf{c}), \tag{1}$$

then, note that under the consistency assumption $Y = XY_1 + (1 - X)Y_0$, the observed conditional mean difference

$$\begin{aligned} &\mathbb{E}(Y|X = 1, Z = z, \mathbf{C} = \mathbf{c}) - \mathbb{E}(Y|X = 0, Z = z, \mathbf{C} = \mathbf{c}) \\ &= \gamma(z, \mathbf{c}) + q(z, \mathbf{c}) \end{aligned}$$

with $q(z, \mathbf{c})$ the selection bias function

$$\mathbb{E}(Y_0|X = 1, Z = z, \mathbf{C} = \mathbf{c}) - \mathbb{E}(Y_0|X = 0, Z = z, \mathbf{C} = \mathbf{c}), \tag{2}$$

which encodes on the additive scale the association between Y_0 and X, conditional on $\mathbf{C}$ and Z, reflecting the degree to which the effect of X on Y is confounded (Robins et al, 1999). Thus, a conventional approach to identify $\gamma(z, \mathbf{c})$ typically entails assuming no unmeasured confounding, or equivalently that the selection bias function is identically zero, i.e. $q(z, \mathbf{c}) = 0$.

An alternative approach is to assume that Z is a valid IV, which requires the existence of the counterfactual outcome Y_{xz} were X and Z set to x and z respectively. We focus until otherwise stated on the important setting of binary IV and exposure and we make the following assumptions.

(IV.1) Unconfounded IV-outcome relation:

$$\mathbb{E}(Y_{0z}|Z = z, \mathbf{C}) = \mathbb{E}(Y_{0z}|\mathbf{C});$$

(IV.2) Exclusion restriction:

$$\mathbb{E}(Y_{0z}|\mathbf{C}) = \mathbb{E}(Y_0|\mathbf{C});$$

(IV.3) Non-null IV-exposure relation:

$$\Pr(X = 1|Z = 1, \mathbf{C}) \neq \Pr(X = 1|Z = 0, \mathbf{C}) \text{ almost surely.}$$

Assumption (IV.1) essentially states that $\mathbf{C}$ includes all common causes of Z and Y, so that for the purpose of inferring the effects of Z on Y, Z essentially behaves as if it were randomized within levels of $\mathbf{C}$. Assumption (IV.2) states that Z has no direct effect on Y, upon setting X to its reference value, within levels of $\mathbf{C}$. Both these assumptions formally encode that upon conditioning on $\mathbf{C}$, any association between Z and Y must be due to an association between Z and X and one between X and Y. The last assumption (IV.3) essentially states that X and Z cannot be independent within levels of $\mathbf{C}$. It is noteworthy that the relation between the IV and the exposure need not be causal even after conditioning on $\mathbf{C}$, although in order for (IV.2) to hold, we require that any unmeasured common cause of X and Z must be independent of Y. These assumptions are somewhat weaker than similar assumptions required for identification of the complier average treatment effect, in that the latter requires that all common causes of Z and X be included in $\mathbf{C}$, and that monotonicity holds for the relation between Z and X.

Robins (1994) established that assumptions (IV.1)-(IV.3) solely do not identify $\gamma(z, \mathbf{c})$ and he makes the following additional assumption to identify the causal effect:

(IV.4) No current treatment value interaction:

$$\gamma\left(Z,\mathbf{C}\right)=\widetilde{\gamma}\left(\mathbf{C}\right)\text{ almost surely.}$$

Assumption (IV.4) states that the effect of treatment is constant for treated individuals with different values of the instrument, upon conditioning on covariates, and therefore it rules out the possibility of effect heterogeneity by IV status. Though the model becomes identified under assumptions (IV.1)-(IV.4), assumptions (IV.1),(IV.2) and (IV.4) should be made with caution, since unlike assumption (IV.3), these three assumptions are not empirically testable. Assumption (IV.4) can be singled out as potentially more problematic than the other three assumptions, because it places a priori restrictions on the causal effect that is the main target of inference. This is a strong assumption and no similar assumption is needed to identify the complier treatment effect.

We briefly review Robins' identification result as a way to introduce the parametrization for the IV model we shall be using throughout. Focusing on binary X and Z, we adopt Robins' re-parametrization of the outcome conditional mean function in terms of $\gamma\left(z,\mathbf{c}\right)$ and $q(z,\mathbf{c})$ (Robins et al, 1999). Under consistency,

$$\begin{aligned}&\mathbb{E}\left(Y|x,z,\mathbf{c}\right)\\&=\mathbb{E}\left(Y_x|x,z,\mathbf{c}\right)-\mathbb{E}\left(Y_0|x,z,\mathbf{c}\right)\\&+\mathbb{E}\left(Y_0|x,z,\mathbf{c}\right)-\mathbb{E}\left(Y_0|X=0,z,\mathbf{c}\right)\\&-\textstyle\sum_{x'}\left\{\mathbb{E}\left(Y_0|x',z,\mathbf{c}\right)-\mathbb{E}\left(Y_0|X=0,z,\mathbf{c}\right)\right\}\Pr(X=x'|z,\mathbf{c})+\mathbb{E}\left(Y_0|z,\mathbf{c}\right)\\&=\gamma\left(z,\mathbf{c}\right)x+q(z,\mathbf{c})\left\{x-\Pr\left(X=1|z,\mathbf{c}\right)\right\}+\mathbb{E}\left(Y_{0z}|z,\mathbf{c}\right)\end{aligned}$$

This parametrization makes clear the lack of identification without additional assumptions, since, for fixed $\mathbf{c}$, the conditional mean function identifies four parameters, but the last equation depends on six unknown parameters, two for $\gamma\left(z,\mathbf{c}\right)$, two for $q\left(z,\mathbf{c}\right)$, and two corresponding to $\mathbb{E}\left(Y_{0z}|z,\mathbf{c}\right)$. Note that $\Pr\left(X=1|z,\mathbf{c}\right)$ is identified from the density of X given $(Z,\mathbf{C})$, and therefore can be considered as known for all practical purposes. Thus, identification of $\gamma\left(z,\mathbf{c}\right)$ requires reducing the number of unknown parameters to four, which is achieved through the IV assumptions. Assumptions (IV.1) and (IV.2) imply $\mathbb{E}\left(Y_{0z}|z,\mathbf{c}\right)=\mathbb{E}\left(Y_0|\mathbf{c}\right)$ which reduces to a single unknown parameter for fixed $\mathbf{c}$, likewise, Robins' assumption (IV.4) reduces $\gamma\left(z,\mathbf{c}\right)$ to $\widetilde{\gamma}\left(\mathbf{c}\right)$, with one unknown parameter. Thus, under (IV.1)-(IV.4) one obtains:

$$\mathbb{E}\left(Y|X,Z,\mathbf{C}\right)=\widetilde{\gamma}\left(\mathbf{C}\right)X+q(Z,C)\left\{X-p\left(Z,\mathbf{C}\right)\right\}+t(\mathbf{C})$$

where $t(\mathbf{C})=\mathbb{E}\left(Y_0|\mathbf{C}\right)$, and $p\left(Z,\mathbf{C}\right)=\Pr\left(X=1|Z,\mathbf{C}\right)$. A parameter count confirms that the right-hand side of the above equation now depends exactly on 4 unknown parameters for every value of $\mathbf{C}$, and therefore is just identified given the observed conditional mean function. Note that under Robins' assumptions, the above parametrization reveals that the selection bias function $q(Z,\mathbf{C})$ remains unrestricted. Robins (1994) proposed semiparametric methods to estimate $\widetilde{\gamma}\left(\mathbf{C}\right)$ that respect this property. G-estimation (Robins, 1994, Robins and Rotnitzky, 2004, Vansteelandt and Goetghebeur, 2003) entails estimating $\widetilde{\gamma}\left(\mathbf{C}\right)$ upon noting that the latter is the unique function to satisfy

$$\mathbb{E}\left[h(\mathbf{C})\left\{X-p\left(Z,\mathbf{C}\right)\right\}\left\{Y-X\widetilde{\gamma}\left(\mathbf{C}\right)\right\}\right]=0$$

for all $h(\mathbf{C})$ with finite variance. A g-estimate of a model for $\widetilde{\gamma}\left(\mathbf{C}\right)$ is readily obtained by an empirical version of the above equation using an estimate of $p\left(Z,\mathbf{C}\right)$ and user specified $h(\mathbf{C})$. Potentially more robust and efficient estimators are obtained by evaluating an empirical version of a doubly robust g-estimating equation

$$\mathbb{E}\left[h(\mathbf{C})\left\{X-p\left(Z,\mathbf{C}\right)\right\}\left\{Y-\widetilde{\gamma}\left(\mathbf{C}\right)-t(\mathbf{C})\right\}\right]=0$$

where $t(\mathbf{C}) = \mathbb{E}[\{Y - \widetilde{\gamma}(\mathbf{C})\}|\mathbf{C})]$ can be obtained without modeling the selection bias function, by regressing $Y - \widetilde{\gamma}(\mathbf{C})$ on $\mathbf{C}$ using a standard regression technique (Robins, 2000, Okui et al, 2012). In the following Section, we present an alternative assumption to (IV.4) which delivers nonparametric identification of $\gamma(Z, \mathbf{C})$, and in subsequent Sections, we develop parametric and semiparametric estimation strategies.

3 Nonparametric Identification of SMM

Suppose that we substitute assumption (IV.4) with

(IV.4') Homogeneous Selection Bias:

$$q(Z, \mathbf{C}) = \widetilde{q}(\mathbf{C}) \text{ almost surely.}$$

Assumption (IV.4') states that the degree of unmeasured confounding is on average balanced with respect to Z conditional on $\mathbf{C}$ on the additive scale. The assumption is made plausible by including enough correlates of (X, Y) in $\mathbf{C}$, to account for any difference in the amount of unmeasured confounding (measured on the additive scale) across levels of the IV; in particular, $q(Z, \mathbf{C})$ becomes zero when $(Z, \mathbf{C})$ is sufficient to control for confounding of the effect of X on Y.

Assumption (IV.4') differs from (IV.4) primarily in that it does not place any restriction on causal effects, but instead restricts the nature of selection bias to be independent of Z on the average additive scale. Note that assumption (IV.4') does not rule out dependence of the CDF $P(Y_0 \leq y|x, z, \mathbf{c})$ on z, and that even if the assumption holds, $\mathbb{E}(Y_0|x, z, \mathbf{c})$ may still depend on z for all x and $\mathbf{c}$. We give our first result.

Lemma 1. *Under assumptions (IV.1)-(IV.3) and assumption (IV.4'), $\gamma(Z, \mathbf{C})$ is nonparametrically identified.*

Proof. *Under assumptions (IV.1)-(IV.3)*

$$\mathbb{E}(Y|x, z, \mathbf{c}) = \gamma(z, \mathbf{c})\, x + q(z, \mathbf{c})\{x - p(z, \mathbf{c})\} + \mathbb{E}(Y_0|\mathbf{c}).$$

and under (IV.4') we have

$$\mathbb{E}(Y|x, z, \mathbf{c}) = \gamma(z, \mathbf{c})\, x + \widetilde{q}(\mathbf{c})\{x - p(z, \mathbf{c})\} + t(\mathbf{c}), \tag{3}$$

so that for fixed $\mathbf{c}$, *we note that* $p(z, \mathbf{c})$ *is identified from the density of* X *given* $(Z, \mathbf{C})$, *and therefore the right hand side of the above equation has 4 unknown parameters, which are identified by the four parameters of the conditional mean function on the left hand side of the equation.*

We learn from Lemma 1 that $\gamma(Z, \mathbf{C})$ is nonparametrically identified under assumptions (IV.1)-(IV.3) and assumption (IV.4'). Note that by assumption (IV.3), we require that $p(z, \mathbf{c})$ is a nontrivial function of z. This is a standard IV assumption, and closer inspection of equation (3) further clarifies its central role. It is clear that if $p(z, \mathbf{c}) = \widetilde{p}(\mathbf{c})$ for a function $\widetilde{p}$, $\widetilde{q}(\mathbf{c})\widetilde{p}(\mathbf{c})$ would become aliased with $t(\mathbf{c})$ and $\widetilde{q}(\mathbf{c})$ and $\gamma(z, \mathbf{c})$ would likewise become aliased, thus leading to loss of identification of $\gamma(z, \mathbf{c})$.

An intuitive explanation of our identification result can be obtained by noting that under consistency and assumption (IV.4'),

$$\begin{aligned}
&\mathbb{E}(Y|x, z, \mathbf{c}) - \mathbb{E}(Y|x = 0, z, \mathbf{c}) \\
&= x\gamma(z, \mathbf{c}) + \widetilde{q}(\mathbf{c})x \\
&= xz\{\gamma(1, \mathbf{c}) - \gamma(0, \mathbf{c})\} + x\{\gamma(0, \mathbf{c}) + \widetilde{q}(\mathbf{c})\}
\end{aligned}$$

This in turn implies that unmeasured confounding biases the main effect given by the observed mean contrast

$$\mathbb{E}(Y|x=1,z=0,\mathbf{c})-\mathbb{E}(Y|x=0,z=0,\mathbf{c})=\gamma(0,\mathbf{c})+\widetilde{q}(\mathbf{c})$$

relative to the main effect of the causal contrast $\gamma(0,\mathbf{c})$, and that confounding bias for this term is equal to $\widetilde{q}(\mathbf{c})$. However, the observed additive interaction

$$\begin{aligned}&\mathbb{E}(Y|x=1,z=1,\mathbf{c})-\mathbb{E}(Y|x=0,z=1,\mathbf{c})-\mathbb{E}(Y|x=1,z=0,\mathbf{c})+\mathbb{E}(Y|x=0,z=0,\mathbf{c})\\&=\gamma(1,\mathbf{c})-\gamma(0,\mathbf{c})\end{aligned}$$

remains unbiased for the additive effect modification by the IV, of the effect of treatment on the treated within levels of $\mathbf{C}$. The foregoing explanation reveals that assumption (IV.4') paired with consistency yields nonparametric identification of the causal contrast $\gamma(1,\mathbf{c})-\gamma(0,\mathbf{c})$ which quantifies heterogeneity of the treatment effect wrt to the IV, even though the main effect $\gamma(0,\mathbf{c})$ may be confounded. We emphasize that this identification result does not require assumptions (IV.1)-(IV.3) and therefore does not rely on the assumption that Z is a valid IV. However, given that $\gamma(1,\mathbf{c})-\gamma(0,\mathbf{c})$ is identified by assumption (IV.4'), assumptions (IV.1)-(IV.3) further non-parametrically identify $\gamma(0,\mathbf{c})$.

4 Statistical Inference

4.1 Inference using nonparametric auxiliary models

In the previous section, we offered a parametrization of $\mathbb{E}(Y|x,z,\mathbf{c})$ in terms of $\gamma(z,\mathbf{c})$ and $\{\widetilde{q}(\mathbf{c}),p(z,\mathbf{c}),t(\mathbf{c})\}$ which we regard as nuisance parameters not directly of scientific interest. Ideally, one may wish to avoid having to rely on parametric models for nuisance parameters and to obtain inferences about $\gamma(\cdot,\cdot)$ under the model $\mathcal{M}_{\text{np}}$, which is defined by assumptions (IV.1), (IV.2) and (IV4') with $\{\widetilde{q}(\cdot),p(\cdot,\cdot),t(\cdot)\}$ and the joint density $f(Z,\mathbf{C})$ of $(Z,\mathbf{C})$ left unrestricted. In this vein, we obtain the following result which describes the first order information about $\gamma(\cdot,\cdot)$ in model $\mathcal{M}_{\text{np}}$, and characterizes the orthogonal complement $\Lambda^{\perp}_{\text{np,nuis}}$ to the nuisance tangent space of model $\mathcal{M}_{\text{np}}$ (Bickel et al, 1993). This subset of L_2^0 (the Hilbert space of all mean zero functions of the observed data with finite variance) contains all influence functions and therefore formally characterizes all regular and asymptotically linear (RAL) estimators of $\gamma(\cdot,\cdot)$.

Theorem 1.The orthocomplement to the nuisance tangent space of $\mathcal{M}_{\text{np}}$ *is*:

$$\Lambda^{\perp}_{\text{np,nuis}}=\{U_{\text{np}}(m,h):m(\mathbf{C}),h(\mathbf{C})\ \textit{unrestricted}\}\cap L_2^0$$

$$\begin{aligned}U_{\text{np}}(m,h)=&\{Y-\gamma(Z,\mathbf{C})X-\widetilde{q}(\mathbf{C})\{X-p(Z,\mathbf{C})\}-t(\mathbf{C})\}\\&\times\left\{\frac{h(\mathbf{C})(-1)^{X+Z}}{f(X,Z|\mathbf{C})}+m(\mathbf{C})\{Z-\Pr(Z=1|\mathbf{C})\}\right\}\\&+m(\mathbf{C})\widetilde{q}(\mathbf{C})\{X-p(Z,\mathbf{C})\}\{Z-\Pr(Z=1|\mathbf{C})\}.\end{aligned}$$

It follows from a standard result of modern semiparametric theory that any RAL estimator $\widehat{\psi}$ of the unknown parameter ψ of a model $\gamma^{\natural}(z,\mathbf{c};\psi)$ of $\gamma(z,\mathbf{c})$ under $\mathcal{M}_{\text{np}}$, satisfies

$$n^{1/2}\left(\widehat{\psi}-\psi\right)=\mathbb{E}\{\partial U_{\text{np}}(\psi;m,h)/\partial\psi\}^{-1}n^{1/2}\mathbb{P}_nU_{\text{np}}(\psi;m,h)+o_p(1),$$

for some functions m and h such that $U_{\text{np}}(\psi; m, h)$ is defined as $U_{\text{np}}(m, h) \in \Lambda^{\perp}_{\text{np,nuis}}$ evaluated at $\gamma(z, \mathbf{c}) = \gamma^{\natural}(z, \mathbf{c}; \psi)$.

Furthermore any RAL estimator of $\widehat{\psi}$ can be obtained (up to asymptotic equivalence) by solving the estimating equation

$$\mathbb{P}_n U_{\text{np}}\left(\widehat{\psi}; m, h\right) = 0,$$

Thus, to construct a RAL estimator of ψ in $\mathcal{M}_{\text{np}}$ requires a consistent estimator of the influence function $U_{\text{np}}(\psi; m, h)$ which in turn requires nonparametric estimates of $\{\widetilde{q}(\cdot), p(\cdot, \cdot), t(\cdot), g(\cdot)\}$, ensuring that inferences remain within $\mathcal{M}_{\text{np}}$, where $g(\mathbf{C}) = \Pr(Z = 1|\mathbf{C})$. Such nonparametric inferences may not be practical at the sample size usually encountered in applications, if, as we generally expect $\mathbf{C}$ includes three or more continuous components, due to the curse of dimensionality. For this reason, we abandon the foregoing strategy and explore in the following sections, three separate dimension reduction strategies that may be more appropriate for high dimensional data.

4.2 Maximum likelihood approach

Suppose we posit parametric models for $\{\widetilde{q}(\cdot), p(\cdot, \cdot), t(\cdot)\}$, say $\{\widetilde{q}(\cdot; \eta), p(\cdot, \cdot; \alpha), t(\cdot; \omega)\}$, and further suppose that we posit a parametric model $f(\cdot|\cdot; \nu)$ for the conditional density of the residual $\varepsilon(\psi, \eta, \alpha, \omega)$ given $(X, Z, \mathbf{C})$, where $\varepsilon(\psi, \eta, \alpha, \omega) = Y - \mathbb{E}(Y|X, Z, \mathbf{C}; \psi, \eta, \alpha, \omega)$. The maximum likelihood estimator $\left(\widehat{\psi}_1, \widehat{\chi}_1\right)$ of ψ and $\chi = (\eta, \alpha, \omega, \nu)$ maximizes

$$\mathbb{P}_n \{\log f(\varepsilon(\psi^*, \eta^*, \alpha^*, \omega^*) | X, Z, \mathbf{C}; \nu^*) + \log f(X|Z, \mathbf{C}; \alpha^*)\} \tag{4}$$

with respect to (ψ^*, χ^*) ,where $f(X = 1|Z, \mathbf{C}; \alpha) = p(Z, \mathbf{C}; \alpha)$. Inference about ψ can then be obtained using standard maximum likelihood theory. Although practical and easy to implement in standard software, say using PROC NLMIXED in SAS which allows one to maximize an arbitrary parametric likelihood model, the validity of the foregoing approach relies almost entirely on the assumption that the likelihood model is correctly specified. Although, the approach is somewhat less sensitive to the choice of a model for the density of the residuals. For instance, suppose that one assumes that the residuals follow a normal distribution, then one can verify that the corresponding score equations for the mean model remain mean zero provided the mean is correctly specified, even if the normality assumption does not hold exactly. To illustrate, consider the model with standard normal residuals, a constant average treatment effect on the treated $\gamma^{\natural}(z, \mathbf{c}; \psi) = \psi$, and a linear model $t(\mathbf{c}; \omega) = (1, \mathbf{c}^T)\omega$. The score equation for (ψ, η, ω) under the model is the normal equation

$$\mathbb{P}_n \left\{\left(X, \Delta(\alpha), 1, \mathbf{C}^T\right)^T \varepsilon(\psi, \eta, \alpha, \omega)\right\} = 0$$

where

$$\varepsilon(\psi, \eta, \alpha, \omega) = Y - \left[\psi X + \eta \Delta(\alpha) + (1, \mathbf{c}^T)\omega\right]$$

and $\Delta(\alpha) = X - p(Z, \mathbf{C}; \alpha)$. Assuming the mean model is correct, the normal equation has mean zero since the residual is mean zero conditional on the exposure, the IV and the covariates. The resulting approach is similar to a standard control function IV method developed in econometrics, whereby a treatment effect is estimated under a linear model, adjusting for the residual $\Delta(\alpha)$ of the regression of the exposure on the IV and the covariates. This approach is typically justified for continuous X and Z, but conditions under which it yields consistent estimates of the treatment effect have not previously been given for binary X and Z. Our proposed framework gives a formal justification for the approach for binary X. An interesting special case of the control function method is obtained using ordinary least squares (OLS) to fit in a first stage, a possibly mis-specified linear model $(\mathbf{1},z,\mathbf{c}^T)\alpha$ for $p(z,\mathbf{c};\alpha)$, which is then used to construct the first stage residual used in the second stage outcome regression. Then, it can be shown that the control function estimator of ψ reduces exactly to the two stage least squares estimator (2SLS). However, although consistent, the resulting estimator will generally be inefficient, even if all working models are correct, including the first stage linear model for the exposure. This is because the OLS fit of α does not generally produce the mle of $p(Z,\mathbf{C})$, since it fails to maximize (4), except perharps, if the first stage regression is saturated. Furthermore, the equivalence between 2SLS and the control function approach described above relies crucially on the constant treatment effect and does not hold in the presence of treatment effect heterogeneity (wrt IV value) since 2SLS may no longer apply.

4.3 Evaluating unmeasured confounding

Similar to the common control function approach prominent in econometrics, the proposed approach allows one to evaluate in a straightforward manner, under assumptions (IV.1)-(IV.3) and assumption (IV.4'), whether unmeasured confounding is present upon adjusting for $\mathbf{C}$, and therefore, whether the IV approach although valid, is strictly necessary. Specifically, a test of the assumption of no unmeasured confounding amounts to a test of the null hypothesis that $\widetilde{q}(\cdot)=0$, which is readily evaluated in the foregoing likelihood framework, either by using a score test, a Wald test or a likelihood ratio test.

4.4 Relaxing identifying assumptions

It is also possible to evaluate using a sensitivity analysis framework, the extent to which a violation of the exclusion restriction assumption (IV.2) might impact inferences about $\gamma(Z,L)$. Focusing on the likelihood framework presented above, the approach entails defining a new intercept function to replace $t(\mathbf{c})$, which allows the latter to depend explicitely on Z. Accordingly, let $t(\mathbf{c},z) = t^*(z,\mathbf{c})+t_0(\mathbf{c})$ such that $t^*(z,\mathbf{c}) = \mathbb{E}(Y_{0z}|\mathbf{c}) - \mathbb{E}(Y_{00}|\mathbf{c})$ encodes the direct effect of Z on Y upon setting X to its reference value within levels of $\mathbf{C}$, and $t_0(\mathbf{c}) = \mathbb{E}(Y_{00}|\mathbf{c})$. The function $t^*(z,\mathbf{c})$ is clearly not identified without an additional assumption, therefore we propose to proceed by obtaining inferences for fixed $t^*(z,\mathbf{c})$ upon substituting $t(z,\mathbf{c})$ for $t(\mathbf{c})$ in the likelihood approach, with $t_0(\mathbf{c})$ estimated from the data under a parametric working model. A sensitivity analysis is thus obtained by varying t^* producing inferences under various forms of violation of assumption (IV.2), in a manner akin to the sensitivity analysis technique described in Robins and Rotnitzky (2004).

It is likewise possible to implement a sensitivity analysis to evaluate the degree to which a violation of assumption (IV.4') might impact inference, by modifying the selection bias function in a manner which allows for a fixed amount of treatment by IV interaction given $\mathbf{C}$. Briefly, this may be achieved, by simply redefining the selection bias function as $q(x,z,\mathbf{c}) = xq_0(\mathbf{c})+xzq^*(\mathbf{c})$ with known interaction function $q^*(\mathbf{c})$ that one may vary to conduct a sensitivity analysis, and $q_0(\mathbf{c})$ may be estimated from the data using a working model.

4.5 Generalized IV g-estimation

In this Section, we propose a different strategy for estimating $\gamma(Z,L)$ that generalizes Robins' g-estimation and that provides an approach for making inferences under assumptions (IV.1)-(IV.3) and assumption (IV.4'), without the need to model the outcome. To proceed, consider the following estimating equation of ψ,

$$\begin{aligned} U_g(\psi;\mu,\alpha,\omega;m,h) &= \{Y-\gamma(Z,\mathbf{C};\psi)X\} \\ &\times\left\{\frac{h(\mathbf{C})(-1)^{X+Z}}{f(X,Z|\mathbf{C};\mu,\alpha)}+m(\mathbf{C})(Z-\Pr(Z=1|\mathbf{C};\mu))\right\}, \end{aligned}$$

with m and h of the same dimension as ψ and where $\Pr(Z=1|\mathbf{C};\mu)$ is a working model for the density of Z given $\mathbf{C}$. The estimating equation U_g does not depend on the outcome regression, nonetheless, assuming that $f(X,Z|\mathbf{C};\mu,\alpha)$ is correctly specified, we have that

$$\begin{aligned} &\mathbb{E}\{U_g(\psi;\mu,\alpha,\omega;m,h)\} \\ &=\mathbb{E}\Big(\mathbb{E}[\{Y-\gamma(Z,\mathbf{C};\psi)X\}|X,Z] \\ &\left\{\frac{h(\mathbf{C})(-1)^{X+Z}}{f(X,Z|\mathbf{C};\mu,\alpha)}+m(\mathbf{C})(Z-\Pr(Z=1|\mathbf{C};\mu))\right\}\Big) \\ &=\mathbb{E}\left(\frac{[\widetilde{q}(\mathbf{C})\{X-p(Z,\mathbf{C})\}+t(\mathbf{C})]h(\mathbf{C})(-1)^{X+Z}}{f(X,Z|\mathbf{C};\mu,\alpha)}\right) \\ &+\mathbb{E}\left([\widetilde{q}(\mathbf{C})\{X-p(Z,\mathbf{C})\}+t(\mathbf{C})]\{m(\mathbf{C})(Z-\Pr(Z=1|\mathbf{C};\mu))\}\right) \\ &=\mathbb{E}\left(\widetilde{q}(\mathbf{C})[-\{1-1\}-\{p(1,\mathbf{C})-p(0,\mathbf{C})\}(1-1)]\right) \\ &+\mathbb{E}\{t(\mathbf{C})(2-2)+m(\mathbf{C})(\Pr(Z=1|\mathbf{C})-\Pr(Z=1|\mathbf{C};\mu))\} \\ &=0, \end{aligned}$$

therefore U_g is an unbiased estimating equation of ψ. An empirical version of the approach produces $\widehat{\psi}_2=\widehat{\psi}_2(h,m)$ which solves

$$0=\mathbb{P}_n\left\{U_g\left(\widehat{\psi}_2;\widehat{\mu}_2,\widehat{\alpha}_2;m,h\right)\right\},$$

where $(\widehat{\mu}_2,\widehat{\alpha}_2)$ maximizes the partial likelihood

$$\mathbb{P}_n\{\log f(X|Z,\mathbf{C};\alpha^*)+\log f(Z|\mathbf{C};\mu^*)\}.$$

Standard Taylor series arguments can be used to obtain the following consistent estimator of the asymptotic variance of $\widehat{\psi}_2$, accounting for estimation of all nuisance parameters

$$\widehat{\Gamma}^{-1}\widehat{\Omega}\widehat{\Gamma}^{-1},$$

with $\widehat{\Gamma}$ a consistent estimator of

$$\Gamma=\mathbb{E}\{\partial U_g(\psi;m,h)/\partial\psi\},$$

and $\widehat{\Omega}$ a consistent estimator of

$$\Omega=\mathbb{E}\left[\left\{U_g(\psi;m,h)-\mathbb{E}\left\{U_g(\psi;m,h)\mathbf{S}^T\right\}\mathbb{E}\left(\mathbf{S}\mathbf{S}^T\right)^{-1}\mathbf{S}\right\}^{\otimes 2}\right],$$

where $\mathbf{S}$ contains the scores of (μ,α) corresponding to the log partial likelihood.

4.6 Doubly Robust Inference

At this juncture, we have two approaches for estimating $\gamma(\cdot,\cdot)$, a maximum likelihood approach that essentially relies for consistency on correct specification of models for $\{\widetilde{q}(\cdot), p(\cdot,\cdot), t(\cdot)\}$, and generalized g-estimation, which relies on correct models for $\{p(\cdot,\cdot), g(\cdot)\}$. In this Section, we consider an inference under a submodeld of the semiparametric union model $\mathcal{M}_{\text{union}}$ which assumes that $p(\cdot,\cdot)$ is unrestricted, and either $g(\cdot)$ is correctly specified or $\{\widetilde{q}(\cdot), t(\cdot)\}$ are correctly specified, but both do not necessarily hold. We construct a class of doubly robust estimators that carefully combines both strategies, and that remain consistent in $\mathcal{M}_{\text{union}}$, under a submodel for $p(\cdot,\cdot)$.

Our result is formalized in the following Theorem. Let $U_{\text{np}}(\psi^*;\mu^*,\eta^*,\alpha^*,\omega^*;m,h)$ denote $U_{\text{np}}(m,h)\in\Lambda^{\perp}_{\text{np,nuis}}$ evaluated at $(\psi^*,\mu^*,\eta^*,\alpha^*,\omega^*)$, and let $\epsilon = Y-\gamma(Z,\mathbf{C})X-t(\mathbf{C})$.

Theorem 2: Suppose that assumptions (IV.1)-(IV.3) and assumption (IV.4') hold, then

$$\mathbb{E}\{U_{\text{np}}(\psi;\mu^*,\omega^*,\eta^*,\alpha^*;m,h)\}=0$$

if either

(i) $\{\widetilde{q}(\cdot;\eta^*), p(\cdot,\cdot;\alpha^*), t(\cdot;\omega^*)\}=\{\widetilde{q}(\cdot), p(\cdot,\cdot), t(\cdot)\}$, *or*

(ii) $\{p(\cdot,\cdot;\alpha^*), g(\cdot;\mu^*)\}=\{p(\cdot,\cdot), g(\cdot)\}$

Furthermore $U_{\text{np}}(\psi;m_{\text{opt}},h_{\text{opt}})$ *is the efficient score of* ψ *in model* $\mathcal{M}_{\text{union}}$, *where*

$$h_{\text{opt}}(\mathbf{C})=\frac{\mathbb{E}(\mathbf{S}_\psi R_1|\mathbf{C})}{\mathbb{E}(R_1^2|\mathbf{C})}-\frac{\mathbb{E}(\mathbf{S}_\psi R_2|\mathbf{C})}{\mathbb{E}(R_2^2|\mathbf{C})}\frac{\mathbb{E}\left[\varepsilon^2\frac{(-1)^{X+Z}}{f(X,Z|\mathbf{C})}\{Z-g(\mathbf{C})\}|\mathbf{C}\right]}{\mathbb{E}\left\{\frac{\varepsilon^2}{f(X,Z|\mathbf{C})^2}|\mathbf{C}\right\}},$$

$$m_{\text{opt}}(\mathbf{C})=\frac{\mathbb{E}(\mathbf{S}_\psi R_2|\mathbf{C})}{\mathbb{E}(R_2^2|\mathbf{C})},$$

$$R_1=\left\{\frac{(-1)^{X+Z}}{f(X,Z|\mathbf{C})}\varepsilon\right\}$$

$$R_2=\left\{\{Z-g(\mathbf{C})\}\epsilon-\frac{\mathbb{E}\left[\varepsilon^2\frac{(-1)^{X+Z}}{f(X,Z|\mathbf{C})}\{Z-g(\mathbf{C})\}|\mathbf{C}\right]}{\mathbb{E}\left\{\frac{\varepsilon^2}{f(X,Z|\mathbf{C})^2}|\mathbf{C}\right\}}\frac{(-1)^{X+Z}\varepsilon}{f(X,Z|\mathbf{C})}\right\}$$

$$\epsilon=\{Y-\gamma(Z,\mathbf{C})X-t(\mathbf{C})\}$$

and $\mathbf{S}_\psi$ *is the score for* ψ.

For doubly robust inference, we propose to use $\widehat{\psi}_3=\widehat{\psi}_3(h,m)$ which solves the estimating equation:

$$\mathbb{P}_n\left\{U_{\text{np}}\left(\widehat{\psi}_3;\widehat{\mu}_2,\widehat{\omega}_1\left(\widehat{\psi}_3,\widehat{\alpha}_2\right),\widehat{\eta}_1\left(\widehat{\psi}_3,\widehat{\alpha}_2\right),\widehat{\alpha}_2;m,h\right)\right\}=0,$$

where $(\widehat{\omega}_1(\psi^*,\alpha_2^*),\widehat{\eta}_1(\psi^*,\alpha_2^*))$ is the mle of (ω,η) for fixed (ψ^*,α_2^*). Then, $\widehat{\psi}_3$ is consistent if condition (i) of Theorem 2 holds, since this implies that $\left\{\widehat{\omega}_1\left(\widehat{\psi}_3,\widehat{\alpha}_2\right),\widehat{\eta}_1\left(\widehat{\psi}_3,\widehat{\alpha}_2\right),\widehat{\alpha}_2\right\}$ is consistent. Likewise $\widehat{\psi}_3$ is consistent, if condition (ii) holds, since then $\{\widehat{\mu}_2,\widehat{\alpha}_2\}$ is consistent, and either condition (i) or (ii) is sufficient for the resulting estimating equation to be asymptotically unbiased. Furthermore, the estimator $\widehat{\psi}_3^{\text{opt}}=\widehat{\psi}_3\left(\widehat{h}_{\text{opt}},\widehat{m}_{\text{opt}}\right)$ achieves the semiparametric efficiency bound of $\mathcal{M}_{\text{union}}$ at the intersection submodel where all models are correctly specified, with $\left(\widehat{h}_{\text{opt}},\widehat{m}_{\text{opt}}\right)$ a consistent estimator of $(h_{\text{opt}},m_{\text{opt}})$.

Note that both (i) and (ii) of Theorem 2 require a consistent estimator of $p(\cdot,\cdot)$. This is not entirely surprising since for a doubly robust estimator to exist, there must exist a consistent estimator for the parameter in view under each of the submodels of the union model. But we have previously shown that the mle is consistent in the submodel with $\{p(\cdot,\cdot), \widetilde{q}(\cdot), t(\cdot)\}$ correctly modeled and likewise, we have also established that generalized g-estimation is consistent in the submodel with $\{p(\cdot,\cdot), g(\cdot)\}$ correctly modeled, and thus, both submodels clearly rely on a correct model for $p(\cdot,\cdot)$.

For inference, one may use a consistent estimate of the asymptotic variance of $n^{1/2}\left(\widehat{\psi}_3 - \psi\right)$, which using standard large sample arguments can be shown to be asymptotically normal with variance equal to the variance of

$$\mathbb{E}\left\{\frac{\partial V\left(\psi;\mu^{\ddagger},\omega^{\ddagger},\eta^{\ddagger},\alpha^{\ddagger};m,h\right)}{\partial\psi}|_{\psi^{\ddagger}}\right\}^{-1} V\left(\psi^{\ddagger};\mu^{\ddagger},\omega^{\ddagger},\eta^{\ddagger},\alpha^{\ddagger};m,h\right)$$

where $\left(\psi^{\ddagger},\mu^{\ddagger},\omega^{\ddagger},\eta^{\ddagger},\alpha^{\ddagger}\right)$ is the probability limit of $\left(\widehat{\psi}_3,\widehat{\mu}_2,\widehat{\omega}_1\left(\widehat{\psi}_3,\widehat{\alpha}_2\right),\widehat{\eta}_1\left(\widehat{\psi}_3,\widehat{\alpha}_2\right)\right)$, and for any value $(\psi^*,\mu^*,\omega^*,\eta^*,\alpha^*)$,

$$\begin{aligned}
&V(\psi^*;\mu^*,\omega^*,\eta^*,\alpha^*;m,h)\\
&= U_{\mathrm{np}}(\psi^*;\mu^*,\omega^*,\eta^*,\alpha^*;m,h)\\
&- \mathbb{E}\left\{\frac{\partial U_{\mathrm{np}}(\psi;\mu,\omega,\eta,\alpha;m,h)}{\partial(\mu,\omega,\eta,\alpha)}\right\}\\
&\times \mathbb{E}\left\{\frac{\partial \mathbf{S}_2(\psi,\mu,\omega,\eta,\alpha)}{\partial(\mu,\omega,\eta,\alpha)}\right\}^{-1}\mathbf{S}_2(\psi^*,\mu^*,\omega^*,\eta^*,\alpha^*)
\end{aligned}$$

with $\mathbf{S}_2(\psi^*,\mu^*,\omega^*,\eta^*,\alpha^*)$ containing the score function for $(\widehat{\mu}_2,\widehat{\alpha}_2)$ and the score function of $\left(\widehat{\omega}_1\left(\widehat{\psi}_3,\widehat{\alpha}_2\right),\widehat{\eta}_1\left(\widehat{\psi}_3,\widehat{\alpha}_2\right)\right)$ evaluated at $(\mu^*,\omega^*,\eta^*,\alpha^*)$. Alternatively, one may use the nonparametric bootstrap for inference.

5 More efficient doubly robust estimation

The union model $\mathcal{M}_{\text{union}}$ considered in the previous Section allowed the law for the exposure to remain unrestricted, and therefore the efficiency bound for the foregoing model may be suboptimal compared to that of the submodel of the union model in which the exposure density is restricted a priori. This is the case, even though for inference, the efficient score for the union model $\mathcal{M}_{\text{union}}$ was eventually evaluated assuming a correctly specified submodel for the law of the exposure. Thus, the foregoing doubly robust approach relies for consistency on correct specification of the exposure model, and yet does not fully exploit this assumption to optimize efficiency. In this section, we address this potential loss of efficiency, and we derive the efficient score for the effect of treatment on the treated in the union submodel $\mathcal{M}_{\text{union}}^{\text{sub}} \subset \mathcal{M}_{\text{union}}$ in which we assume a priori that $f(X|Z,\mathbf{C};\alpha,\tau) \in \mathcal{M}_{x,\text{sub}}$ belongs to a submodel of the nonparametric model for the exposure density, with α a finite dimensional parameter and τ an infinite dimensional parameter, and either $g(\cdot)$ is correctly specified or $\{\widetilde{q}(\cdot), t(\cdot)\}$ is correctly specified. Let $\Lambda_{x,\text{nuis}}$ denote the closed linear span of scores for all regular parametric submodels $f(X|Z,\mathbf{C};\alpha,\tau_s)$ indexed by s under $\mathcal{M}_{\text{union}}^{\text{sub}}$, derived using an individual's contribution to the log partial likelihood $\log\{f_s(X|Z,\mathbf{C})\}$, and let $\Lambda_{x,\text{np}}$ denote the set of all functions of $(X, Z, \mathbf{C})$ with conditional mean zero given $(Z, \mathbf{C})$. Furthermore, let $\Pi(B|\mathcal{D})$ denote the orthogonal projection of a given random variable B onto the subspace $\mathcal{D}$ of L_2^0.

Theorem 3: Under assumptions (IV.1)-(IV.3) and assumption (IV.4'), the set of influence functions of (ψ, α) in the semiparametric model $\mathcal{M}_{\text{union}}^{\text{sub}}$ is given by

$$F = \left\{ \begin{array}{c} U_{\text{np}}(m, h) + P(n) : m(Z, \mathbf{C}), h(\mathbf{C}) \textit{ unrestricted} \\ P(n) = n(Z, \mathbf{C})\Delta \in \Lambda_{x,\text{nuis}}^{\perp} \cap \Lambda_{x,\text{np}} \end{array} \right\}$$

where $\Delta = (X - p(Z, \mathbf{C}))$. *The efficient score of (ψ, α) for this model is* $\Pi(\mathbf{S}_{\psi,\alpha}|F)$, *where* $\mathbf{S}_{\psi,\alpha} = \left(\mathbf{S}_{\psi}^T, \mathbf{S}_{\alpha}^T\right)^T$ *is the score of (ψ, α). In the special case where $f(X|Z, \mathbf{C}; \alpha, \tau) = f(X|Z, \mathbf{C}; \alpha)$ is a parametric model, such that* $\Lambda_{x,\text{nuis}}^{\perp} \cap \Lambda_{x,\text{np}} = \Lambda_{x,\text{np}}$

$$F = \left\{ \begin{array}{c} V_{\text{np}}(m, h, n) = \varepsilon \left\{ \frac{h(\mathbf{C})(-1)^{X+Z}}{f(X,Z|\mathbf{C};\mu,\alpha)} + m(\mathbf{C})(Z - g(C)) \right\} + n(Z, \mathbf{C})\Delta : \\ m(\mathbf{C}), h(\mathbf{C}), n(Z, \mathbf{C}) \textit{ unrestricted} \end{array} \right\}$$

and the efficient score of (ψ, α) is available in closed form

$$\begin{aligned} \Pi(\mathbf{S}_{\psi,\alpha}|F) &= \Pi\left((\mathbf{S}_{\psi}, \mathbf{S}_{\alpha})^T \middle| F \right) \\ &= \left(V_{\text{np}}^{\text{eff},\psi}, V_{\text{np}}^{\text{eff},\alpha}\right)^T = V_{\text{np}}\left(\widetilde{m}_{\text{opt}}, \widetilde{h}_{\text{opt}}, \widetilde{n}_{\text{opt}}\right) \end{aligned}$$

where

$$\begin{aligned} \left(\widetilde{m}_{\text{opt}}(\mathbf{C}), \widetilde{h}_{\text{opt}}(\mathbf{C})\right)^T &= \mathbb{E}\{\mathbf{S}_{\psi,\alpha} U_1 | \mathbf{C}\} \mathbb{E}\left\{U_1 U_1^T | \mathbf{C}\right\}^{-1} U_1 \\ \widetilde{n}_{\text{opt}}(Z, \mathbf{C}) &= \frac{\mathbb{E}\{\mathbf{S}_{\psi,\alpha}\Delta | Z, \mathbf{C}\}}{\mathbb{E}\{\Delta^2 | Z, \mathbf{C}\}} \Delta \\ U_1 &= \left(\frac{(-1)^{X+Z}\varepsilon}{f(X, Z|\mathbf{C}; \mu, \alpha)}, (Z - g(C))\varepsilon \right)^T. \end{aligned}$$

In the special case where a parametric model is used to estimate the exposure law, it is straightforward to verify from Theorem 3 that the efficient score of ψ in $\mathcal{M}_{\text{union}}^{\text{sub}}$ is

$$V_{\text{np}}^{\psi,\text{eff}} - \left\{V_{\text{np}}^{\psi,\text{eff}} V_{\text{np}}^{\alpha,\text{eff}\ T}\right\} \mathbb{E}\left\{V_{\text{np}}^{\alpha,\text{eff}} V_{\text{np}}^{\alpha,\text{eff}T}\right\}^{-1} V_{\text{np}}^{\alpha,\text{eff}}$$

Furthermore, one can also verify that the efficient score function in the above display is in fact a doubly robust locally efficient estimating function of ψ.

6 Theory for more general IV model

We now present a general theory of inference about the treatment effect

$$\gamma(\mathbf{x}, \mathbf{z}, \mathbf{c}) = \mathbb{E}(Y_{\mathbf{x}} - Y_{\mathbf{0}} | \mathbf{X} = \mathbf{x}, \mathbf{Z} = \mathbf{z}, \mathbf{C} = \mathbf{c})$$

where $(\mathbf{X}, \mathbf{Z})$ are possibly vector valued, with both continuous and discrete components, and "$\mathbf{0}$" is a reference value. The consistency assumption now states that $Y = Y_{\mathbf{X},,\mathbf{Z}}$ almost surely. The IV assumptions can be restated to allow for a more general exposure and IV.

(IV.G.1) Unconfounded IV-outcome relation:

$$\mathbb{E}\left(Y_{0\mathbf{z}}|\mathbf{Z}=\mathbf{z},\mathbf{C}\right)=\mathbb{E}\left(Y_{0\mathbf{z}}|\mathbf{C}\right);$$

(IV.G.2) Exclusion restriction:

$$\mathbb{E}\left(Y_{0\mathbf{z}}|\mathbf{C}\right)=\mathbb{E}\left(Y_{0}|\mathbf{C}\right);$$

(IV.G.3) Non-null IV-exposure relation:

$$\mathbf{X}\not\perp\mathbf{Z}\mid\mathbf{C}.$$

Additionally, we assume that the selection bias function

$$q(\mathbf{x},\mathbf{z},\mathbf{c})=\mathbb{E}\left(Y_{0}|\mathbf{X}=\mathbf{x},\mathbf{Z}=\mathbf{z},\mathbf{C}=\mathbf{c}\right)-\mathbb{E}\left(Y_{0}|\mathbf{X}=\mathbf{0},\mathbf{Z}=\mathbf{z},\mathbf{C}=\mathbf{c}\right)$$

follows a known set of restrictions,

(IV.G.4) $q(\cdot,\cdot,\cdot)-\overline{q}(\cdot,\cdot)\in\Gamma_{\text{sub}}\subset\Gamma=\left\{a(\mathbf{x},\mathbf{z},\mathbf{c}):E(a(\mathbf{X},\mathbf{Z},\mathbf{C})|\mathbf{Z},\mathbf{C})=0\right\},$ where for any $q(\cdot,\mathbf{z},\mathbf{c})$, $\overline{q}(\mathbf{z},\mathbf{c})\equiv\mathbb{E}(q(\mathbf{X},\mathbf{z},\mathbf{c})|\mathbf{Z}=\mathbf{z},\mathbf{C}=\mathbf{c})$.

The following result generalizes Theorem 1, and gives the orthocomplement to the nuisance tangent space in the semiparametric model $\mathcal{M}_g$ with sole restrictions (IV.G.1)-(IV.G.4).

Theorem 4:The orthocomplement to the nuisance tangent space in $\mathcal{M}_g$ is given by

$$\left\{U_g(v,m):v\left(\mathbf{X},\mathbf{Z},\mathbf{C}\right)\in\Gamma_{\text{sub}}^{\perp}\cap\ \Gamma,\ m(X,\mathbf{C})\text{ unrestricted}\right\}\cap L_2^0,$$

where

$$\begin{aligned}U_g(v,m)&=\left\{Y-\gamma\left(\mathbf{X},\mathbf{Z},\mathbf{C}\right)-q(\mathbf{X},\mathbf{Z},\mathbf{C})+\overline{q}(\mathbf{Z},\mathbf{C})-t(\mathbf{C})\right\}\\&\times\left\{v\left(\mathbf{X},\mathbf{Z},\mathbf{C}\right)+m(\mathbf{Z},\mathbf{C})-\overline{m}(\mathbf{C})\right\}\\&+\left\{q(\mathbf{X},\mathbf{Z},\mathbf{C})-\overline{q}(\mathbf{Z},\mathbf{C})\right\}\left\{m(\mathbf{Z},\mathbf{C})-\overline{m}(\mathbf{C})\right\},\end{aligned}$$

and $\overline{m}(\mathbf{C})\equiv E(m(\mathbf{Z},\mathbf{C})|\mathbf{C})$.

Theorem 4 characterizes the set of influence functions in model $\mathcal{M}_g$ and may be used as in the previous section, to motivate doubly robust estimators. We illustrate Theorem 4 with an example. Suppose that similar to assumption (IV.4'), we assume that $q(\mathbf{x},\mathbf{z},\mathbf{c})=\widetilde{q}(\mathbf{x},\mathbf{c})$ does not depend on $\mathbf{z}$. This assumption implies that

$$\Gamma_{\text{sub}}=\left\{a(\mathbf{X},\mathbf{C})-\mathbb{E}\left\{a(\mathbf{X},\mathbf{C})|\mathbf{Z},\mathbf{C}\right\}:a(\mathbf{X},\mathbf{C})\text{ unrestricted}\right\}.$$

This further implies that

$$\Gamma_{\text{sub}}^{\perp}=\left\{v\left(\mathbf{X},\mathbf{Z},\mathbf{C}\right):\mathbb{E}\left\{v\left(\mathbf{X},\mathbf{Z},\mathbf{C}\right)|\mathbf{Z},\mathbf{C}\right\}=\mathbb{E}\left\{v\left(\mathbf{X},\mathbf{Z},\mathbf{C}\right)|\mathbf{X},\mathbf{C}\right\}=0\right\}.$$

The set of functions $\Gamma_{\text{sub}}^{\perp}\cap L_2^0$ has previously been characterized by Tchetgen Tchetgen, Robins and Rotnitzky (2010), and their characterization which we give next requires the following definition.

Definition of Admissible Independence Density: Given conditional densities $f_x^{\dagger}\left(\mathbf{X}|\mathbf{C}\right)$ *and* $f_z^{\dagger}\left(\mathbf{Z}|\mathbf{C}\right)$, *the density* $h^{\dagger}\left(\mathbf{X},\mathbf{Z}|\mathbf{C}\right)=f_x^{\dagger}\left(\mathbf{X}|\mathbf{C}\right)f_z^{\dagger}\left(\mathbf{Z}|\mathbf{C}\right)$, *that makes* $\mathbf{X}$ *and* $\mathbf{Z}$ *conditionally independent given* $\mathbf{C}$ *is an admissible independence density if the joint law of* $(\mathbf{X},\mathbf{Z})$ *given* $\mathbf{C}$ *under*

$h^\dagger(\mathbf{X},\mathbf{Z}|\mathbf{C})$ *is absolutely continuous wrt to the true law of* $(\mathbf{X},\mathbf{Z})$ *given* $\mathbf{C}$ *with probability one. Furthermore,* $E^\dagger(\cdot|\cdot,L)$ *denotes conditional expectations with respect to* $h^\dagger(\mathbf{X},\mathbf{Z}|\mathbf{C})$.

Tchetgen Tchetgen et al. (2010) established that, given an admissible independence density $h^\dagger(\mathbf{X},\mathbf{Z}|\mathbf{C})$,

$$\Gamma_{\text{sub}}^{\perp}=\left\{\frac{h^\dagger(\mathbf{X},\mathbf{Z}|\mathbf{C})}{f(\mathbf{X},\mathbf{Z}|\mathbf{C})}\left[v(\mathbf{X},\mathbf{Z},\mathbf{C})-\overline{v}^\dagger(\mathbf{X},\mathbf{Z},\mathbf{C})\right]:v(\mathbf{X},\mathbf{Z},\mathbf{C})\ \text{unrestricted}\right\},$$

where

$$\overline{v}^\dagger(\mathbf{X},\mathbf{Z},\mathbf{C})=\mathbb{E}^\dagger\{v(\mathbf{X},\mathbf{Z},\mathbf{C})|\mathbf{Z},\mathbf{C}\}+\mathbb{E}^\dagger\{v(\mathbf{X},\mathbf{Z},\mathbf{C})|\mathbf{X},\mathbf{C}\}-\mathbb{E}^\dagger\{v(\mathbf{X},\mathbf{Z},\mathbf{C})|\mathbf{C}\}.$$

In the special case where $\mathbf{X}$ and $\mathbf{Z}$ are binary, $\Gamma_{\text{sub}}^{\perp}$ is equivalently characterized as

$$\left\{\frac{h(\mathbf{C})(-1)^{X+Z}}{f(X,Z|\mathbf{C})}:h(\mathbf{C})\ \text{unrestricted}\right\},$$

which is the representation given in Theorem 1. This representation is obtained by taking $f_x^\dagger(X|\mathbf{C})=1/2$ and $f_z^\dagger(Z|\mathbf{C})=1/2$, and $v(X,Z,\mathbf{C})=16\,(X-1/2)\times(Z-1/2)$. In the special case where $\mathbf{X}$ is scalar and $\mathbf{Z}$ is binary, $\Gamma_{\text{sub}}^{\perp}$ is equivalently characterized as

$$\left\{\frac{f(X|\mathbf{C})}{f(X|Z,\mathbf{C})}\left[h(X,\mathbf{C})-\mathbb{E}\{h(X,\mathbf{C})|\mathbf{C}\}\right]\{Z-\mathbb{E}(Z|\mathbf{C})\}:h(X,\mathbf{C})\ \text{unrestricted}\right\}.$$

This representation is obtained by taking $f_x^\dagger(X|\mathbf{C})=f_x(X|\mathbf{C})$ and $f_z^\dagger(Z|\mathbf{C})=f_z(Z|\mathbf{C})$. A third parametrization is obtained by taking $f_x^\dagger(X|\mathbf{C})=f_x(X|Z=0,\mathbf{C})$ and $f_z^\dagger(Z|\mathbf{C})=f_z(Z|X=0,\mathbf{C})$, giving $\Gamma_{\text{sub}}^{\perp}=$

$$\left\{OR(X,Z|\mathbf{C})^{-1}\left[h(X,Z,\mathbf{C})-\int h(x,Z,\mathbf{C})dF_x(x|Z=0,\mathbf{C})-\int h(X,z,\mathbf{C})dF_z(z|X=0,\mathbf{C})\right.\right.$$
$$\left.\left.+\int\int h(x,z,\mathbf{C})dF_z(z|X=0,\mathbf{C})\,dF_x(x|Z=0,\mathbf{C})\right]:h(X,Z,\mathbf{C})\ \text{unrestricted}\right\}$$

where

$$OR(X,Z|\mathbf{C})=\frac{f(X,Z|\mathbf{C})f(X=0,Z=0|\mathbf{C})}{f(X,Z=0|\mathbf{C})f(X=0,Z|\mathbf{C})}$$

is the generalized odds ratio function (Tchetgen Tchetgen et al 2010).

7 Illustration

We illustrate the proposed methodology on a sample of 3010 working men aged between 24 and 34 who were part of the 1976 wave of the US National Longitudinal Survey of Young Men (NLSYM) (Card,1995). In particular, we will estimate the effect of years of education on the log of hourly wages in 1976 (Y). Following Card (1995), we use as an instrumental variable an indicator if the individual lived close to a college that offered 4 year courses in 1966 (Z). All reported analyses are adjusted for covariates ($\mathbf{C}$) years of labour market experience and its square, marital status, an indicator if the individual is black, as well as various measures of geographical location in 1966 and 1976. Twelve years of education was most common (33%) in this study and was therefore used as a reference class by defining X to be the difference between the years of education and 12.

The log of hourly wages is reasonably normally distributed with mean 6.3 (SD 0.44), and is on average 0.075 (95% CI 0.068 to 0.082) higher per extra year of education, after linear regression adjustment for years of labour market experience, marital status, race and geographical location in 1966 and 1976. The partial correlation between education and the instrumental variable is 0.066. In the remainder of this Section, we will report the results from instrumental variables analysis with 95% confidence intervals based on the nonparametric bootstrap with 1000 resamples.

The traditional two-stage least squares (2SLS) analysis, which invokes the no current treatment value interaction assumption, yields an education effect of 0.13 (95% CI 0.0031 to 0.26) on the average log of the hourly wage, corresponding with a one-year increase in education. To allow for current treatment value interaction (i.e., $\gamma(x, z, \mathbf{c}; \psi) = \psi_1 x + \psi_2 xz$) under the assumption of homogeneous selection bias (i.e., $q(x, z, \mathbf{c}; \eta) = \eta x$), we first used the control functions approach of Section 4.2, based on regressing the outcome on X, XZ, $\mathbf{C}$ and the residual from a linear model for X, given Z and $\mathbf{C}$. This yielded estimates of ψ_1 equalling 0.15 (95% CI 0.020 to 0.28) and of ψ_2 equalling -0.0055 (95% CI -0.016 to 0.0047). The maximum likelihood approach of Section 4.2 gave identical results and the doubly robust approach of Section 4.6 gave nearly identical estimates of 0.15 (95% CI 0.020 to 0.28) and -0.00048 (95% CI -0.0026 to 0.0017), respectively. Here, the doubly robust approach was based on the representation given in Section 6 with $h(X, \mathbf{C}) = (X\ X^2)$. Based on these results and the distribution of the instrumental variable by education, we can infer that under the assumption of homogenous selection bias, the average log of hourly wages in 1976 would be 0.60 (95% CI 0.078 to 1.12) higher in men who had 8 years of education, had they received 12 years of education (versus 0.53 via 2SLS). Furthermore, the average log of hourly wages in 1976 would be 0.89 (95% CI 0.11 to 1.67) lower in men who had 18 years of education, had they received 12 years of education (versus 0.79 via TSLS).

We next performed a sensitivity analysis to allow for deviations away from the assumption of homogeneous selection bias. Given the strong similarity between the different estimates, results are reported for the control functions approach only. Assuming that $q(x, z, c; \eta) = (\eta_1 x + \eta_2' x\mathbf{c})(1 + \tau z)$, this involved regressing the outcome on X, XZ, $\mathbf{C}$ and $(1 + \tau Z)$ times the residual from a linear model for X, given Z and $\mathbf{C}$, as well as interactions of this product term with $\mathbf{C}$. We each time repeated the analysis, treating τ as fixed and known, and varying it from -0.25 to 0.25. Results from this sensitivity analysis are reported in Figure 1. They reveal substantial uncertainty in the average effects of 8 and 18 years of education (versus 12 years), although the evidence for an effect remains.

8 Complex Longitudinal Studies

Next, we consider a longitudinal study with J occasions in which one observes data $\overline{\mathbf{O}} = \overline{\mathbf{O}}(J) = \left\{\overline{\mathbf{X}}(J), \overline{\mathbf{Z}}(J), \overline{\mathbf{C}}(J), Y\right\}$ where $\mathbf{H}(j) = \{\mathbf{C}(j), \mathbf{Z}(j), \mathbf{X}(j)\}$ is observed at time j, $\overline{\mathbf{C}}(j) = (\mathbf{C}(1), ..., \mathbf{C}(j))$ and $\overline{\mathbf{H}}(j-1)$ are confounders of the effects of $\{\mathbf{Z}(j), \mathbf{X}(j)\}, ... \{\mathbf{Z}(J), \mathbf{X}(J)\}$ on Y, $\mathbf{Z}(j)$ is a valid IV of the effects of $\mathbf{X}(j), ...\mathbf{X}(J)$ on Y, conditional on $\mathbf{H}(j-1)$. Similar to Robins (1994) causal effects are encoded using a structural nested model:

$$\gamma_j\left(\overline{\mathbf{h}}(j)\right) = \mathbb{E}\left\{Y_{\overline{\mathbf{x}}(j),\overline{\mathbf{0}}} - Y_{\overline{\mathbf{x}}(j-1),\overline{\mathbf{0}}} | \overline{\mathbf{h}}(j)\right\},\ j = 1, ..., J.$$

where $Y_{\overline{\mathbf{x}}(j),\overline{\mathbf{0}}}$ is the potential outcome for an individual with treatment history $\overline{\mathbf{x}}(j)$ up to time j, and the reference treatment value "0" thereafter. Therefore, $\gamma_j\left(\overline{\mathbf{h}}(j)\right)$ describes on the additive scale, the causal effect of one final blip of treatment $\mathbf{x}(j)$ at time j among individuals with observed history $\overline{\mathbf{h}}(j)$. Since this is an average contrast of counterfactuals in a given subset of the population, it amounts to a causal effect. Let $Y_{\overline{\mathbf{z}}(j),(\overline{\mathbf{x}}(j),\overline{\mathbf{0}})}$ denote the potential outcome one would observe under an intervention that sets $\overline{\mathbf{Z}}(j)$ to $\overline{\mathbf{z}}(j)$ and $\overline{\mathbf{X}}(J)$ to $(\overline{\mathbf{x}}(j),\overline{\mathbf{0}})$. Throughout, we assume that consistency holds, i.e.

$$Y = Y_{\overline{\mathbf{Z}}(J),\overline{\mathbf{X}}(J)} \text{ almost surely.}$$

For inference with an IV, we make the following additional assumptions.

(IV.L.1) Unconfounded IV-outcome relation:

$$\mathbb{E}\left\{Y_{\overline{\mathbf{z}}(j),(\overline{\mathbf{x}}(j-1),\overline{\mathbf{0}})}|\overline{\mathbf{h}}(j-1),\mathbf{z}(j),\mathbf{c}(j)\right\} = \mathbb{E}\left\{Y_{\overline{\mathbf{z}}(j),(\overline{\mathbf{x}}(j-1),\overline{\mathbf{0}})}|\overline{\mathbf{h}}(j-1),\mathbf{c}(j)\right\},\ j=1,...,J$$

(IV.L.2) Exclusion restriction:

$$\mathbb{E}\left\{Y_{\overline{\mathbf{z}}(j),(\overline{\mathbf{x}}(j-1),\overline{\mathbf{0}})}|\overline{\mathbf{h}}(j-1),\mathbf{c}(j)\right\} = \mathbb{E}\left\{Y_{\overline{\mathbf{z}}(j-1),(\overline{\mathbf{x}}(j-1),\overline{\mathbf{0}})}|\overline{\mathbf{h}}(j-1),\mathbf{c}(j)\right\},\ j=1,...,J;$$

(IV.L.3) Non-null IV-exposure relation:

$$\mathbf{X}(j') \not\perp \mathbf{Z}(j) \mid \overline{\mathbf{H}}(j-1),\mathbf{C}(j);1\leq\ j\leq j'\leq J$$

These assumptions are a natural longitudinal generalization of similar assumptions previously made for point exposure. Additionally, we assume that the selection bias function

$$q_j\left(\overline{\mathbf{X}}(j),\overline{\mathbf{Z}}(j),\overline{\mathbf{C}}(j)\right) = \mathbb{E}\left\{Y_{\overline{\mathbf{x}}(j-1),\overline{\mathbf{0}}}|\overline{\mathbf{H}}(j)\right\} - \mathbb{E}\left\{Y_{\overline{\mathbf{x}}(j-1),\overline{\mathbf{0}}}|\mathbf{X}(j)=0,\overline{\mathbf{H}}(j-1),\mathbf{Z}(j),\mathbf{C}(j)\right\},\quad j=1,...,J.$$

follows a known set of restrictions. Let

$$\overline{q}_j\left(\overline{\mathbf{X}}(j-1),\overline{\mathbf{Z}}(j),\overline{\mathbf{C}}(j)\right) = \mathbb{E}\left\{q_j\left(\overline{\mathbf{X}}(j),\overline{\mathbf{Z}}(j),\overline{\mathbf{C}}(j)\right)|\overline{\mathbf{X}}(j-1),\overline{\mathbf{Z}}(j),\overline{\mathbf{C}}(j)\right\},$$

then assume

(IV.L.4)

$$q_j(\cdot,\cdot,\cdot)-\overline{q}_j(\cdot,\cdot) \in \Gamma_{\text{sub},j} \subset \Gamma_j = \left\{a(\overline{\mathbf{H}}(j)) : \mathbb{E}\left\{a\left(\overline{\mathbf{X}}(j),\overline{\mathbf{Z}}(j),\overline{\mathbf{C}}(j)\right)|\overline{\mathbf{X}}(j-1),\overline{\mathbf{Z}}(j),\overline{\mathbf{C}}(j)\right\}=0\right\},\quad j=1,...,J.$$

The following result generalizes Theorems 1 and 4, and gives the orthocomplement to the nuisance tangent space in the semiparametric model $\mathcal{M}_L$ with sole restrictions (IV.G.1)-(IV.G.4). To proceed, first we must reparametrize the conditional mean function $\mathbb{E}\left(Y|\overline{\mathbf{H}}\right)$ in terms of the SNM, the selection bias functions and the additional functions b_j defined below encoding associations of the time varying covariates with the outcome

$$\begin{aligned}\mathbb{E}\left(Y|\overline{\mathbf{H}}\right) &= \sum_{j=1}^{J}\gamma_j\left(\overline{\mathbf{H}}(j)\right) + q_j\left(\overline{\mathbf{H}}(j)\right) - \overline{q}_j(\overline{\mathbf{H}}(j-1), \mathbf{Z}\left(j\right), \mathbf{C}\left(j\right)) \\ &\quad + b_j(\overline{\mathbf{H}}(j-1), \mathbf{C}\left(j\right)) - \overline{b}_j(\overline{\mathbf{H}}(j-1)), \\ \varepsilon &= Y - \mathbb{E}\left(Y|\overline{\mathbf{H}}\right),\end{aligned}$$

where

$$\begin{aligned}\overline{b}_1(\overline{\mathbf{H}}(0)) &= 0, \\ b_j(\overline{\mathbf{H}}(j-1), \mathbf{C}\left(j\right)) &= t_j(\overline{\mathbf{H}}(j-1), \mathbf{C}\left(j\right)) - t_j(\overline{\mathbf{H}}(j-1), \mathbf{C}\left(j\right) = 0),\ j > 1.\end{aligned}$$

Theorem 5: The orthocomplement to the nuisance tangent space in $\mathcal{M}_L$ is given by functions

$$\begin{aligned}U_L = \varepsilon\Big\{&h_1\left(\overline{\mathbf{H}}\right) + \textstyle\sum_{j=1}^{J}\left\{m_j(\overline{\mathbf{H}}(j-1), \mathbf{Z}(j), \mathbf{C}\left(j\right)) - \overline{m}_j(\overline{\mathbf{H}}(j-1), \mathbf{C}\left(j\right))\right\}\Big\} \\ &+ \textstyle\sum_{j=1}^{J}\left\{m_j(\overline{\mathbf{H}}(j-1), \mathbf{Z}(j), \mathbf{C}\left(j\right)) - \overline{m}_j(\overline{\mathbf{H}}(j-1), \mathbf{C}\left(j\right))\right\} \\ &\times\left\{q_j(\mathbf{X}\ \ (j), \overline{\mathbf{H}}(j-1), \mathbf{Z}\left(j\right), \mathbf{C}\left(j\right)) - \overline{q}_j(\overline{\mathbf{H}}(j-1), \mathbf{Z}\left(j\right), \mathbf{C}\left(j\right))\right. \\ &\left.+b_j(\overline{\mathbf{H}}(j-1), \mathbf{C}\left(j\right)) - \overline{b}_j(\overline{\mathbf{H}}(j-1))\right\}\end{aligned}$$

where

$$h_1\left(\overline{\mathbf{H}}\right) = \textstyle\sum_{j=1}^{J}\left\{h_{1j}^{*}(\overline{\mathbf{H}}(j)) - \overline{h}_{1j}^{*}(\overline{\mathbf{H}}(j-1), \mathbf{Z}\left(j\right)), \mathbf{C}\left(j\right))\right\} \in \bigcap_j\left\{\Gamma_{\text{sub},j}^{\perp} \cap \Gamma_j\right\}$$

and m_j *is unrestricted.*

Theorem 5 characterizes the set of influence functions in model $\mathcal{M}_L$ and may be used as in the previous sections, to motivate a doubly robust estimator of the parameters of a model for the SNM $\gamma_j, j = 1, ..., J$. To illustrate, suppose that $X(j)$ and $Z(j)$ are binary, and consider the assumption of homogeneous selection bias

$$q_j\left(\overline{\mathbf{x}}\left(j\right), \overline{\mathbf{z}}\left(j\right), \overline{\mathbf{c}}\left(j\right)\right) = q_j\left(\overline{\mathbf{x}}\left(j\right), \overline{\mathbf{z}}\left(j-1\right), \overline{\mathbf{c}}\left(j\right)\right)$$

Then, as in the previous section one can show that

$$\Gamma_{\text{sub},j}^{\perp} \cap\ \Gamma_j = \left\{h_1\left(\overline{\mathbf{H}}; \mathbf{w}_1\right) : \mathbf{w}_1\right\},$$

$$\begin{aligned}h_1\left(\overline{\mathbf{H}}; \mathbf{w}_1\right) &= \sum_{j=1}^{J}\frac{w_{1,j}(\overline{\mathbf{X}}\left(j-1\right), \overline{\mathbf{Z}}\left(j-1\right), \overline{\mathbf{C}}\left(j\right))\left(-1\right)^{X(j)+Z(j)}}{f(X(j), Z(j)|\overline{\mathbf{X}}\left(j-1\right), \overline{\mathbf{Z}}\left(j-1\right), \overline{\mathbf{C}}\left(j\right))}; \\ \mathbf{w}_1 &= (w_{1,1}, ..., w_{1,J})\end{aligned}$$

It is further straightforward to check that U_L has mean zero, and therefore is an unbiased estimating function for the SNM, either if it is evaluated at a correct model for $\mathbb{E}\left(Y|\overline{\mathbf{H}}\right)$, or if it is evaluated at a correct model for $\{f(X(j), Z(j)|\overline{\mathbf{X}}\left(j-1\right), \overline{\mathbf{Z}}\left(j-1\right), \overline{\mathbf{C}}\left(j\right)), j\}$, and both models do not necessarily hold. A generalization of G-estimation is readily obtained by setting m_j, q_j and b_j to zero for all j in U_L, which produces an unbiased estimating equation provided that the joint process for the exposure and the IV is consistently estimated.

9 Closing remarks

This paper presented a general theory of identification and inference for the conditional average additive effect of treatment on the treated in the presence of unobserved confounding, given an instrumental variable and baseline covariates. Although emphasis was mostly given to inference about a conditional treatment effect, the marginal effect of treatment on the treated is readily obtained by averaging the estimated conditional effect with respect to the empirical density of the IV and the covariates in the treated. Thus our methodology can be used to compute the marginal average effect of treatment on the treated with a single additional step.

An important special case arises when $\Pr(X = 0, Z = 0|\mathbf{C}) =1$. This happens, for instance, in randomized trials when there is perfect exclusion of the control group from the treatment. In such cases, the only relevant causal effect is $\mathbb{E}\left(Y_1 - Y_0|X = 1, Z = 1, \mathbf{C}\right)$, which is nonparametrically identified under assumptions (IV.1-IV.3) and therefore neither assumption (IV.4) nor assumption (IV.4') is strictly necessary, and standard g-estimation and double robust g-estimation may be used. Alternatively, since the above condition implies monotonicity, the effect of treatment on the compliers reduces to the effect of treatment on the treated (provided the IV-exposure relation is unconfounded) and thus the methodology developed by Abadie (2003) may also be used in this situation.

The paper focused on a causal effect measured on the additive scale and therefore may be most appropriate for a continuous outcome. Nonetheless, some of the methods described herein may still be appropriate even if the outcome were not continuous. For instance, generalized g-estimation may be used to estimate a causal risk difference encoding the effect of treatment on a binary outcome under assumptions (IV.1)-(IV.4'), without fitting a regression model for the outcome mean. Although, more generally, it may be preferable, particularly for efficiency reasons, to model the outcome using a nonlinear link function for dichotomous or discrete outcomes to ensure that the natural bounds of the model are respected. Whether the methodology described herein can be extended to incorporate a nonlinear link function for the outcome remains an open problem.

APPENDIX

Proof of Theorem 4: *Consider the submodel* $f_s(\mathbf{O}) = f_s(\varepsilon_s|\mathbf{X},\mathbf{Z},\mathbf{C})\, f_s(\mathbf{X},\mathbf{Z},\mathbf{C})$, with $f_0(\varepsilon_0,\mathbf{X},\mathbf{Z},\mathbf{C}) = f(\varepsilon,\mathbf{X},\mathbf{Z},\mathbf{C})$ *and the submodel only varies in the direction of nuisance parameters*

$$\{q_s(\mathbf{X},\mathbf{Z},\mathbf{C}), t_s(\mathbf{C}), f_s(\mathbf{X},\mathbf{Z},\mathbf{C})\},$$

where

$$\varepsilon_s = Y - \gamma(\mathbf{X},\mathbf{Z},\mathbf{C}) - [q_s(\mathbf{X},\mathbf{Z},\mathbf{C}) - \mathbb{E}_s\{q_s(\mathbf{X},\mathbf{Z},\mathbf{C})|\mathbf{Z},\mathbf{C}\}] - t_s(\mathbf{C}),$$
$$\overline{q}_s(\mathbf{Z},\mathbf{C}) = \mathbb{E}_s\{q_s(\mathbf{X},\mathbf{Z},\mathbf{C})|\mathbf{Z},\mathbf{C}\} = \int q_s(\mathbf{X},\mathbf{Z},\mathbf{C})dF_s(\mathbf{X}|\mathbf{Z},\mathbf{C})$$

and $q_s(\cdot,\cdot,\cdot) - \overline{q}_s(\cdot,\cdot) \in \Gamma_{\text{sub}}$. *Let* $\Lambda_{\text{nuis},1}, \Lambda_{\text{nuis},2}, \Lambda_{\text{nuis},3}, \Lambda_{\text{nuis},4}$ *and* $\Lambda_{\text{nuis},5}$ *be the tangent spaces for the nuisance parameters indexing* $f(\varepsilon|\mathbf{X},\mathbf{Z},\mathbf{C})$, $q(\mathbf{X},\mathbf{Z},\mathbf{C}), t(\mathbf{C}), f(\mathbf{X}|\mathbf{Z},\mathbf{C})$ *and* $f(\mathbf{Z},\mathbf{C})$. *The nuisance tangent space is given by*

$$\Lambda_{\text{nuis}} = \Lambda_{\text{nuis},1} \oplus \Lambda_{\text{nuis},2} \oplus \Lambda_{\text{nuis},3} \oplus \Lambda_{\text{nuis},4} \oplus \Lambda_{\text{nuis},5}$$

where

$$\Lambda_{\text{nuis},1} = \left\{ \begin{array}{c} a_1(\varepsilon,\mathbf{X},\mathbf{Z},\mathbf{C}) : \\ \mathbb{E}\{a_1(\varepsilon,\mathbf{X},\mathbf{Z},\mathbf{C})|\mathbf{X},\mathbf{Z},\mathbf{C}\} = \mathbb{E}\{\varepsilon a_1(\varepsilon,\mathbf{X},\mathbf{Z},\mathbf{C})|\mathbf{X},\mathbf{Z},\mathbf{C}\} = 0 \end{array} \right\} \cap L_2^0,$$
$$\Lambda_{\text{nuis},2} = \left\{ \begin{array}{c} \{a_2(\mathbf{X},\mathbf{Z},\mathbf{C}) - \overline{a}_2(\mathbf{Z},\mathbf{C})\} f_\varepsilon(\varepsilon|\mathbf{X},\mathbf{Z},\mathbf{C})/f(\varepsilon|\mathbf{X},\mathbf{Z},\mathbf{C}) \\ a_2(\mathbf{X},\mathbf{Z},\mathbf{C}) - \overline{a}_2(\mathbf{Z},\mathbf{C}) \in \Gamma_{\text{sub}}, \\ f_\varepsilon \text{ the derivative of the density of } \varepsilon \text{ wrt } \varepsilon \end{array} \right\} \cap L_2^0,$$

$$\Lambda_{\text{nuis},3} = \left\{ \begin{array}{c} a_3(\mathbf{C}) f_\varepsilon(\varepsilon|\mathbf{X},\mathbf{Z},\mathbf{C})/f(\varepsilon|\mathbf{X},\mathbf{Z},\mathbf{C}) \\ a_3(\mathbf{C}) \text{ unrestricted} \end{array} \right\} \cap L_2^0,$$
$$\Lambda_{\text{nuis},4} = \left\{ \begin{array}{c} a_4(\mathbf{X},\mathbf{Z},\mathbf{C}) + \left\{\int a_4(\mathbf{X}^*,\mathbf{Z},\mathbf{C})\, q(\mathbf{X}^*,\mathbf{Z},\mathbf{C})dF(\mathbf{X}^*|\mathbf{Z},\mathbf{C})\right\} \\ \times f_\varepsilon(\varepsilon|\mathbf{X},\mathbf{Z},\mathbf{C})/f(\varepsilon|\mathbf{X},\mathbf{Z},\mathbf{C}) \\ E\{a_4(\mathbf{X},\mathbf{Z},\mathbf{C})|\mathbf{Z},\mathbf{C}\} = 0 \end{array} \right\} \cap L_2^0,$$
$$\Lambda_{\text{nuis},5} = \{a_5(\mathbf{Z},\mathbf{C}) : E\{a_5(\mathbf{Z},\mathbf{C})\} = 0\} \cap L_2^0.$$

Consider the set

$$\Lambda_{\text{nuis},1}^\perp \cap \Lambda_{\text{nuis},5}^\perp = \left\{ \begin{array}{c} a_1(\varepsilon,\mathbf{X},\mathbf{Z},\mathbf{C}) + a_5(\mathbf{Z},\mathbf{C}) : E\{a_5(\mathbf{Z},\mathbf{C})\} = 0, \\ E\{a_1(\varepsilon,\mathbf{X},\mathbf{Z},\mathbf{C})|\mathbf{X},\mathbf{Z},\mathbf{C}\} = E\{\varepsilon a_1(\varepsilon,\mathbf{X},\mathbf{Z},\mathbf{C})|\mathbf{X},\mathbf{Z},\mathbf{C}\} = 0 \end{array} \right\}^\perp$$

it is straightforward to verify that

$$\Lambda_{\text{nuis},1}^\perp \cap \Lambda_{\text{nuis},5}^\perp = \{\varepsilon h_1(\mathbf{X},\mathbf{Z},\mathbf{C}) + h_2(\mathbf{X},\mathbf{Z},\mathbf{C}) : h_1, E(h_2(\mathbf{X},\mathbf{Z},\mathbf{C})|\mathbf{Z},\mathbf{C}) = 0\}.$$

Next consider the set $\Lambda_{\text{nuis},1}^\perp \cap \Lambda_{\text{nuis},2}^\perp \cap \Lambda_{\text{nuis},5}^\perp$, *the set of functions in* $\Lambda_{\text{nuis},1}^\perp \cap \Lambda_{\text{nuis},5}^\perp$ *also in* $\Lambda_{\text{nuis},2}^\perp$ *must satisfy :*

$$\begin{aligned} 0 &= \mathbb{E}\left[\{\varepsilon h_1(\mathbf{X},\mathbf{Z},\mathbf{C}) + h_2(\mathbf{X},\mathbf{Z},\mathbf{C})\}\{a_2(\mathbf{X},\mathbf{Z},\mathbf{C}) - \overline{a}_2(\mathbf{Z},\mathbf{C})\} f_\varepsilon(\varepsilon|\mathbf{X},\mathbf{Z},\mathbf{C})/f(\varepsilon|\mathbf{X},\mathbf{Z},\mathbf{C})\right] \\ &= \mathbb{E}\left[\varepsilon h_1(\mathbf{X},\mathbf{Z},\mathbf{C})\{a_2(\mathbf{X},\mathbf{Z},\mathbf{C}) - \overline{a}_2(\mathbf{Z},\mathbf{C})\} f_\varepsilon(\varepsilon|\mathbf{X},\mathbf{Z},\mathbf{C})/f(\varepsilon|\mathbf{X},\mathbf{Z},\mathbf{C})\right] \\ &= \mathbb{E}\left[h_1(\mathbf{X},\mathbf{Z},\mathbf{C})\{a_2(\mathbf{X},\mathbf{Z},\mathbf{C}) - \overline{a}_2(\mathbf{Z},\mathbf{C})\}\right] \\ &= \mathbb{E}\left[\left\{h_1(\mathbf{X},\mathbf{Z},\mathbf{C}) - \overline{h}_1(\mathbf{Z},\mathbf{C}) + m(\mathbf{Z},\mathbf{C})\right\}\{a_2(\mathbf{X},\mathbf{Z},\mathbf{C}) - \overline{a}_2(\mathbf{Z},\mathbf{C})\}\right] \end{aligned}$$

for all $h_1(\mathbf{X},\mathbf{Z},\mathbf{C})-\overline{h}_1(\mathbf{Z},\mathbf{C})\in\Gamma_{\text{sub}}^{\perp}\cap\Gamma$, *and* $m(\mathbf{z},\mathbf{c})$ *unrestricted, which implies that*

$$\Lambda_{\text{nuis},1}^{\perp}\cap\Lambda_{\text{nuis},2}^{\perp}=\left\{\begin{array}{c}\varepsilon\left\{h_1(\mathbf{X},\mathbf{Z},\mathbf{C})-\overline{h}_1(\mathbf{Z},\mathbf{C})+m(\mathbf{Z},\mathbf{C})\right\}+h_2(\mathbf{X},\mathbf{Z},\mathbf{C}):\\ E(h_2(\mathbf{X},\mathbf{Z},\mathbf{C})|\mathbf{Z},\mathbf{C})=0,h_1(\cdot,\cdot,\cdot)-\overline{h}_1(\cdot,\cdot)\in\Gamma_{\text{sub}}^{\perp}\cap\Gamma,m(\cdot,\cdot)\ \textit{unrestricted}\end{array}\right\}$$

Next, consider $\Lambda_{\text{nuis},1}^{\perp}\cap\Lambda_{\text{nuis},2}^{\perp}\cap\Lambda_{\text{nuis},3}^{\perp}\cap\Lambda_{\text{nuis},4}^{\perp}\cap\Lambda_{\text{nuis},5}^{\perp}$. *Since*

$$\varepsilon\left\{h_1(\mathbf{X},\mathbf{Z},\mathbf{C})-\overline{h}_1(\mathbf{Z},\mathbf{C})+m(\mathbf{Z},\mathbf{C})-\overline{m}(\mathbf{C})\right\}+h_2(\mathbf{X},\mathbf{Z},\mathbf{C})\in\Lambda_{\text{nuis},1}^{\perp}\cap\Lambda_{\text{nuis},2}^{\perp}$$

is orthogonal to score functions $a_3(\mathbf{C})f_{\varepsilon}(\varepsilon|\mathbf{X},\mathbf{Z},\mathbf{C})/f(\varepsilon|\mathbf{X},\mathbf{Z},\mathbf{C})$, *such functions must also satisfy*

$$\begin{aligned}0&=\mathbb{E}\left[\left\{\varepsilon\left\{h_1(\mathbf{X},\mathbf{Z},\mathbf{C})-\overline{h}_1(\mathbf{Z},\mathbf{C})+m(\mathbf{Z},\mathbf{C})-\overline{m}(\mathbf{C})\right\}+h_2(\mathbf{X},\mathbf{Z},\mathbf{C})\right\}\right.\\&\quad\times\left.\left\{a_4(\mathbf{X},\mathbf{Z},\mathbf{C})+\left\{\int a_4(\mathbf{X}^*,\mathbf{Z},\mathbf{C})\left\{q(\mathbf{X}^*,\mathbf{Z},\mathbf{C})-\overline{q}(\mathbf{Z},\mathbf{C})\right\}dF(\mathbf{X}^*|\mathbf{Z},\mathbf{C})\right\}f_{\varepsilon}(\varepsilon|\mathbf{X},\mathbf{Z},\mathbf{C})/f(\varepsilon|\mathbf{X},\mathbf{Z},\mathbf{C})\right\}\right]\\&=\mathbb{E}\left[a_4(\mathbf{X},\mathbf{Z},\mathbf{C})\left(h_2(\mathbf{X},\mathbf{Z},\mathbf{C})-\left\{q(\mathbf{X},\mathbf{Z},\mathbf{C})-\overline{q}(\mathbf{Z},\mathbf{C})\right\}\left\{m(\mathbf{Z},\mathbf{C})-\overline{m}(\mathbf{C})\right\}\right)\right]\end{aligned}$$

for all $a_4(\mathbf{X},\mathbf{Z},\mathbf{C})$ *with* $\mathbb{E}\{a_4(\mathbf{X},\mathbf{Z},\mathbf{C})|\mathbf{Z},\mathbf{C}\}=0$, *which implies*

$$h_2(\mathbf{X},\mathbf{Z},\mathbf{C})=\left\{q(\mathbf{X},\mathbf{Z},\mathbf{C})-\overline{q}(\mathbf{Z},\mathbf{C})\right\}\left\{m(\mathbf{Z},\mathbf{C})-\overline{m}(\mathbf{C})\right\},$$

and we can conclude that

$$\begin{aligned}\Lambda_{\text{nuis}}^{\perp}&=\Lambda_{\text{nuis},1}^{\perp}\cap\Lambda_{\text{nuis},2}^{\perp}\cap\Lambda_{\text{nuis},3}^{\perp}\cap\Lambda_{\text{nuis},4}^{\perp}\cap\Lambda_{\text{nuis},5}^{\perp}\\&=\left\{\begin{array}{c}\varepsilon\left[\left\{h_1(\mathbf{X},\mathbf{Z},\mathbf{C})-\overline{h}_1(\mathbf{Z},\mathbf{C})\right\}+m(\mathbf{Z},\mathbf{C})-\overline{m}(\mathbf{C})\right]\\+\left\{q(\mathbf{X},\mathbf{Z},\mathbf{C})-\overline{q}(\mathbf{Z},\mathbf{C})\right\}\left\{m(\mathbf{Z},\mathbf{C})-\overline{m}(\mathbf{C})\right\}\ :\\ h_1(\cdot,\cdot,\cdot)-\overline{h}_1(\cdot,\cdot)\in\Gamma_{\text{sub}}^{\perp}\cap\Gamma,\ m\ \textit{unrestricted.}\end{array}\right\}\end{aligned}$$

□

<u>Proof of Theorem 1:</u> In the special case where X *and* Z *are binary, and under assumption (IV.4'), Theorem 1 of Tchetgen Tchetgen et al (2010) implies that* $\Gamma_{\text{sub}}^{\perp}\cap\Gamma=\left\{h(\mathbf{C})(-1)^{X+Z}/f(X,Z|\mathbf{C}):h\right\}\cap L_2^0$ *which gives the result.*

Proof of Theorem 2: Note that if (ii) holds, we have that

$$\begin{aligned}
&\mathbb{E}\left\{U_{\text{np}}\left(\psi;\mu,\alpha,\omega^*,\eta^*;m,h\right)\right\}\\
&=\mathbb{E}\left(\mathbb{E}\left[\left\{Y-\gamma\left(Z,\mathbf{C};\psi\right)X\right\}|X,Z\right]\right.\\
&\left.\left\{\frac{h(\mathbf{C})\left(-1\right)^{X+Z}}{f(X,Z|\mathbf{C};\mu,\alpha)}+m(Z,\mathbf{C})-\overline{m}(\mathbf{C};\mu)\right\}\right)\\
&-\mathbb{E}\left(\frac{\left[\widetilde{q}(\mathbf{C};\eta^*)\left\{X-p\left(Z,\mathbf{C}\right)\right\}+t\left(\mathbf{C};\omega^*\right)\right]h(\mathbf{C})\left(-1\right)^{X+Z}}{f(X,Z|\mathbf{C};\mu,\alpha)}\right)\\
&+\underbrace{\mathbb{E}\left(\widetilde{q}(\mathbf{C})\left\{X-p\left(Z,\mathbf{C};\alpha\right)\right\}\left\{m(Z,\mathbf{C})-\overline{m}(\mathbf{C};\mu)\right\}\right)}_{=0}\\
&=\mathbb{E}\left(\frac{\left[\widetilde{q}(\mathbf{C})\left\{X-p\left(Z,\mathbf{C}\right)\right\}+t\left(\mathbf{C}\right)\right]h(\mathbf{C})\left(-1\right)^{X+Z}}{f(X,Z|\mathbf{C};\mu,\alpha)}\right)\\
&-\mathbb{E}\left(\frac{\left[\widetilde{q}(\mathbf{C};\eta^*)\left\{X-p\left(Z,\mathbf{C};\alpha\right)\right\}+t\left(\mathbf{C};\omega^*\right)\right]h(\mathbf{C})\left(-1\right)^{X+Z}}{f(X,Z|\mathbf{C};\mu,\alpha)}\right)\\
&+\mathbb{E}\left(\left[\widetilde{q}(\mathbf{C})\left\{X-p\left(Z,\mathbf{C}\right)\right\}+t\left(\mathbf{C}\right)\right]\left\{m(Z,\mathbf{C})-\overline{m}(\mathbf{C};\mu)\right\}\right)\\
&-\mathbb{E}\left(\left[\widetilde{q}(\mathbf{C};\eta^*)\left\{X-p\left(Z,\mathbf{C};\alpha\right)\right\}+t\left(\mathbf{C};\omega^*\right)\right]\left\{m(Z,\mathbf{C})-\overline{m}(\mathbf{C};\mu)\right\}\right)\\
&=\mathbb{E}\left(\widetilde{q}(\mathbf{C})\left[-\left\{1-1\right\}-\left\{p\left(1,\mathbf{C}\right)-p\left(0,\mathbf{C}\right)\right\}\left(1-1\right)\right]\right)\\
&-\mathbb{E}\left(\widetilde{q}(\mathbf{C};\eta^*)\left[-\left\{1-1\right\}-\left\{p\left(1,\mathbf{C};\alpha\right)-p\left(0,\mathbf{C};\alpha\right)\right\}\left(1-1\right)\right]\right)\\
&+\mathbb{E}\left\{t\left(\mathbf{C}\right)\left(2-2\right)+\mathbb{E}\left(m(Z,\mathbf{C})|\mathbf{C};\mu\right)-\overline{m}(\mathbf{C};\mu)\right\}\\
&-\mathbb{E}\left\{t\left(\mathbf{C};\omega^*\right)\left(2-2\right)+\mathbb{E}\left(m(Z,\mathbf{C})|\mathbf{C};\mu\right)-\overline{m}(\mathbf{C};\mu)\right\}\\
&=0.
\end{aligned}$$

Likewise,if (i) holds, we have that

$$\mathbb{E}\left(Y|X,Z,\mathbf{C}\right)=\gamma\left(Z,\mathbf{C};\psi\right)X-\widetilde{q}(\mathbf{C};;\eta)\left\{X-p\left(Z,\mathbf{C};\alpha\right)\right\}-t(\mathbf{C};\omega)$$

therefore

$$\begin{aligned}
&\mathbb{E}\left\{U_{\text{np}}\left(\psi;\mu^*,\alpha,\omega,\eta;m,h\right)\right\}\\
&=\mathbb{E}\left(\underbrace{\mathbb{E}\left[\left\{Y-\mathbb{E}\left(Y|X,Z,\mathbf{C}\right)\right\}|X,Z,\mathbf{C}\right]}_{=0}\right.\\
&\left.\times\left\{\frac{h(\mathbf{C})\left(-1\right)^{X+Z}}{f(X,Z|\mathbf{C};\mu^*,\alpha)}+m(Z,\mathbf{C})-\overline{m}(\mathbf{C};\mu^*)\right\}\right)\\
&+\underbrace{\mathbb{E}\left(\widetilde{q}(\mathbf{C})\left\{X-p\left(Z,\mathbf{C};\alpha\right)\right\}\left\{m(Z,\mathbf{C})-\overline{m}(\mathbf{C};\mu^*)\right\}\right)}_{=0}\\
&=0
\end{aligned}$$

establishing the first part of the result.

To derive the efficient score, note that the set $\left\{U_{\text{np}}\left(m,h\right):m,h\ \textit{unrestricted}\right\}$ *is equal to the set*

$$\begin{aligned}
&\left\{h(\mathbf{C})\left\{\frac{\varepsilon\left(-1\right)^{X+Z}}{f(X,Z|\mathbf{C})}\right\}+m(\mathbf{C})(Z-\Pr(Z=1|\mathbf{C}))\epsilon:m,h\ \textit{unrestricted}\right\}\\
&=\left\{h(\mathbf{C})R_1+m(\mathbf{C})R_2:m,h\ \textit{unrestricted}\right\}
\end{aligned}$$

and the efficient score is given by the population least square projection of the score onto the above set.

□

Proof of Theorem 3: Proceeding as in the proof of Theorem 4 specialized to binary X and Z, *upon replacing* $\Lambda_{\text{nuis},4}$ *with*

$$\Lambda^*_{\text{nuis},4} = \left\{ \begin{array}{c} a_4(\mathbf{x},\mathbf{Z},\mathbf{C}) + \left\{ \int a_4(\mathbf{x}^*,\mathbf{Z},\mathbf{C})\, q(\mathbf{x}^*,\mathbf{Z},\mathbf{C}) dF(\mathbf{x}^*|\mathbf{Z},\mathbf{C}) \right\} f_\varepsilon(\varepsilon|\mathbf{x},\mathbf{Z},\mathbf{C}) / f(\varepsilon|\mathbf{x},\mathbf{Z},\mathbf{C}) \\ a_4(\mathbf{x},\mathbf{Z},\mathbf{C}) \in \Lambda_{x,\text{nuis}} \cap \Lambda_{x,\text{np}} \end{array} \right\},$$

one can verify that the ortho-complement to the nuisance tangent space $\Lambda^\perp_{\text{nuis},1} \cap \Lambda^\perp_{\text{nuis},2} \cap \Lambda^\perp_{\text{nuis},3} \cap \Lambda^{*,\perp}_{\text{nuis},4}$ *is given by*

$$\digamma = \left\{ \begin{array}{c} U_{\text{np}}(m,h) + P(n) : m(Z,\mathbf{C}), h(\mathbf{C}) \textit{ unrestricted} \\ P(n) = n(Z,\mathbf{C})\Delta \in \Lambda^\perp_{x,\text{nuis}} \cap \Lambda_{x,\text{np}} \end{array} \right\}$$

by noting that

$$\varepsilon \left\{ h_1(\mathbf{x},\mathbf{Z},\mathbf{C}) - \overline{h}_1(\mathbf{Z},\mathbf{C}) + m(\mathbf{Z},\mathbf{C}) - \overline{m}(\mathbf{C}) \right\} + h_2(\mathbf{x},\mathbf{Z},\mathbf{C}) \in \Lambda^\perp_{\text{nuis},1} \cap \Lambda^\perp_{\text{nuis},2} \cap \Lambda^\perp_{\text{nuis},3}$$

is orthogonal to $\Lambda^*_{\text{nuis},4}$ *if and only if*

$$h_2(\mathbf{x},\mathbf{Z},\mathbf{C}) = \widetilde{q}(\mathbf{Z}) \left\{ X - p(Z,\mathbf{C};\alpha) \right\} \left\{ m(Z,\mathbf{C}) - \overline{m}(\mathbf{C};\mu^*) \right\} + n(Z,\mathbf{C})\Delta$$

where $n(Z,\mathbf{C})\Delta \in \Lambda^\perp_{x,\text{nuis}} \cap \Lambda_{x,\text{np}}$, *which proves the first result.*

In the special case where $f(X|Z,\mathbf{C};\alpha,\tau) = f(X|Z,\mathbf{C};\alpha)$ *is a parametric model, such that* $\Lambda^\perp_{x,\text{nuis}} \cap \Lambda_{x,\text{np}} = \Lambda_{x,\text{np}}$

$$\digamma = \left\{ \begin{array}{c} V_{\text{np}}(m,h,n) = \varepsilon \left\{ \frac{h(\mathbf{C})(-1)^{X+Z}}{f(X,Z|\mathbf{C})} + m(\mathbf{C})(Z - g(C)) \right\} + n(Z,\mathbf{C})\Delta : \\ m(\mathbf{C}), h(\mathbf{C}), n(Z,\mathbf{C}) \textit{ unrestricted} \end{array} \right\}$$

and the efficient score of (ψ,α) *is obtained by a population least square projection of* $\mathbf{S}_{\psi,\alpha}$ *onto* $\digamma$, *noting that* $n(Z,\mathbf{C})\Delta$ *and* $\varepsilon \left\{ h(\mathbf{C})(-1)^{X+Z} / f(X,Z|\mathbf{C}) + m(\mathbf{C})(Z-g(C)) \right\}$ *are orthogonal for all* n, m, h, and therefore the projection onto $\{ n(Z,\mathbf{C})\Delta : n \}$ and

$$\left\{ \varepsilon \left\{ h(\mathbf{C})(-1)^{X+Z} / f(X,Z|\mathbf{C}) + m(\mathbf{C})(Z-g(C)) \right\} : m, h \right\}$$

can be done separately.

□

Proof of Theorem 5:

Let $\varepsilon = Y - \mathbb{E}(Y|\overline{\mathbf{H}})$ *and consider the submodel* $f_s(\overline{\mathbf{O}}) = f_s(\varepsilon_s|\overline{\mathbf{H}}) f_s(\overline{\mathbf{H}})$, *with* $f_0(\varepsilon_0,\overline{\mathbf{H}}) = f(\varepsilon,\overline{\mathbf{H}})$ *and the submodel varies in all parameters of the model with the exception of the SNM* $\gamma_j, j = 1,...,J$, *which is evaluated at the truth. The nuisance tangent space for the nonparametric model* $\mathcal{M}_L$ *is given*

$$\Lambda_{\text{nuis}} = \Lambda_{\text{nuis},1} \oplus \Lambda_{\text{nuis},2} \oplus \Lambda_{\text{nuis},3} \oplus \Lambda_{\text{nuis},4}$$

where

$$\Lambda_{\text{nuis},1} = \left\{ \begin{array}{c} a_1(\varepsilon,\overline{\mathbf{H}}) : \\ \mathbb{E}\left\{ a_1(\varepsilon,\overline{\mathbf{H}})|\overline{\mathbf{H}} \right\} = \mathbb{E}\left\{ \varepsilon a_1(\varepsilon,\overline{\mathbf{H}})|\overline{\mathbf{H}} \right\} = 0 \end{array} \right\} \cap L^0_2,$$

$$\Lambda_{\text{nuis},2} = \left\{ \begin{array}{c} \sum_{j=1}^J \left\{ a_{2,j}(\overline{\mathbf{H}}(j)) - \overline{a}_{2,j}(\overline{\mathbf{H}}(j-1),\mathbf{Z}(j),\mathbf{C}(j)) \right\} f_\varepsilon(\varepsilon|\overline{\mathbf{H}}) / f(\varepsilon|\overline{\mathbf{H}}) \\ + \sum_{j=1}^J \left\{ a^*_{2,j}(\overline{\mathbf{H}}(j-1),\mathbf{C}(j)) - \overline{a}^*_{2,j}(\overline{\mathbf{H}}(j-1)) \right\} f_\varepsilon(\varepsilon|\overline{\mathbf{H}}) / f(\varepsilon|\overline{\mathbf{H}}) \\ a_{2,j}(\overline{\mathbf{H}}(j)) - \overline{a}_{2,j}(\overline{\mathbf{H}}(j-1),\mathbf{Z}(j),\mathbf{C}(j)) \in \Gamma_{\text{sub},j}, a^*_{2,j} \text{ unrestrited} \end{array} \right\} \cap L^0_2,$$

with $\overline{a}^*_{2,1}(\overline{\mathbf{H}}(0)) = 0$,

$$\Lambda_{\text{nuis},3} = \left\{ \begin{array}{c} \sum_{j=1}^{J} a_{4,j}\left(\overline{\mathbf{H}}(j)\right) - \left\{ \int a_{4,j}\left(\mathbf{X}^*(j), \overline{\mathbf{H}}(j-1), \mathbf{Z}(j), \mathbf{C}(j)\right)\right. \\ \left. q_j(\mathbf{X}^*(j), \overline{\mathbf{H}}(j-1), \mathbf{Z}(j), \mathbf{C}(j)) dF(\mathbf{X}^*(j) | \overline{\mathbf{H}}(j-1), \mathbf{Z}(j), \mathbf{C}(j)) \right\} f_\varepsilon\left(\varepsilon|\overline{\mathbf{H}}\right)/f\left(\varepsilon|\overline{\mathbf{H}}\right) \\ \mathbb{E}\left\{ a_{4,j}\left(\overline{\mathbf{H}}(j)\right) | \overline{\mathbf{H}}(j-1), \mathbf{Z}(j), \mathbf{C}(j) \right\} = 0, j = 1, ...J \end{array} \right\} \cap L_2^0,$$

$$\Lambda_{\text{nuis},4} = \left\{ \begin{array}{c} \sum_{j=1}^{J} a_{5,j}(\mathbf{C}(j), \overline{\mathbf{H}}(j-1)) - \left\{ \int a_{5,j}(\mathbf{C}^*(j), \overline{\mathbf{H}}(j-1)) \right. \\ \left. b_j(\overline{\mathbf{H}}(j-1), \mathbf{C}(j)) dF(\mathbf{C}^*(j) | \overline{\mathbf{H}}(j-1)) \right\} f_\varepsilon\left(\varepsilon|\overline{\mathbf{H}}\right)/f\left(\varepsilon|\overline{\mathbf{H}}\right) \\ + \sum_{j=1}^{J} a_{6,j}(\mathbf{Z}(j), \mathbf{C}(j), \overline{\mathbf{H}}(j-1)) \\ \mathbb{E}\left[a_{5,j}(\mathbf{C}(j), \overline{\mathbf{H}}(j-1)) | \overline{\mathbf{H}}(j-1) \right] = 0, \\ \mathbb{E}\left[a_{6,j}(\mathbf{Z}(j), \mathbf{C}(j), \overline{\mathbf{H}}(j-1)) | \mathbf{C}(j), \overline{\mathbf{H}}(j-1) \right], j = 1, ...J \end{array} \right\} \cap L_2^0.$$

It is straightforward to verify that

$$\Lambda^{\perp}_{\text{nuis},1} = \left\{ \varepsilon h_1\left(\overline{\mathbf{H}}\right) + h_2(\overline{\mathbf{H}}) : h_1, h_2 \right\} \ \cap L_2^0$$

next consider the set $\Lambda^{\perp}_{\text{nuis},1} \cap \Lambda^{\perp}_{\text{nuis},2}$, *the subset of functions contained in* $\Lambda^{\perp}_{\text{nuis},1}$ *also included in* $\Lambda^{\perp}_{\text{nuis},2}$ *must satisfy :*

$$\begin{aligned} 0 &= \mathbb{E}\left[\left\{ \varepsilon h_1\left(\overline{\mathbf{H}}\right) + h_2(\overline{\mathbf{H}}) \right\} \left\{ \sum_{j=1}^{J} \left\{ a_{2,j}(\overline{\mathbf{H}}(j)) - \overline{a}_{2,j}(\overline{\mathbf{H}}(j-1), \mathbf{Z}(j), \mathbf{C}(j)) \right\} \right\} f_\varepsilon\left(\varepsilon|\overline{\mathbf{H}}\right)/f\left(\varepsilon|\overline{\mathbf{H}}\right) \right] \\ &+ \mathbb{E}\left[\left\{ \varepsilon h_1\left(\overline{\mathbf{H}}\right) + h_2(\overline{\mathbf{H}}) \right\} \left\{ \sum_{j=1}^{J} \left\{ a^*_{2,j}(\overline{\mathbf{H}}(j-1), \mathbf{C}(j))) - \overline{a}^*_{2,j}(\overline{\mathbf{H}}(j-1)) \right\} \right\} f_\varepsilon\left(\varepsilon|\overline{\mathbf{H}}\right)/f\left(\varepsilon|\overline{\mathbf{H}}\right) \right] \\ &= \mathbb{E}\left[h_1\left(\overline{\mathbf{H}}\right) \left\{ \sum_{j=1}^{J} \left\{ a_{2,j}(\overline{\mathbf{H}}(j)) - \overline{a}_{2,j}(\overline{\mathbf{H}}(j-1), \mathbf{Z}(j), \mathbf{C}(j)) \right\} \right\} \right] \\ &+ \mathbb{E}\left[h_1\left(\overline{\mathbf{H}}\right) \left\{ \sum_{j=1}^{J} \left\{ a^*_{2,j}(\overline{\mathbf{H}}(j-1), \mathbf{C}(j))) - \overline{a}^*_{2,j}(\overline{\mathbf{H}}(j-1)) \right\} \right\} \right] \\ &= \mathbb{E}\left(\left[\sum_{j=1}^{J} \left\{ h^*_{1j}(\overline{\mathbf{H}}(j)) - \overline{h}^*_{1j}(\overline{\mathbf{H}}(j-1), \mathbf{Z}(j)), \mathbf{C}(j)) \right\} \right.\right. \\ &+ \left. \sum_{j=1}^{J} \left\{ m_j(\overline{\mathbf{H}}(j-1), \mathbf{Z}(j), \mathbf{C}(j)) - \overline{m}_j(\overline{\mathbf{H}}(j-1), \mathbf{C}(j)) \right\} \right] \\ &\times \left. \left\{ \sum_{j=1}^{J} \left\{ a_{2,j}(\overline{\mathbf{H}}(j)) - \overline{a}_{2,j}(\overline{\mathbf{H}}(j-1), \mathbf{Z}(j), \mathbf{C}(j)) \right\} + \left\{ a^*_{2,j}(\overline{\mathbf{H}}(j-1), \mathbf{C}(j))) - \overline{a}^*_{2,j}(\overline{\mathbf{H}}(j-1)) \right\} \right\} \right) \end{aligned}$$

for all $h^*_{1j}(\overline{\mathbf{H}}(j)) - \overline{h}^*_{1j}(\overline{\mathbf{H}}(j-1), \mathbf{Z}(j)), \mathbf{C}(j)) \in \Gamma^{\perp}_{\text{sub},j} \cap \Gamma_j$, *and* m_j *unrestricted. Thus*

$$\begin{aligned} &\Lambda^{\perp}_{\text{nuis},1} \cap \Lambda^{\perp}_{\text{nuis},2} \\ &= \left\{ \begin{array}{c} \varepsilon \left\{ h_1\left(\overline{\mathbf{H}}\right) + \sum_{j=1}^{J} \left\{ m_j(\overline{\mathbf{H}}(j-1), \mathbf{Z}(j), \mathbf{C}(j)) - \overline{m}_j(\overline{\mathbf{H}}(j-1), \mathbf{C}(j)) \right\} \right\} + \sum_{j=1}^{J} h_{2,j}(\overline{\mathbf{H}}(j)) : \\ \mathbb{E}(h_{2,j}(\overline{\mathbf{H}}(j)) | \mathbf{Z}(j), \mathbf{C}(j), \overline{\mathbf{H}}(j-1)) = 0 \\ h_1\left(\overline{\mathbf{H}}\right) = \sum_{j=1}^{J} \left\{ h^*_{1j}(\overline{\mathbf{H}}(j)) - \overline{h}^*_{1j}(\overline{\mathbf{H}}(j-1), \mathbf{Z}(j)), \mathbf{C}(j)) \right\} \in \bigcap_j \left\{ \Gamma^{\perp}_{\text{sub},j} \cap \Gamma_j \right\} \\ m_j(\cdot, \cdot, \cdot) \textit{ unrestricted} \end{array} \right\} \end{aligned}$$

Next, note that elements of the set $\Lambda_{\text{nuis},1}^{\perp} \cap \Lambda_{\text{nuis},2}^{\perp} \cap \Lambda_{\text{nuis},4}^{\perp} \cap \Lambda_{\text{nuis},5}^{\perp}$ must also satisfy

$$
\begin{aligned}
0 = \mathbb{E}\Bigg(\Bigg[\varepsilon \Bigg\{ & \sum_{j=1}^{J} \Big\{ h_{1j}^{*}(\overline{\mathbf{H}}(j)) - \overline{h}_{1j}^{*}(\overline{\mathbf{H}}(j-1), \mathbf{Z}(j)), \mathbf{C}(j)) \Big\} \\
& + \sum_{j=1}^{J} \{ m_j(\overline{\mathbf{H}}(j-1), \mathbf{Z}(j), \mathbf{C}(j)) - \overline{m}_j(\overline{\mathbf{H}}(j-1), \mathbf{C}(j)) \} \Bigg\} + h_2(\overline{\mathbf{H}}) \Bigg] \\
& \times \Bigg[\sum_{j=1}^{J} a_{4,j}\left(\overline{\mathbf{H}}(j)\right) - \Big\{ \int a_{4,j}\left(\mathbf{X}^{*}(j), \overline{\mathbf{H}}(j-1), \mathbf{Z}(j), \mathbf{C}(j)\right) \\
& q_j(\mathbf{X}^{*}(j), \overline{\mathbf{H}}(j-1), \mathbf{Z}(j), \mathbf{C}(j)) dF(\mathbf{X}^{*}(j) | \overline{\mathbf{H}}(j-1), \mathbf{Z}(j), \mathbf{C}(j)) \Big\} f_{\varepsilon}\left(\varepsilon | \overline{\mathbf{H}}\right) / f\left(\varepsilon | \overline{\mathbf{H}}\right) \\
& + \sum_{j=1}^{J} a_{5,j}(\mathbf{C}(j), \overline{\mathbf{H}}(j-1)) - \Big\{ \int a_{5,j}(\mathbf{C}^{*}(j), \overline{\mathbf{H}}(j-1)) \\
& b_j(\overline{\mathbf{H}}(j-1), \mathbf{C}(j)) dF(\mathbf{C}^{*}(j) | \overline{\mathbf{H}}(j-1)) \Big\} f_{\varepsilon}\left(\varepsilon | \overline{\mathbf{H}}\right) / f\left(\varepsilon | \overline{\mathbf{H}}\right) + \sum_{j=1}^{J} a_{6,j}(\mathbf{Z}(j), \mathbf{C}(j), \overline{\mathbf{H}}(j-1)) \Bigg] \Bigg) \\
= \mathbb{E}\Bigg(\Bigg[& \sum_{j=1}^{J} \int a_{4,j}\left(\mathbf{X}^{*}(j), \overline{\mathbf{H}}(j-1), \mathbf{Z}(j), \mathbf{C}(j)\right) \\
& q_j(\mathbf{X}^{*}(j), \overline{\mathbf{H}}(j-1), \mathbf{Z}(j), \mathbf{C}(j)) dF(\mathbf{X}^{*}(j) | \overline{\mathbf{H}}(j-1), \mathbf{Z}(j), \mathbf{C}(j)) + \int a_{5,j}(\mathbf{C}^{*}(j), \overline{\mathbf{H}}(j-1)) \\
& b_j(\overline{\mathbf{H}}(j-1), \mathbf{C}(j)) dF(\mathbf{C}^{*}(j) | \overline{\mathbf{H}}(j-1)) \Bigg] \\
& \times \Bigg[\sum_{j=1}^{J} \Big\{ h_{1j}^{*}(\overline{\mathbf{H}}(j)) - \overline{h}_{1j}^{*}(\overline{\mathbf{H}}(j-1), \mathbf{Z}(j)), \mathbf{C}(j)) \Big\} \\
& + \sum_{j=1}^{J} \{ m_j(\overline{\mathbf{H}}(j-1), \mathbf{Z}(j), \mathbf{C}(j)) - \overline{m}_j(\overline{\mathbf{H}}(j-1), \mathbf{C}(j)) \} \Bigg] \\
& + \Bigg\{ \sum_{j=1}^{J} a_{4,j}\left(\overline{\mathbf{H}}(j)\right) + a_{5,j}(\mathbf{C}(j), \overline{\mathbf{H}}(j-1)) \Bigg\} \\
& \times \left\{ h_2(\overline{\mathbf{H}}) \right\} \Bigg).
\end{aligned}
$$

for all $a_{4,j}\left(\overline{\mathbf{H}}(j)\right)$ with $\mathbb{E}\left\{a_{4,j}\left(\overline{\mathbf{H}}(j)\right) | \overline{\mathbf{H}}(j-1), \mathbf{Z}(j), \mathbf{C}(j)\right\} = 0$ and all $a_{4,j}\left(\overline{\mathbf{H}}(j)\right)$ with $\mathbb{E}\left\{a_{5,j}(\mathbf{C}(j), \overline{\mathbf{H}}(j-\right.$
0. This is equivalently written

$$
\begin{aligned}
0 = \mathbb{E}\Bigg(& \sum_{j=1}^{J} \left\{ a_{4,j}\left((\overline{\mathbf{H}}(j))\right) + a_{5,j}(\mathbf{C}(j), \overline{\mathbf{H}}(j-1)) \right\} \\
& \Bigg[\sum_{j=1}^{J} \{ q_j(\mathbf{X}(j), \overline{\mathbf{H}}(j-1), \mathbf{Z}(j), \mathbf{C}(j)) - \overline{q}_j(\overline{\mathbf{H}}(j-1), \mathbf{Z}(j), \mathbf{C}(j)) + b_j(\overline{\mathbf{H}}(j-1), \mathbf{C}(j)) - \overline{b}_j(\overline{\mathbf{H}}(j-1)) \} \\
& \times \left\{ \{ m_j(\overline{\mathbf{H}}(j-1), \mathbf{Z}(j), \mathbf{C}(j)) - \overline{m}_j(\overline{\mathbf{H}}(j-1), \mathbf{C}(j)) \} \right\} - h_2(\overline{\mathbf{H}}) \Bigg] \Bigg\}
\end{aligned}
$$

which implies that

$$
\begin{aligned}
&h_2(\overline{\mathbf{H}}) \\
&= \sum_{j=1}^{J} \left\{ m_j(\overline{\mathbf{H}}(j-1), \mathbf{Z}(j), \mathbf{C}(j)) + \overline{m}_j(\overline{\mathbf{H}}(j-1), \mathbf{C}(j)) \right\} \\
&\times \left\{ q_j(\mathbf{X}(j), \overline{\mathbf{H}}(j-1), \mathbf{Z}(j), \mathbf{C}(j)) - \overline{q}_j(\overline{\mathbf{H}}(j-1), \mathbf{Z}(j), \mathbf{C}(j)) + b_j(\overline{\mathbf{H}}(j-1), \mathbf{C}(j)) - \overline{b}_j(\overline{\mathbf{H}}(j-1)) \right\}
\end{aligned}
$$

and therefore

$$
\Lambda_{\text{nuis}}^{\perp} = \left\{ \begin{array}{c}
\varepsilon \left\{ h_1\left(\overline{\mathbf{H}}\right) + \sum_{j=1}^{J} \left\{ m_j(\overline{\mathbf{H}}(j-1), \mathbf{Z}(j), \mathbf{C}(j)) - \overline{m}_j(\overline{\mathbf{H}}(j-1), \mathbf{C}(j)) \right\} \right\} \\
+ \sum_{j=1}^{J} \left\{ m_j(\overline{\mathbf{H}}(j-1), \mathbf{Z}(j), \mathbf{C}(j)) + \overline{m}_j(\overline{\mathbf{H}}(j-1), \mathbf{C}(j)) \right\} \\
\times \left\{ q_j(\mathbf{X}\ (j), \overline{\mathbf{H}}(j-1), \mathbf{Z}(j), \mathbf{C}(j)) - \overline{q}_j(\overline{\mathbf{H}}(j-1), \mathbf{Z}(j), \mathbf{C}(j)) \right. \\
\left. + b_j(\overline{\mathbf{H}}(j-1), \mathbf{C}(j)) - \overline{b}_j(\overline{\mathbf{H}}(j-1)) \right\} \\
h_1\left(\overline{\mathbf{H}}\right) = \sum_{j=1}^{J} \left\{ h_{1j}^{*}(\overline{\mathbf{H}}(j)) - \overline{h}_{1j}^{*}(\overline{\mathbf{H}}(j-1), \mathbf{Z}(j)), \mathbf{C}(j)) \right\} \in \bigcap_j \left\{ \Gamma_{\text{sub},j}^{\perp} \cap \Gamma_j \right\} \\
m_j \textit{ unrestricted}
\end{array} \right\}.
$$

□

References

[1] Abadie (2003), "Semiparametric Instrumental Variable Estimation of Treatment Response Models," Journal of Econometrics, 113, 231–263.

[2] Abadie, A., Angrist, J. D., and Imbens, G. W. (2002), "Instrumental Variables Estimates of the Effects of Subsidized Training on the Quantiles of Trainee Earnings," Econometrica , 70, 91–117.

[3] Angrist, J. D., Imbens, G. W., and Rubin, D. B. (1996), "Identification of Causal Effects Using Instrumental Variables" (with discussion), Journal of the American Statistical Association, 91, 444–472.

[4] Barnard, J., Frangakis, C. E., Hill, J. L., and Rubin, D. B. (2003), "Principal Stratification Approach to Broken Randomized Experiments: A Case Study of School Choice Vouchers in New York City" (with discussion), Journal of the American Statistical Association, 98, 299–323.

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

[6] Carneiro, P., Heckman, J. J., and Vytlacil, E. (2003), "Understanding What In-strumental Variables Estimate: Estimating Marginal and Average Returns to Education," working paper, Stanford University, Dept. of Economics.

[7] Frangakis, C. E., Brookmeyer, R. S., Varadhan, R., Safaeian, M., Vlahov, D., and Strathdee, S. A. (2004), "Methodology for Evaluating a Partially Controlled Longitudinal Treatment Using Principal Stratification, With Application to a Needle Exchange Program," Journal of the American Statistical Association, 99, 239–249.

[8] Goldberger, A. S. (1972), "Structural Equation Methods in Social Sciences,"Econometrica, 40, 979–1001.

[9] Hernan, M. A., and Robins, J. M. (2006), "Instruments for Causal Inference: An Epidemiologist's Dream?" Epidemiology, 17, 360–372.

[10] Heckman, J. J. (1997), "Instrumental Variables: A Study of Implicit Behavioral Assumptions Used in Making Program Evaluations," Journal of Human Resources, 32, 441–462.

[11] Hirano, K., Imbens, G. W., Rubin, D. B., and Zhou, X.-H. (2000), "Assessing the Effect of an Influenza Vaccine in an Encouragement Design," Biostatistics, 1, 69–88.

[12] Joffe M, Brensinger C. Weighting in instrumental variables and g-estimation. Statistics in Medicine 2003; 22(8):1285–1303.

[13] Imbens, G. W., and Angrist, J. D. (1994), "Identification and Estimation of Local Average Treatment Effects," Econometrica, 62, 467–476.

[14] Little, R. J., and Yau, L. H. Y. (1998), "Statistical Techniques for Analyzing Data From Prevention Trials: Treatment of No-Shows Using Rubin's Causal Model," Psychological Methods, 3, 147–159.

[15] Richardson TS, Robins JM. (2010). Analysis of the Binary Instrumental Variable Model. Heuristics, Probability and Causality: A Tribute to Judea Pearl H. Geffner and J.Y. Halpern, Editors, College Publications, UK.

[16] Robins, J (1994), "Correcting for Non-Compliance in Randomized Trials UsingStructural Nested Mean Models,"Communications in Statistics, 23, 2379–2412.

[17] 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.

[18] Robins JM (2000). Robust Estimation in Sequentially Ignorable Missing Data and Causal Inference Models, 1999 Proceedings of the American Statistical Association, Alexandria, VA: American Statistical Association, pp. 6–10.

[19] Robins, J. M., and Rotnitzky, A. (2004), "Estimation of Treatment Effects in Randomized Trails With Noncompliance and a Dichotomous Outcome Using Structural Mean Models," Biometrika, 91, 763–783.

[20] Tan, Z. (2006a), "Regression and Weighting Methods for Causal Inference Using Instrumental Variables,"Journal of the American Statistical Association, 101, 1067–1618. (2006b), "A Distributional Approach for Causal Inference Using Propensity Scores,"Journal of the American Statistical Association,101,1619–1637.

[21] Tan, Z. (2010) Marginal and nested structural models using instrumental variables, Journal of the American Statistical Association, 105, 157-169.

[22] Tchetgen Tchetgen E, Robins JM, Rotnitsky A. (2010). On doubly robust estimation in a semiparametric odds ratio model. Biometrika 97(1):171-180.

[23] Okui, R., Small, D., Tan, Z., and Robins, J.M. (2012) Doubly robust instrumental variables regression, Statistica Sinica, 22, 173-205.

[24] Vansteelandt, S., and Goetghebeur, E. (2003), "Causal Inference With Generalized Structural Mean Models," Journal of the Royal Statistical Society, Ser. B, 65, 817–835.

[25] Wooldridge, J. M. (2002),Econometric Analysis of Cross Section and Panel Data, Cambridge, MA: MIT Press.

[26] Wright, S. (1928),The Tariff on Animal and Vegetable Oils ,NewYork:MacMillan, the Appendix.

[27] Wooldridge, J. M. (2002), Econometric Analysis of Cross Section and Panel Data, Cambridge, MA: MIT Press.

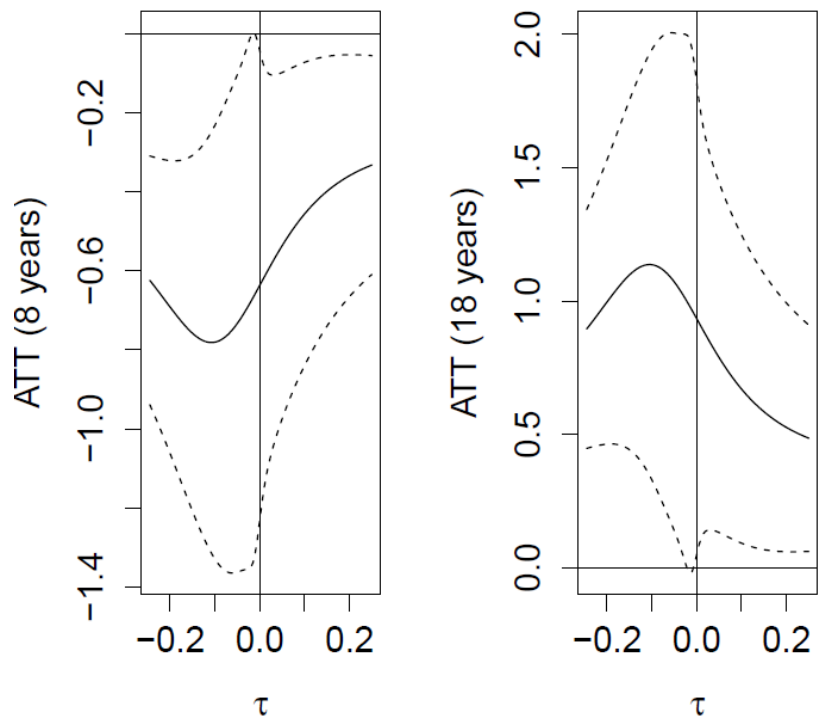

Figure 1. Sensitivity analysis for heterogeneity in the selection bias function in the NLSYM study