Statistica Sinica Preprint No: SS-2021-0354	
Title	Semiparametric Causal Mediation Analysis with Unmeasured Mediator-Outcome Confounding
Manuscript ID	SS-2021-0354
URL	http://www.stat.sinica.edu.tw/statistica/
DOI	10.5705/ss.202021.0354
Complete List of Authors	BaoLuo Sun and Ting Ye
Corresponding Author	BaoLuo Sun
E-mail	stasb@nus.edu.sg

Statistica Sinica: Preprint
doi:10.5705/ss.202021.0354

Statistica Sinica

Semiparametric Causal Mediation Analysis with Unmeasured Mediator-Outcome Confounding

BaoLuo Sun* and and Ting Ye†

* *Department of Statistics and Data Science, National University of Singapore*

†*Department of Biostatistics, University of Washington*

Abstract:

Although exposure can be randomly assigned in studies of mediation effects, direct intervention on the mediator is often infeasible, making unmeasured mediator-outcome confounding possible. We propose a semiparametric identification of natural direct and indirect effects in the presence of unmeasured mediator-outcome confounding by leveraging heteroskedasticity restrictions on the observed data law. For inference, we develop semiparametric estimators that remain consistent under partial misspecifications of the observed data model. We illustrate the proposed estimators using simulations and an application that evaluates the effect of self-efficacy on fatigue among health care workers during the COVID-19 outbreak.

Key words and phrases: Causal Inference, multiple robustness, natural direct

Statistica Sinica: Preprint
doi:10.5705/ss.202021.0354

effect; natural indirect effect; unmeasured confounding.

1. Introduction

Researchers in the health and social sciences often wish to investigate not only the total effect of a point exposure A on an outcome Y, but also the direct and indirect effects operating through a given post-exposure mediating variable M. Since the seminal work of Baron and Kenny (1986) in the context of linear structural equation models, the notions of natural direct effects (NDEs) and natural indirect effects (NIEs) have been formalized in the context of a binary exposure under the potential outcomes framework (Robins and Greenland, 1992; Pearl, 2001). NDEs and NIEs are particularly useful for understanding the causal mediation mechanism, because the sum of these two effects is the average treatment effect of A on Y. Under the sequential ignorability assumption of no unmeasured confounding for the A-M, A-Y, and M-Y relationships (See Section 2 for a more formal treatment), the NDE and NIE can be identified nonparametrically from the observed data distribution based on the so-called mediation formula (Pearl, 2001; VanderWeele and Vansteelandt, 2009; Imai et al., 2010b; Tchetgen Tchetgen and Shpitser, 2014; VanderWeele, 2015). Sequential ignorability is often stated as a conditional version within the strata of a set of measured baseline covariates X not affected by the exposure, in the hope

Statistica Sinica: Preprint
doi:10.5705/ss.202021.0354

that no residual unmeasured confounding remains within the strata of the measured covariates.

A fully parametric approach to evaluating the mediation formula typically entails specifying models for both $E(Y|M, A, X)$ and $E(M|A, X)$ under appropriate link functions (VanderWeele and Vansteelandt, 2009; VanderWeele, 2015), which may be sensitive to model misspecifications. On the other hand, nonparametric inference yields multiply robust estimators that remain consistent and asymptotically normal (CAN) within various strict subsets of the observed data model, which includes the conditional densities $f(M|A, X)$ or $f(A|X)$ (Tchetgen Tchetgen and Shpitser, 2014). As a compromise between the fully parametric and nonparametric approaches, consider the following semiparametric partially linear outcome and mediator models, indexed by $\theta = (\theta_1, \theta_2, \theta_3)^{\mathrm{T}} \in \mathbb{R}^3$:

$$\begin{aligned} E(Y|M, A, X; \theta_1, \theta_2, g) &= \theta_1 M + \theta_2 A + g(X); \\ E(M|A, X; \theta_3, h) &= \theta_3 A + h(X), \end{aligned} \tag{1.1}$$

where the confounding effects of the measured covariates on the outcome and mediator are encoded by the functions $g(\cdot)$ and $h(\cdot)$ respectively, which remain unspecified. If X is sufficiently rich so that sequential ignorability holds, then θ_2 and $\theta_1\theta_3$ capture the NDE and NIE, respectively, per unit change in the exposure (Hines et al., 2021). Let $\pi(X) \equiv E(A|X)$ denote

Statistica Sinica: Preprint
doi:10.5705/ss.202021.0354

the treatment propensity score and $O = (Y, A, M, X)$ the observed data. To estimate θ, Hines et al. (2021) considered the 3×1 vector estimating function $\varphi(O; \theta, \pi, g, h)$ with components

$$\begin{aligned} \varphi_1(O; \theta, \pi, g) &= \{A - \pi(X)\}\{Y - \theta_1 M - \theta_2 A - g(X)\} \\ \varphi_2(O; \theta, h, g) &= \{M - \theta_3 A - h(X)\}\{Y - \theta_1 M - \theta_2 A - g(X)\} \\ \varphi_3(O; \theta, \pi, h) &= \{A - \pi(X)\}\{M - \theta_3 A - h(X)\}. \end{aligned} \tag{1.2}$$

An augmented G-estimator (Robins, 1994) of θ may be constructed as the solution to the empirical moment condition

$$n^{-1} \sum_{i=1}^{n} \varphi(O_i; \theta, \hat{\pi}, \hat{g}, \hat{h}) = 0,$$

where $(\hat{\pi}, \hat{g}, \hat{h})$ is a first-stage estimator of the nuisance parameters under user-specified parametric models. Provided that any pair of nuisance parameters in $\{\pi(x), g(x), h(x)\}$ is correctly modeled, Hines et al. (2021) showed that the resulting augmented G-estimator is CAN for the true value of θ defined under the partially linear model (1.1).

1.1 Motivation and related work

Although, in principle, one can rule out unmeasured confounding of the A-M and A-Y relationships by design when the exposure assignment is randomized (possibly within strata of a known set of measured baseline covariates), it is often infeasible to directly manipulate the mediator. As a

Statistica Sinica: Preprint
doi:10.5705/ss.202021.0354

result, numerous researchers have developed sensitivity analysis (Imai et al., 2010a; VanderWeele, 2010; Tchetgen Tchetgen and Shpitser, 2012; Ding and Vanderweele, 2016) and partial identification approaches (Sjölander, 2009; Robins and Richardson, 2010a) to assess the impact of departures from the no unmeasured M-Y confounding assumption. Identifying causal mediation mechanisms under unmeasured M-Y confounding can sometimes be achieved using the principal stratification approach (Gallop et al., 2009; Mattei and Mealli, 2011) or by leveraging ancillary variables that satisfy certain exclusion restrictions (Imai et al., 2013; Burgess et al., 2015; Frölich and Huber, 2017). Another major strand of work in the health sciences uses baseline covariates interacted with random exposure assignments as instrumental variables for the effect of the mediator on the outcome (Ten Have et al., 2007; Dunn and Bentall, 2007; Albert, 2008; Small, 2012; Zheng and Zhou, 2015); see also the commentary by Ogburn (2012).

Recently, there has been growing interest in econometrics and the health sciences in using higher-order moment restrictions as a source of identification in linear structural models without exclusion restrictions (Rigobon, 2003; Klein and Vella, 2010; Lewbel, 2012; Tchetgen Tchetgen et al., 2021). To the best of our knowledge, the work of Fulcher et al. (2019) was the first to extend this identification framework to causal mediation analysis with

Statistica Sinica: Preprint
doi:10.5705/ss.202021.0354

unmeasured M-Y confounding. They considered identifying and estimating the NIE under structural assumptions that imply the semiparametric partially linear model

$$\begin{aligned} E(Y|M, A, X, U; \theta_1, \theta_2, g) &= \theta_1 M + \theta_2 A + g(X, U); \\ E(M|A, X, U; \theta_3, h) &= \theta_3 A + h(X, U), \end{aligned} \tag{1.3}$$

where U is a set of unmeasured baseline covariates, not affected by the exposure, that confounds the M-Y relationship. Under further assumptions, formalized in Section 2, the parameters $\theta_1\theta_3$ and θ_2 encode the NIE and NDE, respectively, per unit increase in the exposure, which provides a useful summary of the mediation effects. The unspecified functions $g(\cdot)$ and $h(\cdot)$ now encode the confounding effects of both the measured and the unmeasured covariates on the outcome and the mediator, respectively. We extend the results of Fulcher et al. (2019) to identify $\theta = (\theta_1, \theta_2, \theta_3)^{\mathrm{T}}$ (and, hence, both the NDE and the NIE) under the partially linear model (1.3). Furthermore, similarly to Hines et al. (2021), we propose augmented G-estimators that remain CAN for the true value of θ defined by (1.3) if any one of three strict subsets of the nuisance parameters lie in user-specified parametric models, including one in which the parametric models for the nuisance parameters considered by Fulcher et al. (2019) are correctly specified. This marks a significant improvement in robustness to model misspec-

Statistica Sinica: Preprint
doi:10.5705/ss.202021.0354

ification, which is especially useful in observational studies when X contains numerous continuous components.

The rest of the paper is organized as follows. In Section 2, we introduce the formal identification conditions for θ under the partially linear model (1.3), and present multiply robust augmented G-estimation methods in Section 3. We evaluate the finite-sample performance of the proposed methods using simulation studies in Section 4, and illustrate the proposed approach by means of a real-data example in Section 5. We explore several possible extensions in Section 6 including allowing for A-M interactions in the outcome model, before ending with a brief discussion in Section 7.

2. Notation and assumptions

We use the potential outcomes framework (Neyman, 1923; Rubin, 1974) to define the mediation effects of interest. Let M_a denote the mediator value that would be observed had the exposure A been set, possibly contrary to fact, to level a. Similarly, let $Y_{a,m}$ denote the potential outcome that would be observed had A been set to level a, and M to m. The population NDE and NIE of A on Y comparing two exposure levels a and a' are given by $\text{NDE}(a,a') \equiv E(Y_{a,M_{a'}} - Y_{a',M_{a'}})$ and $\text{NIE}(a,a') \equiv E(Y_{a,M_a} - Y_{a,M_{a'}})$, respectively (VanderWeele, 2015). The NDE and NIE are particularly relevant for describing the underlying mechanism by which the exposure op-

Statistica Sinica: Preprint
doi:10.5705/ss.202021.0354

erates, because their sum is equal to the population total effect given by $E(Y_{a,M_a} - Y_{a',M_{a'}})$. Under the sequential ignorability assumption that for any (a, a', m),

$$Y_{a,m} \perp A|X, \quad M_a \perp A|X, \quad Y_{a,m} \perp M|(A, X), \quad Y_{a,m} \perp M_{a'}|X, \tag{2.1}$$

where $B \perp C|D$ indicates the conditional independence of B and C, given D (Dawid, 1979), the mediation effects are nonparametrically identified from the observed data distribution as the functionals

$$\begin{aligned} \text{NDE}(a, a') &= \iint \{E(Y|a, m, x) - E(Y|a', m, x)\} f(m|a', x) f(x) dm dx; \\ \text{NIE}(a, a') &= \iint E(Y|a, m, x) \{f(m|a, x) - f(m|a', x)\} f(x) dm dx, \end{aligned} \tag{2.2}$$

for all a and a' (Pearl, 2001; Imai et al., 2010b; VanderWeele, 2015). Evaluating (2.2) together with the partially linear model (1.1) yields $\text{NDE}(a, a') = \theta_2(a - a')$ and $\text{NIE}(a, a') = \theta_1\theta_3(a - a')$; hence, the NDE and NIE are identified, as long as θ is identified.

2.1 Identification under unmeasured M-Y confounding

In practical settings, it is often infeasible to randomize or intervene on the mediator. Because unmeasured M-Y confounding can seldom be ruled out, we assume that (2.1) holds only conditional on (X, U), that is, for any

Statistica Sinica: Preprint
doi:10.5705/ss.202021.0354

(a, a', m),

$$Y_{a,m} \perp A|(X,U), \quad M_a \perp A|(X,U), \quad Y_{a,m} \perp M|(A,X,U), \quad Y_{a,m} \perp M_{a'}|(X,U). \tag{2.3}$$

In addition, we assume that the exposure is randomly assigned, either by design or through some natural experiments, so that

$$A \perp U|X. \tag{2.4}$$

Figure 1 depicts the causal diagram for such a scenario. Under the latent sequential ignorability assumption (2.3), it is straightward to verify that the mediation effects are now given by the functionals

$$\begin{aligned} \text{NDE}(a,a') &= \iint \{E(Y|a,m,x,u) - E(Y|a',m,x,u)\} f(m|a',x,u) f(x,u) dm dx du; \\ \text{NIE}(a,a') &= \iint E(Y|a,m,x,u) \{f(m|a,x,u) - f(m|a',x,u)\} f(x,u) dm dx du. \end{aligned} \tag{2.5}$$

Evaluating (2.5) together with the partially linear model (1.3) yields $\text{NDE}(a,a') = \theta_2(a-a')$ and $\text{NIE}(a,a') = \theta_1\theta_3(a-a')$, a result given in Fulcher et al. (2019). The partially linear model (1.3) with a randomized exposure that satisfies (2.4) yields the conditional mean independence restrictions

$$\begin{aligned} E\{Y - \theta_1 M - \theta_2 A|A,X\} &= E\{g(X,U)|A,X\} = g^*(X); \\ E\{M - \theta_3 A|A,X\} &= E\{h(X,U)|A,X\} = h^*(X). \end{aligned} \tag{2.6}$$

Statistica Sinica: Preprint
doi:10.5705/ss.202021.0354

The main challenge with identification and estimation based on (2.6) is that there are two restrictions, but three unknown parameters in θ. If U is null, and thus $g(X,U) = g^*(X)$ almost surely, Hines et al. (2021) derives augmented G-estimators of θ based on $E\{Y-\theta_1 M-\theta_2 A|A,M,X\} = g^*(X)$, which is a stronger version of the first restriction in (2.6). However, this restriction fails to hold when U is non-null, because $E\{g(X,U)|A,M,X\}$ remains a function of (A,M,X), owing to collider bias at M within strata of X, as shown in Figure 1. Therefore, we do not impose this restriction, and instead leverage the conditional covariance mean independence restriction

$$E[\{M-\theta_3 A-h^*(X)\}\{Y-\theta_1 M\}|A,X] = E[\text{cov}\{g(X,U),h(X,U)\}|A,X] = \rho(X),$$

which holds under (1.3) and (2.4). We summarize the observed data restrictions below.

Lemma 1. *Under the partially linear model (1.3) and a randomized exposure that satisfies (2.4), the conditional mean independence restriction*

$$E\{\psi(O;\theta,h^*)|A,X\} = E\{\psi(O;\theta,h^*)|X\} \tag{2.7}$$

holds almost surely, where $\psi(O;\theta,h^*)$ is a 3×1 vector function with components $\psi_1(O;\theta) = \{Y-\theta_1 M-\theta_2 A\}$, $\psi_2(O;\theta,h^*) = \{M-\theta_3 A-h^*(X)\}\{Y-\theta_1 M\}$, and $\psi_3(O;\theta) = \{M-\theta_3 A\}$.

Statistica Sinica: Preprint
doi:10.5705/ss.202021.0354

For identification, we also require that (2.7) have a unique solution for θ. This may be partly justified by the linearity of the first and third components of $\psi(O;\theta,h^*)$. The second component is nonlinear in θ and, therefore, requires higher moment restrictions for identification. Following Fulcher et al. (2019), we assume that the observed data distribution satisfies the heteroskedasticity condition that for any pair of exposure values (a,a'),

$$\text{var}(M|A=a,X) \neq \text{var}(M|A=a',X) \text{ if } a \neq a', \tag{2.8}$$

almost surely. We recommend performing the Breusch–Pagan test for heteroskedasticity (Breusch and Pagan, 1979) prior to the analysis using the proposed method. Condition (2.8) may be motivated from the linear structural equation (Pearl, 2000)

$$M = \lambda(A,X,U,\epsilon) = \lambda_0(\epsilon)A + \lambda_1(X,U,\epsilon), \tag{2.9}$$

where $\lambda_0(\cdot)$ and $\lambda_1(\cdot)$ are unspecified functions, and ϵ is a latent error that satisfies $\epsilon \perp (A,X,U)$. Note that (2.9) implies the mediator partially linear model in (1.3). Let $\tilde{\lambda}_0(\epsilon) \equiv \lambda_0(\epsilon) - E\{\lambda_0(\epsilon)\}$ and $\tilde{\lambda}_1(X,U,\epsilon) \equiv \lambda_1(X,U,\epsilon) - E\{\lambda_1(X,U,\epsilon)|X\}$. Then, the conditional variance $\text{var}(M|A,X) = E[\{\tilde{\lambda}_0(\epsilon)A + \tilde{\lambda}_1(X,U,\epsilon)\}^2|A,X]$ depends on A, provided that $\lambda_0(\epsilon)$ depends on the latent source of the effect heterogeneity ϵ. Therefore, condition (2.8) does

Statistica Sinica: Preprint
doi:10.5705/ss.202021.0354

not hold under (2.9) and when the exposure has no effect on the mediator for any individual. However, it does hold when there is a heterogeneous exposure effect on the mediator in the population, a plausible scenario in a variety of health and social sciences settings (Tchetgen Tchetgen et al., 2021) that permits the average exposure effect on the mediator to be zero, that is, $E\{\lambda_0(\epsilon)\} = 0$.

Theorem 1. *Under the partially linear model (1.3) and assumptions (2.4) and (2.8), the parameter θ is identified as the unique solution to*

$$E\{\psi(O;\theta,h^*)|A,X\} = E\{\psi(O;\theta,h^*)|X\}.$$

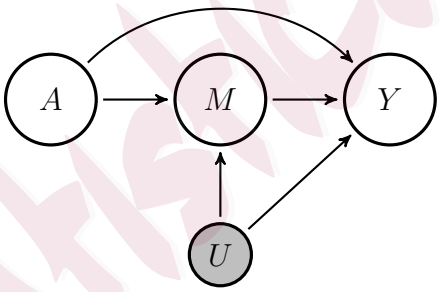

Figure 1: Causal diagram with unmeasured mediator-outcome confounding within strata of X.

3. Semiparametric inference

The conditional mean independence restriction (2.7) implies the follow-

Statistica Sinica: Preprint
doi:10.5705/ss.202021.0354

ing unconditional moment condition for θ;

$$0 = E[\{A - \pi(X)\}\psi(O; \theta, h^*)], \tag{3.1}$$

which depends on the unknown nuisance parameters $\pi(x)$ and $h^*(x)$. In principle, it is possible to estimate $\pi(x)$ and $h^*(x)$ nonparametrically under sufficient smoothness conditions (Ai and Chen, 2003; Newey and Powell, 2003). However, if X contains numerous continuous components, the resulting estimators of θ typically exhibit poor finite-sample behavior in moderately sized samples, because the data are too sparse to be able to conduct a stratified estimation (Robins and Ritov, 1997). This setting is of particular relevance when the analyst considers a broad collection of covariates and their functional forms to render condition (2.4) plausible in observational studies.

As a remedy, following the augmented G-estimation approach (Robins, 1994), we propose estimators of θ that remain CAN if various strict subsets of the nuisance parameters $\eta = \{\pi(x), g^*(x), \rho(x), h^*(x)\}$ are correctly modeled. To this end, we derive the influence function of any regular and asymptotically linear estimator of θ based on (3.1) when $\{\pi(x), h^*(x)\}$ is estimated nonparametrically (Newey, 1994).

Theorem 2. *The influence function of any regular and asymptotically lin-*

Statistica Sinica: Preprint
doi:10.5705/ss.202021.0354

ear estimator of θ *when* $\{\pi(x), h^*(x)\}$ *is estimated nonparametrically is given by* $-\Delta^{-1}\tilde{\varphi}(O;\theta,\eta)$*, where* $\Delta \equiv E[\{A-\pi(X)\}\partial\psi(O;\theta,h^*)/\partial\theta]$*, and* $\tilde{\varphi}(O;\theta,\eta)$ *is a* 3×1 *vector function with components*

$$\begin{aligned}\tilde{\varphi}_1(O;\theta,\pi,g^*) &= \{A-\pi(X)\}\{Y-\theta_1 M-\theta_2 A-g^*(X)\}\\ \tilde{\varphi}_2(O;\theta,\eta) &= \{A-\pi(X)\}\{M-\theta_3 A-h^*(X)\}\{Y-\theta_1 M-\theta_2 A-g^*(X)\}\\ &\quad -\rho(X)\\ \tilde{\varphi}_3(O;\theta,\pi,h^*) &= \{A-\pi(X)\}\{M-\theta_3 A-h^*(X)\}.\end{aligned} \tag{3.2}$$

The first and third components of $\tilde{\varphi}(O;\theta,\eta)$ are the same as those in (1.2); the second component differs because of its reliance on a different conditional mean independence restriction, as discussed in Section 2.1. The proof of Theorem 2 in the Supplementary Material shows that $\tilde{\varphi}(O;\theta,\eta)$ is also equal to the original identifying moment condition (3.1), augmented with the nonparametric influence functions for the estimation of $\{\pi(x), h^*(x)\}$. This yields the so-called orthogonal moment condition, which is locally robust to the nuisance parameters on which it depends (Chernozhukov et al., 2020).

3.1 Multiply robust estimation

We propose a semiparametric estimation of θ based on the estimating

Statistica Sinica: Preprint
doi:10.5705/ss.202021.0354

function $\tilde{\varphi}(O;\theta,\eta)$ evaluated under the working parametric models

$$\{\pi(x;\eta_1), g^*(x;\eta_2), \rho(x;\eta_3), h^*(x;\eta_4) : \eta = (\eta_1^{\mathrm{T}}, \eta_2^{\mathrm{T}}, \eta_3^{\mathrm{T}}, \eta_4^{\mathrm{T}})^{\mathrm{T}} \in \mathbb{R}^{\mathrm{q}}\},$$

with $q < \infty$. Under regularity conditions, the proposed semiparametric estimator of θ is shown to be CAN if one, but not necessarily more than one, of the following model assumptions hold:

$\mathcal{M}_1$: The models for $\{\pi(x), g^*(x)\}$ are correct.

$\mathcal{M}_2$: The models for $\{\pi(x), h^*(x)\}$ are correct.

$\mathcal{M}_3$: The models for $\{g^*(x), \rho(x), h^*(x)\}$ are correct.

For estimation purposes, suppose that $(O_1, ..., O_n)$ are independent and identically distributed (i.i.d.) observations. Let $\hat{E}(\cdot)$ denote the empirical mean operator $\hat{E}\{h(O)\} = n^{-1}\sum_{i=1}^{n} h(O_i)$. We propose a joint estimation of the parameters (θ, η). Here, the estimator $\hat{\theta} = (\hat{\theta}_1, \hat{\theta}_2, \hat{\theta}_3)^{\mathrm{T}}$ solves

$$0 = \hat{E}\{\tilde{\varphi}(O;\theta,\hat{\eta}(\theta))\}, \tag{3.3}$$

with $\hat{\eta}(\theta)$ solving $0 = \hat{E}\{\gamma(O;\theta,\eta)\}$ for a fixed value of θ, and $\gamma(O;\theta,\eta)$ is

Statistica Sinica: Preprint
doi:10.5705/ss.202021.0354

a $q \times 1$ vector function with components

$$\begin{aligned}
\gamma_1(O;\eta) &= \{\partial\pi(X;\eta_1)/\partial\eta_1\}^{\mathrm{T}}\{A - \pi(X;\eta_1)\} \\
\gamma_2(O;\theta,\eta) &= \{\partial g^*(X;\eta_2)/\partial\eta_2\}^{\mathrm{T}}\{Y - \theta_1 M - \theta_2 A - g^*(X;\eta_2)\} \\
\gamma_3(O;\theta,\eta) &= \{\partial\rho(X;\eta_3)/\partial\eta_3\}^{\mathrm{T}}[\{M - \theta_3 A - h^*(X;\eta_4)\}\times \\
&\qquad \{Y - \theta_1 M - \theta_2 A - g^*(X;\eta_2)\} - \rho(X;\eta_3)] \\
\gamma_4(O;\theta,\eta) &= \{\partial h^*(X;\eta_4)/\partial\eta_4\}^{\mathrm{T}}\{M - \theta_3 A - h^*(X;\eta_4)\}.
\end{aligned} \tag{3.4}$$

Lemma 2. *Let $\theta^\dagger$ denote the unique solution to (3.1). Then, under the regularity conditions stated in the Appendix, $n^{1/2}(\hat\theta - \theta^\dagger) \xrightarrow{d} \mathcal{N}(0,\Sigma)$ as $n \to \infty$ in the union model $\{\cup_{j=1}^3 \mathcal{M}_j\}$ (multiple robustness), where*

$$\Sigma = E\left(\left[E\left\{\frac{\partial}{\partial\theta}\tilde\Phi(O;\theta,\bar\eta(\theta^\dagger))\Big|_{\theta=\theta^\dagger}\right\}^{-1}\tilde\Phi(O;\theta^\dagger,\bar\eta(\theta^\dagger))\right]^{\otimes 2}\right),$$

$\bar\eta(\theta)$ denotes the probability limit of $\hat\eta(\theta)$, and

$$\tilde\Phi(O;\theta,\eta) = \tilde\varphi(O;\theta,\eta) - E\left\{\frac{\partial}{\partial\eta}\tilde\varphi(O;\theta,\eta)\right\}E\left\{\frac{\partial}{\partial\eta}\gamma(O;\theta,\eta)\right\}^{-1}\gamma(O;\theta,\eta).$$

As pointed out by a reviewer, $E\left\{\partial\tilde\varphi(O;\theta^\dagger,\eta)/\partial\eta\,|_{\eta=\bar\eta(\theta^\dagger)}\right\} = 0$ at the intersection submodel $\{\cap_{j=1}^3 \mathcal{M}_j\}$, where all of the working models for the nuisance parameters are correctly specified. As such, the estimation of η has no first-order impact on the asymptotic variance of $\hat\theta$. This simplification does not hold in general when one or more of the working models is misspecified (Vermeulen and Vansteelandt, 2015). For inference, a consistent

Statistica Sinica: Preprint
doi:10.5705/ss.202021.0354

estimator $\hat{\Sigma}$ of Σ may be constructed by replacing all expected values with the empirical means evaluated at $\hat{\theta}$ and $\hat{\eta}(\hat{\theta})$. Then, a 95% Wald confidence interval for the NDE per unit change in the exposure is found by calculating $\hat{\theta}_2 \pm 1.96\hat{\sigma}_2$, where $\hat{\sigma}_2$ is the square root of the second component of the diagonal of $n^{-1}\hat{\Sigma}$. We can perform a similar inference for the NIE per unit change in the exposure using the multivariate delta method. Alternatively, a nonparametric bootstrap may also be used to obtain estimates of Σ.

Remark 1. Because the nuisance parameters in $\mathcal{M}_k$, for $k = 1, 2, 3$, are variation independent, the proposed estimation framework provides the analyst with three genuine opportunities, instead of one, to obtain valid inferences about θ and the functionals thereof, even under partial model misspecifications (Robins and Rotnitzky, 2001). Chernozhukov et al. (2018, 2020) established general regularity conditions for the $n^{1/2}$-consistent estimation of finite-dimensional parameters of interest based on orthogonal moment functions, such as $\tilde{\varphi}(O; \theta, \eta)$, even when the complexity of the nuisance parameter space for η is no longer tractable using standard empirical process methods (e.g., Vapnik–Chervonenkis and Donsker classes). We plan to pursue this in future work.

Remark 2. The NIE estimator of $\theta_1\theta_3$ proposed by Fulcher et al. (2019) may be viewed as solving the empirical versions of only the second and third

Statistica Sinica: Preprint
doi:10.5705/ss.202021.0354

components in (3.1), with both $\pi(x)$ and $h^*(x)$ estimated parametrically. It is clear that (3.1) does not have mean zero, and therefore fails to identify (θ_1, θ_3) when either or both parametric models for $\pi(x)$ and $h^*(x)$ are misspecified, that is, the NIE estimator proposed by Fulcher et al. (2019) is CAN only in the semiparametric model $\mathcal{M}_2$. The proposed estimator $\hat{\theta}$ extends the estimation approach of Fulcher et al. (2019) in two main ways, by delivering a $\sqrt{n}$-consistent inference about θ (and, hence, both the NDE and NIE) in the larger semiparametric union model $\{\cup_{j=1}^{3}\mathcal{M}_j\}$.

4. Simulation study

We perform simulations to study the pointwise properties of $\hat{\theta}$ and the associated confidence intervals. We generate the baseline covariates $X_1 \sim \mathcal{N}(0,1)$ and $X_2 \sim \mathcal{N}(0,1)$ independently, followed by

$$U|X \sim \mathcal{N}\{\mu = 1 + X_1 - 0.3X_2, \sigma^2 = \exp(-1.2 + 0.8X_1 - 0.2X_2)\};$$

$$A|X \sim \text{Bernoulli}[p = \{1 + \exp(1 - 1.5X_1 + 0.3X_2)\}^{-1}];$$

$$M = 1 + (1.5 + \epsilon)A + 0.5U, \quad Y = 1 + A + 2M + U.$$

Here, U is an unmeasured factor that induces mediator-outcome dependence. In addition, we introduce latent effect heterogeneity generated independently as $\epsilon \sim N(0,1)$ so that condition (2.8) holds. The true NDE and NIE for a unit change of exposure value are one and three respectively. In addition to the proposed multiply robust estimator `MR`, we implement the

Statistica Sinica: Preprint
doi:10.5705/ss.202021.0354

propensity score-based estimator `PS` of Fulcher et al. (2019) and the product of coefficients estimator `BK` of Baron and Kenny (1986), which does not account for unmeasured M-Y confounding. We evaluate the estimators under the following five scenarios to investigate the impact of a model misspecification on the nuisance parameters: (i) $\{\pi(x), g^*(x), \rho(x), h^*(x)\}$ are all correctly modeled; (ii) only $\{\pi(x), g^*(x)\}$ are correctly modeled; (iii) only $\{\pi(x), h^*(x)\}$ are correctly modeled; (iv) only $\{g^*(x), \rho(x), h^*(x)\}$ are correctly modeled; and (v) none of the nuisance parameters are correctly modeled. Further details on the model specifications are provided in the Supplementary Material. A model is misspecified if the standardized versions of the transformed variables $[\exp(0.5X_1), 10 + X_2/\{1 + \exp(X_1)\}]$ are used as regressors instead of (X_1, X_2). Standard errors are obtained using the empirical sandwich estimator.

We simulate 1000 replicates with sample sizes $n = 400$, 800 for each scenario, and summarize the results in Table 1 for the estimation of NDE and NIE for a change of exposure value from zero to one. The `MR` and `PS` estimators perform similarly in terms of their absolute bias and coverage in scenarios (i) and (iii), but `MR` yields noticeably smaller absolute biases and better coverage than `PS` in scenarios (ii) and (iv), where the model for either $\pi(x)$ or $h^*(x)$ is misspecified. When none of the nuisance parameters

Statistica Sinica: Preprint
doi:10.5705/ss.202021.0354

are correctly modeled, all estimators show bias with a coverage proportion below the nominal value in the estimation of either the NDE or the NIE. In general, `PS` is less efficient than `MR`, because the latter incorporates additional regression models that capture the associations between (Y, M) and (A, X). The estimator `BK` shows large bias and poor coverage across scenarios (i)–(v), supporting the theory.

We also perform simulations with correctly specified models for the nuisance parameters under two additional scenarios: (vi) weaker dependence of $\text{var}(M|A, X)$ on A; and (vii) no unmeasured M-Y confounding. For (vii), we compare the proposed approach with two competing methods under the partially linear model (1.1), namely, the product of coefficients estimator `BK` and the triply robust G-estimator `TG` of Hines et al. (2021). The simulation design and results for these two scenarios are included in the Supplementary Material. Compared with the results in scenario (i), the bias and variance of `MR` and `PS` increase in (vi), while their coverage remains close to the nominal level. In scenario (vii), all estimators yield negligible bias and good coverage, but `MR` and `PS` have larger variances than those of `BK` and `TG`.

Statistica Sinica: Preprint
doi:10.5705/ss.202021.0354

Table 1: Summary of results for the estimation of the NDE and NIE for a change of exposure value from zero to one. The results in the first and second rows for each estimator correspond to the sample sizes $n = 400$ and 800, respectively.

	Scenario (i)			Scenario (ii)			Scenario (iii)			Scenario (iv)			Scenario (v)		
	MR	PS	BK	MR	PS	BK	MR	PS	BK	MR	PS	BK	MR	PS	BK
								NIE							
Bias	−.004	−.008	.740	−.011	−.008	.897	−.006	.073	.881	.069	−.137	.740	.031	.126	1.046
	.004	.000	.741	.003	.083	.885	.002	.000	.898	.079	−.121	.741	.042	.136	1.049
$\sqrt{\text{Var}}$	.257	.266	.287	.297	.266	.303	.258	.268	.310	.262	.303	.287	.268	.273	.338
	.167	.174	.191	.168	.178	.209	.168	.174	.204	.171	.195	.191	.174	.181	.230
$\sqrt{\text{EVar}}$	.244	.256	.274	2.904	.256	.290	.245	.258	.294	.250	.298	.274	.256	.263	.318
	.172	.179	.208	.173	.182	.218	.174	.179	.217	.176	.203	.208	.180	.185	.236
Cov95	.939	.932	.235	.939	.932	.127	.941	.927	.143	.935	.918	.235	.939	.914	.095
	.956	.954	.036	.954	.925	.011	.956	.954	.011	.929	.920	.036	.946	.893	.005
								NDE							
Bias	.015	.019	−.730	.018	.019	−.808	.017	−.062	−.730	.092	.298	−.730	.130	.035	−.808
	.005	.009	−.733	.006	−.074	−.809	.007	.009	−.733	.082	.282	−.733	.118	.024	−.809
$\sqrt{\text{Var}}$	.154	.173	.138	.162	.173	.154	.158	.166	.138	.160	.252	.138	.182	.172	.154
	.102	.112	.092	.104	.110	.103	.107	.112	.092	.105	.162	.092	.122	.110	.103
$\sqrt{\text{EVar}}$	.151	.168	.129	.158	.168	.146	.154	.162	.129	.157	.253	.129	.177	.165	146
	.103	.115	.091	.105	.112	.103	.108	.115	.091	.107	.169	.091	.122	.114	.103
Cov95	.950	.949	.000	.951	.949	.000	.953	.902	.000	.938	.834	.000	.919	.950	.000
	.949	.944	.000	.952	.875	.000	.957	.944	.000	.904	.655	.000	.864	.951	.000

Note: Bias and $\sqrt{\text{Var}}$ are the Monte Carlo bias and standard deviation, respectively, of the point estimates, $\sqrt{\text{EVar}}$ is the square root of the mean of the variance estimates, and Cov95 is the coverage proportion of the 95% Wald confidence interval, based on 1000 replicates.

Statistica Sinica: Preprint
doi:10.5705/ss.202021.0354

5. Application

We apply the proposed methods to reanalyze an observational study that investigated the mediating effect of posttraumatic stress disorder (PTSD) symptoms in the association between self-efficacy and fatigue among health care workers during the COVID-19 outbreak (Hou et al., 2020). The cross-sectional data were collected between March 13 and 20, 2020, from $n = 527$ health care workers in Anqing City, Anhui Province, China, which borders Hubei province, the epicenter of the COVID-19 outbreak. We refer interested readers to Hou et al. (2020) for further details on the study design.

For this illustration, the continuous exposure A is the standardized total score on the General Self-Efficacy Scale. We also consider the binary exposure A, which takes the value one if self-efficacy is above the sample median of total scores on the General Self-Efficacy Scale, and zero otherwise. PTSD symptoms (M) and fatigue (Y) are standardized total scores on the PTSD Checklist-Civilian Version and 14-item Fatigue Scale, respectively. The vector of observed baseline covariates X consists of an intercept, age, gender, marital status, education level, work experience (in years), and seniority, as well as the level of negative coping, dichotomized at the sample median. We specify the working models $\pi(x;\eta_1) = \eta_1^{\mathrm{T}} x$ for continuous exposure or $\pi(x;\eta_1) = \{1+\exp(-\eta_1^{\mathrm{T}} x)\}^{-1}$ for a binary exposure,

Statistica Sinica: Preprint
doi:10.5705/ss.202021.0354

$g^*(x;\eta_2) = \eta_2^{\mathrm{T}} x$, $\rho(x;\eta_3) = \exp(\eta_3^{\mathrm{T}} x)$, and $h^*(x;\eta_4) = \eta_4^{\mathrm{T}} x$. We choose the main effects generalized linear models for the nuisance parameters, owing to their simplicity of illustration. In principle, the goodness-of-fit may be evaluated based on a generalized version of Akaike's information criterion (Konishi and Kitagawa, 1996). Alternatively, one may leverage the multiple robustness property in Lemma 2 to select a model for the nuisance parameters (Robins et al., 2020; Cui and Tchetgen Tchetgen, 2019; Sun et al., 2022). We acknowledge this limitation, and defer model selection to future work. Owing to the limited sample size and because negative coping (X_{nc}) has been hypothesized to be an important effect modifier of the exposure's effects on both the mediator and the outcome (Hou et al., 2020), we further specify $\theta_1(x;\beta_1) = \beta_1^{\mathrm{T}}(1, x_{nc})^{\mathrm{T}}$ and $\theta_3(x;\beta_3) = \beta_3^{\mathrm{T}}(1, x_{nc})^{\mathrm{T}}$. The Breusch–Pagan test for heteroskedasticity (Breusch and Pagan, 1979), based on identical working models for the conditional mediator mean and variance, yields p-values of 8.77×10^{-7} and 0.04 for the continuous and binary exposure, respectively, indicating that the heteroskedasticity condition (2.8) is plausible.

Table 2 shows various estimates of the NDE and NIE of self-efficacy on fatigue mediated through PTSD symptoms. With continuous exposure, the regression approach `BK` of Baron and Kenny (1986) yields an NIE esti-

Statistica Sinica: Preprint
doi:10.5705/ss.202021.0354

Table 2: Estimates ($\pm$ 1.96$\times$standard error) of the NDE and NIE of self-efficacy on fatigue mediated through PTSD symptoms.

	`MR`	`PS`	`BK`
	Continuous exposure		
NDE	$-.352 \pm .119$	$-.455 \pm .171$	$-.234 \pm .083$
NIE	$-.042 \pm .084$	$.062 \pm .130$	$-.159 \pm .053$
	Dichotomized exposure		
NDE	$-.765 \pm .340$	$-.748 \pm .332$	$-.426 \pm .146$
NIE	$.066 \pm .308$	$.049 \pm .297$	$-.273 \pm .084$

mate with the 95% confidence interval $-.159 \pm .053$. This result suggests a significant mediating effect of PTSD symptoms in reducing fatigue, which is consistent with the original findings by Hou et al. (2020). The proposed approach `MR` yields an NIE estimate close to zero, and the concomitant 95% confidence interval $-.042 \pm .084$ includes zero. This suggests that we cannot rule out a null NIE after accounting for possible unmeasured common causes of PTSD and fatigue. Dichotomizing the exposure yields qualitatively similar results.

6. Extensions

6.1 Exposure-mediator interaction

Statistica Sinica: Preprint
doi:10.5705/ss.202021.0354

In the presence of a potential A-M interaction in their effects on the outcome, we may consider the following partially linear models indexed by $\theta = (\theta_1, \theta_2, \zeta, \theta_3)^{\mathrm{T}} \in \mathbb{R}^4$:

$$\begin{aligned} E(Y|M, A, X, U; \theta_1, \theta_2, \zeta, g) &= \theta_1 M + \theta_2 A + \zeta AM + g(X, U); \\ E(M|A, X, U; \theta_3, h) &= \theta_3 A + h(X, U), \end{aligned} \tag{6.1}$$

where ζ is a scalar parameter encoding the interaction. If there is no interaction, so that $\zeta = 0$, then (6.1) reduces to (1.3). Evaluating (2.2) in conjunction with (6.1) yields $\mathrm{NDE}(a, a') = (\theta_2 + \zeta[\theta_3 a' + E\{h^*(X)\}])(a - a')$ and $\mathrm{NIE}(a, a') = \theta_3(\theta_1 + \zeta a)(a - a')$ (VanderWeele, 2015). Because the moment condition

$$0 = E[\omega(X)\{A - \pi(X)\}\psi(O; \theta, h^*)] \tag{6.2}$$

holds for an arbitrary 4×3 matrix-valued function $\omega(\cdot)$, θ is identified from (6.2) if

$$E[\omega(X)\{A - \pi(X)\}\partial\psi(O; \theta, h^*)/\partial\theta]$$

is nonsingular. A multiply robust estimator of θ may then be constructed based on the empirical version of the corresponding orthogonal moment condition. However, because of its dependence on $h^*(x)$, the proposed estimator of $\mathrm{NDE}(a, a')$ can only be doubly robust in the union model $\{\mathcal{M}_2 \cup \mathcal{M}_3\}$.

Statistica Sinica: Preprint
doi:10.5705/ss.202021.0354

6.2 Binary mediator

The proposed semiparametric framework also extends to a binary mediator under the following log-linear model:

$$\begin{aligned} E(Y|M,A,X,U;\theta_1,\theta_2,g) &= \theta_1 M + \theta_2 A + g(X,U); \\ \log\{p(M=1|A,X,U;\theta_3,h)\} &= \theta_3 A + h(X,U). \end{aligned} \tag{6.3}$$

Evaluating (2.5) in conjunction with the log-linear model (6.3) yields $\text{NDE}(a,a') = \theta_2(a-a')$ and $\text{NIE}(a,a') = \theta_1\{\exp(\theta_3 a) - \exp(\theta_3 a')\}E\{\tilde{h}(X)\}$, where

$$\tilde{h}(X) = E[Me^{-\theta_3 A}|X] = E[\exp\{h(X,U)\}|X].$$

Under the log-linear model (6.3) and a randomized exposure that satisfies (2.4), the conditional mean independence restriction

$$E\{\tilde{\psi}(O;\theta,\tilde{h})|A,X\} = E\{\tilde{\psi}(O;\theta,\tilde{h})|X\} \tag{6.4}$$

holds almost surely, where $\tilde{\psi}(O;\theta,\tilde{h})$ is a 3×1 vector function with components $\tilde{\psi}_1(O;\theta) = \{Y-\theta_1 M-\theta_2 A\}$, $\tilde{\psi}_2(O;\theta,\tilde{h}) = \{Y-\theta_1 M\}\{M\exp(-\theta_3 A)-\tilde{h}(X)\}$ and $\tilde{\psi}_3(O;\theta) = \{M\exp(-\theta_3 A)\}$. We can now perform augmented G-estimation of θ based on the identifying restriction (6.4), using methods analogous to those described in Section 3. For a binary outcome, direct extensions of existing methods to identify the NDE and NIE on the risk ratio scale generally require the conditional density $f(M|A,X,U)$ to be normal, with constant variance (VanderWeele and Vansteelandt, 2010; Valeri and

Statistica Sinica: Preprint
doi:10.5705/ss.202021.0354

VanderWeele, 2013; VanderWeele, 2015), in which case the heteroskedasticity condition (2.8) fails to hold. A possible direction for future work is to identify and estimate the NDE and NIE on the risk difference scale under the proposed framework.

7. Conclusion

Unmeasured M-Y confounding is particularly pernicious for credible causal mediation analysis in the health and social sciences, because the mediator can seldom be manipulated directly. The main contribution of this study is to propose a robust inference framework for the NDE and NIE under unmeasured M-Y confounding in partially linear models by leveraging the heteroskedasticity of M with respect to A, a condition that is empirically testable. Note that the fourth condition $Y_{a,m} \perp M_{a^*}|(X, U)$ in (2.3) cannot be guaranteed using experimental interventions, even if we are able to randomize both the exposure and the mediator (Didelez et al., 2006; Imai et al., 2013; Robins and Richardson, 2010b). In the absence of exposure-induced confounding, dropping the fourth condition from (2.3) imbues the functionals in (2.5) with alternative causal interpretations as interventional analogues of the NDE and NIE (VanderWeele et al., 2014; VanderWeele, 2015). Because our results all concern the identification and estimation of the functionals in (2.5), they can be readily applied under

Statistica Sinica: Preprint
doi:10.5705/ss.202021.0354

this alternative interpretation. We conjecture that (2.7) represents all observed data-conditional mean restrictions under the partially linear structural model (1.3) and independence assumption (2.4). The number of observed data restrictions may potentially increase under stronger structural assumptions, for example, by imposing restrictions on the structural distributions $f(Y|M, A, X, U)$ and $f(M|A, X, U)$. The proposed framework can also be extended in several other important directions, including mediation analyses with survival data and multiple mediators (Lin et al., 2017; Huang and Yang, 2017) under unmeasured mediator-outcome confounding. These topics are left to future research.

Supplementary Material

The online Supplementary Material includes proofs of our lemmas and theorems, as well as additional simulation results.

Acknowledgments

BaoLuo Sun was supported by the National University of Singapore Start-Up Grant (R-155-000-203-133). We thank the associate editor and two reviewers for their constructive comments. We also thank Eric Tchetgen Tchetgen and Xu Shi for their helpful suggestions on a previous version of this paper.

Statistica Sinica: Preprint
doi:10.5705/ss.202021.0354

References

Ai, C. and Chen, X. (2003). Efficient estimation of models with conditional moment restrictions containing unknown functions. *Econometrica*, 71(6):1795–1843.

Albert, J. M. (2008). Mediation analysis via potential outcomes models. *Statistics in medicine*, 27(8):1282–1304.

Baron, R. M. and Kenny, D. A. (1986). The moderator–mediator variable distinction in social psychological research: Conceptual, strategic, and statistical considerations. *Journal of personality and social psychology*, 51(6):1173.

Breusch, T. S. and Pagan, A. R. (1979). A simple test for heteroscedasticity and random coefficient variation. *Econometrica: Journal of the Econometric Society*, pages 1287–1294.

Burgess, S., Daniel, R. M., Butterworth, A. S., Thompson, S. G., and Consortium, E.-I. (2015). Network mendelian randomization: using genetic variants as instrumental variables to investigate mediation in causal pathways. *International journal of epidemiology*, 44(2):484–495.

Chernozhukov, V., Chetverikov, D., Demirer, M., Duflo, E., Hansen, C., Newey, W., and Robins, J. (2018). Double/debiased machine learning for treatment and structural parameters. *The Econometrics Journal*, 21(1):C1–C68.

Chernozhukov, V., Escanciano, J. C., Ichimura, H., Newey, W. K., and Robins, J. M. (2020). Locally robust semiparametric estimation.

Statistica Sinica: Preprint
doi:10.5705/ss.202021.0354

REFERENCES

Cui, Y. and Tchetgen Tchetgen, E. (2019). Selective machine learning of doubly robust functionals. *arXiv preprint arXiv:1911.02029*.

Dawid, A. P. (1979). Conditional independence in statistical theory. *Journal of the Royal Statistical Society: Series B (Methodological)*, 41(1):1–15.

Didelez, V., Dawid, A. P., and Geneletti, S. (2006). Direct and indirect effects of sequential treatments. In *Proceedings of the Twenty-Second Conference on Uncertainty in Artificial Intelligence*, UAI'06, pages 138–146, Arlington, Virginia, USA. AUAI Press.

Ding, P. and Vanderweele, T. J. (2016). Sharp sensitivity bounds for mediation under unmeasured mediator-outcome confounding. *Biometrika*, 103(2):483–490.

Dunn, G. and Bentall, R. (2007). Modelling treatment-effect heterogeneity in randomized controlled trials of complex interventions (psychological treatments). *Statistics in Medicine*, 26(26):4719–4745.

Frölich, M. and Huber, M. (2017). Direct and indirect treatment effects-causal chains and mediation analysis with instrumental variables. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 79(5):1645–1666.

Fulcher, I. R., Shi, X., and Tchetgen Tchetgen, E. J. (2019). Estimation of natural indirect effects robust to unmeasured confounding and mediator measurement error. *Epidemiology*, 30(6):825–834.

Gallop, R., Small, D. S., Lin, J. Y., Elliott, M. R., Joffe, M., and Ten Have, T. R. (2009). Mediation analysis with principal stratification. *Statistics in medicine*, 28(7):1108–1130.

Statistica Sinica: Preprint
doi:10.5705/ss.202021.0354

Hines, O., Vansteelandt, S., and Diaz-Ordaz, K. (2021). Robust inference for mediated effects in partially linear models. *Psychometrika*, pages 1–24.

Hou, T., Dong, W., Zhang, R., Song, X., Zhang, F., Cai, W., Liu, Y., and Deng, G. (2020). Self-efficacy and fatigue among health care workers during covid-19 outbreak: A moderated mediation model of posttraumatic stress disorder symptoms and negative coping. *Preprint*.

Huang, Y.-T. and Yang, H.-I. (2017). Causal mediation analysis of survival outcome with multiple mediators. *Epidemiology (Cambridge, Mass.)*, 28(3):370.

Imai, K., Keele, L., and Tingley, D. (2010a). A general approach to causal mediation analysis. *Psychological methods*, 15(4):309.

Imai, K., Keele, L., and Yamamoto, T. (2010b). Identification, inference and sensitivity analysis for causal mediation effects. *Statistical science*, pages 51–71.

Imai, K., Tingley, D., and Yamamoto, T. (2013). Experimental designs for identifying causal mechanisms. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, 176(1):5–51.

Klein, R. and Vella, F. (2010). Estimating a class of triangular simultaneous equations models without exclusion restrictions. *Journal of Econometrics*, 154(2):154–164.

Konishi, S. and Kitagawa, G. (1996). Generalised information criteria in model selection. *Biometrika*, 83(4):875–890.

Lewbel, A. (2012). Using heteroscedasticity to identify and estimate mismeasured and endoge-

Statistica Sinica: Preprint
doi:10.5705/ss.202021.0354

nous regressor models. *Journal of Business & Economic Statistics*, 30(1):67–80.

Lin, S.-H., Young, J. G., Logan, R., and VanderWeele, T. J. (2017). Mediation analysis for a survival outcome with time-varying exposures, mediators, and confounders. *Statistics in medicine*, 36(26):4153–4166.

Mattei, A. and Mealli, F. (2011). Augmented designs to assess principal strata direct effects. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 73(5):729–752.

Newey, W. K. (1994). The asymptotic variance of semiparametric estimators. *Econometrica: Journal of the Econometric Society*, pages 1349–1382.

Newey, W. K. and Powell, J. L. (2003). Instrumental variable estimation of nonparametric models. *Econometrica*, 71(5):1565–1578.

Neyman, J. (1923). Sur les applications de la théorie des probabilités aux experiences agricoles: Essai des principes. *Roczniki Nauk Rolniczych*, 10:1–51.

Ogburn, E. L. (2012). Commentary of "mediation analysis without sequential ignorability: using baseline covariates interacted with random assignment as instrumental variables". *Journal of Statistical Research*, 46:105–111.

Pearl, J. (2000). *Causality: Models, Reasoning and Inference*. Cambridge, UK: Cambridge University Press.

Pearl, J. (2001). Direct and indirect effects. In *Proceedings of the Seventeenth Conference on Uncertainty in Artificial Intelligence*, UAI'01, pages 411–420, San Francisco, CA, USA.

Statistica Sinica: Preprint
doi:10.5705/ss.202021.0354

Morgan Kaufmann Publishers Inc.

Rigobon, R. (2003). Identification through heteroskedasticity. *Review of Economics and Statistics*, 85(4):777–792.

Robins, J., Sued, M., Lei-Gomez, Q., and Rotnitzky, A. (2020). Double-robust and efficient methods for estimating the causal effects of a binary treatment. *arXiv preprint arXiv:2008.00507*.

Robins, J. M. (1994). Correcting for non-compliance in randomized trials using structural nested mean models. *Communications in Statistics-Theory and methods*, 23(8):2379–2412.

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

Robins, J. M. and Richardson, T. (2010a). Alternative graphical causal models and the identification of direct effects. In Shrout, P., Keyes, K., and Ornstein, K., editors, *Causality and Psychopathology: Finding the Determinants of Disorders and Their Cures*, pages 103–58. Oxford University Press, Oxford, UK.

Robins, J. M. and Richardson, T. S. (2010b). Alternative graphical causal models and the identification of direct effects. *Causality and psychopathology: Finding the determinants of disorders and their cures*, pages 103–158.

Robins, J. M. and Ritov, Y. (1997). Toward a curse of dimensionality appropriate (coda) asymptotic theory for semi-parametric models. *Statistics in medicine*, 16(3):285–319.

Statistica Sinica: Preprint
doi:10.5705/ss.202021.0354

REFERENCES

Robins, J. M. and Rotnitzky, A. (2001). Comment on the bickel and kwon article,"inference for semiparametric models: Some questions and an answer". *Statistica Sinica*, 11(4):920–936.

Rubin, D. B. (1974). Estimating causal effects of treatments in randomized and nonrandomized studies. *Journal of educational Psychology*, 66(5):688.

Sjölander, A. (2009). Bounds on natural direct effects in the presence of confounded intermediate variables. *Statistics in Medicine*, 28(4):558–571.

Small, D. S. (2012). Mediation analysis without sequential ignorability: using baseline covariates interacted with random assignment as instrumental variables. *Journal of Statistical Research*, 46:91–103.

Sun, B., Cui, Y., and Tchetgen Tchetgen, E. (2022). Selective machine learning of the average treatment effect with an invalid instrumental variable. *Journal of Machine Learning Research*, in press.

Tchetgen Tchetgen, E., Sun, B., and Walter, S. (2021). The genius approach to robust mendelian randomization inference. *Statistical Science*, 36(3):443–464.

Tchetgen Tchetgen, E. J. and Shpitser, I. (2012). Semiparametric theory for causal mediation analysis: efficiency bounds, multiple robustness, and sensitivity analysis. *Annals of statistics*, 40(3):1816.

Tchetgen Tchetgen, E. J. and Shpitser, I. (2014). Estimation of a semiparametric natural direct effect model incorporating baseline covariates. *Biometrika*, 101(4):849–864.

Statistica Sinica: Preprint
doi:10.5705/ss.202021.0354

Ten Have, T. R., Joffe, M. M., Lynch, K. G., Brown, G. K., Maisto, S. A., and Beck, A. T. (2007). Causal mediation analyses with rank preserving models. *Biometrics*, 63(3):926–934.

Valeri, L. and VanderWeele, T. J. (2013). Mediation analysis allowing for exposure–mediator interactions and causal interpretation: theoretical assumptions and implementation with sas and spss macros. *Psychological methods*, 18(2):137.

VanderWeele, T. J. (2010). Bias formulas for sensitivity analysis for direct and indirect effects. *Epidemiology (Cambridge, Mass.)*, 21(4):540.

VanderWeele, T. J. (2015). *Explanation in causal inference: methods for mediation and interaction*. Oxford University Press.

VanderWeele, T. J. and Vansteelandt, S. (2009). Conceptual issues concerning mediation, interventions and composition. *Statistics and its Interface*, 2(4):457–468.

VanderWeele, T. J. and Vansteelandt, S. (2010). Odds ratios for mediation analysis for a dichotomous outcome. *American journal of epidemiology*, 172(12):1339–1348.

VanderWeele, T. J., Vansteelandt, S., and Robins, J. M. (2014). Effect decomposition in the presence of an exposure-induced mediator-outcome confounder. *Epidemiology (Cambridge, Mass.)*, 25(2):300.

Vermeulen, K. and Vansteelandt, S. (2015). Bias-reduced doubly robust estimation. *Journal of the American Statistical Association*, 110(511):1024–1036.

Zheng, C. and Zhou, X.-H. (2015). Causal mediation analysis in the multilevel intervention

Statistica Sinica: Preprint
doi:10.5705/ss.202021.0354

and multicomponent mediator case. *Journal of the Royal Statistical Society: Series B: Statistical Methodology*, pages 581–615.

Department of Statistics and Data Science, National University of Singapore

E-mail: stasb@nus.edu.sg

Department of Biostatistics, University of Washington

E-mail: tingye1@uw.edu