

---

# Proximal Identification and Estimation in Front-Door Causal Structures with Unobserved Confounding of the Mediator

---

Helen Guo<sup>\*1</sup>

Beatrix Yaxin Wen<sup>2</sup>

Ilya Shpitser<sup>3</sup>

<sup>1</sup>Biostatistics Dept., Johns Hopkins Bloomberg School of Public Health, Baltimore, MD, USA

<sup>2</sup>Applied Mathematics and Statistics Dept., Johns Hopkins Whiting School of Engineering, Baltimore, MD, USA

<sup>3</sup>Computer Science Dept., Johns Hopkins Whiting School of Engineering, Baltimore, MD, USA

## Abstract

Unobserved confounding is a fundamental obstacle in causal inference problems. In the graphical modeling literature, a general theory has been developed that allows identification in the presence of hidden variables, with some limitations. In particular, Pearl’s celebrated front-door criterion allows nonparametric identification in the presence of unobserved common causes of the treatment and the outcome, however it requires the presence of an unconfounded variable that mediates all causal influence from the treatment to the outcome. This stringent requirement limits the applicability of the front-door criterion. We propose proximal generalizations of the front-door criterion, allowing both arbitrary treatment/outcome confounding, and unobserved confounders of the mediator, provided informative proxies for the latter type of confounders are observed. In addition to deriving three new identification strategies in this setting, we provide plug-in and influence function-based estimation strategies for the resulting functionals, and evaluate their performance through simulations.

## 1 INTRODUCTION

Causal effects are not identified in general in the presence of hidden common causes of treatment and outcome variables. A rich body of work has proposed strategies to overcome this obstacle via nonparametric identification theory developed in the graphical modeling literature, leading to the ID algorithm and extensions [Pearl, 1995, Tian and Pearl, 2002, Shpitser and Pearl, 2006, Huang and Valtorta, 2006], or via methods that rely on fortuitous external variables such as instruments, mediators, or proxies (i.e., causal consequences of hidden variables) [Newey and Powell, 2003, Kuroki and Pearl, 2014, Miao et al., 2018]. Recent work [Shpitser et al.,

2023] has combined graphical model-based identification theory and proximal methods to obtain identification in settings where neither approach alone suffices.

We present proximal causal inference results in a novel setting, extending Pearl’s front-door method [Pearl, 1995] to settings where unobserved common causes of the mediator, treatment, and outcome exist. Specifically, we consider a causal model represented by the *composite bow graph* in Fig. 1. This model captures a common scenario in which some hidden variable(s) confound only treatment  $A$  and outcome  $Y$ , while others confound treatment  $A$ , mediator  $\vec{M}$ , and outcome  $Y$ . Such structures arise frequently in practice, as mediators often share unmeasured causes with both treatment and outcome, thereby violating the assumptions required for identification based on the front-door criterion.

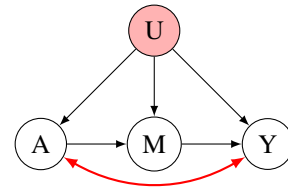


Figure 1: The composite bow graph.

We show that when the latent cause(s) affecting  $A$ ,  $\vec{M}$ , and  $Y$  in the composite bow graph admit proxies, the counterfactual distribution  $p(Y(a))$  (where  $Y(a)$  denotes the *potential outcome* had the treatment variable  $A$  taken value  $a$ ) and thus the causal effect of  $A$  on  $Y$  remains identifiable under mild structural assumptions. We provide three distinct identification strategies, each relying on a different set of structural assumptions. In addition, we develop estimators for functionals obtained by these strategies, including plug-in estimators and an influence function-based estimator. Finally, we illustrate the validity of our identification methods and performance of our proposed estimators through simulation studies.

This paper is organized as follows. Section 2 reviews rele-

vant background and outlines our contributions. Section 3 presents distinct identification strategies under different structural assumptions. Section 4 develops estimators of the causal effect under these strategies. Section 5 evaluates the performance of the proposed estimators via simulation, and Section 6 concludes this work with a discussion.

## 2 BACKGROUND

### 2.1 GRAPHICAL CAUSAL MODELS

We represent fully observed causal structures by means of models associated with directed acyclic graphs (DAG)s. In such graphs, directed edges ( $\rightarrow$ ) encode direct causal relationships between variables. Conditional independence relations implied by the model are given by the graph via *d-separation*. Such causal structures are able to represent responses of outcome variables  $\vec{Y}$  where treatment variables  $\vec{A}$  are set to values  $\vec{a}$  by means of an external intervention operation denoted by the do operator  $\text{do}(\vec{a})$  [Pearl, 2009]; these distributions are also written in the *potential outcome* notation as  $p(\vec{Y}(\vec{a}))$ . Identification of the resulting interventional distributions in causal models of a DAG is obtained by means of the *g-formula*:

$$p(\vec{Y} | \text{do}(\vec{a})) = \sum_{\vec{Y}^* \setminus \vec{Y}} \left( \prod_{V \in \vec{Y}^*} p(V | \text{pa}_{\mathcal{G}}(V)) \right) \Bigg|_{\vec{A}=\vec{a}}$$

Identification of interventional distributions leads to natural estimators of causal parameters of interest, such as the *average causal effect* (ACE):  $\mathbb{E}[Y(A=1) - Y(A=0)]$  for binary  $A$ .

Causal structures with hidden variables are represented using *acyclic directed mixed graphs* (ADMGs) containing both directed ( $\rightarrow$ ) and bidirected ( $\leftrightarrow$ ) edges, and obtained from a hidden variable DAG by means of the latent projection operation [Verma and Pearl, 1990]. In this paper we will also represent hidden variable causal models by *partial latent projections*, where some hidden variables are “projected out” from the graph (replaced by bidirected edges) and other hidden variables are explicitly retained (shown in red). The composite bow graph in Fig. 1 is an example of a partial latent projection. Conditional independence relations in these graphs are given by *m-separation criterion* [Richardson and Spirtes, 2002], the natural extension of d-separation to ADMGs.

In hidden variable models, the task of identification of interventional distributions becomes considerably more complicated. Some interventional distributions are not identified at all, while others are expressible as a functional of the observed data given by the ID algorithm [Tian and Pearl, 2002, Shpitser and Pearl, 2006, Huang and Valtorta, 2006]. A number of criteria have been developed to classify DAGs

or ADMGs where the interventional distribution is identified by a particular simple functional. We now describe two such examples.

### 2.2 THE BACK-DOOR CRITERION

The back-door criterion, introduced by Pearl [1993], provides graphical conditions under which interventional distribution  $p(Y(a))$  and thus the causal effect of  $A$  on  $Y$  is identified for binary  $A$ , along with an identifying formula.

A set of observed variables  $\vec{C}$  satisfies the back-door criterion relative to an ordered treatment-outcome pair  $(A, Y)$  in an ADMG  $\mathcal{G}$  if:

1. No node in  $\vec{C}$  is a descendant of  $A$ ; and
2.  $\vec{C}$  m-separates every back-door path between  $A$  and  $Y$  (e.g. every path from  $A$  to  $Y$  that starts with an arrowhead into  $A$ ).

For example,  $\vec{C}$  satisfies the back-door criterion relative to  $A$  and  $Y$  in the simple case shown in Fig. 2a, where  $\vec{C}$  is a common cause of  $A$  and  $Y$ . If  $\vec{C}$  satisfies the back-door criterion, the interventional distribution  $p(Y(a))$  is identified from the observed data distribution  $p(Y, A, \vec{C})$  by the *adjustment formula*, shown below in Equation 1:

$$p(Y(a) = y) = \sum_{\vec{c}} p(y|a, \vec{c})p(\vec{c}) \quad (1)$$

Had a common cause of  $A$  and  $Y$  not been observed, shown as a red vertex  $U$  in Fig. 2b, the interventional distribution  $p(Y(a))$  is generally not identified from the observed data distribution, which is now  $p(Y, A)$ .

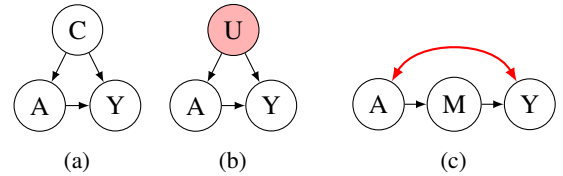


Figure 2: (a) Observed confounding; (b) Hidden confounding; (c) Observed mediation with hidden treatment-outcome confounding.

A recent line of work addressed lack of identifiability in the presence of a hidden common cause  $U$  in Fig. 2b by taking advantage of informative proxy variables for  $U$ . In particular, methods developed by Miao et al. [2018], Kuroki and Pearl [2014], and Allman et al. [2015] show that when proxies for an unobserved confounder are available (e.g.,  $W$  and  $Z$  in Fig. 3a), along with mild structural assumptions, the counterfactual distribution  $p(Y(a))$  can be identified. In particular, the adjustment formula,  $\sum_{\vec{u}} p(Y | a, \vec{u}) p(\vec{u})$ , can be re-expressed in terms of the observed data distribution, even though  $U$  is not observed directly.

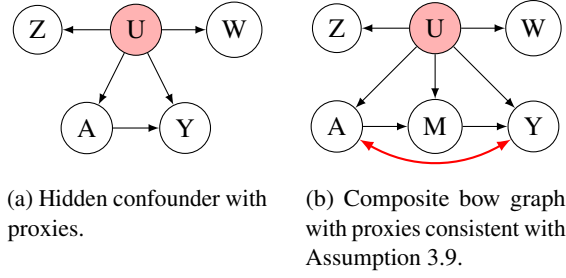


Figure 3: Causal graphs with proxy variables.

### 2.3 THE FRONT-DOOR CRITERION

Leveraging mediator(s)  $\vec{M}$  between  $A$  and  $Y$ , the front-door criterion introduced in Pearl [1995], provides alternative graphical conditions under which interventional distribution  $p(Y(a))$  is identified, while allowing for hidden common causes of treatment  $A$  and outcome  $Y$ .

A set of observed variables  $\vec{M}$  satisfies the front-door criterion relative to an ordered pair  $(A, Y)$  in an ADMG  $\mathcal{G}$  if:

1.  $\vec{M}$  intercepts all directed paths from  $A$  to  $Y$ ;
2. Every back-door path (paths starting with an arrowhead into  $A$ ) between  $A$  and  $\vec{M}$  is marginally m-separated;
3.  $A$  m-separates all back-door paths (paths starting with an arrowhead into  $\vec{M}$ ) between  $\vec{M}$  and  $Y$ .

If  $\vec{M}$  satisfies the front-door criterion, the interventional distribution  $p(Y(a))$  is identified via the *front-door formula*:

$$p(Y(a) = y) = \sum_{\vec{m}, a'} p(y | a', \vec{m}) p(\vec{m} | a) p(a') \quad (2)$$

In the ADMG in Fig. 2c, where the bidirected arc represents a hidden common cause of  $A$  and  $Y$ ,  $\vec{M}$  satisfies the front-door criterion relative to  $(A, Y)$ .

### 2.4 OUTLINE OF OUR CONTRIBUTIONS

A common criticism of the front-door criterion in practice is that the assumptions regarding the mediator variable(s) are rarely met in practical problems where unobserved confounding is a concern. Specifically, while the front-door criterion allows unobserved common causes of  $A$  and  $Y$ , it specifically precludes such causes from influencing the mediator(s)  $\vec{M}$ .

A more natural generalization of the front-door graph, which we call the *composite bow graph*, is shown in Fig. 1. In this graph,  $\vec{M}$  does not satisfy the front-door criterion due to the presence of a hidden common cause  $\vec{U}$  of  $A$ ,  $\vec{M}$ , and  $Y$ . Indeed,  $p(Y(a))$  is generally not identifiable in Fig. 1.

We first note that had  $\vec{U}$  been observed rather than hidden in Fig. 1, the interventional distribution  $p(Y(a))$  would have been identified by a mild generalization of the front-door formula given by Equation 3 [Fulcher et al., 2020]:

$$p(Y(a) = y) = \sum_{\vec{m}, a', \vec{u}} p(y | a', \vec{m}, \vec{u}) p(\vec{m} | a, \vec{u}) p(a', \vec{u}) \quad (3)$$

Next, following the work in Miao et al. [2018], Kuroki and Pearl [2014], Allman et al. [2015], and Dukes et al. [2023], we show that with the existence of proxies  $W$  and  $Z$  for  $\vec{U}$ , along with mild structural assumptions,  $p(Y(a))$  may be identified via proximal generalizations of the formula in Equation 3. We present three different sets of assumption that each guarantee identification in Section 3.

Our first two identification strategies (Sections 3.1 and 3.2) adapt the assumptions similar in spirit to those in Miao et al. [2018]. These approaches rely on distributional variation restrictions together with the existence of solutions to certain Fredholm integral equations of the first kind. This line of work has been extended by Ghassami et al. [2025] to identify causal effects leveraging proxies of hidden mediator(s), and by Dukes et al. [2023] to identify direct and indirect effects in the presence of hidden confounding. The first identification strategy we present has previously appeared in Bai et al. [2025], which gives identification and estimation results for the population indirect effect, a functional identical to that of the causal effect in our model.

Our third identification strategy (Section 3.3) is more closely related to a result from Kuroki and Pearl [2014], and Allman et al. [2015]. This strategy imposes a different set of distributional variation restrictions and enables identification of the full joint distribution, from which  $p(Y(a))$  may be identified via the g-formula. This line of work traces back to work on unique tensor (multidimensional array) decompositions, originating from Kruskal [1977].

## 3 IDENTIFICATION

We now describe the three sets of assumptions we use to allow identification of the interventional distribution  $p(Y(a))$  in the composite bow graph when latent cause(s) affecting  $A$ ,  $\vec{M}$ , and  $Y$  admit two proxies. Throughout this work, whenever we consider causal effects, the treatment variable  $A$  is assumed to be binary, taking values in  $\{0, 1\}$ .

### 3.1 ASSUMPTION SET 1

The first identification strategy may be viewed as a front-door analogue of the proximal result in Dukes et al. [2023] for direct and indirect effects, and has previously appeared in Bai et al. [2025], which established identification and semiparametric estimation results for population indirect effects. The strategy combines conditional independence

relations reflected in Fig. 4a (Assumption 3.1) along with the structural conditions in Assumptions 3.2–3.3.

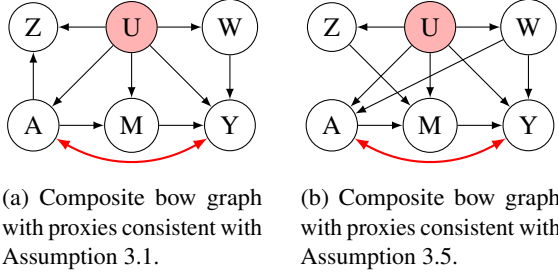


Figure 4: Composite bow graphs under different proxy assumptions.

**Assumption 3.1.** (*Conditional Independences*)

1.  $Y \perp\!\!\!\perp Z \mid \vec{U}, \vec{M}, A$
2.  $W \perp\!\!\!\perp A, \vec{M}, Z \mid \vec{U}$
3.  $W, \vec{M} \perp\!\!\!\perp Z \mid \vec{U}, A$

Assumption 3.1 may be represented graphically via the m-separation criterion in the variant of the composite bow graph shown in Fig. 4a. Note that this assumption allows association between the proxy  $Z$  and the treatment  $A$ , as well as the proxy  $W$  and the outcome  $Y$ .

**Assumption 3.2.** (*Completeness*) For each  $(a, \vec{m})$  and any square integrable  $g$ ,

1.  $\mathbb{E}\{g(\vec{u}) \mid z, a, \vec{m}\} = 0$  a.s. iff  $g(\vec{u}) = 0$  a.s.
2.  $\mathbb{E}\{g(\vec{u}) \mid z, a\} = 0$  a.s. iff  $g(\vec{u}) = 0$  a.s.

Assumption 3.2 imposes distributional variation restrictions referred to as *completeness* assumptions and can be satisfied without parametric distributional restrictions [Newey and Powell, 2003, Hu and Shiu, 2022]. The assumption states that objects in  $\{p(\vec{u}_i \mid Z, a, \vec{m}) : \vec{u}_i\}$  exhibit sufficient variation across values  $\vec{u}_i$  for each  $(a, \vec{m})$ . Applying Bayes' rule, this assumption also implies sufficient variation of distributions in  $\{p(Z \mid \vec{u}_i, a, \vec{m}) : \vec{u}_i\}$ . Note that these latter distributions are not functions of  $\vec{m}$  by Assumption 3.1. Additionally, if the dimension of  $Z$  is greater than that of  $U$ , distributions that fail to satisfy Assumption 3.2 can be made arbitrarily close in total variation distance to distributions that do satisfy it [Cui et al., 2024, Canay et al., 2013]. Since completeness assumptions are not testable, the plausibility of Assumption 3.2 relies on domain knowledge. For a discussion of completeness assumptions sufficient to understand this paper, see Guo et al. [2025].

To build intuition for the meaning of this assumption, consider a finite-support setting in which we form a stochastic matrix which has rows  $i$  corresponding to  $p(\vec{u}_i \mid Z, a, \vec{m})$ . This matrix may be viewed as defining a linear operator that maps a function  $g(\vec{U})$  to its conditional expectation given  $Z$ .

For example, multiplying the matrix by a vector with entries  $g(\vec{u}_i)$  yields

$$\sum_{\vec{u}_i} g(\vec{u}_i) p(\vec{u}_i \mid Z, a, \vec{m}) = \mathbb{E}[g(\vec{U}) \mid Z, a, \vec{m}] \quad (4)$$

If the rows of the matrix are linearly independent, then the matrix has full row rank. Assumption 3.2 imposes a continuous analogue of this full row rank requirement, so that if there exists a function  $h_1$  such that

$$\begin{aligned} \sum_{\vec{u}} p(y \mid \vec{u}, a, \vec{m}, z) p(\vec{u} \mid z, a, \vec{m}) \\ = \sum_{w, \vec{u}} h_1(\cdot) p(w \mid u, a, \vec{m}, z) p(\vec{u} \mid z, a, \vec{m}), \end{aligned} \quad (5)$$

then

$$p(y \mid \vec{u}, a, \vec{m}, z) = \sum_w h_1(\cdot) p(w \mid \vec{u}, a, \vec{m}, z), \quad (6)$$

where independence of the left hand side of  $z$  and right hand side of  $a, \vec{m}, z$  follow by Assumption 3.1.

Completeness enables identification in tandem with the following assumption.

**Assumption 3.3.** (*Fredholm Equation Solutions*) For each  $(a, m)$ ,

1. There exists a function  $h_1$  such that

$$p(y \mid z, a, \vec{m}) = \sum_w h_1(y, a, \vec{m}, w) p(w \mid z, a, \vec{m})$$

2. There exists a function  $h_0$  such that

$$\begin{aligned} \sum_{w, m} h_1(y, \vec{m}, a', w) p(w, \vec{m} \mid z, a) \\ = \sum_w h_0(y, w, a', a) p(w \mid z, a) \end{aligned}$$

Assumption 3.3 posits solutions to Fredholm integral equations of the first kind, which are well studied in functional analysis literature [Kress, 1999]. In finite-support settings, solutions can be constructed easily under relevant matrix invertibility conditions; see Appendix B. These invertibility conditions state that distributions in  $\{p(W \mid z, a, \vec{m}) : z \in \mathcal{Z}\}$  exhibit sufficient variation across values  $z$  in its support  $\mathcal{Z}$  for each  $(a, \vec{m})$ , and that distributions in  $\{p(W \mid z, a) : z \in \mathcal{Z}\}$  likewise vary sufficiently across values  $z$  for each  $a$ .

Given these assumptions, we have the following result.

**Theorem 3.4.** Under Assumptions 3.1–3.3,  $p(Y(a))$  is identified in Fig. 4a by Equation 7 for every value  $y$  as follows

$$p(Y(a) = y) = \sum_{a', w} h_0(y, w, a', a) p(w, a') \quad (7)$$

See Appendix A for a proof.

This result re-expresses Equation 3 such that the first two terms correspond to a bridge function  $h_0$ , and the last term replaces marginalization over the (unobserved)  $U$  with a marginalization over the (observed) proxy  $W$ .

### 3.2 ASSUMPTION SET 2

The second identification strategy relies on the conditional independence relations in Assumption 3.5, reflected in the structure of Fig. 4b, together with the structural conditions in Assumptions 3.6–3.7.

**Assumption 3.5.** (*Conditional Independences*)

1.  $Y \perp\!\!\!\perp Z \mid \vec{U}, \vec{M}, A$
2.  $W \perp\!\!\!\perp \vec{M}, Z \mid \vec{U}, A$
3.  $Z \perp\!\!\!\perp W, A \mid \vec{U}$

Assumption 3.5 may be represented graphically via the m-separation criterion in the variant of the composite bow graph shown in Fig. 4b. Note that Assumptions 3.1 and 3.5 represent non-nested models, meaning neither is more general than the other.

**Assumption 3.6.** (*Completeness*) For each  $(a, \vec{m})$  and any square integrable  $g$ ,

1.  $\mathbb{E}\{g(\vec{u}) \mid z, a, \vec{m}\} = 0$  a.s. iff  $g(\vec{u}) = 0$  a.s.
2.  $\mathbb{E}\{g(\vec{u}) \mid w, a\} = 0$  a.s. iff  $g(\vec{u}) = 0$  a.s.

Just as Assumption 3.2, Assumption 3.6 imposes distributional variation restrictions, so that distributions in  $\{p(Z \mid \vec{u}_i, a, \vec{m}) : \vec{u}_i\}$  exhibit sufficient variation across values  $\vec{u}_i$  for each  $(a, \vec{m})$ , and distributions in  $\{p(W \mid u_i, a) : \vec{u}_i\}$  exhibit sufficient variation across values  $\vec{u}_i$  for each  $a$ .

We also require the following analogue to Assumption 3.3.

**Assumption 3.7.** (*Fredholm Equation Solutions*) For each  $(a, m)$ ,

1. There is a function  $b_1$  such that

$$p(y \mid z, a, \vec{m}) = \sum_w b_1(y, a, \vec{m}, w) p(w \mid z, a, \vec{m})$$

2. There is a function  $b_0$  such that

$$p(\vec{m} \mid w, a) = \sum_z b_0(\vec{m}, a, z) p(z \mid w, a)$$

Given these assumptions, we have the following result.

**Theorem 3.8.** Under Assumptions 3.5–3.7,  $p(Y(a))$  is identified in Fig. 4b for every value  $y$  as follows

$$p(Y(a)=y) = \sum_{\vec{m}, a', w, z} b_1(y, a', \vec{m}, w) b_0(\vec{m}, a, z) p(w, z, a')$$

See Appendix C for a proof.

This results replaces the first two terms in Equation 3 bridge functions  $b_0$  and  $b_1$ , with the resulting functional bearing close structural similarity to Equation 3. Assumptions 3.6 and 3.7 aid in this replacement by ensuring that

$$p(y \mid \vec{u}, a, \vec{m}, z) = \sum_w b_1(y, a, \vec{m}, w) p(w \mid \vec{u}, a, \vec{m}, z),$$

and

$$p(\vec{m} \mid \vec{u}, a, w) = \sum_z b_0(\vec{m}, a, z) p(z \mid \vec{u}, w, a),$$

where the independences on the left and right hand sides follow by Assumption 3.5. With these replacements, summing over  $\vec{U}$  in Equation 3 eliminates its appearance from the expression.

In finite-support settings, solutions to the integral equations in Assumption 3.7 can be constructed easily under invertibility of relevant probability matrices; see Appendix D. These invertibility conditions state that distributions in  $\{p(W \mid z, a, \vec{m}) : z \in \mathcal{Z}\}$  exhibit sufficient variation across values  $z$  in its support  $\mathcal{Z}$  for each  $(a, \vec{m})$ , and that distributions in  $\{p(Z \mid w, a) : w \in \mathcal{W}\}$  likewise vary sufficiently across values  $w$  for each  $a$ . If the cardinalities of  $Z$  and  $W$  are misaligned so that these invertibility conditions fail, a coarsened version of the variables may be employed instead.

### 3.3 ASSUMPTION SET 3

The third identification strategy assumes conditional independence relations in Assumption 3.9, displayed via m-separation in Fig. 3b, and Assumptions 3.10–3.12.

**Assumption 3.9.** (*Conditional Independences*)

1.  $W, Z, \text{ and } Y \text{ are mutually independent} \mid \vec{M}, A, \vec{U}$
2.  $W \perp\!\!\!\perp \vec{M}, A \mid \vec{U}$

These independence assumptions can be displayed graphically via m-separation in the variant of the composite bow graph shown in Fig. 3b.

**Assumption 3.10.** (*Bounded Density*) Random variables  $(A, \vec{M}, Y, W, Z, \vec{U})$  admit a joint density that is bounded.

**Assumption 3.11.** (*Completeness*) For each  $(a, \vec{m})$  and any square integrable  $g$ ,

1.  $\mathbb{E}\{g(\vec{u}) \mid z, a, \vec{m}\} = 0$  a.s. iff  $g(\vec{u}) = 0$  a.s.
2.  $\mathbb{E}\{g(\vec{u}) \mid w, a, \vec{m}\} = 0$  a.s. iff  $g(\vec{u}) = 0$  a.s.

**Assumption 3.12.** (*Distinctness*) For each  $(a, \vec{m})$  and  $i \neq j$ ,  $p(Y \mid \vec{u}_i, a, \vec{m})$  and  $p(Y \mid \vec{u}_j, a, \vec{m})$  differ with positive probability under the marginal distribution of  $Y$ .

Assumption 3.11 imposes distributional variation restrictions, so that distributions in  $\{p(Z \mid \vec{u}_i) : \vec{u}_i, a, \vec{m}\}$  exhibit sufficient variation across values  $\vec{u}_i$  for each  $(a, \vec{m})$ , and that distributions in  $\{p(W \mid u_i, a, \vec{m}) : \vec{u}_i\}$  exhibit sufficient variation across values  $\vec{u}_i$  for each  $(a, \vec{m})$ . Assumption 3.12 states that distributions in  $\{p(Y \mid \vec{u}_i, a, \vec{m}) : \vec{u}_i\}$  differ across values  $\vec{u}_i$ .

**Theorem 3.13.** *Under Assumptions 3.9-3.12,*

$$\{p(A, \vec{M}, Y, \vec{u}_i, W, Z) : \vec{u}_i\}$$

*is identified up to labeling  $i$ . Hence, the counterfactual distribution in Figure 3b is identified via Equation 3.*

See Appendix E for a proof, and Appendix F for an example in a finite-support model that illustrates the main idea. When  $U$  has finite support, Assumptions 3.9–3.12 uniquely determine the size of the support of  $U$  [Derksen, 2013]. A corresponding result holds in real-valued continuous models [Hu and Schennach, 2008]. Note that while the model depicted in Fig. 3b is nested within those depicted in Fig. 4a and 4b, the above theorem allows identification of the entire full law, and thus targets of inference other than the ACE.

This identification strategy extends eigendecomposition-based methods for causal identification under hidden confounding from Kuroki and Pearl [2014] and Allman et al. [2015] by incorporating identification arguments developed by Hu and Schennach [2008] for continuous state spaces.

## 4 ESTIMATION

Having described identifying functionals in Theorems 3.4, 3.8, and 3.13, we now provide plug-in estimation strategies for all three functionals, and an influence function-based estimator for the third functional.

### 4.1 CONSISTENT ESTIMATION UNDER ASSUMPTION SETS 1, 2, AND 3

We construct  $\hat{\psi}^{(S1,PI)}$ ,  $\hat{\psi}^{(S2,PI)}$ ,  $\hat{\psi}^{(S3,PI)}$ , the plug-in estimators for the counterfactual mean  $\psi_a = \mathbb{E}[Y(a)]$  under Assumption Sets 1, 2, and 3, respectively, which are consistent under standard regularity conditions.

Under Assumption Sets 1 and 2, estimation proceeds by first estimating a mean-induced bridge (corresponding to  $Y$ -marginalized solution(s) of integral equation(s) in Equation 43 and 47 in Appendix G), and then taking a sample average over the empirical distribution. Under Assumption Set 3, estimation proceeds by plugging in estimators of each distributional component in Equation 3, obtained by solving a linear system of equations by matrix inversion. Further details for these estimators can be found in Appendix G.

### 4.2 INFLUENCE FUNCTION-BASED ESTIMATION UNDER ASSUMPTION SET 3

Recall by Theorem 3.13, under Assumptions 3.9-3.12,

$$\{p(A, \vec{M}, Y, \vec{u}_i, W, Z) : \vec{u}_i\}$$

is identified up to labeling  $i$ , giving the counterfactual distribution in Fig. 3b via Equation 3.

We propose an influence function-based estimator  $\hat{\psi}_a^{(S3,IF)}$  of the counterfactual mean  $\psi_a = \mathbb{E}[Y(a)]$ . We now briefly review estimation theory in semiparametric models that we use to construct this estimator.

Let  $P$  denote the data-generating distribution,  $\psi(P)$  be the parameter of interest,  $\eta$  be the collection of nuisance functions indexing  $P$ , and  $O$  denote observed variables. An *influence function*  $\varphi(O; \psi(P), \eta_0 = \eta)$  for  $\psi(P)$  at  $P$  satisfies

$$\mathbb{E}_P[\varphi(O; \psi(P), \eta_0)] = 0, \quad (8)$$

and for every regular parametric submodel  $\{P_\epsilon : \epsilon \in \mathbb{R}\}$  through  $P$  with  $P_0 = P$  and score

$$s_\epsilon(O) := \left. \frac{d}{d\epsilon} \log p_\epsilon(O) \right|_{\epsilon=0},$$

it holds that

$$\left. \frac{d}{d\epsilon} \psi(P_\epsilon) \right|_{\epsilon=0} = \mathbb{E}_P[\varphi(O; \psi(P), \eta_0) s_\epsilon(O)] \quad (9)$$

[Tsiatis, 2006, Kennedy, 2024]. An influence function-based estimator  $\hat{\psi}$  solves estimating equation

$$\mathbb{E}_n[\varphi(O; \hat{\psi}, \hat{\eta})] = 0 \quad (10)$$

These estimators can attain  $\sqrt{n}$ -consistency for  $\psi_a$  despite nuisance functions  $\eta$  being estimated at slower than  $\sqrt{n}$  rates [Chernozhukov et al., 2018, Kennedy, 2024].

In addition to this desirable property, the estimator we derived, which we denote  $\hat{\psi}_a^{(S3,IF)}$ , can exhibit *multiple robustness*, meaning it can remain consistent for  $\psi_a$  under misspecification of certain collections of nuisance functions (Theorem 4.2). These properties extend to causal effect estimation (via  $\hat{\psi}^{(S3,IF)} = \hat{\psi}_1^{(S3,IF)} - \hat{\psi}_0^{(S3,IF)}$ ).

Our strategy for obtaining this estimator is to start with the nonparametric full data influence function (e.g. assuming the confounder  $U$  is observed) for the counterfactual mean  $\psi_a = \mathbb{E}[Y(a)]$ , and then transform this object to account for the fact  $U$  is not, in fact, observed.

### 4.3 THE FULL DATA INFLUENCE FUNCTION

In the case where  $A, \vec{M}, \vec{Y}, \vec{U}$  in Fig. 1 are observed, Fulcher et al. [2020] derived the influence function for  $\psi_a$ ,

$$\varphi_{full}((A, \vec{M}, \vec{Y}, \vec{U}); \cdot)$$



shown in Equation 11.

$$\begin{aligned} \varphi_{full}(\cdot) = & \underbrace{\frac{p(\vec{M} | A = a, \vec{U})}{p(\vec{M} | A, \vec{U})} \{Y - \mu(\vec{M}, A, \vec{U})\}}_{\varphi_{full1}(Y, \vec{M}, A, \vec{U})} \quad (11) \\ & + \underbrace{\frac{\mathbb{I}(A = a)}{\pi(A | \vec{U})} \{\xi(\vec{M}, \vec{U}) - \theta(\vec{U})\}}_{\varphi_{full2}(\vec{M}, A, \vec{U})} + \underbrace{\eta(A, \vec{U})}_{\varphi_{full3}(A, \vec{U})} - \psi_a, \end{aligned}$$

where

$$\begin{aligned} \mu(\vec{M}, A, \vec{U}) &:= \mathbb{E}[Y | \vec{M}, A, \vec{U}], \quad \pi(A | \vec{U}) := p(A | \vec{U}), \\ \xi(\vec{M}, \vec{U}) &:= \sum_{a=0}^1 \mu(\vec{M}, a, \vec{U}) \pi(a | \vec{U}), \\ \eta(A, \vec{U}) &:= \sum_{\vec{m}} \mu(\vec{m}, A, \vec{U}) p(\vec{m} | a, \vec{U}) \\ \theta(\vec{U}) &:= \sum_{\vec{m}} \xi(\vec{m}, \vec{U}) p(\vec{m} | a, \vec{U}) \end{aligned}$$

After introducing  $W$  and  $Z$  (as in Fig. 3b),  $\varphi_{full}(\cdot)$  from Equation 11 remains an influence function for  $\psi_a$  in the fully observed model (where  $A, \vec{M}, \vec{Y}, \vec{U}, W, Z$  are observed). See Appendix I for details.

#### 4.4 THE OBSERVED DATA INFLUENCE FUNCTION

Under Assumptions 3.9-3.12 in the model with hidden  $\vec{U}$  in Fig. 3b, we construct an influence function for  $\psi_a$ ,

$$\varphi_{obs}((A, \vec{M}, Y, W, Z); \psi, \eta)$$

shown in Equation 14.

Our construction exploits the fact that the full data distribution, and thus any functional of it, such as  $\varphi_{full}(A, \vec{M}, Y, \vec{U} = \vec{u}_i; \cdot)$ , is identified up to a relabeling of the latent states  $\vec{u}_i$ , and can be estimated. We express  $\varphi_{obs}((A, \vec{M}, Y, W, Z); \cdot)$  as a weighted combination of the terms in  $\varphi_{full}(A, \vec{M}, Y, \vec{U} = \vec{u}_i; \cdot)$  across latent states. The weights are constructed from observed variables and their moments conditional on values of  $\vec{U}$ .

##### 4.4.1 A Simple Example: Binary $U$ , Real-Valued $W, Z$

To illustrate the construction strategy for the weights in our influence function, consider the simple case of binary  $U$  and real-valued  $W, Z$ . Define for  $i \neq j$ ,

$$\begin{aligned} \mathcal{C}_{WZ}(u_i, a) &= \frac{W - \mathbb{E}[W | u_j]}{\mathbb{E}[W | u_i] - \mathbb{E}[W | u_j]} \times \\ & \quad \frac{Z - \mathbb{E}[Z | u_j, a]}{\mathbb{E}[Z | u_i, a] - \mathbb{E}[Z | u_j, a]}. \end{aligned}$$

Define  $\mathcal{C}_{YW}(u_i, a, \vec{m})$  and  $\mathcal{C}_{YZ}(u_i, a, \vec{m})$  similarly. For instance,

$$\begin{aligned} \mathcal{C}_{YW}(u_i, a, \vec{m}) &= \frac{Y - \mathbb{E}[Y | u_j, a, \vec{m}]}{\mathbb{E}[Y | u_i, a, \vec{m}] - \mathbb{E}[Y | u_j, a, \vec{m}]} \times \\ & \quad \frac{W - \mathbb{E}[W | u_j]}{\mathbb{E}[W | u_i] - \mathbb{E}[W | u_j]}, i \neq j. \end{aligned}$$

Lastly, define  $\mathcal{C}_{WZY}(u_i, a, \vec{m})$  as follows:

$$\begin{aligned} \mathcal{C}_{WZY}(u_i, a, \vec{m}) &= \frac{Y - \mathbb{E}[Y | u_j, a, \vec{m}]}{\mathbb{E}[Y | u_i, a, \vec{m}] - \mathbb{E}[Y | u_j, a, \vec{m}]} \times \\ & \quad \frac{W - \mathbb{E}[W | u_j]}{\mathbb{E}[W | u_i] - \mathbb{E}[W | u_j]} \frac{Z - \mathbb{E}[Z | u_j, a]}{\mathbb{E}[Z | u_i, a] - \mathbb{E}[Z | u_j, a]}, i \neq j. \end{aligned}$$

These weights require that conditional means in weight denominators be distinct, not already implied by Assumption 3.11. Minor discrepancies between identification and estimation assumptions are common; e.g., augmented inverse probability weighting assumes a discrete treatment for effect estimation, when it is not required for identification.

Given these weights, which may be viewed as *Lagrange interpolating polynomials*, Theorem 4.1 in the next section gives the desired influence function.

##### 4.4.2 The General Case: Arbitrary Hidden $\vec{U}$ with Finite Support

The simple Lagrange interpolating polynomial weights described in the previous section may be generalized, for a latent variable  $\vec{U}$  with finitely many levels, to weights  $\mathcal{C}_{WZ}(\vec{u}_i, a)$ ,  $\mathcal{C}_{YW}(\vec{u}_i, a, \vec{m})$ ,  $\mathcal{C}_{YZ}(\vec{u}_i, a, \vec{m})$ , and  $\mathcal{C}_{WZY}(\vec{u}_i, a, \vec{m})$  described in Appendix H. These weights are constructed to satisfy Equation 12 and 13.

For each  $X \in \{W, Z\}$ ,

$$\begin{aligned} \mathbb{E}[\mathcal{C}_{WZ}(A, \vec{u}_i) | Y, \vec{M}, A, \vec{u}_i, X] &= k, \quad k \text{ is a constant;} \\ \mathbb{E}[\mathcal{C}_{WZ}(A, \vec{u}_i) | Y, \vec{M}, A, \vec{u}_j, X] &= 0, \quad \vec{u}_i \neq \vec{u}_j. \end{aligned} \quad (12)$$

For each  $X \in \{W, Z, Y\}$ ,

$$\begin{aligned} \mathbb{E}[\mathcal{C}_{WZ}(A, \vec{u}_i) + \mathcal{C}_{YW}(\vec{M}, A, \vec{u}_i) + \mathcal{C}_{YZ}(\vec{M}, A, \vec{u}_i) \\ - 2\mathcal{C}_{WZY}(\vec{M}, A, \vec{u}_i) | \vec{M}, A, \vec{u}_i, X] &= 1; \\ \mathbb{E}[\mathcal{C}_{WZ}(A, \vec{u}_i) + \mathcal{C}_{YW}(\vec{M}, A, \vec{u}_i) + \mathcal{C}_{YZ}(\vec{M}, A, \vec{u}_i) \\ - 2\mathcal{C}_{WZY}(\vec{M}, A, \vec{u}_i) | \vec{M}, A, \vec{u}_j, X] &= 0, \quad \vec{u}_i \neq \vec{u}_j. \end{aligned} \quad (13)$$

Developing the analogous construction of these weights beyond the setting where  $U$  has finite support remains an open problem, which we leave for future work. Nevertheless, if Equations 12 and 13 hold, we have the following result.

**Theorem 4.1.** In the model in Fig. 3b (where  $\vec{U}$  is unobserved), under Assumptions 3.9–3.12, a nonparametric observed data influence function for  $\psi_a$  is given by

$$\begin{aligned} & \varphi_{\text{obs}}((A, \vec{M}, Y, W, Z); \psi_a, \eta) \\ &= \sum_i \mathcal{C}_{WZ}(A, \vec{u}_i) \varphi_{\text{full}_1}(Y, \vec{M}, A, \vec{u}_i) + \sum_i \left\{ \mathcal{C}_{WZ}(A, \vec{u}_i) \right. \\ &+ \mathcal{C}_{YW}(\vec{M}, A, \vec{u}_i) + \mathcal{C}_{YZ}(\vec{M}, A, \vec{u}_i) - 2\mathcal{C}_{WZY}(\vec{M}, A, \vec{u}_i) \left. \right\} \\ &\cdot \left\{ \varphi_{\text{full}_2}(\vec{M}, A, \vec{u}_i) + \varphi_{\text{full}_3}(A, \vec{u}_i) \right\} - \psi_a, \end{aligned} \quad (14)$$

where the terms  $\varphi_{\text{full}_1}(Y, \vec{M}, A, \vec{u}_i)$ ,  $\varphi_{\text{full}_2}(\vec{M}, A, \vec{u}_i)$ , and  $\varphi_{\text{full}_3}(A, \vec{u}_i)$  are defined in Equation 11, and the weights  $\mathcal{C}_{WZ}(A, \vec{u}_i)$ ,  $\mathcal{C}_{YW}(\vec{M}, A, \vec{u}_i)$ ,  $\mathcal{C}_{YZ}(\vec{M}, A, \vec{u}_i)$ , and  $\mathcal{C}_{WZY}(\vec{M}, A, \vec{u}_i)$  are constructed to satisfy Equation 12 and 13.

The proof appears in Appendix J.

As is the case for many influence function based estimators, the estimator  $\hat{\psi}_a^{(S3,IF)}$  obtained by solving Equation 10 using the influence function in Equation 14 can exhibit multiple robustness. Specifically, we have the following result (proved in Appendix L).

**Theorem 4.2.** Let  $\varphi(O; \psi_a, \eta)$  be defined as in Theorem 4.1. The estimator  $\hat{\psi}_a^{(S3,IF)}$ , obtained as the solution to Equation 10, is consistent for  $\psi_a$  if any one of the nuisance sets (1)–(5) is correctly specified.

1.  $p(\vec{M} | A, \vec{U})$ ,  $\{\mathbb{E}[h_j(W) | \vec{U}]\}_j$ ,  $\{\mathbb{E}[f_j(Z) | \vec{U}, A]\}_j$ ;
2.  $p(\vec{M} | A, \vec{U})$ ,  $\{\mathbb{E}[h_j(W) | \vec{U}]\}_j$ ,  $\{\mathbb{E}[t_j(Y) | \vec{M}, A, \vec{U}]\}_j$ ;
3.  $f(\vec{M} | A, \vec{U})$ ,  $\{\mathbb{E}[f_j(Z) | \vec{U}, A]\}_j$ ,  $\{\mathbb{E}[t_j(Y) | \vec{M}, A, \vec{U}]\}_j$ ;
4.  $\{\mathbb{E}[t_j(Y) | \vec{M}, A, \vec{U}]\}_j$ ,  $\pi(A | \vec{U})$ ,  $\{\mathbb{E}[h_j(W) | \vec{U}]\}_j$ ;
5.  $\{\mathbb{E}[t_j(Y) | \vec{M}, A, \vec{U}]\}_j$ ,  $\pi(A | \vec{U})$ ,  $\{\mathbb{E}[f_j(Z) | \vec{U}, A]\}_j$ ,

where  $\{h_j(W)\}_j$ ,  $\{f_j(Z)\}_j$ , and  $\{t_j(Y)\}_j$  are collections defined in Appendix H. By definition, these collections include the identity functions on  $W$ ,  $Z$ , and  $Y$ . When  $\vec{U}$  is binary, the collections reduce to the identity functions.

Moreover,  $\hat{\psi}_a^{(S3,IF)}$  is asymptotically normal and  $\sqrt{n}$ -consistent despite slower than parametric rates of convergence for nuisance functions. We have the following result (proved in Appendix L).

**Theorem 4.3.** Let  $\varphi(O; \psi_a, \eta)$  be defined as in Theorem 4.1. Then if (1)–(5) below hold (for all  $j$  where applicable), under regularity conditions including sample splitting and consistency of all nuisance functions, the estimator  $\hat{\psi}_a^{(S3,IF)}$  obtained as the solution to Equation 10 has the property

$$\sqrt{n}(\hat{\psi}_a - \psi_a) \xrightarrow{d} \mathcal{N}\left(0, \mathbb{E}[\varphi^2(O; \psi_a, \eta)]\right).$$

1.  $\|\hat{\mu}(\vec{M}, A, \vec{U}) - \mu(\vec{M}, A, \vec{U})\|$   
 $\|\hat{p}(\vec{M} | A, \vec{U}) - p(\vec{M} | A, \vec{U})\| = o_p(n^{-1/2})$ ,
2.  $\|\hat{\pi}(A | \vec{U}) - \pi(A | \vec{U})\|$   
 $\|\hat{p}(\vec{M} | A, \vec{U}) - p(\vec{M} | A, \vec{U})\| = o_p(n^{-1/2})$ ;
3.  $\|\hat{\mathbb{E}}[t(Y)_j | \vec{M}, A, \vec{U}] - \mathbb{E}[t(Y)_j | \vec{M}, A, \vec{U}]\|$   
 $\|\hat{\mathbb{E}}[h(W)_j | \vec{U}] - \mathbb{E}[h(W)_j | \vec{U}]\| = o_p(n^{-1/2})$ ,
4.  $\|\hat{\mathbb{E}}[t(Y)_j | \vec{M}, A, \vec{U}] - \mathbb{E}[t(Y)_j | \vec{M}, A, \vec{U}]\|$   
 $\|\hat{\mathbb{E}}[f(Z)_j | \vec{U}, A] - \mathbb{E}[f(Z)_j | \vec{U}, A]\| = o_p(n^{-1/2})$ ,
5.  $\|\hat{\mathbb{E}}[f(Z)_j | \vec{U}, A] - \mathbb{E}[f(Z)_j | \vec{U}, A]\|$   
 $\|\hat{\mathbb{E}}[h(W)_j | \vec{U}] - \mathbb{E}[h(W)_j | \vec{U}, A]\| = o_p(n^{-1/2})$ ,

where  $\{h_j(W)\}_j$ ,  $\{f_j(Z)\}_j$ , and  $\{t_j(Y)\}_j$  are collections defined in Appendix H. By definition, these collections include the identity functions on  $W$ ,  $Z$ , and  $Y$ . When  $\vec{U}$  is binary, the collections reduce to the identity functions.

## 5 EXPERIMENTS

We demonstrate consistency of our proposed estimators using a simulation study for a finite-support data-generating process (DGP) with binary variables in Table 1, and also for a mixed DGP in Table 2. To ensure the datasets generated satisfy all three sets of assumptions in Section 3, we selected models in their intersection. See Appendix M and Appendix N for details of the finite-support DGP and mixed DGP, respectively.

We consider the following estimators (definitions can be found in Appendix G and O):

1. The oracle plug-in estimator  $\hat{\psi}^{(\text{Oracle}, \text{PI})}$  that has access to the full data, where variables  $A, \vec{M}, Y, \vec{U}, W, Z$  are all observed (Equation 74);
2. A front-door plug-in estimator  $\hat{\psi}^{(\text{FD}, \text{PI})}$  that (incorrectly) assumes  $\vec{M}$  satisfies the front-door criterion relative to  $(A, Y)$  (Equation 75);
3. The plug-in estimators  $\hat{\psi}^{(S1, \text{PI})}$ ,  $\hat{\psi}^{(S2, \text{PI})}$ ,  $\hat{\psi}^{(S3, \text{PI})}$  (Equations 46, 50, and 51), and the influence function-based estimator  $\hat{\psi}^{(S3, \text{IF})}$  (obtained by solving Equation 10 using influence function in Equation 14);
4. The estimators  $\hat{\psi}^{(S3, \text{PI}, \text{MIS})}$ ,  $\hat{\psi}^{(S3, \text{IF}, \text{MIS})}$  which mirror the formulation of  $\hat{\psi}^{(S3, \text{PI})}$  and  $\hat{\psi}^{(S3, \text{IF})}$  but misspecify  $p(M|a, u)$  for each  $a$  value by labeling  $u_i$  incorrectly.

Results for a finite-support DGP are shown in Table.1. At relatively small sample size ( $\leq 1000$ ), all non-oracle estimators perform poorly. Additionally,  $\hat{\psi}^{(S3, \text{IF})}$  exhibits instability. Estimator instability is a known issue with many inverse weighted estimators arising in semi-parametric theory, even in cases where proximal inference is not involved. In Table 1, the result for  $\hat{\psi}^{(S3, \text{IF}, \text{Clipped})}$  illustrates how clipping can improve estimator stability at low sample size. At a moderately large sample size ( $\geq 3000$ ), the front-door estimator  $\hat{\psi}^{(\text{FD}, \text{PI})}$  still exhibits substantial bias, due to



Table 1: Simulation results for the binary DGP\*; Ground truth ACE = 0.258.

| n    | method                                   | mean    | bias     | variance |
|------|------------------------------------------|---------|----------|----------|
| 1000 | $\hat{\psi}^{(Oracle)}$                  | 0.254   | -2.5e-03 | 7.0e-04  |
| 1000 | $\hat{\psi}^{(FD,PI)}$                   | 0.159   | -9.7e-02 | 2.3e-04  |
| 1000 | $\hat{\psi}^{(S1,PI)}$                   | 0.231   | -2.6e-02 | 1.5e-02  |
| 1000 | $\hat{\psi}^{(S2,PI)}$                   | 0.237   | -2.0e-02 | 9.5e-03  |
| 1000 | $\hat{\psi}^{(S3,PI)}$                   | 0.163   | -9.3e-02 | 8.7e-03  |
| 1000 | $\hat{\psi}^{(S3,IF)}$                   | 2.6e+11 | 2.6e+11  | 3.3e+23  |
| 1000 | $\hat{\psi}^{(S3,IF,Clipped)^{\dagger}}$ | 0.333   | 7.6e-02  | 1.6e-01  |
| 3000 | $\hat{\psi}^{(Oracle)}$                  | 0.257   | 4.0e-04  | 2.0e-04  |
| 3000 | $\hat{\psi}^{(FD,PI)}$                   | 0.161   | -9.6e-02 | 6.0e-05  |
| 3000 | $\hat{\psi}^{(S1,PI)}$                   | 0.259   | 2.4e-03  | 7.2e-04  |
| 3000 | $\hat{\psi}^{(S2,PI)}$                   | 0.259   | 1.9e-03  | 6.2e-04  |
| 3000 | $\hat{\psi}^{(S3,PI)}$                   | 0.242   | -1.5e-02 | 3.7e-03  |
| 3000 | $\hat{\psi}^{(S3,IF)}$                   | 0.261   | 3.9e-03  | 7.2e-03  |
| 3000 | $\hat{\psi}^{(S3,PI,MIS)}$               | -0.181  | -4.4e-01 | 4.0e-03  |
| 3000 | $\hat{\psi}^{(S3,IF,MIS)}$               | 0.243   | -1.4e-02 | 9.2e-03  |
| 6000 | $\hat{\psi}^{(Oracle)}$                  | 0.257   | 6.0e-04  | 9.7e-05  |
| 6000 | $\hat{\psi}^{(FD,PI)}$                   | 0.161   | -9.6e-02 | 3.1e-05  |
| 6000 | $\hat{\psi}^{(S1,PI)}$                   | 0.257   | 1.5e-04  | 4.2e-04  |
| 6000 | $\hat{\psi}^{(S2,PI)}$                   | 0.258   | 1.1e-03  | 4.2e-04  |
| 6000 | $\hat{\psi}^{(S3,PI)}$                   | 0.249   | -8.1e-03 | 2.7e-03  |
| 6000 | $\hat{\psi}^{(S3,IF)}$                   | 0.249   | -7.6e-03 | 2.5e-03  |
| 6000 | $\hat{\psi}^{(S3,PI,MIS)}$               | -0.198  | -4.5e-01 | 3.2e-03  |
| 6000 | $\hat{\psi}^{(S3,IF,MIS)}$               | 0.253   | -4.1e-03 | 8.1e-03  |

\* Data-generating process can be found in Appendix M.

$\dagger$  Clipping was applied to  $\frac{Y - E[Y|u_j, a, \tilde{m}]}{E[Y|u_i, a, \tilde{m}] - E[Y|u_j, a, \tilde{m}]}$ , where the denominator's magnitude was lower bounded by 0.01.

model misspecification, while  $\hat{\psi}^{(S1,PI)}, \hat{\psi}^{(S2,PI)}, \hat{\psi}^{(S3,PI)}$ , and  $\hat{\psi}^{(S3,IF)}$  all provide consistent estimates.

To demonstrate the robustness property of the influence function-based estimator  $\hat{\psi}^{(S3,IF)}$ , we misspecified  $p(M | a, u)$  for each  $a$  value by labeling  $u_i$  incorrectly and recording the resulting estimates as  $\hat{\psi}^{(S3,PI,MIS)}$  and  $\hat{\psi}^{(S3,IF,MIS)}$ . Unlike plug-in estimator  $\hat{\psi}^{(S3,PI,MIS)}$ , which outputs a biased estimate, the influence function-based estimator  $\hat{\psi}^{(S3,IF,MIS)}$  remains consistent.

Results for the mixed DGP are shown in Table 2. The estimators exhibit patterns similar to those observed in Table 1, although small-sample bias is reduced.

## 6 DISCUSSION

The core contribution of our work is development of methods that address a central difficulty with the front-door crite-

Table 2: Simulation results for the mixed DGP\*; Ground truth ACE = 0.169.

| n    | method                  | mean  | bias     | variance |
|------|-------------------------|-------|----------|----------|
| 1000 | $\hat{\psi}^{(Oracle)}$ | 0.167 | -2.1e-03 | 1.0e-03  |
| 1000 | $\hat{\psi}^{(FD,PI)}$  | 0.303 | 1.3e-01  | 1.4e-03  |
| 1000 | $\hat{\psi}^{(S1,PI)}$  | 0.174 | 5.2e-03  | 3.1e-02  |
| 1000 | $\hat{\psi}^{(S2,PI)}$  | 0.175 | 6.0e-03  | 1.9e-02  |
| 1000 | $\hat{\psi}^{(S3,PI)}$  | 0.197 | 2.8e-02  | 1.9e-03  |
| 1000 | $\hat{\psi}^{(S3,IF)}$  | 0.161 | -7.7e-03 | 5.9e-03  |
| 3000 | $\hat{\psi}^{(Oracle)}$ | 0.170 | 7.0e-04  | 3.0e-04  |
| 3000 | $\hat{\psi}^{(FD,PI)}$  | 0.303 | 1.3e-01  | 7.0e-04  |
| 3000 | $\hat{\psi}^{(S1,PI)}$  | 0.173 | 3.8e-03  | 9.0e-04  |
| 3000 | $\hat{\psi}^{(S2,PI)}$  | 0.168 | -1.3e-03 | 1.3e-03  |
| 3000 | $\hat{\psi}^{(S3,PI)}$  | 0.180 | 1.1e-02  | 1.2e-03  |
| 3000 | $\hat{\psi}^{(S3,IF)}$  | 0.179 | 1.1e-02  | 2.8e-03  |
| 6000 | $\hat{\psi}^{(Oracle)}$ | 0.169 | -2.0e-04 | 2.0e-04  |
| 6000 | $\hat{\psi}^{(FD,PI)}$  | 0.301 | 1.3e-01  | 3.0e-04  |
| 6000 | $\hat{\psi}^{(S1,PI)}$  | 0.164 | -4.6e-03 | 4.0e-04  |
| 6000 | $\hat{\psi}^{(S2,PI)}$  | 0.167 | -1.1e-03 | 5.0e-04  |
| 6000 | $\hat{\psi}^{(S3,PI)}$  | 0.168 | -9.0e-04 | 5.0e-04  |
| 6000 | $\hat{\psi}^{(S3,IF)}$  | 0.166 | -2.7e-03 | 1.5e-03  |

\* Data-generating process can be found in Appendix N.

tion: that mediators often share unmeasured causes with the treatment and outcome. Specifically, we develop three distinct identification strategies based on proxies which allow identification of causal effects in cases where the front-door criterion fails. The first two strategies accommodate continuous hidden confounders and identify causal effects via solutions to Fredholm integral equations, relying on non-nested conditional independence assumptions that permit different causal structures. The third strategy permits a more restricted class of causal structures, but recovers the full law, allowing any functional of the full data distribution to be identified.

In addition, we develop estimation strategies for the obtained identifying functionals based on the plug-in principle. Finally, we develop an influence function based estimator for the functional obtained via the third strategy, based on estimation theory in semi-parametric models.

Simulation experiments demonstrate that all three approaches effectively debias causal effect estimation in the presence of hidden confounding when the front-door criterion fails. Moreover, the simulations illustrate that the influence function-based estimator for functionals obtained under the third strategy remains consistent even under certain forms of model misspecification.

## Author Contributions

Helen Guo, Beatrix Wen, and Ilya Shpitser developed the identification results. Helen Guo developed the semiparametric estimation results. Beatrix Wen and Helen Guo designed and conducted the simulation studies.

## Acknowledgements

This work was supported by the Office of Naval Research (Grant No. N000142412701), the National Institutes of Health (Grant No. 1R56AI191526-01), the National Science Foundation CAREER Award (Grant No. 1942239), and the National Library of Medicine, (Grant No. 1R01LM014800-01A1).

## References

- E. Allman, J. Rhodes, E. Stanghellini, and M. Valtorta. Parameter identifiability of discrete bayesian networks with hidden variables. *Journal of Causal Inference*, 3(2):189–205, 2015. doi: 10.1515/jci-2014-0021.
- Yang Bai, Yifan Cui, and Baoluo Sun. Proximal inference on population intervention indirect effect. *arXiv*, 2025.
- Ivan A. Canay, Andres Santos, and Azeem M. Shaikh. On the testability of identification in some nonparametric models with endogeneity. *Econometrica*, 81(6):2535–2559, 2013. doi: <https://doi.org/10.3982/ECTA10851>.
- Victor Chernozhukov, Denis Chetverikov, Mert Demirer, Esther Duflo, Christian Hansen, Whitney Newey, and James Robins. Double/debiased machine learning for treatment and structural parameters. *The Econometrics Journal*, 21(1):C1–C68, 01 2018. ISSN 1368-4221. doi: 10.1111/ectj.12097.
- Yifan Cui, Hongming Pu, Xu Shi, Wang Miao, and Eric Tchetgen Tchetgen. Semiparametric proximal causal inference. *Journal of the American Statistical Association*, 119(546):1348–1359, 2024.
- Harm Derksen. Kruskal’s uniqueness inequality is sharp. *Linear Algebra and its Applications*, 438(2):708–712, 2013. ISSN 0024-3795. doi: 10.1016/j.laa.2011.05.041. Tensors and Multilinear Algebra.
- Oliver Dukes, Ilya Shpitser, and Eric J Tchetgen Tchetgen. Proximal mediation analysis. *Biometrika*, 110(4):973–987, 2023.
- Isabel R Fulcher, Ilya Shpitser, Stella Marealle, and Eric J Tchetgen Tchetgen. Robust inference on population indirect causal effects: the generalized front door criterion. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 82(1):199–214, 2020.
- Amiremad Ghassami, Alan Yang, Ilya Shpitser, and Eric Tchetgen Tchetgen. Causal inference with hidden mediators. *Biometrika*, 112(1):asae037, 2025.
- Anna Guo, David Benkeser, and Razieh Nabi. Flexible nonparametric inference for causal effects under the front-door model. *arXiv*, 2026.
- Helen Guo, Elizabeth L. Ogburn, and Ilya Shpitser. Comparing two proxy methods for causal identification. *arXiv*, 2025.
- Yingyao Hu and Susanne M Schennach. Instrumental variable treatment of nonclassical measurement error models. *Econometrica*, 76(1):195–216, 2008.
- Yingyao Hu and Ji-Liang Shiu. A simple test of completeness in a class of nonparametric specification. *Econometric Reviews*, 41(4):373–399, 2022. doi: 10.1080/07474938.2021.1957285.
- Yimin Huang and Marco Valtorta. Pearl’s calculus of intervention is complete. In *Twenty Second Conference On Uncertainty in Artificial Intelligence*, 2006.
- Edward H. Kennedy. Semiparametric doubly robust targeted double machine learning: A review. In *Handbook of Statistical Methods for Precision Medicine*, chapter 10, pages 207–236. Chapman and Hall/CRC, 1 edition, 2024.
- Edward H Kennedy, Sivaraman Balakrishnan, and Max G’Sell. Sharp instruments for classifying compliers and generalizing causal effects. *The Annals of Statistics*, 48(4):2008–2030, 2020.
- Rainer Kress. *Linear Integral Equations*. Springer, 1999.
- J. Kruskal. Three-way arrays: rank and uniqueness of trilinear decompositions, with applications to arithmetic complexity and statistics. *Linear Algebra and its Applications*, 18:95–138, 1977. doi: 10.1016/0024-3795(77)90069-6.
- Manabu Kuroki and Judea Pearl. Measurement bias and effect restoration in causal inference. *Biometrika*, 101(2):423–437, 2014.
- Wang Miao, Zhi Geng, and Eric J Tchetgen Tchetgen. Identifying causal effects with proxy variables of an unmeasured confounder. *Biometrika*, 105(4):987–993, 2018.
- W. K. Newey and J. L. Powell. Instrumental variable estimation of nonparametric models. *Econometrica*, 71:1565–1578, 2003.
- Judea Pearl. Comment: Graphical models, causality, and intervention. *Statistical Science*, 8(3):266–269, 1993.
- Judea Pearl. Causal diagrams for empirical research. *Biometrika*, 82(4):669–688, 1995.
- Judea Pearl. *Causality*. Cambridge university press, 2009.

Thomas Richardson and Peter Spirtes. Ancestral graph Markov models. *Annals of Statistics*, 30:962–1030, 2002.

Ilya Shpitser and Judea Pearl. Identification of joint interventional distributions in recursive semi-Markovian causal models. In *Proceedings of the Twenty-First National Conference on Artificial Intelligence (AAAI-06)*. AAAI Press, Palo Alto, 2006.

Ilya Shpitser, Zach Wood-Doughty, and Eric J Tchetgen Tchetgen. The proximal id algorithm. *Journal of Machine Learning Research*, 24(188):1–46, 2023.

Jin Tian and Judea Pearl. On the Identification of Causal Effects. Technical Report R-290-L, UCLA, 2002.

Anastasios Tsiatis. *Semiparametric Theory and Missing Data*. Springer-Verlag New York, 1st edition edition, 2006.

Thomas S. Verma and Judea Pearl. Equivalence and synthesis of causal models. Technical Report R-150, Department of Computer Science, University of California, Los Angeles, 1990.

---

## Supplementary Material

---

Helen Guo<sup>\*1</sup>

Beatrix Yaxin Wen<sup>2</sup>

Ilya Shpitser<sup>3</sup>

<sup>1</sup>Biostatistics Dept., Johns Hopkins Bloomberg School of Public Health, Baltimore, MD, USA

<sup>2</sup>Applied Mathematics and Statistics Dept., Johns Hopkins Whiting School of Engineering, Baltimore, MD, USA

<sup>3</sup>Computer Science Dept., Johns Hopkins Whiting School of Engineering, Baltimore, MD, USA

### A IDENTIFICATION PROOF FOR ASSUMPTION SET 1

*Proof.* Assumption 3.1 and first bridge solution in Assumption 3.3 gives

$$\sum_{\vec{u}} p(y \mid \vec{u}, a', \vec{m}, z) p(\vec{u} \mid z, a', \vec{m}) = \sum_{w, \vec{u}} h_1(y, a', \vec{m}, w) p(w \mid u, \cancel{a', \vec{m}}, z) p(\vec{u} \mid z, a', \vec{m}). \quad (15)$$

Then, the first completeness condition in Assumption 3.2 gives

$$p(y \mid \vec{u}, a', \vec{m}) = \sum_w h_1(y, a', \vec{m}, w) p(w \mid \vec{u}). \quad (16)$$

Similarly, Assumption 3.1 and the second bridge solution in Assumption 3.3 gives

$$\sum_{w, \vec{m}, \vec{u}} h_1(y, \vec{m}, a', w) p(w, \vec{m} \mid \vec{u}, z, a) p(\vec{u} \mid z, a) = \sum_{w, \vec{u}} h_0(y, w, a', a) p(w \mid \vec{u}, \cancel{z, a}) p(\vec{u} \mid z, a). \quad (17)$$

Then, the second completeness condition in Assumption 3.2 gives

$$\sum_{w, \vec{m}} h_1(y, \vec{m}, a', w) p(w, \vec{m} \mid \vec{u}, a) = \sum_w h_0(y, w, a', a) p(w \mid \vec{u}). \quad (18)$$

Hence, we have that

$$p(Y(a) = y) = \sum_{\vec{u}, \vec{m}, a'} p(y \mid \vec{m}, a', \vec{u}) p(\vec{m} \mid a, \vec{u}) p(a' \mid \vec{u}) p(\vec{u}) \quad (19)$$

$$= \sum_{\vec{u}, \vec{m}, a', w} h_1(y, \vec{m}, a', w) p(w \mid \vec{u}) p(\vec{m} \mid a, \vec{u}) p(a' \mid \vec{u}) p(\vec{u}) \text{ by Equation 16} \quad (20)$$

$$= \sum_{\vec{u}, \vec{m}, a', w} h_1(y, \vec{m}, a', w) p(w \mid \vec{u}, a) p(\vec{m} \mid a, \vec{u}) p(a' \mid \vec{u}) p(\vec{u}) \text{ by Assumption 3.1} \quad (21)$$

$$= \sum_{\vec{u}, a', \vec{m}, w} h_1(y, \vec{m}, a', w) p(w, \vec{m} \mid a, \vec{u}) p(a' \mid \vec{u}) p(\vec{u}) \quad (22)$$

$$= \sum_{\vec{u}, a', w} h_0(y, w, a', a) p(w \mid \vec{u}) p(a' \mid \vec{u}) p(\vec{u}) \text{ by Equation 18} \quad (23)$$

$$= \sum_{\vec{u}, a', w} h_0(y, w, a', a) p(w \mid \vec{u}, a') p(a' \mid \vec{u}) p(\vec{u}) \text{ by Assumption 3.1} \quad (24)$$

$$= \sum_{\vec{u}, a', w} h_0(y, w, a', a) p(w, a' \mid \vec{u}) p(\vec{u}) \quad (25)$$

$$= \sum_{a', w} h_0(y, w, a', a) p(w, a') \quad (26)$$

□

## B BRIDGE SOLUTIONS FOR ASSUMPTION 3.3 IN FINITE-SUPPORT MODELS

For each  $(a, \vec{m})$ , let

$$\mathbf{P}_{Y|Z,a,\vec{m}} = \begin{bmatrix} \Pr(y_1 \mid z_1, a, \vec{m}) & \cdots & \Pr(y_1 \mid z_{|\mathcal{Z}|}, a, \vec{m}) \\ \Pr(y_2 \mid z_1, a, \vec{m}) & \cdots & \Pr(y_2 \mid z_{|\mathcal{Z}|}, a, \vec{m}) \\ \vdots & \ddots & \vdots \\ \Pr(y_{|\mathcal{Y}|} \mid z_1, a, \vec{m}) & \cdots & \Pr(y_{|\mathcal{Y}|} \mid z_{|\mathcal{Z}|}, a, \vec{m}) \end{bmatrix}$$

and

$$\mathbf{P}_{W|Z,a,\vec{m}} = \begin{bmatrix} \Pr(w_1 \mid z_1, a, \vec{m}) & \cdots & \Pr(w_1 \mid z_{|\mathcal{Z}|}, a, \vec{m}) \\ \Pr(w_2 \mid z_1, a, \vec{m}) & \cdots & \Pr(w_2 \mid z_{|\mathcal{Z}|}, a, \vec{m}) \\ \vdots & \ddots & \vdots \\ \Pr(w_{|\mathcal{W}|} \mid z_1, a, \vec{m}) & \cdots & \Pr(w_{|\mathcal{W}|} \mid z_{|\mathcal{Z}|}, a, \vec{m}) \end{bmatrix}.$$

Then if distributions in  $\{p(W \mid z, a, \vec{m}) : z \in \mathcal{Z}\}$  vary sufficiently across values  $z$  in its support  $\mathcal{Z}$  for each  $(a, \vec{m})$  so that  $\mathbf{P}_{W|Z,a,\vec{m}}$  is left invertible with left inverse  $\mathbf{P}_{W|Z,a,\vec{m}}^\dagger$ , we have that

$$\mathbf{P}_{Y|Z,a,\vec{m}} = \mathbf{P}_{Y|Z,a,\vec{m}} \mathbf{P}_{W|Z,a,\vec{m}}^\dagger \mathbf{P}_{W|Z,a,\vec{m}} \quad (27)$$

Hence,  $\mathbf{P}_{Y|Z,a,\vec{m}} \mathbf{P}_{W|Z,a,\vec{m}}^\dagger$  defines  $h_1(y, a, \vec{m}, w)$  in Assumption 3.3 such that

$$p(y \mid z, a, \vec{m}) = \sum_w h_1(y, a, \vec{m}, w) p(w \mid z, a, \vec{m}).$$

Similarly, for each  $a$ , let

$$\mathbf{P}_{h_1(Y)|Z,a} = \begin{bmatrix} \sum_{w,\vec{m}} h_1(y_1, a, \vec{m}, w) p(w, \vec{m} \mid z_1, a) & \cdots & \sum_{w,\vec{m}} h_1(y_1, a, \vec{m}, w) p(w, \vec{m} \mid z_{|\mathcal{Z}|}, a) \\ \sum_{w,\vec{m}} h_1(y_2, a, \vec{m}, w) p(w, \vec{m} \mid z_2, a) & \cdots & \sum_{w,\vec{m}} h_1(y_2, a, \vec{m}, w) p(w, \vec{m} \mid z_{|\mathcal{Z}|}, a) \\ \vdots & \ddots & \vdots \\ \sum_{w,\vec{m}} h_1(y_{|\mathcal{Y}|}, a, \vec{m}, w) p(w, \vec{m} \mid z_1, a) & \cdots & \sum_{w,\vec{m}} h_1(y_{|\mathcal{Y}|}, a, \vec{m}, w) p(w, \vec{m} \mid z_{|\mathcal{Z}|}, a) \end{bmatrix}$$

Then if  $\mathbf{P}_{W|Z,a}$  is left invertible with left inverse  $\mathbf{P}_{W|Z,a}^\dagger$ , we have that

$$\mathbf{P}_{h_1(Y)|Z,a'} = \mathbf{P}_{h_1(Y)|Z,a'} \mathbf{P}_{W|Z,a}^\dagger \mathbf{P}_{W|Z,a} \quad (28)$$

Hence,  $\mathbf{P}_{h_1(Y)|Z,a'} \mathbf{P}_{W|Z,a}^\dagger$  defines  $h_0(y, w, a', a)$  in Assumption 3.3 such that

$$\sum_{w, \vec{m}} h_1(y, \vec{m}, a', w) p(w, \vec{m} | z, a) = \sum_w h_0(y, w, a', a) p(w | z, a).$$

## C IDENTIFICATION PROOF FOR ASSUMPTION SET 2

*Proof.* Assumption 3.5 and the first bridge solution in Assumption 3.7 gives

$$\sum_{\vec{u}} p(y | u, a, \vec{m}, \vec{z}) p(\vec{u} | a, \vec{m}, z) = \sum_{w, \vec{u}} b_1(y, a, \vec{m}, w) p(\vec{u} | \vec{u}, a, \vec{m}, \vec{z}) p(\vec{u} | a, \vec{m}, z). \quad (29)$$

Then, the first completeness condition in Assumption 3.6 gives

$$p(y | \vec{u}, a, \vec{m}) = \sum_w b_1(y, a, \vec{m}, w) p(w | \vec{u}, a). \quad (30)$$

Similarly, Assumption 3.5 and the second bridge solution in Assumption 3.7 gives

$$\sum_{\vec{u}} p(\vec{m} | \vec{u}, a, \vec{z}) p(\vec{u} | w, a) = \sum_{z, w} b_0(\vec{m}, a, z) p(z | \vec{u}, w, a) p(\vec{u} | w, a). \quad (31)$$

Then, the second completeness condition in Assumption 3.6 gives

$$p(\vec{m} | \vec{u}, a) = \sum_z b_0(\vec{m}, a, z) p(z | \vec{u}). \quad (32)$$

Hence, we have that

$$p(Y(a) = y) = \sum_{\vec{u}, \vec{m}, a'} p(y | \vec{u}, a', \vec{m}) p(a' | \vec{u}) p(\vec{m} | a, \vec{u}) p(\vec{u}) \quad (33)$$

$$= \sum_{\vec{u}, \vec{m}, a', w, z} b_1(\cdot) p(w | \vec{u}, a') p(a' | \vec{u}) b_0(\cdot) p(z | \vec{u}) p(\vec{u}) \text{ by Equation 30 and Assumption 3.5} \quad (34)$$

$$= \sum_{\vec{u}, \vec{m}, a', w, z} b_1(\cdot) p(w, a' | \vec{u}) b_0(\cdot) p(z | \vec{u}) p(\vec{u}) \text{ by Equation 32} \quad (35)$$

$$= \sum_{\vec{u}, \vec{m}, a', w, z} b_1(\cdot) p(w, z, a' | \vec{u}) b_0(\cdot) p(\vec{u}) \text{ by Assumption 3.5} \quad (36)$$

$$= \sum_{\vec{m}, a', w, z} b_1(y, a', \vec{m}, w) b_0(\vec{m}, a, z) p(w, z, a') \quad (37)$$

□



## D BRIDGE SOLUTIONS FOR ASSUMPTION 3.7 IN FINITE-SUPPORT MODELS

For each  $(a, \vec{m})$ , let

$$\mathbf{P}_{Y|Z,a,\vec{m}} = \begin{bmatrix} \Pr(y_1 | z_1, a, \vec{m}) & \cdots & \Pr(y_1 | z_{|\mathcal{Z}|}, a, \vec{m}) \\ \Pr(y_2 | z_1, a, \vec{m}) & \cdots & \Pr(y_2 | z_{|\mathcal{Z}|}, a, \vec{m}) \\ \vdots & \ddots & \vdots \\ \Pr(y_{|\mathcal{Y}|} | z_1, a, \vec{m}) & \cdots & \Pr(y_{|\mathcal{Y}|} | z_{|\mathcal{Z}|}, a, \vec{m}) \end{bmatrix}$$

and

$$\mathbf{P}_{W|Z,a,\vec{m}} = \begin{bmatrix} \Pr(w_1 | z_1, a, \vec{m}) & \cdots & \Pr(w_1 | z_{|\mathcal{Z}|}, a, \vec{m}) \\ \Pr(w_2 | z_1, a, \vec{m}) & \cdots & \Pr(w_2 | z_{|\mathcal{Z}|}, a, \vec{m}) \\ \vdots & \ddots & \vdots \\ \Pr(w_{|\mathcal{W}|} | z_1, a, \vec{m}) & \cdots & \Pr(w_{|\mathcal{W}|} | z_{|\mathcal{Z}|}, a, \vec{m}) \end{bmatrix}.$$

Then if distributions in  $\{p(W | z, a, \vec{m}) : z \in \mathcal{Z}\}$  vary sufficiently across values  $z$  in its support  $\mathcal{Z}$  for each  $(a, \vec{m})$  so that  $\mathbf{P}_{W|Z,a,\vec{m}}$  is left invertible with left inverse  $\mathbf{P}_{W|Z,a,\vec{m}}^\dagger$ , we have that

$$\mathbf{P}_{Y|Z,a,\vec{m}} = \mathbf{P}_{Y|Z,a,\vec{m}} \mathbf{P}_{W|Z,a,\vec{m}}^\dagger \mathbf{P}_{W|Z,a,\vec{m}} \quad (38)$$

Hence,  $\mathbf{P}_{Y|Z,a,\vec{m}} \mathbf{P}_{W|Z,a,\vec{m}}^\dagger$  defines  $b_1(y, a, \vec{m}, w)$  in Assumption 3.7 such that

$$p(y | z, a, \vec{m}) = \sum_w b_1(y, a, \vec{m}, w) p(w | z, a, \vec{m}).$$

Similarly, for each  $(a, \vec{m})$ , define  $\mathbf{P}_{\vec{M}|W,a,\vec{m}}$  and  $\mathbf{P}_{Z|W,a,\vec{m}}$ .

Then if  $\mathbf{P}_{Z|W,a,\vec{m}}$  is left invertible with left inverse  $\mathbf{P}_{Z|W,a,\vec{m}}^\dagger$ , we have that

$$\mathbf{P}_{\vec{M}|W,a,\vec{m}} = \mathbf{P}_{\vec{M}|W,a,\vec{m}} \mathbf{P}_{Z|W,a,\vec{m}}^\dagger \mathbf{P}_{Z|W,a,\vec{m}} \quad (39)$$

Hence,  $\mathbf{P}_{\vec{M}|W,a,\vec{m}} \mathbf{P}_{Z|W,a,\vec{m}}^\dagger$  defines  $b_0(\vec{m}, a, z)$  in Assumption 3.7 such that

$$p(\vec{m} | w, a) = \sum_z b_0(\vec{m}, a, z) p(z | w, a).$$

## E IDENTIFICATION PROOF FOR ASSUMPTION SET 3

*Proof.* By Theorem 1 in Hu and Schennach [2008], under Assumptions 3.9- 3.12,  $p(Y, \vec{u}_i, W, Z | a, \vec{m})$  is identified up to label  $i$  for each  $(a, \vec{m})$ . Then, by  $W \perp\!\!\!\perp A, \vec{M} \perp\!\!\!\perp \vec{U}$  in Assumption 3.9,  $p(W | u_i, a, \vec{m}) = p(W | u_i)$ . Consequently, the label  $u_i$  can be matched across the conditional distributions  $p(Y, \vec{u}_i, W, Z | a, \vec{m})$  for different values of  $(a, \vec{m})$ , thereby recovering the full-data law  $p(A, \vec{M}, Y, \vec{u}_i, W, Z)$  up to label  $i$ . Hence,  $p(Y(a) = y) = \sum_{\vec{u}, \vec{m}, a'} p(y | \vec{u}_i, a', \vec{m}) p(a' | \vec{u}_i) p(\vec{m} | a, \vec{u}_i) p(\vec{u}_i)$  is identified in Fig. 3b.  $\square$

## F FINITE-SUPPORT IDENTIFICATION EXAMPLE UNDER ASSUMPTION SET 3

*Proof.* Suppose  $A, \vec{M}, Y, \vec{U}, W, Z$  are have finite support, and furthermore that  $W, Z$ , and  $\vec{U}$  have the same number of categories.

Define

$$\mathbf{P}_{W|Z} = \begin{bmatrix} \Pr(w_1 | z_1) & \cdots & \Pr(w_1 | z_n) \\ \Pr(w_2 | z_1) & \cdots & \Pr(w_2 | z_n) \\ \vdots & \ddots & \vdots \\ \Pr(w_n | z_1) & \cdots & \Pr(w_n | z_n) \end{bmatrix},$$

and analogously define  $\mathbf{P}_{W|U}$ ,  $\mathbf{P}_{U|Z}$ , and  $\mathbf{P}_{y,W|Z,a,\vec{m}}$ . In other words, the  $(k, j)^{\text{th}}$  entry of  $\mathbf{P}_{y,W|Z,a,\vec{m}}$  is

$$\mathbf{P}_{y,W|Z,a,\vec{m}}(k, j) = \Pr(Y = y, W = w_k \mid Z = z_j, A = a, \vec{M} = \vec{m}).$$

Additionally, define  $\text{diag}(\mathbf{P}_{y|U,a,\vec{m}})$  as the diagonal matrix with  $i^{\text{th}}$  diagonal entry  $\Pr(Y = y \mid \vec{U} = \vec{u}_i, A = a, \vec{M} = \vec{m})$ .

By Assumption 3.9,

$$\mathbf{P}_{W|Z} = \mathbf{P}_{W|U} \mathbf{P}_{U|Z}, \text{ and} \quad (40)$$

$$\mathbf{P}_{y,W|Z,a,\vec{m}} = \mathbf{P}_{W|U} \text{diag}(\mathbf{P}_{y|U,a,\vec{m}}) \mathbf{P}_{U|Z}. \quad (41)$$

By Assumption 3.11,  $\mathbf{P}_{U|Z}$  and  $\mathbf{P}_{W|U}$  are invertible. Then we have the following:

$$\mathbf{P}_{W|Z}^{-1} = \mathbf{P}_{U|Z}^{-1} \mathbf{P}_{W|U}^{-1}.$$

Hence,

$$\mathbf{P}_{y,W|Z,a,\vec{m}} \mathbf{P}_{W|Z}^{-1} = \mathbf{P}_{W|U} \text{diag}(\mathbf{P}_{y|U,a,\vec{m}}) \mathbf{P}_{W|U}^{-1}. \quad (42)$$

Eigendecomposition of  $\mathbf{P}_{y,W|Z,a,\vec{m}} \mathbf{P}_{W|Z}^{-1}$  gives us the column vectors of  $\mathbf{P}_{W|U}$  under Assumption 3.12. In other words, we recover conditional distributions  $p(W \mid \vec{u}_i)$  [up to label  $i$ ]. Inverting  $p(Z, A, \vec{M}, Y, W) = \sum_i p(Z, A, \vec{M}, Y, \vec{u}_i) p(W \mid \vec{u}_i)$ , which holds by Assumption 3.9, we recover  $p(Z, A, \vec{M}, Y \mid \vec{u}_i)$  and  $p(\vec{u}_i)$ .

Hence, we recover  $p(A, \vec{M}, Y, \vec{u}_i, W, Z) = p(W \mid \vec{u}_i) p(Z, A, \vec{M}, Y \mid \vec{u}_i) p(\vec{u}_i)$ .

This proof extends to settings in which  $W$  and  $Z$  have higher cardinality than  $\vec{U}$  by appropriately coarsening variables.  $\square$

## G PLUG-IN ESTIMATORS FOR ASSUMPTION SETS 1, 2, AND 3

### G.1 PLUG-IN ESTIMATOR FOR ASSUMPTION SET 1

Under Assumptions 3.1–3.3, the counterfactual distribution in Figure 4a is identified as in Theorem 3.4.

Define the mean-induced bridge:

$$\bar{h}_0(w, a', a) := \sum_y y h_0(y, w, a', a). \quad (43)$$

Then the counterfactual mean satisfies

$$\mathbb{E}[Y(a)] = \sum_{a', w} \bar{h}_0(w, a', a) p(w, a') = \mathbb{E}[\bar{h}_0(W, A', a)], \quad (44)$$

and hence the average causal effect can be written as

$$\mathbb{E}[Y(1) - Y(0)] = \mathbb{E}[\bar{h}_0(W, A', 1)] - \mathbb{E}[\bar{h}_0(W, A', 0)]. \quad (45)$$

We estimate this quantity using the plug-in estimator

$$\hat{\psi}^{(\text{S1,PI})} = \mathbb{E}_n[\bar{h}_0(W, A', 1)] - \mathbb{E}_n[\bar{h}_0(W, A', 0)]. \quad (46)$$

Provided that  $\bar{h}_0(W, A', a_1) \xrightarrow{P} \bar{h}_0(W, A', a; \eta_1)$  and standard regularity conditions hold, consistency of  $\hat{\psi}^{(\text{S1})}$  follows.

## G.2 PLUG-IN ESTIMATOR FOR ASSUMPTION SET 2

Under Assumptions 3.5–3.7, the counterfactual distribution in Figure 4b is identified as in Theorem 3.8.

Define the mean-induced bridge:

$$\bar{b}_2(a', a, w, z) := \sum_{y, \vec{m}} y b_1(y, a', \vec{m}, w) b_0(\vec{m}, a, z). \quad (47)$$

Then the counterfactual mean satisfies

$$\mathbb{E}[Y(a)] = \sum_{a', w, z} \bar{b}_2(a', a, w, z) p(w, z, a'), \quad (48)$$

and hence the average causal effect can be written as

$$\mathbb{E}[Y(1) - Y(0)] = \mathbb{E}[\bar{b}_2(A', 1, W, Z)] - \mathbb{E}[\bar{b}_2(A', 0, W, Z)]. \quad (49)$$

We estimate this quantity using the estimator

$$\hat{\psi}^{(S2, PI)} = \mathbb{E}_n[\bar{b}_2(A', 1, W, Z_2)] - \mathbb{E}_n[\bar{b}_2(A', 0, W, Z_2)]. \quad (50)$$

Provided that  $\bar{b}_2(A', 1, W, Z_2) \xrightarrow{P} \bar{b}_2(A', 0, W, Z; \eta_2)$  and standard regularity conditions hold, consistency of  $\hat{\psi}^{(S2)}$  follows.

## G.3 PLUG-IN ESTIMATOR FOR ASSUMPTION SET 3

By Theorem 3.13, under Assumptions 3.9-3.12, the counterfactual distribution in Fig. 3b can be derived from the full law

$$\{p(A, \vec{M}, Y, \vec{u}_i, W, Z) : \vec{u}_i\}$$

up to relabeling of  $u_i$ . Distributional components  $p(Y|\vec{m}, a', \vec{u}_i)$ ,  $p(a'|\vec{u}_i)$ , and  $p(\vec{m}|a, \vec{u}_i)$  can be estimated for each level of  $\vec{u}_i$ , e.g. via eigendecomposition tasks details in Appendix F. Then, we construct the plug-in estimator

$$\hat{\psi}^{(S3, PI)} = \sum_{\vec{u}_i, \vec{m}, a'} \hat{E}[Y | \vec{m}, a', \vec{u}_i] \hat{p}(a' | \vec{u}_i) [\hat{p}(\vec{m} | a = 1, \vec{u}_i) - \hat{p}(\vec{m} | a = 0, \vec{u}_i)] \hat{p}(\vec{u}_i). \quad (51)$$

## H PROXY WEIGHT CONSTRUCTIONS FOR FINITE-SUPPORT HIDDEN VARIABLE $U$

For each  $(\vec{u}_i, a)$ , define

$$\mathcal{C}_{WZ}(\vec{u}_i, a) = \prod_{j \neq i} \left\{ \frac{h_j(W) - \mathbb{E}[h_j(W) | \vec{U} = \vec{u}_j]}{\mathbb{E}[h_j(W) | \vec{U} = \vec{u}_i] - \mathbb{E}[h_j(W) | \vec{U} = \vec{u}_j]} \cdot \frac{f_j(Z) - \mathbb{E}[f_j(Z) | \vec{U} = \vec{u}_j, a]}{\mathbb{E}[f_j(Z) | \vec{U} = \vec{u}_i, a] - \mathbb{E}[f_j(Z) | \vec{U} = \vec{u}_j, a]} \right\}, \quad (52)$$

and for each  $(\vec{u}_i, a, \vec{m})$ , define

$$\mathcal{C}_{YW}(\vec{u}_i, a, \vec{m}) = \prod_{j \neq i} \left\{ \frac{t_j(Y) - \mathbb{E}[t_j(Y) | \vec{U} = \vec{u}_j, a, \vec{m}]}{\mathbb{E}[t_j(Y) | \vec{U} = \vec{u}_i, a, \vec{m}] - \mathbb{E}[t_j(Y) | \vec{U} = \vec{u}_j, a, \vec{m}]} \cdot \frac{h_j(W) - \mathbb{E}[h_j(W) | \vec{U} = \vec{u}_j]}{\mathbb{E}[h_j(W) | \vec{U} = \vec{u}_i] - \mathbb{E}[h_j(W) | \vec{U} = \vec{u}_j]} \right\}, \quad (53)$$

$\mathcal{C}_{YZ}(\vec{u}_i, a, \vec{m}) =$

$$\prod_{j \neq i} \left\{ \frac{t_j(Y) - \mathbb{E}[t_j(Y) | \vec{U} = \vec{u}_j, a, \vec{m}]}{\mathbb{E}[t_j(Y) | \vec{U} = \vec{u}_i, a, \vec{m}] - \mathbb{E}[t_j(Y) | \vec{U} = \vec{u}_j, a, \vec{m}]} \cdot \frac{f_j(Z) - \mathbb{E}[f_j(Z) | \vec{U} = \vec{u}_j, a]}{\mathbb{E}[f_j(Z) | \vec{U} = \vec{u}_i, a] - \mathbb{E}[f_j(Z) | \vec{U} = \vec{u}_j, a]} \right\}, \quad (54)$$

$$\begin{aligned} \mathcal{C}_{YZW}(\vec{u}_i, a, \vec{m}) = & \prod_{j \neq i} \left\{ \frac{t_j(Y) - \mathbb{E}[t_j(Y) \mid \vec{U} = \vec{u}_j, a, \vec{m}]}{\mathbb{E}[t_j(Y) \mid \vec{U} = \vec{u}_i, a, \vec{m}] - \mathbb{E}[t_j(Y) \mid \vec{U} = \vec{u}_j, a, \vec{m}]} \cdot \frac{h_j(W) - \mathbb{E}[h_j(W) \mid \vec{U} = \vec{u}_j]}{\mathbb{E}[h_j(W) \mid \vec{U} = \vec{u}_i] - \mathbb{E}[h_j(W) \mid \vec{U} = \vec{u}_j]} \right. \\ & \left. \cdot \frac{f_j(Z) - \mathbb{E}[f_j(Z) \mid \vec{U} = \vec{u}_j, a]}{\mathbb{E}[f_j(Z) \mid \vec{U} = \vec{u}_i, a] - \mathbb{E}[f_j(Z) \mid \vec{U} = \vec{u}_j, a]} \right\}, \end{aligned} \quad (55)$$

where the collections of random functions  $\{h_j(W)\}_j$ ,  $\{f_j(Z)\}_j$ , and  $\{t_j(Y)\}_j$  are chosen such that for  $j' \neq j \neq i$ ,

$$\mathbb{E} \left[ \prod_{j \neq i} h_j(W) \mid \vec{U} = \vec{u}_i \right] = \prod_{j \neq i} \mathbb{E} [h_j(W) \mid \vec{U} = \vec{u}_i]; \quad (56)$$

$$\mathbb{E} \left[ \prod_{j \neq i} f_j(Z) \mid \vec{U} = \vec{u}_i, a \right] = \prod_{j \neq i} \mathbb{E} [f_j(Z) \mid \vec{U} = \vec{u}_i, a]; \quad (57)$$

$$\mathbb{E} \left[ \prod_{j \neq i} t_j(Y) \mid \vec{U} = \vec{u}_i, a, \vec{m} \right] = \prod_{j \neq i} \mathbb{E} [t_j(Y) \mid \vec{U} = \vec{u}_i, a, \vec{m}]. \quad (58)$$

We include the identity functions on  $W$ ,  $Z$ , and  $Y$  in the collections  $\{h_j(W)\}_j$ ,  $\{f_j(Z)\}_j$ , and  $\{t_j(Y)\}_j$ , respectively. When  $\vec{U}$  is binary, these collections reduce to the identity functions on  $W$ ,  $Z$ , and  $Y$ .

## I INFLUENCE FUNCTION FOR $\psi_a$ IN FIGURE 3b WITHOUT $\vec{U}$ HIDDEN

*Proof.* We aim to show  $\varphi_{\text{full}}(A, \vec{M}, Y, \vec{U}; \cdot)$  from Equation 11 is an influence function for  $\psi_a$  in Figure 3b without  $\vec{U}$  hidden (variables  $A, \vec{M}, Y, \vec{U}, W, Z$  are all observed). In other words, that

$$\left. \frac{d}{dt} \psi(P_\epsilon) \right|_{\epsilon=0} = \mathbb{E}_{A, \vec{M}, Y, \vec{U}, W, Z} \left[ \varphi_{\text{full}}(A, \vec{M}, Y, \vec{U}; \cdot) s_\epsilon(A, \vec{M}, Y, \vec{U}, W, Z) \right],$$

where score for regular parametric submodel  $p_\epsilon(A, \vec{M}, Y, \vec{U}, W, Z)$  is defined as

$$s_\epsilon(A, \vec{M}, Y, \vec{U}, W, Z) = \left. \frac{d}{d\epsilon} \log p_\epsilon(A, \vec{M}, Y, \vec{U}, W, Z) \right|_{\epsilon=0}$$

i.e.,

$$p_\epsilon(A, \vec{M}, Y, \vec{U}, W, Z) = p(A, \vec{M}, Y, \vec{U}, W, Z) (1 + \epsilon s_\epsilon(A, \vec{M}, Y, \vec{U}, W, Z)) + o(\epsilon).$$

We can construct a regular parametric submodel for the marginal law through marginalization of a regular parametric submodel for the full data law.

$$\begin{aligned} p_\epsilon(A, \vec{M}, Y, \vec{U}) &= \sum_{w, z} [p(A, \vec{M}, Y, \vec{U}, w, z) (1 + \epsilon s_\epsilon(A, \vec{M}, Y, \vec{U}, w, z)) + o(\epsilon)] \\ &= p(A, \vec{M}, Y, \vec{U}) \left[ 1 + \epsilon \mathbb{E}[s_\epsilon(A, \vec{M}, Y, \vec{U}, W, Z) \mid A, \vec{M}, Y, \vec{U}] \right] + o(\epsilon) \\ &= p(A, \vec{M}, Y, \vec{U}) (1 + \epsilon s_\epsilon(A, \vec{M}, Y, \vec{U})) + o(\epsilon), \end{aligned}$$

i.e.,

$$s_\epsilon(A, \vec{M}, Y, \vec{U}) = \left. \frac{d}{d\epsilon} \log p_\epsilon(A, \vec{M}, Y, \vec{U}) \right|_{\epsilon=0}.$$

Fulcher et al. [2020] already showed that

$$\left. \frac{d}{dt} \psi(P_\epsilon) \right|_{\epsilon=0} = \mathbb{E}_{A, \vec{M}, Y, \vec{U}} \left[ \varphi_{\text{full}}(A, \vec{M}, Y, \vec{U}; \cdot) s_\epsilon(A, \vec{M}, Y, \vec{U}) \right].$$

Then by Assumption 3.9, which also implies the factorization

$$p(A, \vec{M}, Y, \vec{U}, W, Z) = p(W \mid \vec{M}, A, \vec{U}) p(Z \mid \vec{M}, A, \vec{U}) p(A, \vec{M}, Y, \vec{U}),$$

we have

$$s_\epsilon(A, \vec{M}, Y, \vec{U}, W, Z) = s_\epsilon(A, \vec{M}, Y, \vec{U}) + s_\epsilon(W \mid \vec{M}, A, \vec{U}) + s_\epsilon(Z \mid \vec{M}, A, \vec{U}).$$

Finally by iterated expectations,

$$\begin{aligned} \left. \frac{d}{dt} \psi(P_\epsilon) \right|_{\epsilon=0} &= \mathbb{E}_{A, \vec{M}, Y, \vec{U}} \left[ \varphi_{\text{full}}(A, \vec{M}, Y, \vec{U}; \cdot) s_\epsilon(A, \vec{M}, Y, \vec{U}) \right] + 0 + 0 \\ &= \mathbb{E}_{A, \vec{M}, Y, \vec{U}} \left[ \varphi_{\text{full}}(A, \vec{M}, Y, \vec{U}; \cdot) s_\epsilon(A, \vec{M}, Y, \vec{U}) \right] \\ &\quad + \mathbb{E}_{A, \vec{M}, Y, \vec{U}} \left[ \varphi_{\text{full}}(A, \vec{M}, Y, \vec{U}; \cdot) \mathbb{E}[s_\epsilon(W \mid \vec{M}, A, \vec{U}) \mid A, \vec{M}, Y, \vec{U}] \right] \\ &\quad + \mathbb{E}_{A, \vec{M}, Y, \vec{U}} \left[ \varphi_{\text{full}}(A, \vec{M}, Y, \vec{U}; \cdot) \mathbb{E}[s_\epsilon(Z \mid \vec{M}, A, \vec{U}) \mid A, \vec{M}, Y, \vec{U}] \right] \\ &= \mathbb{E}_{A, \vec{M}, Y, \vec{U}, W, Z} \left[ \varphi_{\text{full}}(A, \vec{M}, Y, \vec{U}; \cdot) s_\epsilon(A, \vec{M}, Y, \vec{U}, W, Z) \right]. \end{aligned}$$

□

## J PROOF OF THEOREM 4.1

*Proof.* A regular parametric submodel for the observed data law is induced by a regular parametric submodel for the full data law through marginalization.

$$\begin{aligned} p_\epsilon(A, \vec{M}, Y, W, Z) &= \sum_u p_\epsilon(A, \vec{M}, Y, \vec{U}, W, Z) \\ &= \sum_u [p(A, \vec{M}, Y, u, W, Z) (1 + \epsilon s_\epsilon(A, \vec{M}, Y, u, W, Z)) + o(\epsilon)] \\ &= p(A, \vec{M}, Y, W, Z) \left[ 1 + \epsilon \mathbb{E}[s_\epsilon(A, \vec{M}, Y, \vec{U}, W, Z) \mid A, \vec{M}, Y, W, Z] \right] + o(\epsilon) \\ &= p(A, \vec{M}, Y, W, Z) (1 + \epsilon s_\epsilon(A, \vec{M}, Y, W, Z)) + o(\epsilon), \end{aligned}$$

i.e.,

$$s_\epsilon(A, \vec{M}, Y, W, Z) = \left. \frac{d}{d\epsilon} \log p_\epsilon(A, \vec{M}, Y, W, Z) \right|_{\epsilon=0}.$$

We aim to show  $\varphi_{\text{obs}}(A, \vec{M}, Y, W, Z; \cdot)$  from Equation 14 is an influence function for  $\psi_a$  in Figure 3b with  $\vec{U}$  hidden ( $A, \vec{M}, Y, W, Z$  observed):

$$\left. \frac{d}{dt} \psi(P_\epsilon) \right|_{\epsilon=0} = \mathbb{E}_{A, \vec{M}, Y, W, Z} \left[ \varphi_{\text{obs}}(A, \vec{M}, Y, W, Z; \cdot) s_\epsilon(A, \vec{M}, Y, W, Z) \right].$$

We already know from Appendix I that

$$\left. \frac{d}{dt} \psi(P_\epsilon) \right|_{\epsilon=0} = \mathbb{E}_{A, \vec{M}, Y, \vec{U}, W, Z} \left[ \varphi_{\text{full}}(A, \vec{M}, Y, \vec{U}; \cdot) s_\epsilon(A, \vec{M}, Y, \vec{U}, W, Z) \right].$$

Breaking up  $s_\epsilon(A, \vec{M}, Y, \vec{U}, W, Z)$ ,

$$\begin{aligned}
\left. \frac{d}{dt} \psi(P_\epsilon) \right|_{\epsilon=0} &= \mathbb{E} \left[ \varphi_{\text{full}}(A, \vec{M}, Y, \vec{U}; \cdot) s_\epsilon(A, \vec{M}, Y, \vec{U}) \right] \\
&\quad + \mathbb{E} \left[ \varphi_{\text{full}}(A, \vec{M}, Y, \vec{U}; \cdot) s_\epsilon(W \mid \vec{M}, A, \vec{U}) \right] \\
&\quad + \mathbb{E} \left[ \varphi_{\text{full}}(A, \vec{M}, Y, \vec{U}; \cdot) s_\epsilon(Z \mid \vec{M}, A, \vec{U}) \right] \\
&= \mathbb{E} \left[ \mathbb{E} \left\{ \varphi_{\text{full}}(A, \vec{M}, Y, \vec{U}; \cdot) \mid A, \vec{M}, Y, \vec{U} \right\} s_\epsilon(A, \vec{M}, Y, \vec{U}) \right] \\
&\quad + \mathbb{E} \left[ \mathbb{E} \left\{ \varphi_{\text{full}}(A, \vec{M}, Y, \vec{U}; \cdot) \mid \vec{M}, A, \vec{U}, W \right\} s_\epsilon(W \mid \vec{M}, A, \vec{U}) \right] \\
&\quad + \mathbb{E} \left[ \mathbb{E} \left\{ \varphi_{\text{full}}(A, \vec{M}, Y, \vec{U}; \cdot) \mid \vec{M}, A, \vec{U}, Z \right\} s_\epsilon(Z \mid \vec{M}, A, \vec{U}) \right].
\end{aligned}$$

Hence, it suffices to verify that the following conditional expectations coincide:

1.  $\mathbb{E} \left[ \varphi_{\text{obs}}(A, \vec{M}, Y, W, Z; \cdot) \mid A, \vec{M}, Y, \vec{U} \right] = \mathbb{E} \left[ \varphi_{\text{full}}(A, \vec{M}, Y, \vec{U}; \cdot) \mid A, \vec{M}, Y, \vec{U} \right]$
2.  $\mathbb{E} \left[ \varphi_{\text{obs}}(A, \vec{M}, Y, W, Z; \cdot) \mid \vec{M}, A, \vec{U}, W \right] = \mathbb{E} \left[ \varphi_{\text{full}}(A, \vec{M}, Y, \vec{U}; \cdot) \mid \vec{M}, A, \vec{U}, W \right]$
3.  $\mathbb{E} \left[ \varphi_{\text{obs}}(A, \vec{M}, Y, W, Z; \cdot) \mid \vec{M}, A, \vec{U}, Z \right] = \mathbb{E} \left[ \varphi_{\text{full}}(A, \vec{M}, Y, \vec{U}; \cdot) \mid \vec{M}, A, \vec{U}, Z \right]$

because then

$$\begin{aligned}
\left. \frac{d}{dt} \psi(P_\epsilon) \right|_{\epsilon=0} &= \mathbb{E}_{A, \vec{M}, Y, \vec{U}, W, Z} \left[ \varphi_{\text{obs}}(A, \vec{M}, Y, W, Z; \cdot) s_\epsilon(A, \vec{M}, Y, \vec{U}, W, Z) \right] \\
&= \mathbb{E}_{A, \vec{M}, Y, W, Z} \left[ \varphi_{\text{obs}}(A, \vec{M}, Y, W, Z; \cdot) \mathbb{E} [s_\epsilon(A, \vec{M}, Y, \vec{U}, W, Z) \mid A, \vec{M}, Y, W, Z] \right] \\
&= \mathbb{E}_{A, \vec{M}, Y, W, Z} \left[ \varphi_{\text{obs}}(A, \vec{M}, Y, W, Z; \cdot) s_\epsilon(A, \vec{M}, Y, W, Z) \right].
\end{aligned}$$

Indeed, (1)-(3) hold noting Equation 12 and 13, as well as Assumption 3.9:

1.  $\mathbb{E} \left[ \varphi_{\text{obs}}(A, \vec{M}, Y, W, Z; \cdot) \mid A, \vec{M}, Y, \vec{U} \right] = \mathbb{E} \left[ \varphi_{\text{full}}(A, \vec{M}, Y, \vec{U}; \cdot) \mid A, \vec{M}, Y, \vec{U} \right] = \varphi_{\text{full}}(A, \vec{M}, Y, \vec{U}; \cdot)$
2.  $\mathbb{E} \left[ \varphi_{\text{obs}}(A, \vec{M}, Y, W, Z; \cdot) \mid \vec{M}, A, \vec{U}, W \right] = \mathbb{E} \left[ \varphi_{\text{full}}(A, \vec{M}, Y, \vec{U}; \cdot) \mid \vec{M}, A, \vec{U}, W \right] = \varphi_{\text{full}_2}(\vec{M}, A, \vec{U}) + \varphi_{\text{full}_3}(A, \vec{U}) - \psi_a$
3.  $\mathbb{E} \left[ \varphi_{\text{obs}}(A, \vec{M}, Y, W, Z; \cdot) \mid \vec{M}, A, \vec{U}, Z \right] = \mathbb{E} \left[ \varphi_{\text{full}}(A, \vec{M}, Y, \vec{U}; \cdot) \mid \vec{M}, A, \vec{U}, Z \right] = \varphi_{\text{full}_2}(\vec{M}, A, \vec{U}) + \varphi_{\text{full}_3}(A, \vec{U}) - \psi_a$

□

## K PROOF OF THEOREM 4.2

*Proof.* Solving Equation 10 with  $\varphi_{\text{obs}}(A, \vec{M}, Y, W, Z; \cdot)$  in Equation 14 gives

$$\begin{aligned}
\hat{\psi}_a^{(S3, IF)} &= \mathbb{E}_n \left[ \sum_i \mathcal{C}_{WZ}(A, \vec{u}_i) \varphi_{\text{full}_1}(Y, \vec{M}, A, \vec{u}_i) \right] \\
&\quad + \mathbb{E}_n \left[ \sum_i \left\{ \mathcal{C}_{WZ}(A, \vec{u}_i) + \mathcal{C}_{YW}(\vec{M}, A, \vec{u}_i) + \mathcal{C}_{YZ}(\vec{M}, A, \vec{u}_i) - 2 \mathcal{C}_{WZY}(\vec{M}, A, \vec{u}_i) \right\} \right. \\
&\quad \left. \cdot \left\{ \varphi_{\text{full}_2}(\vec{M}, A, \vec{u}_i) + \varphi_{\text{full}_3}(A, \vec{u}_i) \right\} \right].
\end{aligned}$$

In the following arguments, we use  $\mathbb{E}_{n,\text{sub}}[\cdot \mid \cdot]$  to denote the empirical average taken only over the subsample of observations satisfying the conditioning event appearing after the vertical bar  $\mid$ . Standard regularity conditions are assumed.

**Case 1.**  $p(\vec{M} \mid A, \vec{U})$ ,  $\{\mathbb{E}[h_j(W) \mid \vec{U}]\}_j$ ,  $\{\mathbb{E}[f_j(Z) \mid \vec{U}, A]\}_j$  are correctly specified.

Since  $\{\mathbb{E}[h_j(W) \mid \vec{U}]\}_j$  and  $\{\mathbb{E}[f_j(Z) \mid \vec{U}, A]\}_j$  are correctly specified, and Assumption 3.9 holds,

$$\begin{aligned} \mathbb{E}_{n,\text{sub}}[\mathcal{C}_{WZ}(A, \vec{u}_i) \mid Y, \vec{M}, A, \vec{U} = \vec{u}_i] &\xrightarrow{P} \mathbb{E}[\mathcal{C}_{WZ}(A, \vec{u}_i) \mid Y, \vec{M}, A, \vec{U} = \vec{u}_i] = 1, \text{ and} \\ \mathbb{E}_{n,\text{sub}}[\mathcal{C}_{WZ}(A, \vec{u}_j) \mid Y, \vec{M}, A, \vec{U} = \vec{u}_i] &\xrightarrow{P} \mathbb{E}[\mathcal{C}_{WZ}(A, \vec{u}_j) \mid Y, \vec{M}, A, \vec{U} = \vec{u}_i] = 0, j \neq i. \end{aligned}$$

Additionally,

$$\begin{aligned} \mathbb{E}_{n,\text{sub}}[\mathcal{C}_{WZ}(A, \vec{u}_i) + \mathcal{C}_{YW}(\vec{M}, A, \vec{u}_i) + \mathcal{C}_{YZ}(\vec{M}, A, \vec{u}_i) - 2\mathcal{C}_{WZY}(\vec{M}, A, \vec{u}_i) \mid \vec{M}, A, \vec{U} = \vec{u}_i] &\xrightarrow{P} \\ \mathbb{E}[\mathcal{C}_{WZ}(A, \vec{u}_i) + \mathcal{C}_{YW}(\vec{M}, A, \vec{u}_i) + \mathcal{C}_{YZ}(\vec{M}, A, \vec{u}_i) - 2\mathcal{C}_{WZY}(\vec{M}, A, \vec{u}_i) \mid \vec{M}, A, \vec{U} = \vec{u}_i] &= 1, \text{ and} \\ \mathbb{E}_{n,\text{sub}}[\mathcal{C}_{WZ}(A, \vec{u}_j) + \mathcal{C}_{YW}(\vec{M}, A, \vec{u}_j) + \mathcal{C}_{YZ}(\vec{M}, A, \vec{u}_j) - 2\mathcal{C}_{WZY}(\vec{M}, A, \vec{u}_j) \mid \vec{M}, A, \vec{U} = \vec{u}_i] &\xrightarrow{P} \\ \mathbb{E}[\mathcal{C}_{WZ}(A, \vec{u}_j) + \mathcal{C}_{YW}(\vec{M}, A, \vec{u}_j) + \mathcal{C}_{YZ}(\vec{M}, A, \vec{u}_j) - 2\mathcal{C}_{WZY}(\vec{M}, A, \vec{u}_j) \mid \vec{M}, A, \vec{U} = \vec{u}_i] &= 0, j \neq i. \end{aligned}$$

Fulcher et al. [2020] show that, provided  $p(\vec{M} \mid A, \vec{U})$  is correctly specified,

$$\mathbb{E}_n[\varphi_{\text{full}_1}(Y, \vec{M}, A, \vec{U}) + \varphi_{\text{full}_2}(\vec{M}, A, \vec{U}) + \varphi_{\text{full}_3}(A, \vec{U})] \xrightarrow{P} \psi_a.$$

Combining these results, it follows that

$$\begin{aligned} \hat{\psi}_a^{(S3, IF)} &= \mathbb{E}_n[\mathbb{E}_{n,\text{sub}}[\sum_i \mathcal{C}_{WZ}(A, \vec{u}_i) \varphi_{\text{full}_1}(Y, \vec{M}, A, \vec{u}_i)] \mid Y, \vec{M}, A, \vec{U}] \\ &\quad + \mathbb{E}_n[\mathbb{E}_{n,\text{sub}}[\sum_i \left\{ \mathcal{C}_{WZ}(A, \vec{u}_i) + \mathcal{C}_{YW}(\vec{M}, A, \vec{u}_i) + \mathcal{C}_{YZ}(\vec{M}, A, \vec{u}_i) - 2\mathcal{C}_{WZY}(\vec{M}, A, \vec{u}_i) \right\} \\ &\quad \cdot \left\{ \varphi_{\text{full}_2}(\vec{M}, A, \vec{u}_i) + \varphi_{\text{full}_3}(A, \vec{u}_i) \right\}] \mid \vec{M}, A, \vec{U}] \xrightarrow{P} \psi_a. \end{aligned}$$

**Case 2.**  $p(\vec{M} \mid A, \vec{U})$ ,  $\{\mathbb{E}[h_j(W) \mid \vec{U}]\}_j$ ,  $\{\mathbb{E}[t_j(Y) \mid \vec{M}, A, \vec{U}]\}_j$  are correctly specified.

Since  $\{\mathbb{E}[h_j(W) \mid \vec{U}]\}_j$  is correctly specified, and Assumption 3.9 holds,

$$\begin{aligned} \mathbb{E}_{n,\text{sub}}[\mathcal{C}_{WZ}(A, \vec{u}_i) \mid Y, \vec{M}, A, \vec{U} = \vec{u}_i] &\xrightarrow{P} k, \text{ where } k \text{ is a constant, and} \\ \mathbb{E}_{n,\text{sub}}[\mathcal{C}_{WZ}(A, \vec{u}_j) \mid Y, \vec{M}, A, \vec{U} = \vec{u}_i] &\xrightarrow{P} 0, j \neq i. \end{aligned}$$

Then, since  $\{\mathbb{E}[t_j(Y) \mid \vec{M}, A, \vec{U}]\}_j$  (thereby  $\mathbb{E}[Y \mid \vec{M}, A, \vec{U}]$ ) is correctly specified,

$$\mathbb{E}_n[\mathbb{E}_{n,\text{sub}}[\sum_i \mathcal{C}_{WZ}(A, \vec{u}_i) \varphi_{\text{full}_1}(Y, \vec{M}, A, \vec{u}_i)] \mid Y, \vec{M}, A, \vec{U}] \xrightarrow{P} 0.$$

Since  $\{\mathbb{E}[h_j(W) \mid \vec{U}]\}_j$ ,  $\{\mathbb{E}[t_j(Y) \mid \vec{M}, A, \vec{U}]\}_j$  are correctly specified, and Assumption 3.9 holds,

$$\begin{aligned} \mathbb{E}_n[\mathcal{C}_{WZ}(A, \vec{u}_i) + \mathcal{C}_{YW}(\vec{M}, A, \vec{u}_i) + \mathcal{C}_{YZ}(\vec{M}, A, \vec{u}_i) - 2\mathcal{C}_{WZY}(\vec{M}, A, \vec{u}_i) \mid \vec{M}, A, \vec{U} = \vec{u}_i] &\xrightarrow{P} \\ \mathbb{E}[\mathcal{C}_{WZ}(A, \vec{u}_i) + \mathcal{C}_{YW}(\vec{M}, A, \vec{u}_i) + \mathcal{C}_{YZ}(\vec{M}, A, \vec{u}_i) - 2\mathcal{C}_{WZY}(\vec{M}, A, \vec{u}_i) \mid \vec{M}, A, \vec{U} = \vec{u}_i] &= 1, \text{ and} \end{aligned}$$

$$\begin{aligned} & \mathbb{E}_{n,\text{sub}}[\mathcal{C}_{WZ}(A, \vec{u}_j) + \mathcal{C}_{YW}(\vec{M}, A, \vec{u}_j) + \mathcal{C}_{YZ}(\vec{M}, A, \vec{u}_j) - 2\mathcal{C}_{WZY}(\vec{M}, A, \vec{u}_j) \mid \vec{M}, A, \vec{U} = \vec{u}_i] \xrightarrow{P} \\ & \mathbb{E}[\mathcal{C}_{WZ}(A, \vec{u}_j) + \mathcal{C}_{YW}(\vec{M}, A, \vec{u}_j) + \mathcal{C}_{YZ}(\vec{M}, A, \vec{u}_j) - 2\mathcal{C}_{WZY}(\vec{M}, A, \vec{u}_j) \mid \vec{M}, A, \vec{U} = \vec{u}_i] = 0, j \neq i. \end{aligned}$$

Fulcher et al. [2020] show that, provided  $p(\vec{M} \mid A, \vec{U})$  is correctly specified,

$$\mathbb{E}_n[\varphi_{\text{full}_2}(\vec{M}, A, \vec{U}) + \varphi_{\text{full}_3}(A, \vec{U})] \xrightarrow{P} \psi_a.$$

Combining these results, it follows that

$$\begin{aligned} \hat{\psi}_a^{(S3, IF)} &= \mathbb{E}_n \left[ \mathbb{E}_{n,\text{sub}} \left[ \sum_i \mathcal{C}_{WZ}(A, \vec{u}_i) \varphi_{\text{full}_1}(Y, \vec{M}, A, \vec{u}_i) \right] \mid Y, \vec{M}, A, \vec{U} \right] \\ &+ \mathbb{E}_n \left[ \mathbb{E}_{n,\text{sub}} \left[ \sum_i \left\{ \mathcal{C}_{WZ}(A, \vec{u}_i) + \mathcal{C}_{YW}(\vec{M}, A, \vec{u}_i) + \mathcal{C}_{YZ}(\vec{M}, A, \vec{u}_i) - 2\mathcal{C}_{WZY}(\vec{M}, A, \vec{u}_i) \right\} \right. \right. \\ &\quad \left. \left. \cdot \left\{ \varphi_{\text{full}_2}(\vec{M}, A, \vec{u}_i) + \varphi_{\text{full}_3}(A, \vec{u}_i) \right\} \right] \mid \vec{M}, A, \vec{U} \right] \xrightarrow{P} 0 + \psi_a = \psi_a. \end{aligned}$$

**Case 3.**  $p(\vec{M} \mid A, \vec{U})$ ,  $\{\mathbb{E}[f_j(Z) \mid \vec{U}, A]\}_j$ ,  $\{\mathbb{E}[t_j(Y) \mid \vec{M}, A, \vec{U}]\}_j$  are correctly specified.

The same reasoning as in Case 2 applies, replacing  $\{\mathbb{E}[h_j(W) \mid \vec{U}]\}_j$  with  $\{\mathbb{E}[f_j(Z) \mid \vec{U}, A]\}_j$ .

**Case 4.**  $\{\mathbb{E}[t_j(Y) \mid \vec{M}, A, \vec{U}]\}_j$ ,  $\pi(A \mid \vec{U})$ ,  $\{\mathbb{E}[h_j(W) \mid \vec{U}]\}_j$  are correctly specified.

The same reasoning as Case 2 applies, noting Fulcher et al. [2020] show that, provided  $\{\mathbb{E}[t_j(Y) \mid \vec{M}, A, \vec{U}]\}_j$  and  $\pi(A \mid \vec{U})$  are correctly specified,

$$\mathbb{E}_n[\varphi_{\text{full}_2}(\vec{M}, A, \vec{U}) + \varphi_{\text{full}_3}(A, \vec{U})] \xrightarrow{P} \psi_a.$$

**Case 5.**  $\{\mathbb{E}[t_j(Y) \mid \vec{M}, A, \vec{U}]\}_j$ ,  $\pi(A \mid \vec{U})$ ,  $\{\mathbb{E}[f_j(Z) \mid \vec{U}, A]\}_j$  are correctly specified.

The same reasoning as in Case 4 applies, replacing  $\{\mathbb{E}[h_j(W) \mid \vec{U}]\}_j$  with  $\{\mathbb{E}[f_j(Z) \mid \vec{U}, A]\}_j$ .

□

## L PROOF OF THEOREM 4.3

*Proof.* Let  $\varphi = \varphi(O; \psi_a, \eta)$  from Theorem 4.1. Expand

$$\hat{\psi}_a - \psi_a = \mathbb{P}_n \hat{\varphi} - \mathbb{P} \varphi = (\mathbb{P}_n - \mathbb{P})(\hat{\varphi} - \varphi) + \mathbb{P}(\hat{\varphi} - \varphi) + (\mathbb{P}_n - \mathbb{P})\varphi = R_1 + R_2 + (\mathbb{P}_n - \mathbb{P})\varphi,$$

where  $R_1 = (\mathbb{P}_n - \mathbb{P})(\hat{\varphi} - \varphi)$ ,  $R_2 = \mathbb{P}(\hat{\varphi} - \varphi)$ .

$R_1 = o_p(n^{-1/2})$  under consistency of nuisance functions and sample splitting by Lemma 2 in Kennedy et al. [2020].

Below we show  $R_2 = o_p(n^{-1/2})$  under consistency of nuisance functions and conditions (1)-(5), therefore

$$\hat{\psi}_a - \psi_a = (\mathbb{P}_n - \mathbb{P})\varphi + o_p(n^{-1/2}) = \mathbb{P}_n \varphi + o_p(n^{-1/2})$$

which is equivalent to  $\sqrt{n}(\hat{\psi}_a - \psi_a) \xrightarrow{d} \mathcal{N}(0, \mathbb{E}[\varphi^2])$ .

Note that

$$\begin{aligned} R_2 &= \mathbb{P}[\hat{\varphi} - \varphi] \\ &= \sum_i \mathbb{E} \left[ \hat{\mathcal{C}}_{WZ}(A, \vec{u}_i) \hat{\varphi}_{\text{full},1}(Y, \vec{M}, A, \vec{u}_i) \right] \\ &+ \sum_i \mathbb{E} \left[ \left\{ \hat{\mathcal{C}}_{WZ}(A, \vec{u}_i) + \hat{\mathcal{C}}_{YW}(\vec{M}, A, \vec{u}_i) + \hat{\mathcal{C}}_{YZ}(\vec{M}, A, \vec{u}_i) - 2\hat{\mathcal{C}}_{WZY}(\vec{M}, A, \vec{u}_i) \right\} \left\{ \hat{\varphi}_{\text{full},2}(\vec{M}, A, \vec{u}_i) + \hat{\varphi}_{\text{full},3}(A, \vec{u}_i) \right\} \right] - \psi_a. \end{aligned}$$



Consider the first term

$$\begin{aligned}
& \sum_i \mathbb{E} \left[ \hat{\mathcal{C}}_{WZ}(A, \vec{u}_i) \hat{\varphi}_{\text{full},1}(Y, \vec{M}, A, \vec{u}_i) \right] \\
&= \sum_i \mathbb{E} \left\{ \frac{\hat{p}(\vec{M} \mid A = a, \vec{u}_i)}{\hat{p}(\vec{M} \mid A, \vec{u}_i)} \mathbb{E}[\hat{\mathcal{C}}_{WZ}(A, \vec{u}_i) \{Y - \hat{\mu}(\vec{M}, A, \vec{u}_i)\} \mid \vec{M}, A, \vec{U}] \right\} \\
&= \sum_i \mathbb{E} \left\{ \frac{\hat{p}(\vec{M} \mid A = a, \vec{u}_i)}{\hat{p}(\vec{M} \mid A, \vec{u}_i)} \mathbb{E}[\hat{\mathcal{C}}_{WZ}(A, \vec{u}_i) \mid \vec{M}, A, \vec{U}] [\mu(\vec{M}, A, \vec{U}) - \hat{\mu}(\vec{M}, A, \vec{u}_i)] \right\}.
\end{aligned}$$

Recall the following from Appendix H:

$$\begin{aligned}
& \mathbb{E}[\hat{\mathcal{C}}_{WZ}(A, \vec{u}_i) \mid \vec{M}, A, \vec{U} = \vec{u}_j] \\
&= \prod_{j \neq i} \left\{ \frac{\mathbb{E}[h_j(W) \mid \vec{U} = \vec{u}_j] - \hat{\mathbb{E}}[h_j(W) \mid \vec{U} = \vec{u}_j]}{\hat{\mathbb{E}}[h_j(W) \mid \vec{U} = \vec{u}_i] - \hat{\mathbb{E}}[h_j(W) \mid \vec{U} = \vec{u}_j]} \cdot \frac{\mathbb{E}[g_j(Z) \mid \vec{U} = \vec{u}_j, A] - \hat{\mathbb{E}}[g_j(Z) \mid \vec{U} = \vec{u}_j, A]}{\hat{\mathbb{E}}[g_j(Z) \mid \vec{U} = \vec{u}_i, A] - \hat{\mathbb{E}}[g_j(Z) \mid \vec{U} = \vec{u}_j, A]} \right\}.
\end{aligned}$$

Then by (5),  $\mathbb{E}[\hat{\mathcal{C}}_{WZ}(\vec{u}_j) \mid \vec{M}, A, \vec{u}_j] = o_p(n^{-1/2})$ .

Additionally we have that

$$\begin{aligned}
& \mathbb{E}[\hat{\mathcal{C}}_{WZ}(A, \vec{u}_i) \mid \vec{M}, A, \vec{U} = \vec{u}_i] \\
&= \prod_{j \neq i} \left\{ \frac{\mathbb{E}[h_j(W) \mid \vec{U} = \vec{u}_i] - \hat{\mathbb{E}}[h_j(W) \mid \vec{U} = \vec{u}_j]}{\hat{\mathbb{E}}[h_j(W) \mid \vec{U} = \vec{u}_i] - \hat{\mathbb{E}}[h_j(W) \mid \vec{U} = \vec{u}_j]} \cdot \frac{\mathbb{E}[g_j(Z) \mid \vec{U} = \vec{u}_i, A] - \hat{\mathbb{E}}[g_j(Z) \mid \vec{U} = \vec{u}_j, A]}{\hat{\mathbb{E}}[g_j(Z) \mid \vec{U} = \vec{u}_i, A] - \hat{\mathbb{E}}[g_j(Z) \mid \vec{U} = \vec{u}_j, A]} \right\} \\
&= \prod_{j \neq i} \left\{ \frac{\mathbb{E}[h_j(W) \mid \vec{U} = \vec{u}_i] - \hat{\mathbb{E}}[h_j(W) \mid \vec{U} = \vec{u}_i] + \hat{\mathbb{E}}[h_j(W) \mid \vec{U} = \vec{u}_i] - \hat{\mathbb{E}}[h_j(W) \mid \vec{U} = \vec{u}_j]}{\hat{\mathbb{E}}[h_j(W) \mid \vec{U} = \vec{u}_i] - \hat{\mathbb{E}}[h_j(W) \mid \vec{U} = \vec{u}_j]} \right. \\
&\quad \cdot \left. \frac{\mathbb{E}[g_j(Z) \mid \vec{U} = \vec{u}_i, A] - \hat{\mathbb{E}}[g_j(Z) \mid \vec{U} = \vec{u}_i, A] + \hat{\mathbb{E}}[g_j(Z) \mid \vec{U} = \vec{u}_i, A] - \hat{\mathbb{E}}[g_j(Z) \mid \vec{U} = \vec{u}_j, A]}{\hat{\mathbb{E}}[g_j(Z) \mid \vec{U} = \vec{u}_i, A] - \hat{\mathbb{E}}[g_j(Z) \mid \vec{U} = \vec{u}_j, A]} \right\} \\
&= \prod_{j \neq i} \left\{ \frac{\mathbb{E}[h_j(W) \mid \vec{U} = \vec{u}_i] - \hat{\mathbb{E}}[h_j(W) \mid \vec{U} = \vec{u}_i]}{\hat{\mathbb{E}}[h_j(W) \mid \vec{U} = \vec{u}_i] - \hat{\mathbb{E}}[h_j(W) \mid \vec{U} = \vec{u}_j]} + 1 \right\} \cdot \left\{ \frac{\mathbb{E}[g_j(Z) \mid \vec{U} = \vec{u}_i, A] - \hat{\mathbb{E}}[g_j(Z) \mid \vec{U} = \vec{u}_i, A]}{\hat{\mathbb{E}}[g_j(Z) \mid \vec{U} = \vec{u}_i, A] - \hat{\mathbb{E}}[g_j(Z) \mid \vec{U} = \vec{u}_j, A]} + 1 \right\}.
\end{aligned}$$

By (3) and (4),  $\mathbb{E}[\hat{\mathcal{C}}_{WZ}(A, \vec{u}_i) \mid \vec{M}, A, \vec{u}_i] [\mu(\vec{M}, A, \vec{u}_i) - \hat{\mu}(\vec{M}, A, \vec{u}_i)] = o_p(n^{-1/2})$ . Hence, the first term is  $\sqrt{n}$ -consistent, i.e.,  $\sum_i \mathbb{E}[\hat{\mathcal{C}}_{WZ}(A, \vec{u}_i) \hat{\varphi}_{\text{full},1}(Y, \vec{M}, A, \vec{u}_i)] = o_p(n^{-1/2})$ .

Now consider the second term

$$\begin{aligned}
& \sum_i \mathbb{E} \left[ \{ \hat{\mathcal{C}}_{WZ}(A, \vec{u}_i) + \hat{\mathcal{C}}_{YW}(\vec{M}, A, \vec{u}_i) + \hat{\mathcal{C}}_{YZ}(\vec{M}, A, \vec{u}_i) - 2\hat{\mathcal{C}}_{YZ}(\vec{M}, A, \vec{u}_i) \} \{ \hat{\varphi}_{\text{full},2}(\vec{M}, A, \vec{u}_i) + \hat{\varphi}_{\text{full},3}(A, \vec{u}_i) \} \right] - \psi_a \\
&= \sum_i \mathbb{E} \left[ \{ \hat{\varphi}_{\text{full},2}(\vec{M}, A, \vec{u}_i) + \hat{\varphi}_{\text{full},3}(A, \vec{u}_i) \} \{ \mathbb{E}[\hat{\mathcal{C}}_{WZ}(A, \vec{u}_i) \mid \vec{M}, A, \vec{U}] + \mathbb{E}[\hat{\mathcal{C}}_{YW}(\vec{M}, A, \vec{u}_i) \mid \vec{M}, A, \vec{U}] \right. \\
&\quad \left. + \mathbb{E}[\hat{\mathcal{C}}_{YZ}(\vec{M}, A, \vec{u}_i) \mid \vec{M}, A, \vec{U}] - 2\mathbb{E}[\hat{\mathcal{C}}_{YZ}(\vec{M}, A, \vec{u}_i) \mid \vec{M}, A, \vec{U}] \} \right] - \psi_a.
\end{aligned}$$

By (3)-(5),  $\mathbb{E}[\hat{\mathcal{C}}_{WZ}(A, \vec{u}_j) \mid \vec{M}, A, \vec{u}_j] = o_p(n^{-1/2})$ ,  $\mathbb{E}[\hat{\mathcal{C}}_{YW}(\vec{M}, A, \vec{u}_j) \mid \vec{M}, A, \vec{u}_j] = o_p(n^{-1/2})$ ,  $\mathbb{E}[\hat{\mathcal{C}}_{YZ}(\vec{M}, A, \vec{u}_j) \mid \vec{M}, A, \vec{u}_j] = o_p(n^{-1/2})$ , and  $\mathbb{E}[\hat{\mathcal{C}}_{YZ}(\vec{M}, A, \vec{u}_j) \mid \vec{M}, A, \vec{u}_j] = o_p(n^{-1/2})$ .

Additionally, we have that

$$\begin{aligned} & \mathbb{E}[\hat{\mathcal{C}}_{WZ}(A, \vec{u}_i) \mid \vec{M}, A, \vec{U} = \vec{u}_i] \\ &= \prod_{j \neq i} \left\{ \frac{\mathbb{E}[h_j(W) \mid \vec{U} = \vec{u}_i] - \hat{\mathbb{E}}[h_j(W) \mid \vec{U} = \vec{u}_i]}{\hat{\mathbb{E}}[h_j(W) \mid \vec{U} = \vec{u}_i] - \hat{\mathbb{E}}[h_j(W) \mid \vec{U} = \vec{u}_j]} + 1 \right\} \cdot \left\{ \frac{\mathbb{E}[g_j(Z) \mid \vec{U} = \vec{u}_i, A] - \hat{\mathbb{E}}[g_j(Z) \mid \vec{U} = \vec{u}_i, A]}{\hat{\mathbb{E}}[g_j(Z) \mid \vec{U} = \vec{u}_i, A] - \hat{\mathbb{E}}[g_j(Z) \mid \vec{U} = \vec{u}_j, A]} + 1 \right\}, \end{aligned}$$

and likewise for  $\mathbb{E}[\hat{\mathcal{C}}_{YW}(\vec{M}, A, \vec{u}_i) \mid \vec{M}, A, \vec{U} = \vec{u}_i]$  and  $\mathbb{E}[\hat{\mathcal{C}}_{YZ}(\vec{M}, A, \vec{u}_i) \mid \vec{M}, A, \vec{U} = \vec{u}_i]$ .

Similarly,

$$\begin{aligned} & \mathbb{E}[\hat{\mathcal{C}}_{YWZ}(\vec{M}, A, \vec{u}_i) \mid \vec{M}, A, \vec{U} = \vec{u}_i] \\ &= \prod_{j \neq i} \left\{ \frac{\mathbb{E}[t_j(Y) \mid \vec{M}, A, \vec{U} = \vec{u}_i] - \hat{\mathbb{E}}[t_j(Y) \mid \vec{M}, A, \vec{U} = \vec{u}_i]}{\hat{\mathbb{E}}[t_j(Y) \mid \vec{M}, A, \vec{U} = \vec{u}_i] - \hat{\mathbb{E}}[t_j(Y) \mid \vec{M}, A, \vec{U} = \vec{u}_j]} + 1 \right\} \cdot \left\{ \frac{\mathbb{E}[h_j(W) \mid \vec{U} = \vec{u}_i] - \hat{\mathbb{E}}[h_j(W) \mid \vec{U} = \vec{u}_i]}{\hat{\mathbb{E}}[h_j(W) \mid \vec{U} = \vec{u}_i] - \hat{\mathbb{E}}[h_j(W) \mid \vec{U} = \vec{u}_j]} + 1 \right\} \\ & \cdot \left\{ \frac{\mathbb{E}[g_j(Z) \mid \vec{U} = \vec{u}_i, A] - \hat{\mathbb{E}}[g_j(Z) \mid \vec{U} = \vec{u}_i, A]}{\hat{\mathbb{E}}[g_j(Z) \mid \vec{U} = \vec{u}_i, A] - \hat{\mathbb{E}}[g_j(Z) \mid \vec{U} = \vec{u}_j, A]} + 1 \right\}. \end{aligned}$$

By (3)–(5),

$$\begin{aligned} & \mathbb{E}[\hat{\mathcal{C}}_{WZ}(A, \vec{u}_i) \mid \vec{M}, A, \vec{u}_i] + \mathbb{E}[\hat{\mathcal{C}}_{YW}(\vec{M}, A, \vec{u}_i) \mid \vec{M}, A, \vec{u}_i] \\ & + \mathbb{E}[\hat{\mathcal{C}}_{YZ}(\vec{M}, A, \vec{u}_i) \mid \vec{M}, A, \vec{u}_i] - 2\mathbb{E}[\hat{\mathcal{C}}_{YWZ}(\vec{M}, A, \vec{u}_i) \mid \vec{M}, A, \vec{u}_i] = 1 + o_p(n^{-1/2}). \end{aligned}$$

Thus, we have for the second term

$$\begin{aligned} & \sum_i \mathbb{E} \left[ \{ \hat{\mathcal{C}}_{WZ}(A, \vec{u}_i) + \hat{\mathcal{C}}_{YW}(\vec{M}, A, \vec{u}_i) + \hat{\mathcal{C}}_{YZ}(\vec{M}, A, \vec{u}_i) - 2\hat{\mathcal{C}}_{YWZ}(\vec{M}, A, \vec{u}_i) \} \{ \hat{\varphi}_{\text{full},2}(\vec{M}, A, \vec{u}_i) + \hat{\varphi}_{\text{full},3}(A, \vec{u}_i) \} \right] - \psi_a \\ &= \sum_i \mathbb{E} \left[ \{ \hat{\varphi}_{\text{full},2}(\vec{M}, A, \vec{u}_i) + \hat{\varphi}_{\text{full},3}(A, \vec{u}_i) \} \{ \mathbb{E}[\hat{\mathcal{C}}_{WZ}(A, \vec{u}_i) \mid \vec{M}, A, \vec{U}] + \mathbb{E}[\hat{\mathcal{C}}_{YW}(\vec{M}, A, \vec{u}_i) \mid \vec{M}, A, \vec{U}] \right. \\ & \quad \left. + \mathbb{E}[\hat{\mathcal{C}}_{YZ}(\vec{M}, A, \vec{u}_i) \mid \vec{M}, A, \vec{U}] - 2\mathbb{E}[\hat{\mathcal{C}}_{YWZ}(\vec{M}, A, \vec{u}_i) \mid \vec{M}, A, \vec{U}] \} \right] - \psi_a \\ &= \sum_i \mathbb{E} \left[ \{ \hat{\varphi}_{\text{full},2}(\vec{M}, A, \vec{u}_i) + \hat{\varphi}_{\text{full},3}(A, \vec{u}_i) \} \mathbf{1}\{\vec{U} = \vec{u}_i\} \right] - \psi_a + o_p(n^{-1/2}) \\ &= \mathbb{E} \left[ \hat{\varphi}_{\text{full},2}(\vec{M}, A, \vec{U}) + \hat{\varphi}_{\text{full},3}(A, \vec{U}) \right] - \psi_a + o_p(n^{-1/2}). \end{aligned}$$

Theorem 5.1 from Guo et al. [2026] showed that under (1) and (2),  $\mathbb{E} \left[ \hat{\varphi}_{\text{full},2}(\vec{M}, A, \vec{U}) + \hat{\varphi}_{\text{full},3}(A, \vec{U}) \right] - \psi_a = o_p(n^{-1/2})$ .

Hence, the second term is also  $\sqrt{n}$ -consistent.

Altogether,  $R_2 = o_p(n^{-1/2})$  as desired.  $\square$

## M EXPERIMENTS: FINITE-SUPPORT DGP DATA GENERATION

The setup is based on the causal model reflected in Fig.3b:

$$U \sim \text{Ber}(0.1) \tag{59}$$

$$H \sim \text{Ber}(0.8) \tag{60}$$

$$A \mid UH \sim \text{Ber}(0.4 + 0.2 * U + 0.1 * H) \tag{61}$$

$$M \mid AU \sim \text{Ber}(0.8 - 0.5 * \text{xor}(U, A) + 0.1 * A) \tag{62}$$

$$Y \mid MUH \sim \text{Ber}(\max(\min( \tag{63}$$

$$0.5 - 0.4 * U + 0.6 * M - 0.1 * H, 0.9), 0.1)) \tag{64}$$

$$Z \mid U \sim \text{Ber}(0.3 + 0.5 * U) \tag{65}$$

$$W \mid U \sim \text{Ber}(0.9 - 0.7 * U) \tag{66}$$

## N EXPERIMENTS: MIXED DGP DATA GENERATION

The setup is based on the causal model reflected in Fig. 3b:

$$U_1 \sim Ber(0.4) \quad (67)$$

$$U_2 \sim N(0, 1) \quad (68)$$

$$A \mid U_1, U_2 \sim Ber((-0.5 + 0.8U_1 + 0.6U_2)) \quad (69)$$

$$M \mid A, U_1 \sim Ber((-0.3 + 1.3U_1 + 0.9A)) \quad (70)$$

$$Y \mid M, U_1, U_2 \sim N(1 + 2U_1 + 0.8U_2 + 0.9M, 1) \quad (71)$$

$$W \mid U_1 \sim N(0.4 + U_1, 1) \quad (72)$$

$$Z \mid U_1 \sim N(-0.2 + 1.1U_1, 1). \quad (73)$$

All exogenous noise terms are mutually independent.

## O THE BENCHMARK ESTIMATORS

We used the following benchmark estimators:

An oracle plug-in estimator  $\hat{\psi}^{(\text{Oracle}, \text{PI})}$  (Equation 74) that has access to the full data, had  $U$  not been hidden:

$$\hat{\psi}^{(\text{Oracle}, \text{PI})} = \mathbb{E}_n \left[ \sum_{a'} \hat{E}[Y \mid M, a', U] \hat{p}(a' \mid U) [\hat{p}(M \mid a = 1, U) - \hat{p}(M \mid a = 0, U)] \right] \quad (74)$$

A front-door plug-in estimator  $\hat{\psi}^{(\text{FD}, \text{PI})}$  (Equation 75) that falsely assumes  $\vec{M}$  satisfies the front-door criterion.

$$\hat{\psi}^{(\text{FD}, \text{PI})} = \mathbb{E}_n \left[ \sum_{a'} \hat{E}[Y \mid M, a'] \hat{p}(a') [\hat{p}(M \mid a = 1) - \hat{p}(M \mid a = 0)] \right] \quad (75)$$