Robust Weighted Triangulation of Causal Effects Under Model Uncertainty

Rohit Bhattacharya[1,*] **Ina Ocelli**[1,*] **Ted Westling**[2]

rb17@williams.edu io1@williams.edu twestling@umass.edu

[1,*] Equal contributors. Dept. of Computer Science, Williams College
[2] Dept. of Mathematics and Statistics, University of Massachusetts Amherst

Abstract

A fundamental challenge in causal inference with observational data is correct specification of a causal model. When there is model uncertainty, analysts may seek to use estimates from multiple candidate models that rely on distinct, and possibly partially overlapping, sets of identifying assumptions to infer the causal effect, a process known as *triangulation*. Principled methods for triangulation, however, remain underdeveloped. Here, we develop a framework for causal effect triangulation that combines model testability methods from causal discovery with statistical inference methods from semiparametric theory, while avoiding explicit model selection and post-selection inference problems. We propose a triangulation functional that combines identified functionals from each model with data-driven measures of model validity. We provide a bound on the distance of the functional from the true causal effect along with conditions under which this distance can be taken to zero. Finally, we derive valid statistical inference for this functional. Our framework formalizes robustness under causal pluralism without requiring agreement across models or commitment to a single specification. We demonstrate its performance through simulations and an empirical application.

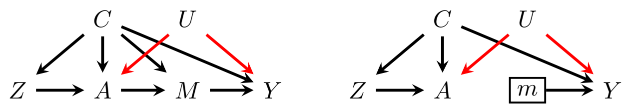
1 INTRODUCTION

Causal inference with observational data is rarely conducted under a single, indisputable set of assumptions. For a single observed dataset, there may be multiple plausible causal models that identify the target causal parameter under different sets of assumptions, some of which may be untestable. A common response to model uncertainty is to combine evidence from each of these models with the hope of reaching more robust conclusions than reliance on just a single one, a process commonly referred to as *triangulation* [Thurmond, 2001, Farmer et al., 2006, Lawlor et al., 2016].

Despite its intuitive appeal, triangulated effect estimation remains underdeveloped. A prominent line of work formalizes triangulation through *evidence factors*, which combine p-values across multiple analyses to test causal hypotheses [Rosenbaum, 2010, 2011, Karmakar et al., 2019]. While powerful, these approaches are fundamentally geared toward hypothesis testing rather than estimation, and rely on assumptions that are often difficult to justify in observational settings. In particular, they require that different analyses do not share sources of bias and that the joint distribution of p-values satisfies certain stochastic dominance properties under the null. Yang et al. [2024] leveraged the joint convergence properties of semiparametric estimators to design a robust test of the causal null hypothesis that remains valid as long as at least one model is correct. However, Yang et al. [2024] also focused on hypothesis testing rather than effect estimation.

In this paper, we develop a general framework for triangulating causal effect estimates across multiple candidate models without requiring explicit model selection or correctness of a plurality of models, as required by voting-based triangulation procedures. Our approach instead assigns data-driven weights to each model based on testable implications of its identifying assumptions, such as conditional independence or generalized equality constraints. These weights are used to form a smooth aggregation of model-specific effect estimates, yielding an estimator that is consistent for a *weighted triangulation functional*. We show that the absolute difference between this functional and the true causal parameter can be bounded as a function of (i) the maximal bias among incorrect models and (ii) our ability to separate correct and incorrect models using observed data. We use this bound to provide conditions under which the bias of the triangulation functional is small, yielding a form of robustness to misspecification of some causal models through triangulation. In particular, we show the bias can be controlled if at least one

Proceedings of the 42nd Conference on Uncertainty in Artificial Intelligence (UAI 2026), PMLR 337:567–584.



(a) $\mathcal{G}(V \cup U)$ (b) $\mathcal{G}(V \cup U)_{\mathrm{do}(m)}$

Figure 1: (a) A hidden variable causal DAG. (b) Graph representing intervention on M from which we can read the Verma constraint $Z \perp\!\!\!\perp Y \mid C$ in $P(V)/P(M \mid A, Z, C)$.

candidate model is correct and testable from observed data. We demonstrate the effectiveness of the proposed method through simulations and an empirical application.

Contributions. Our contributions can be viewed from two complementary perspectives. First, we advance the triangulation literature by providing a principled and quantitative method for combining causal effect estimates that achieves robustness to model misspecification by leveraging testable implications in the observed data. Second, we contribute to the literature on post-selection inference, which studies valid inference after data-driven model selection. Rather than selecting a single model, our approach avoids explicit selection by using smooth weights, thereby mitigating post-selection bias while still incorporating information from model diagnostics. Unlike recent work in causal discovery [Gradu et al., 2025, Chang et al., 2026], which focuses on learning the full causal graph, our framework targets a more modest but practically important goal: testing a minimal set of assumptions required to identify a causal parameter and combining estimators of the identified functionals.

Related work. The methods in Rakshit et al. [2025], Kang et al. [2016], Sun et al. [2021], Yao et al. [2024] allow for some models to be misspecified, but require that a plurality of models be correctly specified. This can be a strong assumption in observational settings, where several models may fail for similar reasons. Moreover, these methods are typically tailored to specific classes of causal models, such as proxy-variable approaches or instrumental variables. Bayesian model averaging approaches [Horii, 2021, Steiner and Steel, 2025] provide another avenue for combining estimates, but often rely on parametric assumptions such as linearity or are similarly restricted to specific model classes.

2 CAUSAL GRAPH PRELIMINARIES

Although the candidate causal models need not be graphical, we frame our discussion with causal directed acyclic graphs (causal DAGs), as they provide a transparent way to represent identifying assumptions and their testable implications.

The causal model of a DAG $\mathcal{G}(V)$ can be understood as distributions generated by a system of structural equations equipped with the $\mathrm{do}(\cdot)$ operator [Pearl, 2009a]. Specifically, for each variable $V_i \in V$, there is an equation of the form $V_i \leftarrow f_i\big(\mathrm{pa}_{\mathcal{G}}(V_i), \epsilon_i\big)$, where $\mathrm{pa}_{\mathcal{G}}(V_i)$ denotes a set of values for the parents of V_i in $\mathcal{G}$ and ϵ_i is an exogenous (latent) error term. A joint distribution $P(V)$ induced by such a system is said to be Markov with respect to $\mathcal{G}$, meaning that it factorizes as $P(V) = \prod_{V_i \in V} P\big(V_i \mid \mathrm{Pa}_{\mathcal{G}}(V_i)\big)$. Equivalently, P satisfies the global Markov property with respect to $\mathcal{G}$ stated in terms of the well-known d-separation criterion: $X \perp\!\!\!\perp_{\text{d-sep}} Y \mid Z \implies X \perp\!\!\!\perp Y \mid Z$ in P [Pearl, 2009a]. The distribution P is considered *faithful* to $\mathcal{G}$ if the converse is also true, i.e., $X \perp\!\!\!\perp_{\text{d-sep}} Y \mid Z \iff X \perp\!\!\!\perp Y \mid Z$ in P.

In fully observed causal DAG models, counterfactual distributions arising from interventions on a set $A \subset V$, written as $P(V \setminus A \mid \mathrm{do}(a))$, are identified via a truncated factorization known as the g-formula [Robins, 1986, Pearl, 2009a]:

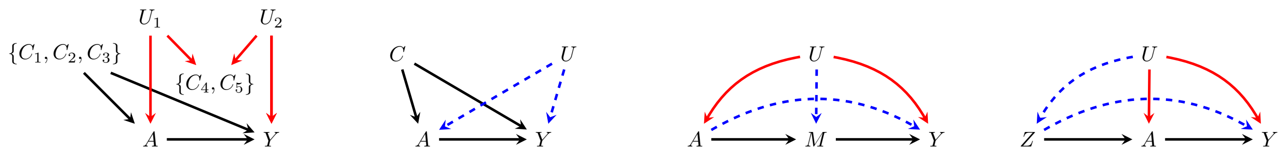
$$P(V \setminus A \mid \mathrm{do}(a)) = \prod_{V_i \in V \setminus A} P\big(V_i \mid \mathrm{Pa}_{\mathcal{G}}(V_i)\big)\Big|_{A=a}. \quad (1)$$

When some variables U are unobserved, identification theory becomes more complex (as in Scenario 2 of Section 3). In addition, the observed distribution $P(V)$ consists of not only conditional independences from d-separation, but also generalized equality constraints, known as *Verma constraints* [Robins, 1986, Verma and Pearl, 1990]. Verma constraints are obtained via d-separation in conditional DAGs corresponding to identifiable post-intervention distributions. As an example, consider the DAG $\mathcal{G}(V \cup U)$ in Figure 1(a). By d-separation, $\mathcal{G}$ implies $Z \perp\!\!\!\perp M \mid A, C$. However, there is no set $X \subset V \setminus \{Z, Y\}$ such that $Z \perp\!\!\!\perp_{\text{d-sep}} Y \mid X$. That is, there is no ordinary conditional independence between Z and Y in the observed joint $P(V)$. However, it is well known that there exists a Verma constraint $Z \perp\!\!\!\perp Y \mid C$ in the Markov kernel $P(V)/P(M \mid A, Z, C)$ [Verma and Pearl, 1990]. Under a causal interpretation of $\mathcal{G}(V \cup U)$, for any fixed value of m, $P(V)/P(M = m \mid A, Z, C)|_{M=m} = P(Z, A, C, Y \mid \mathrm{do}(m))$ by the g-formula in (1). That is, the kernel $P(V)/P(M \mid A, Z, C)$ factorizes according to the conditional DAG in Figure 1(b), where M is now a fixed node with all incoming edges removed, and from which the Verma constraint can be read using d-separation—notice that Z and Y are d-separated given C in $\mathcal{G}(V \cup U)_{\mathrm{do}(m)}$.

When the measures of model validity used rely on Verma constraints, we will require $P(V)$ to be *Verma constraint faithful* to $\mathcal{G}(V \cup U)$. That is, $X \perp\!\!\!\perp Y \mid Z$ in $P(V)/P(A|B)$ implies that edges in $\mathcal{G}$ are such that $P(V \setminus A \mid \mathrm{do}(a)) = P(V)/P(A|B)|_{A=a}$ and $X \perp\!\!\!\perp_{\text{d-sep}} Y \mid Z$ in $\mathcal{G}_{\mathrm{do}(a)}$.

3 MOTIVATING SCENARIOS

Suppose an analyst is interested in estimating the average causal effect $\theta \equiv \mathbb{E}[Y \mid \mathrm{do}(A = 1)] - \mathbb{E}[Y \mid \mathrm{do}(A = 0)]$. In this section, we present two classic scenarios where proper triangulation of effect estimates from multiple plausible causal models would lead to more robust causal inference.



(a) M-bias example (b) Backdoor model (c) Frontdoor model (d) IV model

Figure 2: Causal DAGs used for motivating robust triangulation. (a) A causal DAG where uncertainty about what variables to adjust for may yield M-bias. (b, c, d) Under model uncertainty, analysts may wish to triangulate effects from each of these models that rely on qualitatively distinct assumptions—violations of these assumptions are shown via blue dashed edges.

Scenario 1. Adjust for all pre-treatment covariates or not? When a set of pre-treatment covariates L blocks all backdoor paths between A and Y we obtain a well-known instance of the g-formula, known as the backdoor formula:

$$\theta = \sum_l P(l) \times \big(\mathbb{E}[Y \mid A=1, l] - \mathbb{E}[Y \mid A=0, l]\big). \quad (2)$$

A longstanding debate in causal inference is whether analysts should include *all* observed pre-treatment covariates in the backdoor formula (2). Some researchers [Rosenbaum, 2002, Ding and Miratrix, 2015, Rubin, 2009] argue that conditional ignorability (blockage of all backdoor paths) is likely to hold only when adjusting for a large set of variables, and thus advocate adjusting for all pre-treatment covariates C associated with the treatment and outcome. Other researchers emphasize that adjustment for all such variables can induce collider bias, such as M-bias [Pearl, 2009b, Shrier, 2009, Sjölander, 2009]. Thus, when the true DAG is unknown—as is typical in observational studies—the analyst may face significant uncertainty about whether excluding certain covariates risks confounding bias or including them risks collider bias. This setting naturally motivates triangulation across multiple adjustment strategies.

As a concrete example, let Figure 2(a) be the true (but unknown) DAG. Here, adjusting for all observed pre-treatment covariates C yields invalid effect estimates due to the colliders at C_4 and C_5. Due to model uncertainty, the analyst might partition the observed covariates into two groups: those that certainly do not give M-bias and are essential for adjustment, and those about which there is some uncertainty. Say they correctly deem $\{C_1, C_2, C_3\}$ to be essential, but are uncertain about $\{C_4, C_5\}$. They may then wish to triangulate estimates from multiple overlapping adjustment sets, e.g., $\{C_1, C_2, C_3\}$ that excludes all uncertain covariates, $\{C_1, C_2, C_3, C_4\}$ that includes some but not others, and $\{C_1, C_2, C_3, C_4, C_5\}$ that includes all pre-treatment covariates. In this example, only the first set leads to valid effect estimates, underscoring the importance of a triangulation procedure that is robust to incorrect identifying assumptions in multiple models as long as at least one is correct.

Scenario 2. Use backdoor, frontdoor, or an instrument? In many observational settings, there is uncertainty about whether unmeasured confounding between A and Y is present. A natural response to this is to triangulate estimates obtained from a backdoor adjustment model (Figure 2(b)) with those from alternative identification strategies that explicitly allow for unmeasured A–Y confounding but impose different structural assumptions. One such alternative is the frontdoor model (Figure 2(c)). The frontdoor model permits unmeasured confounding between A and Y, but assumes (i) that A affects Y only through a mediator set M (i.e., no direct $A \to Y$ edge), and (ii) that the unmeasured variable U does not confound the A–M or M–Y relationships (i.e., no $U \to M$ edge). Under these assumptions, the causal effect is identified by the frontdoor formula [Pearl, 1995]. A conditional version that adjusts for baseline covariates L is

$$\theta = \sum_{m,l} \Big\{ \Big(\sum_{a'} P(a', l) \cdot \mathbb{E}[Y \mid a', m, l] \Big) \cdot \big(P(m \mid A=1, l) - P(m \mid A=0, l) \big) \Big\}. \quad (3)$$

Alternatively, one may use instrumental variable (IV) models (Figure 2(d)). An instrument Z typically satisfies three structural assumptions: (i) conditional independence from the unmeasured confounder U (no $U \to Z$ edge), (ii) an exclusion restriction (no $Z \to Y$ edge), and (iii) instrument relevance ($Z \to A$ exists). Under additional non-graphical assumptions, such as effect homogeneity, the causal effect is point-identified as [Angrist et al., 1996]:

$$\theta = \frac{\sum_l P(l) \cdot \big(\mathbb{E}[Y \mid Z=1, l] - \mathbb{E}[Y \mid Z=0, l]\big)}{\sum_l P(l) \cdot \big(\mathbb{E}[A \mid Z=1, l] - \mathbb{E}[A \mid Z=0, l]\big)}. \quad (4)$$

Unlike Scenario 1, this setting involves three qualitatively distinct causal models. Each has their own unique drawbacks, but may imply testable restrictions in $P(V)$ under some assumptions of faithfulness and causal ordering of variables [Entner et al., 2013, Bhattacharya and Nabi, 2022]. A natural approach then would be to test such implications, retain non-rejected models, and aggregate their estimates. However, this two-step strategy introduces a post-selection inference problem, particularly when models share data, have overlapping adjustment sets, or shared sources of bias. Further, standard remedies for this, such as sample splitting [Hansen, 2000, Newey and Robins, 2018], can substantially reduce effective sample size when multiple models are considered. These scenarios and associated challenges motivate our new triangulation functional and inference procedure.

4 NEW TRIANGULATION METHOD

We now present the general setup that we consider. Let θ denote the causal parameter of interest, such as the average or conditional average causal effect. We assume the observed data consists of n IID realizations $O_1, \ldots, O_n$ drawn from some unknown distribution P. For a finite integer $K > 1$, let $\mathcal{M}_1 \ldots, \mathcal{M}_K$ denote the different candidate causal models, and let ψ_k (we suppress the dependence on P for these parameters and ones that follow for notational brevity) denote the identifying functional of $\mathcal{M}_k$. That is, if the assumptions of $\mathcal{M}_k$ are true, then $\theta = \psi_k$. Let $\beta_1, \ldots, \beta_K$ denote observed data parameters and $\mathcal{A}$ be a set of assumptions such that if $\mathcal{A}$ is true and $\beta_k = 0$, then the identifying assumptions of $\mathcal{M}_k$ are true. We do not require, however, that if $\mathcal{A}$ holds and $\beta_k \neq 0$, then the assumptions of $\mathcal{M}_k$ must be incorrect; we allow this scenario to also correspond to the model $\mathcal{M}_k$ being untestable using the observed data.

The role of each β_k is to encode a testable implication of the identifying assumptions of model $\mathcal{M}_k$. Examples of such testable implications from the causal discovery literature include conditional independence constraints for backdoor models [Entner et al., 2013], Verma constraints for front-door models [Bhattacharya and Nabi, 2022], and tetrad constraints for proxy-based models [Xie et al., 2024]. That is, each β_k is an observed data parameter constructed so that it is zero if and only if a particular equality constraint, such as the aforementioned ones, holds in the observed distribution P. More specifically, β_k may be a regression coefficient in a parametric model, a log-odds ratio in a semiparametric model [Chen, 2007], or a generalized covariance measure [Shah and Peters, 2020, He et al., 2025, Bergen et al., 2026]. Under assumptions $\mathcal{A}$ (typically including a faithfulness condition and partial knowledge of temporal ordering of variables), $\beta_k = 0$ implies the absence of specific edges in $\mathcal{G}$, and absence of these edges is sufficient for the identifying assumptions of $\mathcal{M}_k$ to hold. Thus, whether $\beta_k = 0$ provides empirical evidence about the validity of $\mathcal{M}_k$.

4.1 TRIANGULATION FUNCTIONAL

The above setup motivates the following naive triangulation functional: $\psi_{\text{naive}} = \frac{\sum_{k=1}^K \mathbb{I}(\beta_k=0)\cdot\psi_k}{\sum_{j=1}^K \mathbb{I}(\beta_j=0)}$, where $\mathbb{I}(\cdot)$ is the indicator function. That is, ψ_{naive} is an average over all ψ_k derived from models $\mathcal{M}_k$ for which $\beta_k = 0$. If assumptions $\mathcal{A}$ used to test model correctness hold, ψ_{naive} is *causally robust*: $\psi_{\text{naive}} = \theta$ if at least one model is correct and testable, as it reduces to an average of only those ψ_k such that $\psi_k = \theta$.

We consider this to be naive as filtering based on $\mathbb{I}(\beta_k = 0)$ is unlikely to succeed for any model $\mathcal{M}_k$ when β_k is unknown and must be estimated from finite samples. Instead, we propose a smooth approximation of the naive functional by replacing $\mathbb{I}(\beta_k = 0)$ with Gaussian kernels of the form $\delta_a(\beta_k) = \frac{1}{|a|\sqrt{\pi}} e^{-(\beta_k/a)^2}$, where $a > 0$ is a constant that controls the sharpness of the approximation. As $a \to 0$, the function becomes concentrated around $\beta_k = 0$. Our proposed triangulation functional is then given by

$$\psi = \sum_{k=1}^K w_k \psi_k, \quad \text{where} \quad w_k = \frac{\delta_a(\beta_k)}{\sum_{j=1}^K \delta_a(\beta_j)}. \tag{5}$$

This functional exhibits an *approximate* causal robustness property due to smoothing via the kernels, as stated below.

Theorem 1. *Suppose assumptions $\mathcal{A}$ used to justify the tests of model correctness hold, and let $\mathcal{C} \subseteq \{1, \ldots, K\}$ and $\mathcal{I} = \{1, \ldots, K\} \setminus \mathcal{C}$ be the subsets of indices k such that model $\mathcal{M}_k$ is correct and incorrect, respectively. Then*

$$|\psi - \theta| \leq \frac{\max_k |\psi_k - \theta|}{1 + D_a} \tag{6}$$

where $D_a = \left(\sum_{k\in\mathcal{C}} \delta_a(\beta_k)\right) / \left(\sum_{k\in\mathcal{I}} \delta_a(\beta_k)\right)$. Further, if at least one model $\mathcal{M}_k$ is correct and testable using the observed data, then $D_a \geq e^{\varepsilon^2/a^2}/|\mathcal{I}|$, where $\varepsilon = \min_{k\in\mathcal{I}} |\beta_k|$.

The proof is in Appendix A. Theorem 1 demonstrates that the absolute difference between the triangulation functional ψ and the true causal parameter of interest θ is bounded by the maximal bias of the functionals ψ_k from the incorrect causal models divided by a "discrimination factor" $1 + D_a$. The discrimination factor depends both on the weighting function chosen and the true values of the testing functionals $\beta_1, \ldots, \beta_K$. Roughly speaking, when the β_k's from correct causal models are small in magnitude compared to those from incorrect causal models on the scale determined by δ_a, then the discrimination factor is large, and the absolute difference between ψ and θ is reduced. We note that the second statement of Theorem 1 implies that if at least one model $\mathcal{M}_k$ is both correct and testable using the observed data and all testing functionals β_k in incorrect models are non-zero, then $\lim_{a\to 0} D_a = \infty$, so that $\lim_{a\to 0} \psi = \theta$.

To make the robustness property more concrete, consider setting the Gaussian kernel parameter to $a = 0.1$ and suppose there is one correct model $\mathcal{C} = \{1\}$ and two incorrect models $\mathcal{I} = \{2, 3\}$. Let $\varepsilon = \min_{k\in\mathcal{I}} |\beta_k| = 0.2$, indicating that violations of the incorrect models are only weakly detectable since the value of β_k is close to 0 even when $\mathcal{M}_k$ is incorrect. From Theorem 1, this yields $1 + D_a \geq 1 + e^{(0.2/0.1)^2}/2 \approx 30$, implying that the bias of the triangulation functional is at most the maximal bias among incorrect models divided by 30.

While smaller values of a reduce the absolute difference between the triangulated functional ψ and the target causal effect θ, in practice we cannot set a too small because we only have estimates of ψ_k and β_k. Thus, we must balance the variability of these estimators against the bias induced by $a > 0$. We return to this discussion and provide concrete recommendations for setting a in Section 4.3.

We also note that if *no* candidate model is both correct and testable, then the triangulation functional does not converge to θ as $a \to 0$. Fortunately, this behavior is straightforward to diagnose in practice: the estimated values of ψ often diverge due to $\sum_{j=1}^{K} \delta_a(\beta_j) \approx 0$. The normalizing term, however, also poses challenges in finite samples even when a correct and testable model $\mathcal{M}_k$ exists in theory, e.g., due to sampling variability. We briefly address this issue of numerical stability in the following subsection on inference.

4.2 INFERENCE PROCEDURE

We now discuss our approach to estimation and inference for the triangulation functional ψ. Suppose that for each k, we have estimators $\psi_{k,n}$ and $\beta_{k,n}$ of each identifying functional ψ_k and each measure of model correctness β_k, respectively. Our triangulation estimator is then

$$\psi_n = \sum_{k=1}^{K} w_{k,n}\psi_{k,n}, \text{ for } w_{k,n} = \frac{\delta_a(\beta_{k,n})}{\lambda_n + \sum_j \delta_a(\beta_{j,n})}, \quad (7)$$

where $\lambda_n > 0$ is an additional term used to ensure numerical stability of the weights $w_{k,n}$ even when the normalizing function in the denominator is close to zero. To avoid affecting first-order asymptotics of the estimator, we require $\lambda_n = o(n^{-1/2})$. A simple choice is to fix $\lambda_n = 1/n$. Note that λ_n in the triangulation functional is optional. In practice, we find in our simulations and data application that we do not need it and that using the well-known log-sum-exp trick [Blanchard et al., 2021] to compute the normalizing function is sufficient to prevent numerical instability. However, we keep it in our exposition for the sake of completeness.

Since $\psi_{1,n}, \ldots, \psi_{K,n}, \beta_{1,n}, \ldots, \beta_{K,n}$ are constructed using a common observed dataset, they are typically dependent, even asymptotically. We suppose that under a set of statistical regularity conditions $\mathcal{S}$ (e.g., rates of convergence and complexity constraints on the nuisance estimators), we can construct asymptotically linear estimators of both β_k and ψ_k with influence functions ϕ_{β_k} and ϕ_{ψ_k}, respectively, for each k. That is, under $\mathcal{S}$ we have $\beta_{k,n} - \beta_k = \frac{1}{n}\sum_{i=1}^{n} \phi_{\beta_k}(O_i) + o_p(n^{-1/2})$ and $\psi_{k,n} - \psi_k = \frac{1}{n}\sum_{i=1}^{n} \phi_{\psi_k}(O_i) + o_p(n^{-1/2})$. Here ϕ_{β_k} and ϕ_{ψ_k} are assumed to satisfy $\mathbb{E}[\phi_{\beta_k}] = \mathbb{E}[\phi_{\psi_k}] = 0$, $\mathbb{E}[\phi^2_{\beta_k}] < \infty$, and $\mathbb{E}[\phi^2_{\psi_k}] < \infty$. These estimators could be parametric or semiparametric in nature; we will discuss specific approaches to estimation more below.

Asymptotic linearity of any finite collection of estimators implies joint convergence to a multivariate normal distribution [Van der Vaart, 2000]. Let $\kappa = [\beta_1, \ldots, \beta_K, \psi_1, \ldots, \psi_K]$ and κ_n denote the corresponding vector of estimates. Then, $n^{1/2}(\kappa_n - \kappa) \xrightarrow{d} N(0, \Sigma)$, where Σ is a $2K \times 2K$ covariance matrix of the influence functions of each ψ_k and β_k. Each entry of Σ is of the form $\mathbb{E}[\phi_{\beta_j}, \phi_{\beta_k}], \mathbb{E}[\phi_{\psi_j}, \phi_{\psi_k}]$, or $\mathbb{E}[\phi_{\beta_j}, \phi_{\psi_k}]$. Applying the delta method then gives the asymptotic distribution of the triangulation estimator as,

$$n^{1/2}(\psi_n - \psi) \xrightarrow{d} N(0, \gamma^T \Sigma \gamma), \quad (8)$$

where γ is a vector of partial derivatives defined as,

$$\gamma = \left[\frac{\partial \psi}{\partial \beta_1}, \ldots, \frac{\partial \psi}{\partial \beta_K}, \frac{\partial \psi}{\partial \psi_1}, \ldots, \frac{\partial \psi}{\partial \psi_K} \right]. \quad (9)$$

The following lemma also gives us closed form expressions for computing the vector γ. The proof is in Appendix B.

Lemma 1. *The partial derivatives of ψ are*

$$\frac{\partial \psi}{\partial \psi_k} = w_k \quad \textit{and} \quad \frac{\partial \psi}{\partial \beta_k} = \frac{2\beta_k w_k}{a^2}(\psi - \psi_k).$$

We now suggest three approaches to constructing asymptotically linear estimators and associated approaches to valid inference. Our preferred strategy is influence function-based estimation. In this approach, the influence functions ϕ_{β_k} and ϕ_{ψ_k} are derived explicitly and used to construct $\beta_{k,n}$ and $\psi_{k,n}$ using, e.g., the one-step construction [Bickel, 1982] or estimating equations [Chernozhukov et al., 2018]. A benefit to this approach is that machine learning estimators can be used for nuisance functions, and asymptotic linearity follows under sufficient rates of convergence of these nuisance estimators.

When the influence functions are available in closed form—several of which have been derived in the context of causal graphs Jung et al. [2021], Bhattacharya et al. [2022], Guo and Nabi [2024]—we can construct consistent estimators $\phi_{\beta_k,n}$ and $\phi_{\psi_k,n}$ of the influence functions. A consistent estimator Σ_n of Σ is then simply the sample covariance matrix of the influence function estimators. An estimator γ_n of γ can be obtained by plugging in $\psi_{k,n}$ and $\beta_{k,n}$ into the form of γ provided in Lemma 1. A variance estimator of ψ_n is then given by $\widehat{\text{var}}(\psi_n) = \gamma_n^T \Sigma_n \gamma_n / n$, which can be used to construct Wald-type confidence intervals for ψ.

For some functionals ψ_k or β_k, explicit influence functions might not be available in the current literature. In this case, one option is to use plug-in estimators $\psi_{k,n}$ and $\beta_{k,n}$. However, plug-in estimators are often only asymptotically linear when correctly-specified parametric models are employed for nuisance estimators. In this case, the empirical bootstrap yields asymptotically valid inference under mild smoothness conditions on the parametric models [Efron and Tibshirani, 1994]. Specifically, on each bootstrap sample, the parametric nuisance estimators are re-estimated and plugged in to obtain bootstrap estimates $\psi^*_{k,n}$ and $\beta^*_{k,n}$. These estimates are then substituted as in (7) to obtain a bootstrap estimate

ψ_n^*. The distribution of bootstrap estimates can then be used to construct a confidence interval in the standard ways.

Finally, if plug-in estimators with nuisance estimators based on machine learning are used, the resulting estimators may not be asymptotically linear due to excess bias, and hence the empirical bootstrap is not guaranteed to work. In this case, *subsampling* may still be asymptotically valid [Politis et al., 2001]. Subsampling involves drawing b random subsamples of size $m < n$ from the full dataset, computing estimates $\psi_m^{(1)}, \dots, \psi_m^{(b)}$ for each subsample, and constructing a $(1-\alpha)$ confidence interval based on the $\alpha/2$ and $1-\alpha/2$ quantiles of these estimates. A sufficient condition for this procedure to yield valid inference is that $\tau_n(\psi_n - \psi)$ converges to a non-degenerate distribution, which can be true even when the estimators are not asymptotically linear.

We summarize our recommendations below. The order in the if-elif-else logic reflects our preferred inference strategies.

Inference Procedure for Triangulation

1. Construct estimators $\psi_{k,n}$ and $\beta_{k,n}$ for each model $\mathcal{M}_k$ and obtain the point estimate ψ_n using (7).
2. If every $\psi_{k,n}$ and $\beta_{k,n}$ is influence-function-based, let $\widehat{\mathrm{SE}}(\psi_n) = \sqrt{\gamma_n^\top \Sigma_n \gamma_n / n}$ and construct a $(1-\alpha)$ CI as $\psi_n \pm z_{1-\alpha/2}\,\widehat{\mathrm{SE}}(\psi_n)$.
3. Else, if $\psi_{k,n}$ and $\beta_{k,n}$ are plug-in estimators based on parametric nuisance models: Construct a $(1-\alpha)$ CI as $[2\psi_n - q_{1-\alpha/2},\, 2\psi_n - q_{\alpha/2}]$, where q_p is the p^{th} empirical quantile of estimates $\{\psi_n^{(1)}, \dots, \psi_n^{(b)}\}$ obtained via empirical bootstrap and re-fitting of nuisance models.
4. Else, construct a $(1-\alpha)$ CI as $[2\psi_n - q_{1-\alpha/2},\, 2\psi_n - q_{\alpha/2}]$, based on the empirical quantiles of estimates $\{\psi_m^{(1)}, \dots, \psi_m^{(b)}\}$ obtained via subsampling with $m = n^{4/5}$ and re-fitting of nuisance models.

4.3 SETTING THE KERNEL BANDWIDTH

Finally, we address setting the kernel bandwidth parameter a. Our recommendations in this section apply when estimators that converge at the rate $n^{-1/2}$, which includes influence function-based estimators and plug-in estimators based on parametric nuisance models (options 2 and 3 of the display in Section 4.2). Setting a when using estimators with slowers rates of convergence, such as many plug-in estimators based on flexible nuisance models, is an interesting problem left to future work.

We recommend choosing a such that (1) $n^{1/2}a \to \infty$ and (2) $a\sqrt{\log(n)} \to 0$. Examples of this include setting $a = n^{-1/3}$ and $a = 1/\log(n)$. In our numerical experiments and data application we choose $a = n^{-1/3}$, and also test this setting against other kernel bandwidths that lie outside of our recommended range. Our recommendation is based on the following theoretical analysis.

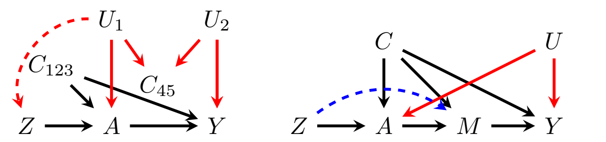


Figure 3: Causal DAG used in our simulations to demonstrate robustness to (a) M-bias in some adjustment sets, and (b) misspecification of backdoor, frontdoor, or IV models.

From Theorem 1 and Section 4.2, the bias of the triangulation estimator ψ_n for the true parameter θ decays at a rate proportional to $e^{-1/a^2} + o_P(n^{-1/2})$, while its standard deviation (SD) goes to zero at rate $n^{-1/2}$ when using influence function-based or parametric estimators. If $a\sqrt{\log(n)} \to 0$ (i.e., a goes to zero faster than $1/\sqrt{\log(n)}$), then $e^{-1/a^2} = o(1/\sqrt{n})$, so that the bias goes to zero faster than the SD. Hence, as long as at least one model is correct and testable and regularity conditions for nuisance estimation hold, then our methods produce asymptotically valid inference for the true causal effect if $a\sqrt{\log(n)} \to 0$.

Now let $\mathcal{C}'$ denote the set of indices of all models $\mathcal{M}_k$ that are both correct and testable. When the estimators $\beta_{k,n}$ are asymptotically linear, then $\sqrt{n}\beta_{k,n}$ is asymptotically normal if model k is correct and testable, and $c_n\beta_{k,n} \to_p \infty$ for any sequence $c_n \to \infty$ otherwise. This implies the following: (i) If a goes to zero faster than than $n^{-1/2}$, then asymptotically ψ_n is equal to $\psi_{k,n}$ for some k randomly selected from $\mathcal{C}'$; (ii) If a goes to zero at the rate $n^{-1/2}$, then asymptotically ψ_n is a random weighted combination of all $\psi_{k,n}$ for $k \in \mathcal{C}'$, with the joint distribution of weights depending on the asymptotic covariance of $\sqrt{n}\beta_{k,n}$ for $k \in \mathcal{C}'$; (iii) If a goes to zero slower than $n^{-1/2}$, then asymptotically ψ_n is an equally-weighted average of all $\psi_{k,n}$ for $k \in \mathcal{C}'$.

In our view, scenario (iii) is the most desirable behavior. Asymptotically, scenario (i) will result in focusing exclusively on the model whose testing functional estimator is closest to zero in any particular sample, which is overly narrow. The asymptotic covariance of the testing functionals is not necessarily related to the asymptotic variance of $\psi_{k,n}$, so there is not a good reason to weight estimators as in scenario (ii). Thus, (iii) arguably reflects the most reasonable behavior. It follows that we recommend choosing a such that $n^{1/2}a \to \infty$ (i.e., a goes to zero slower than $n^{-1/2}$).

5 APPLYING THE TRIANGULATION PROCEDURE IN PRACTICE

We now illustrate practical applications of our general method, along with numerical experiments and an empirical data application. The first subsection outlines a general framework for triangulating causal effect estimates derived from backdoor models with different candidate adjustment sets. The second subsection describes triangulation with backdoor, frontdoor model, and IV models. Explicit descriptions of the data generating processes and estimators are in Appendix D and E respectively, with proofs in Appendix C.

5.1 MULTIPLE CANDIDATE ADJUSTMENT SETS

We first establish a set of assumptions $\mathcal{A}$ under which we can use observed data parameters β_k to test the validity of each proposed adjustment set corresponding to different causal models $\mathcal{M}_k$. Let C denote all pre-treatment covariates under consideration for adjustment and Z denote an "anchor" variable, such that the following assumptions hold:

$$\begin{aligned}&\mathcal{A}_1 : P \text{ is faithful wrt a causal DAG } \mathcal{G}(V \cup U)\\&\mathcal{A}_2 : \mathcal{G} \text{ satisfies the causal ordering } \{Z, C\} < A < Y\\&\mathcal{A}_3 : Z \to A \to Y \text{ exists in } \mathcal{G}\end{aligned} \tag{10}$$

Note that $\mathcal{A}_1$ above is ordinary faithfulness as we do not rely on Verma constraints for this particular application, $\mathcal{A}_2$ simply imposes some weak background knowledge on the causal ordering of the variables, and $\mathcal{A}_3$ is a relevance assumption that ensures the testable implications are not trivial and actually rule out non-identifying edges in $\mathcal{G}$.[1]

Proposition 1. *Let $\mathcal{M}_k$ be a backdoor model that proposes $W \subseteq C$ as an adjustment set and β_k be any observed data parameter such that $\beta_k = 0 \iff Y \perp\!\!\!\perp Z \mid A, W$. Under assumptions $\mathcal{A}$ in* (10)*, $\beta_k = 0$ implies $\mathcal{M}_k$ is correct, i.e., $\psi_k = \theta$, where ψ_k is the backdoor formula* (2) *with $L = W$.*

A natural choice for β_k above is the log-odds ratio. For distinct sets of variables A, B, C and a choice of reference values a_0, b_0, the odds ratio function $\text{OR}(A, B \mid C)$ is a non-parametric measure of association given by [Chen, 2007],

$$\text{OR}(a, b \mid c) = \frac{P(a \mid b, c)}{P(a_0 \mid b, c)} \times \frac{P(a_0 \mid b_0, c)}{P(a \mid b_0, c)},$$

and $\log(\text{OR}(A, B \mid C)) = 0 \iff A \perp\!\!\!\perp B \mid C$. Though the log-odds ratio is a function, it is most commonly treated as a single scalar parameter in a semiparametric model; see for e.g., [Chen, 2007, Tchetgen Tchetgen et al., 2010, Malinsky et al., 2019]. We adopt this setup, computing log-odds ratios of the form $\log(\text{OR}(Y, Z \mid A, W))$ for different choices of adjustment sets W as given by $\mathcal{M}_1, \ldots, \mathcal{M}_k$, and use these as our β_k in the triangulation functional (5).

We return to the M-bias scenario in Section 3 as a concrete application. The candidate adjustment sets are $\mathcal{M}_1 : \{C_1, C_2, C_3\}$, $\mathcal{M}_2 : \{C_1, C_2, C_3, C_4\}$, and $\mathcal{M}_3 : \{C_1, C_2, C_3, C_4, C_5\}$, where only $\mathcal{M}_1$ is correct. Figure 3(a) shows a DAG with an anchor variable Z satisfying assumptions $\mathcal{A}$ in (10) with or without the red dashed edge. Note that choices for anchor variables in the causal discovery literature are often candidate IVs, such as Z in Figure 3(a). While one could include IV-based estimates in the triangulation functional, when a valid adjustment set exists, backdoor estimators are more efficient and avoid homogeneity assumptions.

By Proposition 1, $\beta_1 = \log(\text{OR}(Y, Z \mid A, C_{123})) = 0$, while β_2 and β_3 are non-zero. We evaluate two settings, one where $\varepsilon = \min_{k \in \mathcal{I}} = |\beta_k| = 0.71$, and another where $\varepsilon = 0.36$. We achieve the latter by excluding $U_1 \to Z$ in Figure 3(a). Per Theorem 1, the second setting is more challenging for our triangulation estimator, as the absolute bias $|\psi - \theta|$ is higher, since it is harder to detect the incorrect models. Here we use influence function-based estimators of the triangulation functional. As shown in Figure 4 (top row), the estimator remains causally robust in both cases, despite the plurality of models being incorrect. We require more samples, however, to obtain reliable estimates when $\varepsilon = 0.36$, as predicted by our theory. To test our variance computation proposal and Wald-type CI construction, we also compute the coverage of the estimator for both ψ and the target parameter θ. We report coverage for both, as these parameters are different in general, albeit with a small bounded difference. At $n = 5000$, the estimator achieves the nominal coverage of 95% for both ψ and θ in both settings.

5.2 BACKDOOR, FRONTDOOR, AND IV

Similar to the previous subsection, we first establish assumptions under which we can use observed data to test each $\mathcal{M}_k$, where $\mathcal{M}_k$ could be a backdoor, frontdoor, or IV model. As before, let C denote pre-treatment covariates used for adjustment and Z an anchor variable. As stated earlier, the anchor Z is often a candidate IV, and we will treat it as such in this subsection. Further, let M be a mediator set such that:

$$\begin{aligned}&\mathcal{A}_1 : P \text{ is Verma faithful wrt a causal DAG } \mathcal{G}(V \cup U)\\&\mathcal{A}_2 : \mathcal{G} \text{ satisfies the ordering } \{Z, C\} < A < M < Y\\&\mathcal{A}_3 : Z \to A \to M \to Y \text{ exist in } \mathcal{G}\\&\mathcal{A}_4 : \text{If } U_i \in U \text{ causes } M \text{ it also causes } Y\end{aligned} \tag{11}$$

Note that $\mathcal{A}_1$ includes ordinary faithfulness as a special case when the intervention set is empty, $\mathcal{A}_2$ imposes a causal ordering of the variables as before, and $\mathcal{A}_3, \mathcal{A}_4$ are relevance assumptions that ensure the testable implications are not trivial and actually rule out non-identifying edges in $\mathcal{G}$.

Proposition 2. *Let $\mathcal{M}_1, \mathcal{M}_2, \mathcal{M}_3$ be backdoor, frontdoor, and IV models respectively and $\beta_1, \beta_2, \beta_3$ be observed data parameters that are zero iff the independence $Y \perp\!\!\!\perp Z \mid A, C$, the Verma constraint $Y \perp\!\!\!\perp Z \mid C$ in $p(V)/p(M|A, Z, C)$, and the independence $M \perp\!\!\!\perp Z \mid$*

[1]Since $A \to Y$ exists, the sharp causal null is assumed false. For robust causal null hypothesis tests, see Yang et al. [2024].

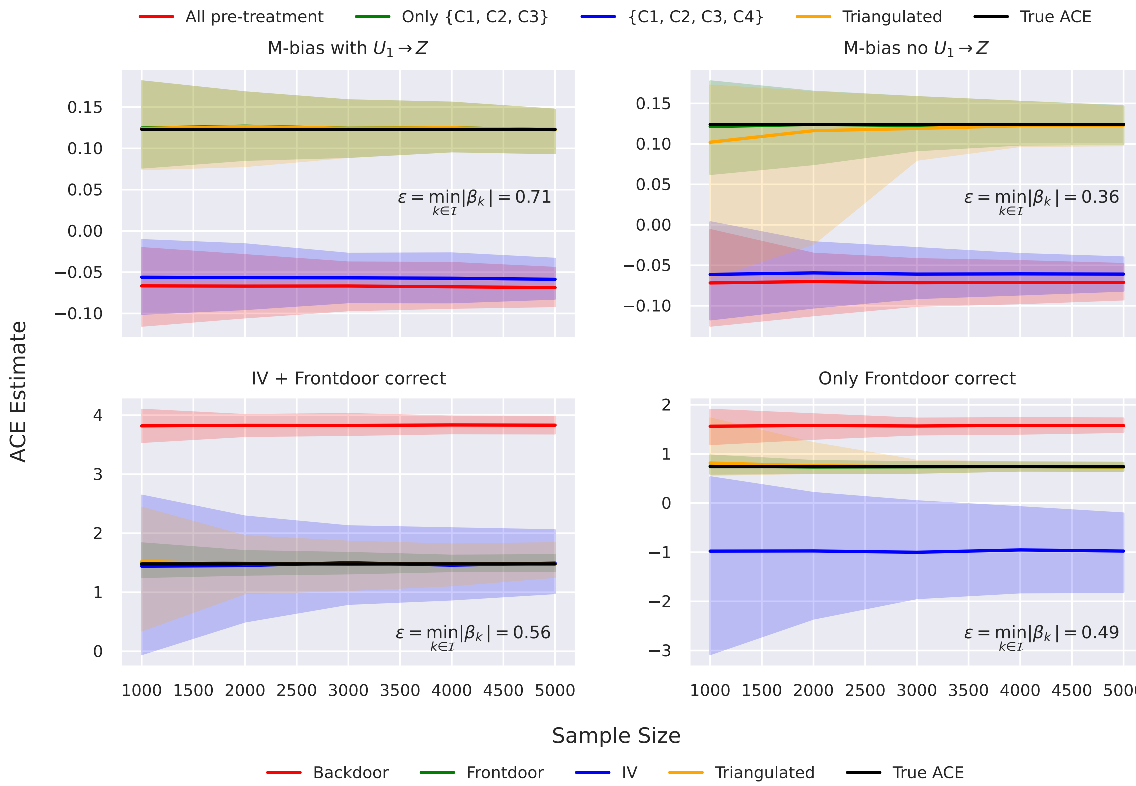


Figure 4: Point estimates averaged over 200 trials; shaded bands correspond to 2.5 and 97.5 percentiles of the estimates.

A, C hold in P respectively. Under assumptions $\mathcal{A}$ in (11)*,*

$$\beta_1 = 0 \implies \theta = \psi_1 \text{ in (2) with } L = C,$$
$$\beta_2 = 0 \implies \theta = \psi_2 \text{ in (3) with } M = M, L = C \cup \{Z\},$$
$$\beta_2 = \beta_3 = 0 \implies \theta = \psi_3 \text{ in (4) with } Z = Z, L = C.$$

We will use $\log(\text{OR}(Y, Z \mid A, C))$ and $\log(\text{OR}(M, Z \mid A, C))$ as β_1 and β_3 respectively. For β_2, we use a reweighted log-odds $\log(\text{OR}_{\widetilde{P}}(Y, Z \mid C))$ defined with respect to the distribution $\widetilde{P} = p(V)/P(M \mid A, Z, C)$. This is estimated using a reweighted regression procedure, similar to those described in Robins and Wasserman [1997], Bhattacharya and Nabi [2022], Robins et al. [2000].

The validity of the IV model also relies on assumptions of homogeneity of the treatment effect—we assume such conditions hold apriori and focus only on structural assumptions of the causal models. Further, per Proposition 2, testability of the IV model relies on testability of the frontdoor model, since β_2 must be zero as well—other known tests of the IV assumptions are also in over-identified models [Kitagawa, 2015]. Thus, to test the IV model we use $\widetilde{\beta_3} = \beta_2 + \beta_3$ under the assumption that these parameters do not exactly cancel out, which can be taken to be a form of faithfulness.

We return to Scenario 2 in Section 3. Figure 3(b) is an example of a DAG that satisfies assumptions $\mathcal{A}$ in (11). Since influence function-based estimators of parameters encoding Verma constraints, such as $\log(\text{OR}_{\widetilde{P}}(Y, Z \mid C))$, are underdeveloped, we rely on plug-in estimators for each ψ_k, β_k with parametric nuisance estimators. This also serves as a test of the bootstrapping branch of our triangulation procedure. We test two scenarios, one in which frontdoor and IV are correct and testable while backdoor is incorrect, and the other in which only frontdoor is correct and testable (using the blue dashed edge in Figure 3(b)). Figure 4 shows that our estimator maintains robustness in both scenarios. When both frontdoor and IV are correct and testable, our triangulation estimator also exhibits lower variance than relying solely on the IV model. Thus, the estimator has benefits even when an analyst may be certain of some identifying assumptions if these do not yield the most efficient estimator. At $n = 5000$ and when both frontdoor and IV are correct, we achieve coverage of 97% and 96% for ψ and θ respectively; when only the frontdoor model is correct, the coverage is just below nominal coverage—93% for both ψ and θ.

We also use Scenario 2 to test our recommendations in Section 4.3 for setting a. In particular, we use the scenario where the frontdoor and IV models are correct, and try four different settings of the bandwidth parameter: $a = n^{-1/3}$, $1/\log(n)$, $1/\sqrt{\log(n)}$, and $1/n$. The results are shown in Figure 5. The choices $a = n^{-1/3}$ and $a = 1/\log(n)$ lie within our recommended range and have roughly similar

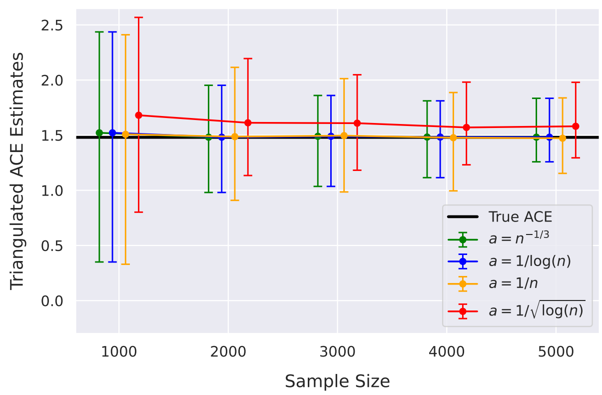


Figure 5: Triangulated point estimates using different settings of a over 200 trials where frontdoor and IV are correct.

performance. Setting $a = 1/\sqrt{\log(n)}$ takes a to zero too slowly, resulting in excess bias and invalid inference for the true parameter θ. Finally, setting $a = 1/n$ takes a to zero faster than our recommendation and results in larger variance.

5.3 FRAMINGHAM DATA APPLICATION

Method	$\widehat{\text{ACE}}$	$\beta_{k,n}, w_{k,n}$
Backdoor	$0.086\ (0.048, 0.115)$	$-0.04, 0.45$
Frontdoor	$0.011\ (0.006, 0.016)$	$-0.02, 0.55$
IV	$-3.12\ (-14.7, 11.09)$	$-0.41, \approx 0$
Triangulation	$0.044(0.026, 0.062)$	–

Table 1: Results for the empirical data application.

We use data from the Framingham Heart Study [Kannel and Gordon, 1968] to estimate the effect of blood glucose levels on coronary heart disease. We treat blood glucose levels as a binary treatment variable, where $A = 1$ corresponds to higher than average levels. The outcome Y is a binary indicator of coronary heart disease. Similar to our setup in Section 5.2, we consider backdoor, frontdoor, and IV models, where Z = educational attainment, C = sex, and M = hypertension. We intentionally adjust for only one confounder C, so that the assumptions of backdoor in particular are difficult to justify, and so triangulation with other estimates from frontdoor or IV may be important.

Results are shown in Table 1, where the last column reports model weights. The frontdoor model receives the highest weight and the IV model the lowest. The backdoor model also retains substantial weight—likely because sex is one of the strongest confounders for both heart disease and blood glucose, making its assumptions approximately valid. The triangulated estimate lies between the backdoor and frontdoor estimates, largely discounting the IV estimate based on diagnostics, and indicates a 4.4% increase in coronary heart disease under intervention to raise blood glucose. As a check, we confirm including additional confounders (e.g., age) does indeed increase the weight applied to the backdoor model, while its point estimate drops to about 0.05.

6 DISCUSSION AND CONCLUSION

In this work we proposed a general framework for triangulating causal effects that uses data-driven weights based on measures of model validity. We showed that our triangulation functional trades robustness to causal model misspecification in exchange for some modest bias, as well as additional assumptions like faithfulness and background knowledge. That is, while robustness through triangulation is desirable, our framework is not without limitations.

Faithfulness, in particular, is an assumption that is often contested in the causal discovery literature. While violations of faithfulness are rare in a measure-theoretic sense [Meek, 1995, Boeken et al., 2024], near violations of faithfulness can be fairly frequent [Uhler et al., 2013]. In our framework, near violations of faithfulness can result in small ε in Theorem 1 and thus large bias. Analysis for cases where ε shrinks with n would provide deeper understanding of the impact of near violations of faithfulness. This is an area of potential future research.

We also developed inference strategies with frequentist guarantees for our proposed triangulation functional, and demonstrated their performance through numerical studies and a data application. This included challenging scenarios in which the plurality of models vote similarly on an incorrect causal effect. Extensions to this may focus on sensitivity analysis, deriving other practical scenarios in which the framework can be used, and data-driven selection of the kernel parameter a. Another interesting avenue for future research is to incorporate ideas from Bayesian paradigms for model averaging into our framework, particularly those which provide robustness to unfaithfulness and the need for point identification [Silva and Evans, 2016].

Acknowledgements

The authors would like to thank Daniel Malinsky for helpful discussions on influence function-based estimation of odds ratios, Hyunseung Kang and Oliver Dukes for helpful discussions on subsampling, and five anonymous reviewers for their helpful peer review. RB would like to thank the Isaac Newton Institute for Mathematical Sciences for the support and hospitality during the programme *Causal inference: From theory to practice and back again* when some of the work on this paper was undertaken. This work was supported by: EPSRC grant EP/Z000580/1 (RB), NSF CRII grant 2348287 (RB), and NSF DMS 2113171 (TW).

References

Joshua D Angrist, Guido W Imbens, and Donald B Rubin. Identification of causal effects using instrumental variables. *Journal of the American Statistical Association*, 91 (434):444–455, 1996.

Heejung Bang and James M Robins. Doubly robust estimation in missing data and causal inference models. *Biometrics*, 61(4):962–973, 2005.

Luca Bergen, Dino Sejdinovic, and Vanessa Didelez. The generalised kernel covariance measure. In *Fifth Conference on Causal Learning and Reasoning*, 2026.

Rohit Bhattacharya and Razieh Nabi. On testability of the front-door model via Verma constraints. In *Uncertainty in Artificial Intelligence*, pages 202–212. PMLR, 2022.

Rohit Bhattacharya, Razieh Nabi, and Ilya Shpitser. Semiparametric inference for causal effects in graphical models with hidden variables. *Journal of Machine Learning Research*, 23(295):1–76, 2022.

P. J. Bickel. On adaptive estimation. *The Annals of Statistics*, 10(3):647–671, 1982.

Pierre Blanchard, Desmond J Higham, and Nicholas J Higham. Accurately computing the log-sum-exp and softmax functions. *IMA Journal of Numerical Analysis*, 41(4):2311–2330, 10 2021.

Philip Boeken, Patrick Forré, and Joris M Mooij. Are Bayesian networks typically faithful? *arXiv preprint arXiv:2410.16004*, 2024.

Ting-Hsuan Chang, Zijian Guo, and Daniel Malinsky. Post-selection inference for causal effects after causal discovery. *Biometrika*, 113(1):asaf073, 2026.

Hua Yun Chen. A semiparametric odds ratio model for measuring association. *Biometrics*, 63(2):413–421, 2007.

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.

Stephen R Cole, Alexander Breskin, Bonnie E Shook-Sa, Paul N Zivich, Michael G Hudgens, and Jessie K Edwards. Five facts about influence functions. *Epidemiology*, 36 (4):467–472, 2025.

Peng Ding and Luke W Miratrix. To adjust or not to adjust? sensitivity analysis of M-bias and butterfly-bias. *Journal of Causal Inference*, 3(1):41–57, 2015.

Bradley Efron and Robert J Tibshirani. *An Introduction To The Bootstrap*. Chapman and Hall/CRC, 1994.

Doris Entner, Patrik Hoyer, and Peter Spirtes. Data-driven covariate selection for nonparametric estimation of causal effects. In *Artificial Intelligence and Statistics*, pages 256–264. PMLR, 2013.

Tracy Farmer, Kerry Robinson, Susan J Elliott, and John Eyles. Developing and implementing a triangulation protocol for qualitative health research. *Qualitative Health Research*, 16(3):377–394, 2006.

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.

Paula Gradu, Tijana Zrnic, Yixin Wang, and Michael I Jordan. Valid inference after causal discovery. *Journal of the American Statistical Association*, 120(550):1127–1138, 2025.

Anna Guo and Razieh Nabi. Average causal effect estimation in dags with hidden variables: Beyond back-door and front-door criteria. *arXiv preprint arXiv:2409.03962*, 2024.

Bruce E Hansen. Sample splitting and threshold estimation. *Econometrica*, 68(3):575–603, 2000.

Zheng He, Roman Pogodin, Yazhe Li, Namrata Deka, Arthur Gretton, and Danica J. Sutherland. On the hardness of conditional independence testing in practice. In *The Thirty-ninth Annual Conference on Neural Information Processing Systems*, 2025.

Shunsuke Horii. Bayesian model averaging for causality estimation and its approximation based on Gaussian scale mixture distributions. In *International Conference on Artificial Intelligence and Statistics*, pages 955–963. PMLR, 2021.

Yonghan Jung, Jin Tian, and Elias Bareinboim. Estimating identifiable causal effects through double machine learning. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 35, pages 12113–12122, 2021.

Hyunseung Kang, Anru Zhang, T Tony Cai, and Dylan S Small. Instrumental variables estimation with some invalid instruments and its application to Mendelian randomization. *Journal of the American statistical Association*, 111(513):132–144, 2016.

William B Kannel and Tavia Gordon. *The Framingham Study: An epidemiological investigation of cardiovascular disease*. Number 9-13. Department of Health, Education, and Welfare, Public Health Service . . . , 1968.

Bikram Karmakar, Benjamin French, and Dylan S Small. Integrating the evidence from evidence factors in observational studies. *Biometrika*, 106(2):353–367, 2019.

Toru Kitagawa. A test for instrument validity. *Econometrica*, 83(5):2043–2063, 2015.

Debbie A Lawlor, Kate Tilling, and George Davey Smith. Triangulation in aetiological epidemiology. *International Journal of Epidemiology*, 45(6):1866–1886, 2016.

Daniel Malinsky, Ilya Shpitser, and Thomas Richardson. A potential outcomes calculus for identifying conditional path-specific effects. In *The 22nd International Conference on Artificial Intelligence and Statistics*, pages 3080–3088. PMLR, 2019.

Christopher Meek. Strong completeness and faithfulness in Bayesian networks. In *Proceedings of the Eleventh conference on Uncertainty in artificial intelligence*, pages 411–418, 1995.

Whitney K Newey and James M Robins. Cross-fitting and fast remainder rates for semiparametric estimation. *arXiv preprint arXiv:1801.09138*, 2018.

Judea Pearl. Causal diagrams for empirical research. *Biometrika*, 82(4):669–710, 1995.

Judea Pearl. *Causality*. Cambridge University Press, 2009a.

Judea Pearl. Remarks on the method of propensity score. *Statistics in Medicine*, 28:1415–1416, 2009b.

Dimitris N Politis, Joseph P Romano, and Michael Wolf. On the asymptotic theory of subsampling. *Statistica Sinica*, pages 1105–1124, 2001.

Prabrisha Rakshit, Xu Shi, and Eric Tchetgen Tchetgen. Adaptive proximal causal inference with some invalid proxies. *arXiv preprint arXiv:2507.19623*, 2025.

James Robins. A new approach to causal inference in mortality studies with a sustained exposure period—application to control of the healthy worker survivor effect. *Mathematical modelling*, 7(9-12):1393–1512, 1986.

James M Robins and Larry Wasserman. Estimation of effects of sequential treatments by reparameterizing directed acyclic graphs. In *Proceedings of the Thirteenth Conference on Uncertainty in Artificial Antelligence*, pages 409–420, 1997.

James M Robins, Miguel Angel Hernán, and Babette Brumback. Marginal structural models and causal inference in epidemiology. *Epidemiology*, 11(5):550–560, 2000.

Paul R Rosenbaum. Observational studies. In *Observational studies*, pages 1–17. Springer, 2002.

Paul R Rosenbaum. Evidence factors in observational studies. *Biometrika*, 97(2):333–345, 2010.

Paul R Rosenbaum. Some approximate evidence factors in observational studies. *Journal of the American Statistical Association*, 106(493):285–295, 2011.

Donald B Rubin. Should observational studies be designed to allow lack of balance in covariate distributions across treatment groups? *Statistics in Medicine*, 28(9):1420–1423, 2009.

Rajen D Shah and Jonas Peters. The hardness of conditional independence testing and the generalized covariance measure. *The Annals of Statistics*, 48(3):1514–1538, 2020.

Ian Shrier. Letter to the editor. *Statistics in Medicine*, 28: 2740–2741, 2009.

Ricardo Silva and Robin Evans. Causal inference through a witness protection program. *Journal of Machine Learning Research*, 17(56):1–53, 2016.

Arvid Sjölander. Propensity scores and M-structures. *Statistics in medicine*, 28(9):1416–1420, 2009.

Gregor Steiner and Mark Steel. Bayesian model averaging in causal instrumental variable models. *arXiv preprint arXiv:2504.13520*, 2025.

Baoluo Sun, Zhonghua Liu, and Eric Tchetgen Tchetgen. On multiply robust Mendelian randomization (MR2) with many invalid genetic instruments. *medRxiv*, pages 2021–10, 2021.

Zhiqiang Tan. On doubly robust estimation for logistic partially linear models. *Statistics & Probability Letters*, 155:108577, 2019.

Eric J Tchetgen Tchetgen, James M Robins, and Andrea Rotnitzky. On doubly robust estimation in a semiparametric odds ratio model. *Biometrika*, 97(1):171–180, 2010.

Veronica A Thurmond. The point of triangulation. *Journal of Nursing Scholarship*, 33(3):253–258, 2001.

Jin Tian and Judea Pearl. A general identification condition for causal effects. In *Eighteenth National Conference on Artificial Intelligence*, pages 567–573, 2002.

Caroline Uhler, Garvesh Raskutti, Peter Bühlmann, and Bin Yu. Geometry of the faithfulness assumption in causal inference. *The Annals of Statistics*, pages 436–463, 2013.

Aad W Van der Vaart. *Asymptotic Statistics*, volume 3. Cambridge University Press, 2000.

Thomas Verma and Judea Pearl. Equivalence and synthesis of causal models. In *Proceedings of the Sixth Annual Conference on Uncertainty in Artificial Intelligence*, pages 255–270, 1990.

Feng Xie, Zhengming Chen, Shanshan Luo, Wang Miao, Ruichu Cai, and Zhi Geng. Automating the selection of proxy variables of unmeasured confounders. In *Forty-first International Conference on Machine Learning*, 2024.

Junhui Yang, Rohit Bhattacharya, Youjin Lee, and Ted Westling. Statistical and causal robustness for causal null hypothesis tests. In *The 40th Conference on Uncertainty in Artificial Intelligence*, 2024.

Minhao Yao, Gary W Miller, Badri N Vardarajan, Andrea A Baccarelli, Zijian Guo, and Zhonghua Liu. Deciphering proteins in Alzheimer's disease: A new Mendelian randomization method integrated with AlphaFold3 for 3D structure prediction. *Cell Genomics*, 4(12), 2024.

Robust Weighted Triangulation of Causal Effects Under Model Uncertainty (Supplementary Material)

Rohit Bhattacharya[1,*] **Ina Ocelli**[1,*] **Ted Westling**[2]

rb17@williams.edu io1@williams.edu twestling@umass.edu

[1,*] Equal contributors. Dept. of Computer Science, Williams College
[2] Dept. of Mathematics and Statistics, University of Massachusetts Amherst

In this supplement we provide proofs that were omitted from the main paper for space, as well as details of data generating processes and the specific estimators used in our numerical experiments and data application.

A PROOF OF THEOREM 1

Proof. Since $\sum_k w_k = 1$, $w_k \geq 0$, and $\psi_k = \theta$ for any $k \in \mathcal{C}$, we have

$$\begin{aligned}
|\psi - \theta| &= \left| \sum_{k=1}^{K} w_k \psi_k - \theta \right| \\
&= \frac{\left| \sum_{k=1}^{K} \delta_a(\beta_k) \left[\psi_k - \theta\right] \right|}{\sum_{k=1}^{K} \delta_a(\beta_k)} \\
&= \frac{\left| \sum_{k \in \mathcal{I}} \delta_a(\beta_k) \left[\psi_k - \theta\right] \right|}{\sum_{k \in \mathcal{I}} \delta_a(\beta_k) + \sum_{k \in \mathcal{C}} \delta_a(\beta_k)} \\
&\leq \frac{\sum_{k \in \mathcal{I}} \delta_a(\beta_k) \max_k |\psi_k - \theta|}{\sum_{k \in \mathcal{I}} \delta_a(\beta_k) + \sum_{k \in \mathcal{C}} \delta_a(\beta_k)} \\
&= \frac{\max_k |\psi_k - \theta|}{1 + D_a}.
\end{aligned}$$

Now by the definition of δ_a,

$$D_a = \frac{\sum_{k \in \mathcal{C}} \delta_a(\beta_k)}{\sum_{k \in \mathcal{I}} \delta_a(\beta_k)} = \frac{\sum_{k \in \mathcal{C}} e^{-\beta_k^2/a^2}}{\sum_{k \in \mathcal{I}} e^{-\beta_k^2/a^2}}.$$

Under our assumptions, $\beta_k = 0$ for any model that is correct and testable. If there is at least one correct and testable model, then $e^{-\beta_k^2/a^2} = 1$ for this model, and so $\sum_{k \in \mathcal{C}} e^{-\beta_k^2/a^2} \geq 1$. For the denominator, by monotonicity of $x \mapsto e^{-(x/a)^2}$ for $x \geq 0$, we get $\sum_{k \in \mathcal{I}} e^{-\beta_k^2/a^2} \leq |\mathcal{I}| e^{-\varepsilon^2/a^2}$. The result follows. $\square$

B DERIVING THE PARTIAL DERIVATIVE VECTOR IN LEMMA 1

To obtain the variance of the combined estimator using the delta method, we require the following vector of partial derivatives,

$$\gamma = \left[\frac{\partial \psi}{\partial \beta_1}, \ldots, \frac{\partial \psi}{\partial \beta_K}, \frac{\partial \psi}{\partial \psi_1}, \ldots, \frac{\partial \psi}{\partial \psi_K} \right]. \tag{12}$$

We divide the task of deriving the partial derivatives $\frac{\partial \psi}{\partial \beta_k}$ and $\frac{\partial \psi}{\partial \psi_k}$ for all $k \in \{1, \ldots, K\}$ into two subsections as follows.

B.1 PARTIAL DERIVATIVE OF ψ WITH RESPECT TO β_k

First note that by plugging in the definition of ψ and by linearity of differentiation we have,

$$\frac{\partial\psi}{\partial\beta_k} = \frac{\partial}{\partial\beta_k}\left(\sum_{i=1}^{K} w_i\psi_i\right) = \sum_{i=1}^{K}\left(\frac{\partial(w_i\psi_i)}{\partial\beta_k}\right). \tag{13}$$

For each term in (15), we can apply the product rule to get,

$$\frac{\partial\psi}{\partial\beta_k} = \sum_{i=1}^{K}\left(w_i\frac{\partial\psi_i}{\partial\beta_k} + \psi_i\frac{\partial w_i}{\partial\beta_k}\right). \tag{14}$$

The estimators ψ_i are not functions of β_k (even when $i = k$). Thus, $\frac{\partial\psi_i}{\partial\beta_k}$ is always 0. So the derivative simplifies to,

$$\frac{\partial\psi}{\partial\beta_k} = \sum_{i=1}^{K}\psi_i\frac{\partial w_i}{\partial\beta_k}. \tag{15}$$

Plugging in the definition for weights w_i we get,

$$\frac{\partial\psi}{\partial\beta_k} = \sum_{i=1}^{K}\psi_i \times \frac{\partial}{\partial\beta_k}\left(\delta(\beta_i)/\big(\lambda_n + \sum_{j=1}^{K}\delta(\beta_j)\big)\right). \tag{16}$$

We now separate the above terms in the summation based on whether $i = k$ or not, as these are the only two cases that lead to substantively different partial derivates with respect to β_k:

$$\frac{\partial\psi}{\partial\beta_k} = \psi_k \times \frac{\partial}{\partial\beta_k}\left(\delta(\beta_k)/\big(\lambda_n + \sum_{j=1}^{K}\delta(\beta_j)\big)\right) + \sum_{i\neq k}\psi_i \times \frac{\partial}{\partial\beta_k}\left(\delta(\beta_i)/\big(\lambda_n + \sum_{j=1}^{K}\delta(\beta_j)\big)\right). \tag{17}$$

We first deal with,

$$\frac{\partial}{\partial\beta_k}\left(\delta(\beta_k)/\big(\lambda_n + \sum_{j=1}^{K}\delta(\beta_j)\big)\right).$$

By the quotient rule,

$$\frac{\partial}{\partial\beta_k}\left(\delta(\beta_k)/\big(\lambda_n + \sum_{j=1}^{K}\delta(\beta_j)\big)\right) = \frac{\frac{\partial\delta(\beta_k)}{\partial\beta_k}\big(\lambda_n + \sum_{j=1}^{K}\delta(\beta_j)\big) - \delta(\beta_k)\frac{\partial\sum_{j=1}^{K}\delta(\beta_j)}{\partial\beta_k}}{\left(\lambda_n + \sum_{j=1}^{K}\delta(\beta_j)\right)^2}. \tag{18}$$

We introduce two helper derivatives of the Dirac delta function to simplify (18):

$$\frac{\partial\,\delta(\beta_j)}{\partial\beta_k} = \frac{\partial}{\partial\beta_k}\left(\frac{1}{|a|\sqrt{\pi}}e^{-\left(\frac{\beta_j}{a}\right)^2}\right) = \begin{cases} -\frac{2\beta_k}{a^2}\delta(\beta_k) & \text{when } j = k \\ 0 & \text{when } j \neq k. \end{cases} \tag{19}$$

Plugging these into (18) and simplifying gives,

$$\frac{\partial}{\partial\beta_k}\left(\delta(\beta_k)/\big(\lambda_n + \sum_{j=1}^{K}\delta(\beta_j)\big)\right) = \frac{\left(-\frac{2\beta_k}{a^2}\delta(\beta_k)\right)\big(\lambda_n + \sum_{j=1}^{K}\delta(\beta_j)\big) - \delta(\beta_k)\left(-\frac{2\beta_k}{a^2}\delta(\beta_k)\right)}{\left(\lambda_n + \sum_{j=1}^{K}\delta(\beta_j)\right)^2} \tag{20}$$

$$= \frac{\left(-\frac{2\beta_k}{a^2}\delta(\beta_k)\right)\left(\lambda_n + \sum_{j\neq k}\delta(\beta_j)\right)}{\left(\lambda_n + \sum_{j=1}^{K}\delta(\beta_j)\right)^2} \tag{21}$$

$$= -\frac{2\beta_k w_k}{a^2}\frac{\left(\lambda_n + \sum_{j\neq k}\delta(\beta_j)\right)}{\left(\lambda_n + \sum_{j=1}^{K}\delta(\beta_j)\right)}. \tag{22}$$

In the above, the first equality comes from plugging in the helper derivatives, the second follows from factorizing a common term, and the third comes from merging terms that correspond to w_k. Now we tackle the second set of terms in (17) involving,

$$\frac{\partial}{\partial \beta_k}\left(\delta(\beta_i)/\big(\lambda_n + \sum_{j=1}^{K}\delta(\beta_j)\big)\right),$$

where $i \neq k$. By the quotient rule again,

$$\frac{\partial}{\partial \beta_k}\left(\delta(\beta_i)/\big(\lambda_n + \sum_{j=1}^{K}\delta(\beta_j)\big)\right) = \frac{\frac{\partial \delta(\beta_i)}{\partial \beta_k}\big(\lambda_n + \sum_{j=1}^{K}\delta(\beta_j)\big) - \delta(\beta_i)\frac{\partial \sum_{j=1}^{K}\delta(\beta_j)}{\partial \beta_k}}{\Big(\lambda_n + \sum_{j=1}^{K}\delta(\beta_j)\Big)^2}. \tag{23}$$

Plugging in the helper derivatives in (19) and simplifying gives,

$$\frac{\partial}{\partial \beta_k}\left(\delta(\beta_i)/\big(\lambda_n + \sum_{j=1}^{K}\delta(\beta_j)\big)\right) = \frac{0 - \delta(\beta_i)\left(-\frac{2\beta_k}{a^2}\delta(\beta_k)\right)}{\Big(\lambda_n + \sum_{j=1}^{K}\delta(\beta_j)\Big)^2} \tag{24}$$

$$= \frac{2\beta_k}{a^2}\frac{\delta(\beta_k)\delta(\beta_i)}{\Big(\lambda_n + \sum_{j=1}^{K}\delta(\beta_j)\Big)^2} \tag{25}$$

$$= \frac{2\beta_k w_k w_i}{a^2}. \tag{26}$$

Plugging (22) and (26) back into (17) gives us the following expression for the partial derivative of the combined estimator ψ with respect to β_k:

$$\frac{\partial \psi}{\partial \beta_k} = -\psi_k\left(\frac{2\beta_k w_k}{a^2}\frac{\Big(\lambda_n + \sum_{j\neq k}\delta(\beta_j)\Big)}{\Big(\lambda_n + \sum_{j=1}^{K}\delta(\beta_j)\Big)}\right) + \sum_{i\neq k}\psi_i\left(\frac{2\beta_k w_k w_i}{a^2}\right). \tag{27}$$

This can be further simplified to yield the final expression as,

$$\frac{\partial \psi}{\partial \beta_k} = \frac{2\beta_k w_k}{a^2}\left(\sum_{i\neq k} w_i\psi_i - \psi_k\frac{\Big(\lambda_n + \sum_{j\neq k}\delta(\beta_j)\Big)}{\Big(\lambda_n + \sum_{j=1}^{K}\delta(\beta_j)\Big)} + w_k\psi_k - w_k\psi_k\right) \tag{28}$$

$$= \frac{2\beta_k w_k}{a^2}\left(\psi - \psi_k\left[\frac{\Big(\lambda_n + \sum_{j\neq k}\delta(\beta_j)\Big)}{\Big(\lambda_n + \sum_{j=1}^{K}\delta(\beta_j)\Big)} + w_k\right]\right) \tag{29}$$

$$= \frac{2\beta_k w_k}{a^2}\left(\psi - \psi_k\right). \tag{30}$$

B.2 PARTIAL DERIVATIVE OF ψ WITH RESPECT TO ψ_k

Here we again apply the linearity of differentiation to get,

$$\frac{\partial \psi}{\partial \psi_k} = \frac{\partial}{\partial \psi_k}\left(\sum_{i=1}^{K} w_i\psi_i\right) = \sum_{i=1}^{K}\left(\frac{\partial(w_i\psi_i)}{\partial \psi_k}\right). \tag{31}$$

Applying the product rule gives us,

$$\frac{\partial \psi}{\partial \psi_k} = \sum_{i=1}^{K}\left(w_i\frac{\partial \psi_i}{\partial \psi_k} + \psi_i\frac{\partial w_i}{\partial \psi_k}\right). \tag{32}$$

Similar to the previous subsection, notice that none of the weights w_i are a function of any of the estimators ψ_k (even when $i = k$). Thus, $\frac{\partial w_i}{\partial \psi_k}$ is always 0. So the derivative simplifies to,

$$\frac{\partial \psi}{\partial \psi_k} = \sum_{i=1}^{K} w_i \frac{\partial \psi_i}{\partial \psi_k}. \tag{33}$$

Finally, this simplifies nicely as,

$$\frac{\partial \psi_i}{\partial \psi_k} = \begin{cases} 1 & \text{when } \; i = k \\ 0 & \text{when } i \neq k. \end{cases} \tag{34}$$

Plugging these in gives us the final expression for the partial derivative of ψ with respect to ψ_i,

$$\frac{\partial \psi}{\partial \psi_k} = w_k. \tag{35}$$

C PROOFS OF PROPOSITIONS 1 AND 2

Proof of Proposition 1

Proof. Entner et al. [2013] show under assumptions $\mathcal{A}_1$ and $\mathcal{A}_2$ in (10) that $W \subseteq C$ is a valid backdoor adjustment if $Y \not\perp\!\!\!\perp Z \mid W$ and $Y \perp\!\!\!\perp Z \mid A, W$. The additional assumption $\mathcal{A}_3$ we make in (10) ensures that $Y \not\perp\!\!\!\perp Z \mid W$ is already true due to the existence of the path $Z \to A \to Y$. Thus, under assumptions $\mathcal{A}_1, \mathcal{A}_2, \mathcal{A}_3$, the independence $Y \perp\!\!\!\perp Z \mid A, W$ alone is sufficient to ensure that model $\mathcal{M}_k$ is correct (i.e., W is a valid backdoor adjustment set). The conclusion then follows, as we suppose that β_k is an observed data parameter such that $\beta_k = 0 \iff Y \perp\!\!\!\perp Z \mid A, W$. □

Proof of Proposition 2

Proof. The argument for $\beta_1 = 0$ implying correctness of the backdoor model with adjustment set C is essentially the same as the proof of Proposition 1 with $W = C$, since the assumptions used in Proposition 1 are a superset of those in (10).

For the second implication, we make an argument similar to the one used in Bhattacharya and Nabi [2022]. Under the Verma faithfulness assumption $\mathcal{A}_1$, we know that if $\beta_2 = 0$, the causal DAG $\mathcal{G}(V \cup U)$ must support identification of $P(A, Z, C, Y \mid \mathrm{do}(m))$ by the g-formula and $Y \perp\!\!\!\perp_{\text{d-sep}} Z \mid C$ in $\mathcal{G}_{\mathrm{do}(m)}$. By assumption $\mathcal{A}_3$ we know that $Z \to A$ exists in $\mathcal{G}$. We now show that the existence of $A \to Y$ in $\mathcal{G}$ contradicts existence of the Verma constraint under faithfulness. Suppose $A \to Y$ does exist in $\mathcal{G}$. Then, $Y \not\perp\!\!\!\perp_{\text{d-sep}} Z \mid C$ in $\mathcal{G}_{\mathrm{do}(m)}$ due to the open path $Z \to A \to Y$, which is a contradiction. Thus, the frontdoor exclusion restriction of no $A \to Y$ is satisfied. Per Tian and Pearl [2002], the distribution $P(V \setminus \{A\} \mid \mathrm{do}(a))$ where A is a single treatment variable is identified if and only if there is no path from A to any child X of A of the form $A \leftarrow \cdots \to X$ such that every collider on the path is an observed variable $V_i \in V$ and every non-collider on the path is an unmeasured variable in $U_i \in U$. By the causal ordering assumption $\mathcal{A}_2$ and the previous argument ruling out the existence of $A \to Y$, the only child of A is M, so we only need to show that no such path exists from A to M. The existence of such a path between A and M also contradicts the presence of the Verma constraint, as it would imply the existence of some $U_i \in U$ that causes M and thus also causes Y (by assumption $\mathcal{A}_4$), preventing identification of $p(A, Z, C, Y \mid \mathrm{do}(m))$. Thus, the second frontdoor restriction is satisfied and $P(Z, C, M, Y \mid \mathrm{do}(a))$ is identified as [Tian and Pearl, 2002]

$$\left(\sum_{a'} P(a', Z, C) \cdot \mathbb{E}[Y \mid a', M, C, Z] \right) \cdot p(M \mid A = a, Z, C) \cdot$$

It is straightforward then to obtain the frontdoor functional in (3) for θ by summing over Z, C to obtain $P(Y \mid \mathrm{do}(a))$.

Finally for the third implication, Z already satisfies the IV relevance condition by assumption $\mathcal{A}_3$. The second condition for IV validity is that Z should be d-separated from Y given C in a graph where we delete the outgoing edges from A. This is satisfied when $M \perp\!\!\!\perp Z \mid A, C$ and the Verma constraint holds per Corollary 1.1 in Bhattacharya and Nabi [2022]. □

D DETAILS OF DATA GENERATING PROCESSES

In this subsection we describe the data generating processes (DGPs) underlying each of our numerical experiments. In our descriptions, we define $\mathrm{expit}(x) \coloneqq 1/(1 + \exp(-x))$ for $x \in \mathbb{R}$. We use $\mathrm{Bern}(p)$ as shorthand for the Bernoulli distribution with probability p and $N(\mu, \sigma^2)$ as shorthand for the normal distribution with mean μ and variance σ^2.

D.1 DGPS FOR SECTION 5.1

When the $U_1 \to Z$ edge is absent, the data are generated according to Figure 3(a) as,

$$\begin{aligned}
Z &\sim \text{Bern}(0.5)\\
U_1, U_2, C_1, C_2, C_3 &\sim N(0,1)\\
C_4 &\sim N\big(-2.5 \cdot U_1 + 2 \cdot U_2,\ 1\big)\\
C_5 &\sim N\big(-2.5 \cdot U_1 + 2 \cdot U_2,\ 1\big)\\
A &\sim \text{Bern}\big(\text{expit}(2.75 \cdot Z - 3 \cdot U_1 + C_1 + C_2 + C_3)\big)\\
Y &\sim \text{Bern}\big(\text{expit}(1.5 \cdot A + 2 \cdot U2 + C_3 + C_4 + C_5)\big)
\end{aligned}$$

When the $U_1 \to Z$ edge is present, the data are generated according to Figure 3(a) as,

$$\begin{aligned}
U_1, U_2, C_1, C_2, C_3 &\sim N(0,1)\\
Z &\sim \text{Bern}\big(\text{expit}(U_1)\big)\\
C_4 &\sim N\big(-2.5 \cdot U_1 + 2 \cdot U_2,\ 1\big)\\
C_5 &\sim N\big(-2.5 \cdot U_1 + 2 \cdot U_2,\ 1\big)\\
A &\sim \text{Bern}\big(\text{expit}(2 \cdot Z - 3 \cdot U_1 + C_1 + C_2 + C_3)\big)\\
Y &\sim \text{Bern}\big(\text{expit}(A + 2 \cdot U2 + C_3 + C_4 + C_5)\big)
\end{aligned}$$

D.2 DGPS FOR SECTION 5.2

When both frontdoor and IV models are correct, data are generated according to Figure 3(b) without the blue dashed edge as,

$$\begin{aligned}
Z &\sim \text{Bern}(0.5)\\
C &\sim N(0,1)\\
U &\sim N(0,1)\\
A &\sim \text{Bern}\big(\text{expit}(2 \cdot Z + 2 \cdot C + 2 \cdot U)\big)\\
M &\sim \text{Bern}\big(\text{expit}(-2 + 4 \cdot A - 0.5 \cdot C)\big)\\
Y &\sim N\big(2 \cdot M + 2 \cdot C + 2 \cdot U,\ 1\big)
\end{aligned}$$

When only the frontdoor model is correct, data are generated according to Figure 3(b) with the blue dashed edge as,

$$\begin{aligned}
Z &\sim \text{Bern}(0.5)\\
C &\sim N(0,1)\\
U &\sim N(0,1)\\
A &\sim \text{Bern}\big(\text{expit}(Z + C - 0.5 \cdot U)\big)\\
M &\sim \text{Bern}\big(\text{expit}(-1 + 2 \cdot A - Z + C)\big)\\
Y &\sim N\big(2 \cdot M - 0.75 \cdot C - 2 \cdot U,\ 1\big)
\end{aligned}$$

E ESTIMATORS USED IN NUMERICAL EXPERIMENTS AND DATA APPLICATION

Below we describe the specific estimators used in our numerical experiments and data application in more detail.

E.1 ESTIMATORS USED IN SECTION 5.1

Estimators for β_k

For each log-odds ratio β_k, we use an influence function-based estimator of $\log(\text{OR}(Y, Z|A, W))$ from Tchetgen Tchetgen et al. [2010] and Tan [2019]. An R implementation of their method is publicly available from Wu and Malinsky at https://github.com/chaoqiw0324/ortest; we translate this to Python for our purposes. First, let $\zeta(a, w) \coloneqq \mathbb{E}[Y \mid z_0, a, w]$ and $\eta(a, w) \coloneqq \mathbb{E}[Z \mid y_0, a, w]$. Let $O = \{Y, A, Z, W\}$. Then, an unbiased estimating function of β_k is

$$g(o, \zeta, \eta, \beta_k) = (y - \zeta(a, w)) \cdot (z - \eta(a, w)) \cdot e^{-\beta_k \cdot y \cdot z}.$$

Since g is an estimating function, we can construct a point estimate $\beta_{k,n}$ of the log-odds ratio by first constructing estimators ζ_n and η_n of ζ and η respectively, and using these to obtain the value of β_k that solves the equation $\sum_{i=1}^{n} g(o_i, \zeta_n, \eta_n, \beta_k) = 0$. This estimator is asymptotically linear under doubly robust conditions on the nuisance estimators ζ_n and η_n. Further, by standard Z-estimator theory, the influence function $\phi_{\beta_k} = B^{-1} g(o, \zeta, \eta, \beta_k)$, where $B^{-1} = \mathbb{E}[-\partial g / \partial \beta_k]$; see for example the review in Cole et al. [2025]. We can then construct an estimator of the influence function $\phi_{\beta_k, n}$ as

$$\frac{1}{n} \sum_{i=1}^{n} \left(\frac{-\partial g(o_i, \zeta_n, \eta_n, \beta_k)}{\partial \beta_k} \right)^{-1} \cdot g(o_i, \zeta_n, \eta_n, \beta_k) \bigg|_{\beta_k = \beta_{k,n}}.$$

Estimators for ψ_k

For each $\mathcal{M}_k$ that uses an adjustment set W, we use the AIPW estimator [Bang and Robins, 2005] of the backdoor formula (2) with $L = W$. Let $\pi(w) \coloneqq P(A = 1 \mid W = w)$ and $\mu(a, w) \coloneqq \mathbb{E}[Y \mid A = a, W = w]$. Then the AIPW estimator is an asymptotically linear estimator of ψ_k under doubly robust conditions on the nuisance estimators π_n and μ_n of π and μ respectively, with influence function ϕ_{ψ_k} given by

$$\phi_{\psi_k} = \{y - \mu(a, w)\} \left\{ \frac{a - \pi(w)}{\pi(w)(1 - \pi(w))} \right\} + \{\mu(1, w) - \mu(0, w)\} - \psi_k.$$

E.2 ESTIMATORS USED IN SECTION 5.2 AND DATA APPLICATION

For these experiments we construct plug-in estimators of $\beta_{k,n}$ and $\psi_{k,n}$ based on parametric nuisance estimation.

Estimators for β_k

For β_1 and β_3, we construct parametric nuisance estimators ζ_n and η_n of $\zeta(y, c, a, z) \coloneqq P(Y \mid A, C, Z)$ and $\eta(m, a, z, c) \coloneqq P(M \mid A, Z, C)$ respectively. In the parametric models we consider, estimates of the log-odds ratio $\log(\text{OR}(Y, Z \mid A, C))$ and $\log(\text{OR}(M, Z \mid A, C))$ are given by the coefficients of Z in ζ_n and η_n respectively.

For β_2, the parameter used to test the Verma constraint, let $\alpha \coloneqq P(Y \mid Z, C)$ and $g(y, z, c, \alpha)$ be any unbiased estimating function used to construct a parametric nuisance estimator α_n of α. The reweighted log-odds ratio $\log(\text{OR}_{\widetilde{P}}(Y, Z \mid C))$ is obtained as the coefficient of Z, one of the parameters estimated in α_n, where α_n is the solution to a set of reweighted estimating equations $\sum_{i=1}^{n} g(y_i, z_i, c_i, \alpha) / \eta_n(m_i, a_i, z_i, c_i) = 0$ and η_n is a parametric nuisance estimator of η as before. In terms of practical implementation, $\log(\text{OR}_{\widetilde{P}}(Y, Z \mid C))$ can be treated as the coefficient of Z in a reweighted (linear or logistic) regression of the outcome on the covariates C and anchor variable Z, where the weights are given by $1/\eta(m, a, z, c)$.

Estimators for ψ_k

For ψ_1, the backdoor functional (2) with $L = C$, we construct a plug-in estimator under parametric specification of the outcome regression model $\mu \coloneqq \mathbb{E}[Y \mid A, C]$. For ψ_2, the frontdoor functional (3) with $L = C \cup \{Z\}$, we use a plug-in estimator of the more convenient dual IPW functional instead under parametric specification of the mediator model $\eta \coloneqq P(M \mid A, Z, C)$ [Fulcher et al., 2020, Bhattacharya et al., 2022]

$$\psi_{2,\text{Dual IPW}} = \mathbb{E}\left[\frac{P(M \mid A = 1, Z, C)}{P(M \mid A, Z, C)} \times Y \right] - \mathbb{E}\left[\frac{P(M \mid A = 0, Z, C)}{P(M \mid A, Z, C)} \times Y \right].$$

Finally, for ψ_3, the IV functional (4) with $L = C$, we construct a plug-in estimator under parametric specification of the nuisance function in the numerator $\nu \coloneqq \mathbb{E}[Y \mid Z, C]$ and denominator $\xi \coloneqq \mathbb{E}[A \mid Z, C]$.