

Comparing Two Proxy Methods for Causal Identification

Helen Guo, Elizabeth L. Ogburn, and Ilya Shpitser

Abstract. Identifying causal effects in the presence of unmeasured variables is a fundamental challenge in causal inference, for which proxy variable methods have emerged as a powerful solution. We contrast two major approaches in this framework: (1) bridge equation methods, which leverage solutions to integral equations to recover causal targets, and (2) array decomposition methods, which recover latent factors composing counterfactual quantities by exploiting unique determination of eigenspaces. We compare the model restrictions underlying these two approaches and provide insight into implications of the underlying assumptions, clarifying the scope of applicability for each method.

1. INTRODUCTION

Identifying causal effects in the presence of unmeasured variables is a fundamental challenge in causal inference. Interest in proxy variable methods that leverage observed consequences of latent variables (i.e., proxies), has grown rapidly in both applied and methodological domains. Within this landscape, two foundational lines of work underpin nonparametric proxy variable methods for causal effect identification: (1) the bridge equation approach from [Miao, Geng and Tchetgen Tchetgen \(2018\)](#); and (2) the array decomposition approach, introduced to the causal literature by [Kuroki and Pearl \(2014\)](#) and [Allman et al. \(2015\)](#), building on ideas for full law identification in discrete latent variable models which trace back to [Kruskal \(1977\)](#). Both approaches admit generalizations, to varying degrees, to continuous settings.

The bridge equation approach was developed for settings in which proxies are available for an unmeasured confounder, allowing the counterfactual distribution to be recovered directly through re-expression of the adjustment formula, without identification of the full law ([Miao, Geng and Tchetgen Tchetgen, 2018](#)). The method leverages solutions to bridge equations—a generalization of solutions to linear systems—to recover the causal target. This method was later adapted to settings where proxies are available for an unmeasured mediator in the absence of hidden confounder proxies, recovering causal effects through re-expression of the front-door adjustment formula ([Ghassami et al., 2024](#)).

Department of Biostatistics, Johns Hopkins Bloomberg School of Public Health , (e-mail: hguo51@jhu.edu; eogburn@jhu.edu) Department of Computer Science, Johns Hopkins Whiting School of Engineering , (e-mail: ilyas@cs.jhu.edu)

The array decomposition approach imposes similar invertibility conditions, but rather than solving directly for a causal target, it recovers the full law of latent and observed variables, from which causal effects follow. Originating from the work of [Kruskal \(1977\)](#), this framework allows causal effect identification in models with arbitrary hidden variables, such as in settings with unmeasured confounders ([Kuroki and Pearl, 2014](#)) or unmeasured exposures ([Zhou and Tchetgen, 2024](#)). Though initially formulated for effect identification in fully discrete models ([Kuroki and Pearl, 2014; Allman, Matias and Rhodes, 2009](#)), its identifying assumptions extend to continuous settings via the results of [Hu and Schennach \(2008\)](#), as shown by [Deaner \(2023\)](#).

Despite their shared ability to identify causal effects in the presence of unmeasured variables, these two approaches exhibit non-overlapping model restrictions. Prior work from [Deaner \(2023\)](#) has shown that the conditional independence assumptions (i.e., Markov restrictions) underlying the two approaches for hidden confounding are distinct and the models non-nested.

Until now, no work has characterized the relationship between the downstream assumptions required by each approach under shared conditional independences. This leaves unclear when a methodologist or practitioner should prefer one framework over the other. Hence, our work fills an important gap by providing a transparent comparison of the structural assumptions required by each method under shared Markov restrictions, clarifying how the two frameworks differ and the settings in which each is most appropriate.

1.1 Causal Graphical Models

Bayesian networks, represented as directed acyclic graphs (DAGs), offer a structured framework for statistical modeling, where nodes denote random variables

and edges encode conditional independence assumptions (i.e., Markov restrictions) (Pearl, 1988). DAGs may be interpreted causally (Pearl, 2009; Spirtes, Glymour and Scheines, 2001), encoding counterfactual relationships which enable formal reasoning about causal effects and their identifiability from observed data (Richardson and Robins, 2013). For example, Figure 1 depicts a simple confounding structure in which a common cause C influences both treatment A and outcome Y . This structure entails the counterfactual independence statement $Y(a) \perp\!\!\!\perp A | C$, known as conditional ignorability, where $Y(a)$ denotes a *potential outcome* random variable, interpreted to mean “ Y had A , possibly contrary to fact, been set to value a ” (Richardson and Robins, 2013).

The methods we discuss rely on conditional independences and counterfactual assumptions reflected in causal graphical models referenced throughout the paper. In these figures, dashed edges (\dashrightarrow) indicate causal relationships that are optional and do not contradict identifying assumptions imposed.

1.2 Adjustment Formula

Before introducing proximal methods for identification of causal targets, we first review the standard covariate adjustment formula (Equation 1).

Consider the causal graphical model in Figure 1, where confounders C between binary treatment A and outcome Y are fully observed. The counterfactual distribution $f_{Y(a)}(y)$ is identified by the adjustment formula (Equation 1) if the following hold for each $a \in \mathcal{A}$ and $c \in \mathcal{C}$:

ASSUMPTION 1 (Conditional Ignorability).

$$Y(a) \perp\!\!\!\perp A | C.$$

ASSUMPTION 2 (Consistency).

$$Y = Y(a) \text{ when } A = a.$$

ASSUMPTION 3 (Positivity).

$$f_{A|C}(a | c) > 0.$$

$$(1) \quad f_{Y(a)}(y) = \int f_{Y|A,C}(y | a, c) f_C(c) dc.$$

Consequently, the average causal effect in the case of binary A is identified by Equation 2.

$$(2) \quad \begin{aligned} \mathbb{E}[Y(1)] - \mathbb{E}[Y(0)] = \\ \int (\mathbb{E}[Y | A = 1, C = c] - \mathbb{E}[Y | A = 0, C = c]) f_C(c) dc. \end{aligned}$$

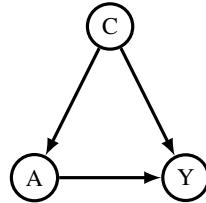


Fig 1: Observed confounding;
Supports $\mathcal{A}, \mathcal{Y}, \mathcal{C}$

In general, Equation 1 holds if C satisfies conditional ignorability assumption $Y(a) \perp\!\!\!\perp A | C$. This assumption is justified when, by the rules of d-separation, conditioning on C blocks all backdoor paths (which start with an arrowhead into A) from A to Y (Pearl, 2009; Richardson and Robins, 2013).

We provide an analogous review of the front-door adjustment formula in Appendix A.

2. PROXIMAL CAUSAL LEARNING VIA BRIDGE EQUATIONS

We now present the assumptions of the proximal causal learning approach via bridge equations, introduced by Miao, Geng and Tchetgen Tchetgen (2018) as a framework for addressing hidden confounding using confounder proxies, which enables recovery of the counterfactual distribution via the adjustment formula. Ghassami et al. (2024) develop analogous assumptions leveraging proxies of a hidden mediator in the absence of hidden confounder proxies, enabling identification causal effects via the frontdoor adjustment formula in Equation 7.

If confounders C between treatment A and outcome Y are unobserved, the counterfactual distribution $f_{Y(a)}(y)$, and consequently average causal effects, are not identifiable without further assumptions.

Consider proxies Z and W for an unmeasured confounder (now denoted U), as illustrated in the causal graphical model in Figure 2. Note had U been observed, the adjustment formula would yield the identifying expression:

$$(3) \quad f_{Y(a)}(y) = \int f_{Y|U,A}(y | u, a) f_U(u) du.$$

Miao, Geng and Tchetgen Tchetgen (2018) rewrite Equation 3 in terms of the observed law $f_{A,Y,W,Z}(a, y, w, z)$, leveraging the conditional independence restrictions listed in Assumption 4 and 5.

ASSUMPTION 4. $W \perp\!\!\!\perp Z, A | U$.

ASSUMPTION 5. $Z \perp\!\!\!\perp Y | U, A$.

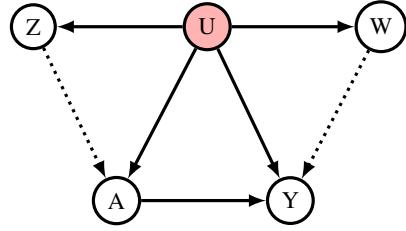


Fig 2: Miao et al. (2018);
Supports $\mathcal{A}, \mathcal{Y}, \mathcal{U}, \mathcal{W}, \mathcal{Z}$

In addition to Assumption 4 and 5, Miao, Geng and Tchetgen Tchetgen (2018) presume Assumptions 6 and 7, where $\mathbb{E}[g(u) | z, a]$ is a vector indexed by values $z \in \mathcal{Z}$. A detailed proof of identification is given in their paper. We will elaborate upon these additional two assumptions and their meanings in Sections 2.0.1 and 2.0.2 below.

ASSUMPTION 6 (Completeness). For each $a \in \mathcal{A}$ and g square-integrable,

$$\mathbb{E}[g(u) | z, a] = 0 \text{ a.s.} \iff g(u) = 0 \text{ a.s.}$$

ASSUMPTION 7 (Solution to Bridge Equation). For each $a \in \mathcal{A}$, there exists $h(w, a, y)$ such that

$$f_{Y|Z,A}(y | z, a) = \int h(w, a, y) f_{W|Z,A}(w | z, a) dw < +\infty.$$

It is important to emphasize that while this approach identifies the interventional distribution $f_{Y(a)}(y)$, it does not, in general, identify factors of the full law of hidden and observed variables—for instance, the conditional distribution $f_{Y|U,A}(y | u, a)$ or the marginal distribution $f_U(u)$ of the unobserved variable. This stands in subtle contrast to the array decomposition approach discussed in Section 3.

2.0.1 Completeness Conditions Assumption 6 generalizes the linear algebra notion of linear independence. In discrete models, Assumption 6 means that, for each $a \in \mathcal{A}$, the conditional laws $f_{Z|U,a}(z | u, a)$ span linearly independent directions across different values of U .

To illustrate this idea, suppose for the time being that the model depicted in Figure 2 is fully discrete. Define for each $a \in \mathcal{A}$,

$$\mathbf{P}_{U|Z,a} = [f_{U|Z,A}(u | z_1, a), \dots, f_{U|Z,A}(u | z_{|\mathcal{Z}|}, a)],$$

where columns are distribution vectors of the form

$$f_{U|Z,a}(u | z, a) = \begin{bmatrix} \Pr(U = u_1 | Z = z, A = a) \\ \vdots \\ \Pr(U = u_{|\mathcal{U}|} | Z = z, A = a) \end{bmatrix}.$$

Assumption 6 is equivalent to right invertibility of matrix $\mathbf{P}_{U|Z,a}$. To see this, note the expansion of Assumption 6 below:

$$(4) \quad \begin{bmatrix} \mathbb{E}[g(u) | z_1, a] \\ \mathbb{E}[g(u) | z_2, a] \\ \vdots \\ \mathbb{E}[g(u) | z_{|\mathcal{Z}|}, a] \end{bmatrix} = \underbrace{\begin{bmatrix} \Pr(u_1 | z_1, a) & \cdots & \Pr(u_{|\mathcal{U}|} | z_1, a) \\ \Pr(u_1 | z_2, a) & \cdots & \Pr(u_{|\mathcal{U}|} | z_2, a) \\ \vdots & \ddots & \vdots \\ \Pr(u_1 | z_{|\mathcal{Z}|}, a) & \cdots & \Pr(u_{|\mathcal{U}|} | z_{|\mathcal{Z}|}, a) \end{bmatrix}}_{\mathbf{P}_{U|Z,a}^T} \begin{bmatrix} g(u_1) \\ g(u_2) \\ \vdots \\ g(u_{|\mathcal{U}|}) \end{bmatrix}.$$

Right invertibility of $\mathbf{P}_{U|Z,a}$ implies left invertibility of

$$\mathbf{P}_{Z|U,a} = [f_{Z|U,A}(z | u_1, a), \dots, f_{Z|U,A}(z | u_{|\mathcal{U}|}, a)]$$

since elementary row operations preserve matrix rank. Equivalently, for each $a \in \mathcal{A}$, the conditional laws $f_{Z|U,a}(z | u, a)$ span linearly independent directions across different values of U . Notably, Z must have at least as many categories as U in order for Assumption 6 to hold, as noted in Cui et al. (2023). Assumption 6 extends this idea to continuous settings.

Assumption 6 is often satisfied under widely used parametric specifications. For example, it holds when, for each treatment level a , the conditional distribution $f_{U|Z,a}(u | z, a)$ belongs to an exponential family indexed by z , provided that z has support on an open interval of the real line (Brown, 1986; Newey and Powell, 2003). Since this setup may feel more natural when Z acts as an instrumental variable rather than a proxy, it is worth noting that no identifying assumptions are violated under the modification $Z \rightarrow U$ in Figure 2. In addition, under standard conjugate-prior relationships for exponential families, $f_{U|Z,a}(u | z, a)$ may lie in the same exponential family as $f_{Z|U,a}(z | u, a)$.

Hu and Shiu (2022) provide several characterizations of nonparametric families of complete distributions, where U is generated as an arbitrary function of Z plus independent additive noise. In general, Assumption 6 is not testable without imposing further conditions (Canay, Santos and Shaikh, 2013).

2.0.2 Solution to a Bridge Equation The integral equation in Assumption 7, referred to as a bridge equation, enables recovery of the counterfactual distribution $f_{Y(a)}(y)$ alongside prior assumptions. Since the existence of a solution is usually assumed point blank, we unpack sufficient conditions for and implications of this assumption.

Define for each $a \in \mathcal{A}$,

$$\mathbf{P}_{Y|Z,a} = [f_{Y|Z,A}(y | z_1, a), \dots, f_{Y|Z,A}(y | z_{|\mathcal{Z}|}, a)],$$

$$\mathbf{P}_{W|Z,a} = [f_{W|Z,A}(w | z_1, a), \dots, f_{W|Z,A}(w | z_{|\mathcal{Z}|}, a)].$$

In the discrete case, Assumption 7 is equivalent to the existence of a matrix \mathbf{H}_a such that

$$\mathbf{P}_{Y|Z,a} = \mathbf{H}_a \mathbf{P}_{W|Z,a}.$$

A new completeness Assumption 7a shown below—i.e. that $\mathbf{P}_{W|Z,a}$ is left invertible in a discrete model—is sufficient for Assumption 7 in the discrete setting since

$$\mathbf{P}_{Y|Z,a} = \mathbf{P}_{Y|Z,a} \mathbf{P}_{W|Z,a}^\dagger \mathbf{P}_{W|Z,a},$$

where $\mathbf{P}_{W|Z,a}^\dagger$ denotes a left inverse of $\mathbf{P}_{W|Z,a}$. This requires W to have at least as many categories as Z , and therefore at least as many as the latent variable U under Assumption 6.

In continuous models, Assumptions 7b–7d in addition to Assumption 7a are sufficient to ensure that Assumption 7 holds (Miao, Geng and Tchetgen Tchetgen, 2018).

ASSUMPTION 7a. For each $a \in \mathcal{A}$ and g square-integrable,

$$\mathbb{E}[g(z) | w, a] = 0 \text{ a.s.} \iff g(z) = 0 \text{ a.s.}$$

ASSUMPTION 7b. For each $a \in \mathcal{A}$, $\iint f_{W|Z,A}(w | z, a) f_{Z|W,A}(z | w, a) dw dz < +\infty$.

ASSUMPTION 7c. For each $a \in \mathcal{A}$,

$$\int f_{Y|Z,A}^2(y | z, a) f_{Z|A}(z | a) dz < +\infty.$$

ASSUMPTION 7d. For each $a \in \mathcal{A}$,

$$\sum_{n=1}^{\infty} \lambda_n^{-2} |\langle f_{Y|Z,A}(y | z, a), \psi_n \rangle|^2 < +\infty,$$

where $(\lambda_n, \varphi_n, \psi_n)_{n=1}^{\infty}$ is the singular value decomposition of the operator

$$\mathbf{K} : L^2\{F(w | a)\} \mapsto L^2\{F(z | a)\},$$

defined by $\mathbf{K}h = \mathbb{E}\{h(w) | z, a\}$ for $h \in L^2\{F(w | a)\}$.

Assumption 7b ensures that the operator \mathbf{K} is compact, exhibiting behavior similar to finite matrices. Assumption 7c places a square-integrability condition on $f_{Y|Z,A}(y | z, a)$, and Assumption 7d controls the behavior of $f_{Y|Z,A}(y | z, a)$ to lie in the range of \mathbf{K} . Further discussion of these analytic conditions can be found in Kress (1989).

3. CAUSAL TARGET IDENTIFICATION WITH PROXIES VIA ARRAY DECOMPOSITION

Another prominent strategy for identifying counterfactual quantities with proxies of hidden variables employs array decomposition to recover the full law of hidden and observed variables, up to the labels of hidden variable categories (e.g., Kuroki and Pearl (2014); Allman et al. (2015); Zhou and Tchetgen (2024)). We now review the assumptions of this approach in the context of the discrete hidden confounding model depicted in Figure 3, from Kuroki and Pearl (2014). In Appendix B, we provide analogous assumptions which leverage proxies of a hidden mediator in the absence of hidden confounder proxies, enabling identification causal effects via the frontdoor adjustment formula in Equation 7.

In the model depicted in Figure 3, taken to be fully discrete for now, where Z and W function as proxy variables for the unobserved confounder U , Kuroki and Pearl (2014) show that the latent components $f_{Y|U,A}(y | u_i, a)$ and $f_U(u_i)$ are identifiable jointly up to label of the latent categories of U . With these components identified, Kuroki and Pearl (2014) recover $f_{Y(a)}(y)$ from the adjustment formula in Equation 1. We focus on the fully discrete version of the model for the time being, deferring discussion of continuous models to a later point.

Kuroki and Pearl (2014) employ the following assumptions:

Define for each $a \in \mathcal{A}$,

$$\mathbf{P}_{Z|U,a} = [f_{Z|U,A}(z | u_1, a), \dots, f_{Z|U,A}(z | u_{|\mathcal{U}|}, a)],$$

$$\mathbf{P}_{W|U} = [f_{W|U}(w | u_1), \dots, f_{W|U}(w | u_{|\mathcal{U}|})],$$

$$\mathbf{P}_{Y|U,a} = [f_{Y|U,A}(y | u_1, a), \dots, f_{Y|U,A}(y | u_{|\mathcal{U}|}, a)].$$

ASSUMPTION 8. $W \perp\!\!\!\perp A | U$.

ASSUMPTION 9.

W, Z, Y , are mutually independent $| U, A$.

ASSUMPTION 10. For each $a \in \mathcal{A}$, the matrices $\mathbf{P}_{Z|U,a}$ and $\mathbf{P}_{W|U}$ are invertible.

ASSUMPTION 11. For each $a \in \mathcal{A}$, some row of $\mathbf{P}_{Y|U,a}$ has distinct entries.

A detailed proof of identification is given in Kuroki and Pearl (2014). For Assumption 10 to be satisfied, Z and W must have the same number of levels. Assumption 10 can be relaxed to left invertibility of $\mathbf{P}_{Z|U,a}$ and $\mathbf{P}_{W|U}$, allowing cases where Z and W each have at least as many levels as U . Perhaps surprisingly, Assumption 10 can even be relaxed to allow settings where both proxies Z and W have fewer levels than the latent variable. These generalizations are discussed in Section 3.1.

Techniques for identifying causal targets via array decomposition, presented in works including Kuroki and

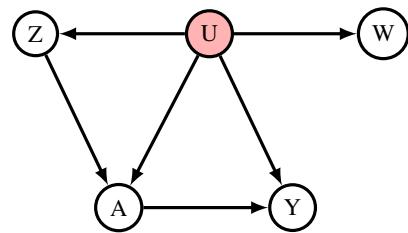


Fig 3: Kuroki and Pearl (2014);
Supports $\mathcal{A}, \mathcal{Y}, \mathcal{U}, \mathcal{W}, \mathcal{Z}$

Pearl (2014), Allman et al. (2015), and Zhou and Tchetgen (2024) can be viewed as special cases of Kruskal's uniqueness theorem for the Candecomp/Parafac (CP) decomposition (Kruskal, 1977), which we discuss in Section 3.1. As a consequence, any factor of the full law of hidden and observed variables can be recovered in these settings up to latent category labels. A natural extension of these ideas to continuous settings is explored in Section 3.2.

3.1 Discrete Hidden Variable Models

The array decomposition approach in discrete models is a special case of Kruskal's uniqueness theorem for the Candecomp/Parafac (CP) decomposition of a three-way array (or data tensor, informally), as noted in Allman et al. (2015). The theorem provides conditions under which any factor of the full law of hidden and observed variables is identified up to latent state labels, and rules out any lower-dimensional representation of the latent variable (Kruskal, 1977).

The remainder of this section presents the theorem and examines in detail how its assumptions relate to those of Kuroki and Pearl (2014). A detailed and concise proof of the theorem is provided in Rhodes (2010).

DEFINITION 3.1 (CP Decomposition). The CP decomposition of a three-way array \mathcal{T} with dimensions $M \times N \times J$ consists of three factor matrices $\mathbf{A} \in \mathbb{R}^{M \times R}$, $\mathbf{B} \in \mathbb{R}^{N \times R}$, and $\mathbf{C} \in \mathbb{R}^{J \times R}$ such that the j^{th} slice of \mathcal{T} can be written as (Stegeman and Sidiropoulos, 2007; Kruskal, 1977):

$$(5) \quad \mathcal{T}_j = \mathbf{A}\mathbf{C}_j\mathbf{B}^T, \text{ where } \mathbf{C}_j \text{ is diagonal} \\ \text{with its diagonal given by the } j^{\text{th}} \text{ row of } \mathbf{C}.$$

DEFINITION 3.2 (Essential Uniqueness). Three-way array \mathcal{T} is essentially unique if its CP decomposition is uniquely determined up to permutation and scaling of the columns of \mathbf{A} , \mathbf{B} , and \mathbf{C} .

DEFINITION 3.3 (Three-Way Rank). The three-way rank of \mathcal{T} is the smallest integer R for which its CP decomposition in equation (5) holds.

DEFINITION 3.4 (Kruskal Rank (k -rank)). The Kruskal rank, or k -rank, of a matrix is the largest integer k such that any subset of k columns is linearly independent.

THEOREM 3.5 (Kruskal's Uniqueness Theorem). A three-way array \mathcal{T} is essentially unique and has three-way rank R if the following condition holds:

$$(6) \quad 2R + 2 \leq k_A + k_B + k_C,$$

where k_A , k_B , and k_C are the k -ranks of \mathbf{A} , \mathbf{B} , and \mathbf{C} , respectively.

Under the approach of Kuroki and Pearl (2014) described at the beginning of Section 3, let

$$\mathbf{A} = \mathbf{P}_{W|U}, \quad \mathbf{B} = \mathbf{P}_{Z,U|a} = \mathbf{P}_{Z|U,a} \text{ diag}(f_{U|a}), \\ \text{and } \mathbf{C} = \mathbf{P}_{Y|U,a},$$

in Definition 3.3, so that $\mathcal{T}_j = \mathbf{A}\mathbf{C}_j\mathbf{B}^T = \mathbf{P}_{y_j,W,Z|a} = [f_{Y,W,Z|A}(y_j, w, z_1 | a), \dots, f_{Y,W,Z|A}(y_j, w, z_{|\mathcal{Z}|} | a)]$.

By Assumption 10, both \mathbf{A} and \mathbf{B} are invertible, which implies $k_A = k_B = R = |\mathcal{U}|$. Furthermore, Assumption 11 ensures that $k_C \geq 2$, since the columns of \mathbf{C} cannot be collinear; the conditional distributions $f_{Y|U,A}(y | u, a)$ differ across values of latent variable U . Hence $2R + 2 \leq k_A + k_B + k_C$.

By Theorem 3.5, identification of latent factors

$f_{W|U,A}(w | u_i, a)$, $f_{Y|U,A}(y | u_i, a)$, and $f_{U,Z|A}(u_i, z | a)$ reduces to solving the decomposition for each $a \in \mathcal{A}$. Further by $W \perp U | A$, latent factors for label i recovered across values of $a \in \mathcal{A}$ can be aligned. Hence, the full law $f_{A,Y,U,W,Z}(a, y, u_i, w, z)$ is identified up to labels i .

The original proof of Theorem 3.5 is nonconstructive, and can be viewed in terms of uniquely determined eigenspaces (Rhodes, 2010). In the special setting considered by Kuroki and Pearl (2014), the authors showed latent components can be obtained directly via eigendecomposition tasks. Indeed, when $k_A = k_B = R = |\mathcal{U}|$ and $k_C \geq 2$, latent components can be constructed from applying the argument from Kuroki and Pearl (2014) to submatrices of \mathcal{T}_j , as also noted by Allman et al. (2015). Modern methods for recovering latent factors under the general assumptions of Theorem 3.5 often rely on iterative procedures such as alternating least squares or gradient-based optimization (Uschmajew, 2012; Kolda and Hong, 2020).

Notably, Theorem 3.5 leverages constraints across all slices of the three-way array \mathcal{T} . Since the inequality treats W , Z , and Y symmetrically, variation in the conditional laws $f_{Y|U,A}(y | u_i, a)$ across levels of U can offset limited variation in laws $f_{W|U}(w | u_i)$ or $f_{Z|U,A}(z | u_i, a)$. In this manner, the inequality in Theorem 3.5 allows for situations where Z and/or W have fewer levels than U , as long as Y has enough levels. A necessary requirement for the inequality in Theorem 3.5 to hold is that variables W , Z , and Y collectively have at least $2|\mathcal{U}| + 2$ distinct categories.

Note the CP decomposition framework is typically introduced in the simpler case of a single latent variable with three proxy variables that are mutually independent conditional upon the latent, depicted in Figure 4. The resulting three-way array formed by the proxy variables uniquely determines the full law up to latent labels, given the condition in Theorem 3.5 holds (Kruskal, 1977). We now turn to how this framework generalizes for continuous models.

3.2 Continuous Hidden Variable Models

For continuous hidden variable models, [Hu and Schennach \(2008\)](#) provide a natural extension of the array decomposition framework for unique determination of the full law up to latent state labels. A key requirement is Assumption 12, which states that three observed variables are mutually independent conditional on the latent variable L , as illustrated in Figure 4. [Hu and Schennach \(2008\)](#) also invoke the bounded density conditions in Assumption 13.

ASSUMPTION 12.

$$W, Z, Y \text{ are mutually independent } | L.$$

ASSUMPTION 13. W, Z, Y , and L admit bounded, nonzero densities with respect to the Lebesgue measure on $\mathcal{W} \times \mathcal{Z} \times \mathcal{L}$, and with respect to some dominating measure μ on \mathcal{Y} . All marginal and conditional densities are assumed to be bounded as well.

Additionally, the following two assumptions are imposed, paralleling Assumptions 10 and 11 in [Kuroki and Pearl \(2014\)](#), or the inequality required by Theorem 3.5 from [Kruskal \(1977\)](#).

ASSUMPTION 14. For g square-integrable, $\mathbb{E}[g(l) | w] = 0$ a.s. $\iff g(l) = 0$ a.s. and $\mathbb{E}[g(l) | z] = 0$ a.s. $\iff g(l) = 0$ a.s.

ASSUMPTION 15. For all $i \neq j$, the conditional distributions $f_{Y|L}(y | l_i)$ and $f_{Y|L}(y | l_j)$ differ with positive probability under the marginal distribution of Y .

[Hu and Schennach \(2008\)](#) prove that the full law $f_{Y,W,Z,U}(y, w, z, l)$ is uniquely determined up to labeling of latent states [Assumptions 19 or 20 give the labelings]. For ease of comparison, we restate Assumptions 13–15 in a form that ensures unique identification of the counterfactual distribution $f_{Y(a)}(y)$ in the setting depicted in Figure 3 from [Kuroki and Pearl \(2014\)](#), instead allowing variables to be continuous. This was proven by [Deaner \(2023\)](#).

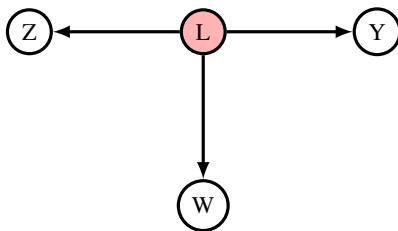


Fig 4: Latent Variable with Proxies;
Supports $\mathcal{Y}, \mathcal{W}, \mathcal{Z}, \mathcal{L}$

ASSUMPTION 16. $W, Z, Y, U | A = a$ have bounded, nonzero densities with respect to the Lebesgue measure on $\mathcal{W} \times \mathcal{Z} \times \mathcal{U}$, and with respect to some dominating measure μ on \mathcal{Y} . All marginal and conditional factors of $f_{W,Z,Y,U|a}(w, z, y, u | a)$ are bounded as well.

ASSUMPTION 17. For each $a \in \mathcal{A}$ and g square-integrable, $\mathbb{E}[g(u) | w, a] = 0$ a.s. $\iff g(u) = 0$ a.s. and $\mathbb{E}[g(u) | z, a] = 0$ a.s. $\iff g(u) = 0$ a.s.,

ASSUMPTION 18. For each $a \in \mathcal{A}$ and $i \neq j$, the conditional distributions $f_{Y|U,A}(y | u_i, a)$ and $f_{Y|U,A}(y | u_j, a)$ differ with positive probability under the marginal distribution of $Y | A = a$.

Assumption 17 is the continuous analogue of Assumption 10 from [Kuroki and Pearl \(2014\)](#), or more generally $k_A = k_B = R = |\mathcal{U}|$, while Assumption 18 corresponds to Assumption 11 from [Kuroki and Pearl \(2014\)](#), or more generally $k_C \geq 2$. Since Theorem 3.5 permits situations where both Z and W have fewer categories than the latent variable U in discrete models, whether the result from [Hu and Schennach \(2008\)](#) admits a similar relaxation remains an open question.

3.3 Identifying Latent Labels

Outside of the hidden confounding context, it is often necessary to impose additional conditions in order to recover the labels of the latent variable. When the treatment is unobserved but proxies of it are available, label recovery for A is essential for identifying a specific counterfactual distribution $f_{Y(a_i)}(y) = \int f_{Y|A,C}(y | a_i, c) f_C(c) dc$ ([Zhou and Tchetgen, 2024](#)).

3.3.1 Unbiasedness Conditions for Label Recovery
[Hu and Schennach \(2008\)](#) show that in the model depicted in Figure 4 under Assumptions 12, 14, and 17, latent state labels can be recovered under an unbiasedness condition, stating that a measure of central tendency of a proxy, conditional on each latent value, must equal the latent value itself. This condition is relayed in Assumption 19.

ASSUMPTION 19. L is ordinal and there exists a known functional M such that for some $V \in \{W, Z, Y\}$ and for each $l \in \mathcal{L}$, $M[f_{V|L}(v | l)] = l$.

3.3.2 Monotonicity Conditions for Label Recovery As an alternative to Assumption 19, labels may also be recovered under a monotonicity condition relayed in Assumption 20, stating that a measure of central tendency of a proxy is strictly ordered across latent categories.

ASSUMPTION 20. L is ordinal and there exists a known functional M such that, for some $V \in \{W, Z, Y\}$ and all $l_i < l_j$, either $M[f_{V|L}(v | l_i)] < M[f_{V|L}(v | l_j)]$ or $M[f_{V|L}(v | l_i)] > M[f_{V|L}(v | l_j)]$.

4. MODEL COMPARISON

The array decomposition approach in Section 3 can be used to recover counterfactual quantities across a broad class of hidden variable models. For example, Zhou and Tchetgen (2024) apply the method to hidden treatment models. This flexibility underscores a key strength of the array decomposition framework. Moreover, the two proximal strategies for hidden confounding specifically invoke distinct Markov restrictions, which are non-nested, allowing the eigendecompositon approach to swap the roles of Y and A (Deaner, 2023).

We now compare the two approaches under a unified set of Markov restrictions compatible with both frameworks. Specifically, we consider the hidden confounding model that is Markov to Figure 3, i.e., reflecting the conditional independences in Assumptions 8 and 9.

4.0.1 Discrete Model Comparison For discrete models, we compare Assumption 6 and 7 from the bridge equation approach in Miao et al. (2018) with two alternative assumption sets. Set 1 is from Kuroki and Pearl (2014); Set 2 is from Kruskal (1977), the generalization of Kuroki and Pearl (2014) via Theorem 3.5.

Comparing Set 1, all model overlap occurs under the assumptions of Kuroki and Pearl (2014). In particular, the invertibility conditions in Assumption 10 from Kuroki and Pearl (2014)—namely, that $\mathbf{P}_{Z|U,a}^{-1}$ and $\mathbf{P}_{W|U,a}^{-1}$ exist—also imply the existence of $\mathbf{P}_{W|Z,a}^{-1}$. This directly satisfies Assumptions 6 and 7 in Miao, Geng and Tchetgen Tchetgen (2018), since one can define

$$\mathbf{H}_a = \mathbf{P}_{Y|Z,a} \mathbf{P}_{W|Z,a}^{-1} \quad \text{so that} \quad \mathbf{P}_{Y|Z,a} = \mathbf{H}_a \mathbf{P}_{W|Z,a}.$$

Thus, the model from Kuroki and Pearl (2014) is contained within that of Miao, Geng and Tchetgen Tchetgen (2018).

To compare Set 2, again let

$$\mathbf{A} = \mathbf{P}_{W|U}, \quad \mathbf{B} = \mathbf{P}_{Z,U|a}, \quad \text{and} \quad \mathbf{C} = \mathbf{P}_{Y|U,a},$$

in Definition 3.3, so that $\mathcal{T}_j = \mathbf{A} \mathbf{C}_j \mathbf{B}^\top = \mathbf{P}_{y_j, W, Z|a} = [f_{Y,W,Z|A}(y_j, w, z_1 | a), \dots, f_{Y,W,Z|A}(y_j, w, z_{|\mathcal{Z}|} | a)]$. Define k_A , k_B , and k_C as the k -ranks of \mathbf{A} , \mathbf{B} , and \mathbf{C} , respectively.

Assumptions 6 and 7 from Miao, Geng and Tchetgen Tchetgen (2018) do not imply that $2|\mathcal{U}| + 2 \leq k_A + k_B + k_C$, nor does the converse hold. In other words, the models are non-nested.

4.0.2 Continuous Model Comparison Table 2 reviews the assumptions of Miao, Geng and Tchetgen Tchetgen (2018) alongside those of Hu and Schennach (2008) for continuous models. We conjecture that the two models are non-nested in the model Markov to Figure 3. We give an example satisfying the assumptions of both approaches, which appears in the latter half of Kuroki and Pearl (2014) and is revisited in Miao, Geng and Tchetgen Tchetgen (2018).

5. A BRIEF NOTE ON ESTIMATION

While our primary focus in this paper is on comparing model assumptions for identification, it is worth noting that Miao, Geng and Tchetgen Tchetgen (2018) provide an illustrative example in which $f_{Y(a)}(y)$ can be derived analytically under location-scale family assumptions, yielding a tractable framework for consistent estimation of causal effects. In their work, the efficient influence function has also been established under further model assumptions, giving rise to regular estimation of causal effects achieving \sqrt{n} -consistency and semiparametric efficiency (Cui et al., 2023). In contrast, Hu and Schennach (2008) rely on sieve methods to approximate factors of the full law. Some semiparametric theory has also been developed for estimating causal effects recovered from the array decomposition approach in the context of proxies for hidden treatment, though this literature remains relatively nascent (Zhou and Tchetgen, 2024).

6. CONCLUSION AND FUTURE DIRECTIONS

Proxy-based approaches have emerged as a promising solution to identifying causal targets when hidden variables pose an obstacle. Within this framework, two complementary strategies stand out. The bridge equation method identifies causal targets by solving functional equations, while the array decomposition approach leverages unique determination of eigenspaces to uncover latent structures which constitute those targets. Although both methods hinge on assumptions regarding the informativeness of proxy variables for the latent structure, comparing their respective assumptions and areas of overlap offers clearer insight into the contexts in which each method is most appropriately applied.

Looking ahead, several directions for future research present themselves. One avenue is to explore how proximal methods can extend to more complicated scenarios that are generally non-identifiable. A further direction is to establish general semiparametric theory and influence function-based estimators for causal effects identified through these proximal approaches. Such advances would expand the scope of settings in which causal targets are identifiable and provide robust tools for researchers working with complex data to estimate them.

APPENDIX A: FRONTDOOR ADJUSTMENT FORMULA

Consider the hidden variable model in Figure 5, where confounders U between the treatment A and outcome Y are unobserved, however a mediator M between A and Y is observed. Here, the counterfactual distribution $f_{Y(a)}(y)$ is identified by the frontdoor formula in Equation 7, provided the following conditions hold for each $a \in \mathcal{A}$ and $m \in \mathcal{M}$:

TABLE 1
Comparison for discrete hidden confounding model Markov to Figure 3.*

Bridge Equation Approach [†]	Array Decomposition Approach [‡]	Overlap
Assumption Set 1		
<ul style="list-style-type: none"> – $\mathbf{P}_{Z U,a}$ is left invertible – $\exists \mathbf{H}_a$ s.t. $\mathbf{P}_{Y Z,a} = \mathbf{H}_a \mathbf{P}_{W Z,a}$ 	<ul style="list-style-type: none"> – $\mathbf{P}_{Z U,a}^{-1}$ and $\mathbf{P}_{W U}^{-1}$ exist – Some row of $\mathbf{P}_{Y U,a}$ has distinct entries 	<ul style="list-style-type: none"> – $\mathbf{P}_{Z U,a}^{-1}$ and $\mathbf{P}_{W U}^{-1}$ exist – Some row of $\mathbf{P}_{Y U,a}$ has distinct entries
Assumption Set 2		
<ul style="list-style-type: none"> – $\mathbf{P}_{Z U,a}$ is left invertible – $\exists \mathbf{H}_a$ s.t. $\mathbf{P}_{Y Z,a} = \mathbf{H}_a \mathbf{P}_{W Z,a}$ 	<ul style="list-style-type: none"> – $2 \mathcal{U} + 2 \leq k_A + k_B + k_C$ 	<ul style="list-style-type: none"> – $\mathbf{P}_{Z U,a}$ is left invertible $\implies k_B = \mathcal{U}$ – $\exists \mathbf{H}_a$ s.t. $\mathbf{P}_{Y Z,a} = \mathbf{H}_a \mathbf{P}_{W Z,a}$ – $\mathcal{U} + 2 \leq k_A + k_C$
Identified Targets		
<ul style="list-style-type: none"> – $f_{Y(a)}(y)$ 	<ul style="list-style-type: none"> – $f_{A,Y,U,W,Z}(a, y, u, w, z)$ up to latent labels – $f_{Y(a)}(y)$ 	<ul style="list-style-type: none"> – $f_{A,Y,U,W,Z}(a, y, u, w, z)$ up to latent labels – $f_{Y(a)}(y)$

* All assumptions are stated for each $a \in \mathcal{A}$.

† Bridge equation approach assumptions are from [Miao, Geng and Tchetgen Tchetgen \(2018\)](#).

‡ Array decomposition approach assumptions are as follows: Set 1 from [Kuroki and Pearl \(2014\)](#); Set 2 from [Kruskal \(1977\)](#).

ASSUMPTION 21 (Counterfactual Independences).

$$M(a) \perp\!\!\!\perp A \text{ and } Y(m) \perp\!\!\!\perp M | A.$$

ASSUMPTION 22 (Exclusion restriction).

$$Y(a, m) = Y(m).$$

ASSUMPTION 23 (Consistency).

$$M = M(a) \text{ when } A = a, \text{ and}$$

$$Y = Y(a, m) \text{ when } A = a, M = m.$$

ASSUMPTION 24 (Positivity).

$$f_{M|A}(m | a) > 0 \text{ and } f_A(a) > 0.$$

(7)

$$f_{Y(a)}(y) = \iint f_{Y|A,M}(y | a', m) f_{M|A}(m | a) f_A(a') da' dm.$$

APPENDIX B: MEDIATOR PROXY ANALOGUE FOR KUROKI AND PEARL (2018)

We provide analogous assumptions to Assumption 8-11 in [Kuroki and Pearl \(2014\)](#), instead leveraging proxies of a hidden mediator depicted in Figure 6.

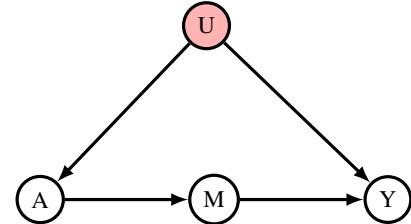


Fig 5: Hidden confounding with observed mediation;
Supports $\mathcal{A}, \mathcal{M}, \mathcal{Y}, \mathcal{U}$

Define for each $a \in \mathcal{A}$,

$$\mathbf{P}_{Z|M,a} = [f_{Z|M,A}(z | m_1, a), \dots, f_{Z|M,A}(z | m_{|\mathcal{M}|}, a)],$$

$$\mathbf{P}_{W|M} = [f_{W|M}(w | m_1), \dots, f_{W|M}(w | m_{|\mathcal{M}|})],$$

$$\mathbf{P}_{Y|M,a} = [f_{Y|M,A}(y | m_1, a), \dots, f_{Y|M,A}(y | m_{|\mathcal{M}|}, a)].$$

ASSUMPTION 25. $W \perp\!\!\!\perp A | M$.

ASSUMPTION 26.

W, Z, Y , are mutually independent $| M, A$.

ASSUMPTION 27. For each $a \in \mathcal{A}$, the matrices $\mathbf{P}_{Z|M,a}$ and $\mathbf{P}_{W|M}$ are invertible.

TABLE 2
Comparison for continuous hidden confounding model Markov to Figure 3.*

Bridge Equation Approach [†]	Array Decomposition Approach [‡]	Overlap Example
Assumptions		
<ul style="list-style-type: none"> Completeness: For g square-integrable, $\mathbb{E}[g(u) z, a] = 0$ a.s. $\iff g(u) = 0$ a.s. Bridge equation: $\exists h(w, a, y)$ s.t. $f_{Y Z,A}(y z, a) = \int h(w, a, y) f_{W Z,A}(w z, a) dw < \infty$ 	<ul style="list-style-type: none"> Bounded densities: $W, Z, Y, U A = a$ admit bounded, nonzero densities wrt Lebesgue measure on $\mathcal{W} \times \mathcal{Z} \times \mathcal{U}$ and a dominating measure μ on \mathcal{Y}; all marginals and conditionals are bounded Completeness: For g square-integrable, $\mathbb{E}[g(u) z, a] = 0$ a.s. $\iff g(u) = 0$ a.s. and $\mathbb{E}[g(u) w, a] = 0$ a.s. $\iff g(u) = 0$ a.s. Non-collinearity: $f_{Y U,A}(y u_i, a) \neq f_{Y U,A}(y u_j, a)$ with positive probability under $Y A = a, i \neq j$ 	$U = \mu_U + \varepsilon_U,$ $Z = \beta_0 Z + \alpha_{UZ} U + \varepsilon_Z,$ $A = \beta_0 A + \alpha_{UA} U + \alpha_{ZA} Z + \varepsilon_A,$ $W = \beta_0 W + \alpha_{UW} U + \varepsilon_W,$ $Y = \beta_0 Y + \alpha_{AY} A + \alpha_{UY} U + \varepsilon_Y,$ with all disturbances mean-zero Gaussian and mutually independent
Identified Targets		
<ul style="list-style-type: none"> $f_{Y(a)}(y)$ 	<ul style="list-style-type: none"> $f_{A,Y,U,W,Z}(a, y, u, w, z)$ up to latent labels $f_{Y(a)}(y)$ 	<ul style="list-style-type: none"> $f_{A,Y,U,W,Z}(a, y, u, w, z)$ up to latent labels $f_{Y(a)}(y)$

* All assumptions are stated for each $a \in \mathcal{A}$.

† Bridge equation approach assumptions are from Miao, Geng and Tchetgen Tchetgen (2018).

‡ Array decomposition approach assumptions are from Hu and Schennach (2008), from which Deaner (2023) shows $f_{Y(a)}(y)$ is identified.

ASSUMPTION 28. For each $a \in \mathcal{A}$, some row of $\mathbf{P}_{Y|M,a}$ has distinct entries.

We prove $f_{Y(a)}(y)$ is identified, relying on Lemma B.1.

LEMMA B.1. Let $\mathbf{M} \in \mathbb{R}^{n \times n}$ be an invertible square matrix. Suppose \mathbf{M} admits the eigendecomposition

$$\mathbf{M} = \mathbf{V}\Lambda\mathbf{V}^{-1},$$

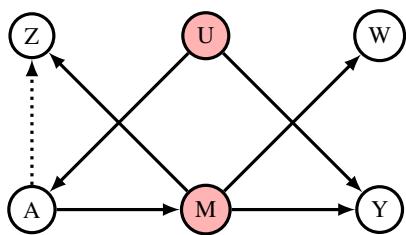


Fig 6: Analogue of Kuroki and Pearl (2014);
Supports $\mathcal{A}, \mathcal{M}, \mathcal{Y}, \mathcal{U}, \mathcal{W}, \mathcal{Z}$

where $\mathbf{V} \in \mathbb{R}^{n \times n}$ is the matrix whose columns are the eigenvectors of \mathbf{M} , and $\Lambda \in \mathbb{R}^{n \times n}$ is the diagonal matrix of the corresponding eigenvalues. Then the eigendecomposition of \mathbf{M} is uniquely determined up to rescaling of the eigenvectors if and only if the eigenvalues of \mathbf{M} (i.e., the diagonal entries of Λ) are distinct.

PROOF. By Assumption 25 and Assumption 26, for each $a \in \mathcal{A}$ and $y \in \mathcal{Y}$,

$$\mathbf{P}_{W|Z,a} = \mathbf{P}_{W|M}\mathbf{P}_{M|Z,a}, \text{ and}$$

$$\mathbf{P}_{y,W|Z,a} = \mathbf{P}_{W|M} \text{diag}(\mathbf{P}_{y|M,a}) \mathbf{P}_{M|Z,a}.$$

By Assumption 27, $\mathbf{P}_{M,Z|a}^{-1}$ exists and

$$\mathbf{P}_{W|Z,a}^{-1} = \mathbf{P}_{M|Z,a}^{-1} \mathbf{P}_{W|M}^{-1}.$$

Hence, $\mathbf{P}_{y,W|Z,a} \mathbf{P}_{W|Z,a}^{-1} = \mathbf{P}_{W|M} \text{diag}(\mathbf{P}_{y|M,a}) \mathbf{P}_{W|M}^{-1}$.

Lemma B.1 allows us to identify the columns of $\mathbf{P}_{W|M}$ under Assumptions 27 and 28. In other words, we recover conditional distributions $f_{W|M}(w | m_i)$ up to labels i . Inverting the integral equation $f_{Y,W|A}(y, w | a) =$

$\sum_i f_{Y,M|A}(y, m_i | a) f_{W|M}(w | m_i)$, we can solve for $f_{Y,M|A}(y, m_i | a)$ for each $a \in \mathcal{A}$ up to labeling. Further by $W \perp M | A$, we can align latent factors for label i across values of $a \in \mathcal{A}$, enabling reconstruction of the frontdoor adjustment formula in Equation 7. \square

ACKNOWLEDGMENTS

The authors are grateful to Eric Tchetgen Tchetgen, Trung Phung, and Beatrix Wen for helpful discussions.

FUNDING

The authors were supported by ONR Grant N000142412701.

REFERENCES

- ALLMAN, E., MATIAS, C. and RHODES, J. (2009). Identifiability of Parameters in Latent Structure Models with Many Observed Variables. *The Annals of Statistics* **37** 3099–3132.
- ALLMAN, E., RHODES, J., STANGHELLINI, E. and VALTORTA, M. (2015). Parameter Identifiability of Discrete Bayesian Networks with Hidden Variables. *Journal of Causal Inference* **3** 189–205. <https://doi.org/10.1515/jci-2014-0021>
- BROWN, L. D. (1986). *Fundamentals of Statistical Exponential Families. Lecture Notes–Monograph Series*. Institute of Mathematical Statistics, Hayward, CA.
- CANAY, I. A., SANTOS, A. and SHAIKH, A. M. (2013). On the Testability of Identification in Some Nonparametric Models With Endogeneity. *Econometrica* **81** 2535–2559. <https://doi.org/10.3982/ECTA10851>
- CUI, Y., PU, H., SHI, X., MIAO, W. and TCETGEN TCETGEN, E. (2023). Semiparametric proximal causal inference. *Journal of the American Statistical Association*. <https://doi.org/10.1080/01621459.2023.2191817>
- DEANER, B. (2023). Controlling for Latent Confounding with Triple Proxies.
- GHASSAMI, A., YANG, A., SHPITSER, I. and TCETGEN TCETGEN, E. (2024). Causal Inference with Hidden Mediators. *Biometrika*.
- HU, Y. and SCHENNACH, S. (2008). Instrumental Variable Treatment of Nonclassical Measurement Error Models. *Econometrica* **76** 195–216. <https://doi.org/10.1111/j.0012-9682.2008.00823.x>
- HU, Y. and SHIU, J.-L. (2022). A simple test of completeness in a class of nonparametric specification. *Econometric Reviews* **41** 373–399. <https://doi.org/10.1080/07474938.2021.1957285>
- KOLDA, T. G. and HONG, D. (2020). Stochastic Gradients for Large-Scale Tensor Decomposition. *SIAM Journal on Mathematics of Data Science* **2** 1066–1095. <https://doi.org/10.1137/19M1266265>
- KRESS, R. (1989). *Linear Integral Equations*. Springer, Berlin.
- KRUSKAL, J. (1977). Three-way arrays: rank and uniqueness of trilinear decompositions, with applications to arithmetic complexity and statistics. *Linear Algebra and its Applications* **18** 95–138. [https://doi.org/10.1016/0024-3795\(77\)90069-6](https://doi.org/10.1016/0024-3795(77)90069-6)
- KUROKI, M. and PEARL, J. (2014). Measurement bias and effect restoration in causal inference. *Biometrika* **101** 423–437.
- MIAO, W., GENG, Z. and TCETGEN TCETGEN, E. J. (2018). Identifying causal effects with proxy variables of an unmeasured confounder. *Biometrika* **105** 987–993.
- NEWHEY, W. K. and POWELL, J. L. (2003). Instrumental variable estimation of nonparametric models. *Econometrica* **71** 1565–1578.
- PEARL, J. (1988). *Probabilistic Reasoning in Intelligent Systems*. Morgan and Kaufmann, San Mateo.
- PEARL, J. (2009). *Causality: Models, Reasoning, and Inference*, 2 ed. Cambridge University Press.
- RHODES, J. A. (2010). A concise proof of Kruskal's theorem on tensor decomposition. *Linear Algebra and its Applications* **432** 1818–1824. <https://doi.org/10.1016/j.laa.2009.11.033>
- RICHARDSON, T. S. and ROBINS, J. M. (2013). Single World Intervention Graphs (SWIGs): A Unification of the Counterfactual and Graphical Approaches to Causality. *preprint: http://www.css.washington.edu/Papers/wp128.pdf*.
- SPIRTES, P., GLYmour, C. and SCHEINES, R. (2001). *Causation, Prediction, and Search*, 2 ed. Springer Verlag, New York.
- STEGEMAN, A. and SIDIROPOULOS, N. (2007). On Kruskal's uniqueness condition for the Candecomp/Parafac decomposition. *Linear Algebra and its Applications* **420** 540–552. <https://doi.org/10.1016/j.laa.2006.08.010>
- USCHMAJEW, A. (2012). Local Convergence of the Alternating Least Squares Algorithm for Canonical Tensor Approximation. *SIAM Journal on Matrix Analysis and Applications* **33** 639–652. <https://doi.org/10.1137/110843587>
- ZHOU, Y. and TCETGEN, E. T. (2024). Causal Inference for a Hidden Treatment.