Beware of Hinges in Proximal Variables Regression: Adjusting for Colliding-Mediators with Nuisance PV

Hubert Drazkowski HUBERT.DRAZKOWSKI@DI.KU.DK
University of Copenhagen

Editors: Bijan Mazaheri and Niels Richard Hansen

Abstract

Conditional proximal variables regression (PV) enables estimation of causal effects in the presence of unobserved confounding by using treatment-side and outcome-side proxies that are conditionally separated given observed covariates. We study a structural violation of this separation caused by hinges—observed variables that act simultaneously as mediators and colliders between the two proxy blocks—so that no single conditioning choice blocks all bias-inducing paths. Ignoring a hinge can introduce mediator leakage into proximal moment equations, while conditioning on it can activate collider paths and induce cross-side noise endogeneity. We formalize hinges in a structural causal model for continuous, possibly multivariate treatments and proxies, and develop complementary identification results via a outcome bridge and a novel regression based treatment bridge. We express identification and uniqueness through transparent moment equations and rank requirements on conditional covariance operators, enabling direct comparison with instrumental variables regression. Finally, we propose Nuisance Proximal Variables (NPV): a hinge-robust correction that augments the moment system with a nuisance block and identifies the target effect after projecting out the resulting nuisance span. Synthetic experiments illustrate the two hinge failure modes and show that NPV recovers the causal effect when nuisance-span conditions hold.

Keywords: Proximal causal inference, Proximal variables regression, Unobserved confounding, Causal identification, Graphical models

1. Introduction

Proximal variables regression (PV) estimates causal effects from observational data in the presence of unobserved confounding (e.g., Miao et al., 2024; Tchetgen et al., 2020; Cui et al., 2023). PV leverages two noisy—yet conditionally independent—proxy blocks for the latent confounder: a treatment-side block and an outcome-side block. Typically, PV rests on three assumptions: (i) each proxy block is informative about the confounder (*relevance*); (ii) the two proxy blocks satisfy *cross-side exclusion* after conditioning on observed covariates; and (iii) there exists a *bridge* that uses one proxy block to represent (and thus remove) the confounding component revealed by the other (see Kallus et al., 2021; Guo et al., 2025, for a full discussion on the assumptions). But what if the proxies cannot be conditionally separated in the first place?

We argue that a common driver of such failures can be the presence of *hinges*. A hinge is an observed variable that simultaneously behaves as (a) a *mediator* transmitting information across the treatment- and outcome-side proxy blocks and (b) a *collider* that can be activated by conditioning. In this collider–mediator configuration, no single conditioning choice yields valid cross-side exclusions: if one ignores it, mediator pathways can induce cross-side dependence; if one conditions on it to block mediation, collider activation can induce dependence through otherwise blocked paths. In PV this dependence can be especially damaging because the bridge residual generally contains proxy noise, so collider activation can manifest as *cross-side noise endogeneity* in the proximal moment equations. This makes hinges qualitatively different from generic proxy-violation discussions based on partial identification or sensitivity/bounds (Ghassami et al., 2023; Zhang and Su, 2024), and motivates dedicated identification tools.

Hinges connect to, but differ from, "collider-mediator" phenomena studied in instrumental variables (IV) settings (Thams et al., 2024). IV requires a single exclusion and treats the instrument as exogenous, whereas proximal identification relies on two cross-side exclusions and proxies that are deliberately informative about the latent confounder. In this perspective, our goal is therefore not to extend IV assumptions, but to characterize when proximal assumptions fail structurally and how to restore identification in a hinge-aware way. Additional related work is discussed in Appendix A.

Examples Figure 1 illustrates examples of hinge patterns together with representative interpretations. Throughout, $\mathcal{U}$ is an unobserved confounder, $\mathcal{X}$ is the treatment of interest (not assumed binary), Y is the outcome, $\mathcal{Z}$ and $\mathcal{W}$ are treatment- and outcome-side proxies, and $\mathcal{H}$ is the hinge. Additional canonical hinge patterns appear in the Appendix C.1.
(a) *Emergency care (triage).* Let $\mathcal{U}$ capture latent frailty/comorbidities/health-seeking, $\mathcal{X}$ be the intensity or type of emergency intervention, Y a health outcome, $\mathcal{Z}$ prior utilization, and $\mathcal{W}$ physiology at admission; the hinge $\mathcal{H}$ is on-site triage. Triage aggregates information from both utilization history and current physiology (collider) while also mediating access to downstream care pathways and thus influencing both treatment allocation and outcomes (mediator).
(b) *Education (timing/administration);* Let $\mathcal{U}$ be ability/motivation, $\mathcal{X}$ placement on a math track in year t–1, Y math exam performance in t–1, $\mathcal{Z}$ involvement reflected by attendance in an unrelated enrichment course in $t-1$ (e.g. philosophy), and $\mathcal{W}$ a benchmark/diagnostic assessment score recorded in year t. The hinge $\mathcal{H}$ is the timing and administration conditions of that evaluation in year t (e.g., early vs late testing window, session assignment, or evaluation form). These timing/administration decisions can depend on both prior achievement and engagement signals (collider), and they can also mediate how those signals translate into the realized benchmark score through differences in instructional exposure, test conditions, or form difficulty (mediator).
(c) *Manufacturing (process temperature).* Let $\mathcal{U}$ denote impurities, $\mathcal{X}$ press force, Y on-time completion, $\mathcal{Z}$ exposure grade, and $\mathcal{W}$ material hardness; the hinge $\mathcal{H}$ is peak temperature. The peak temperature is jointly determined by the upstream choices and material properties (collider) and mediates how upstream factors manifest in downstream quality measurements such as exposure grade (mediator).

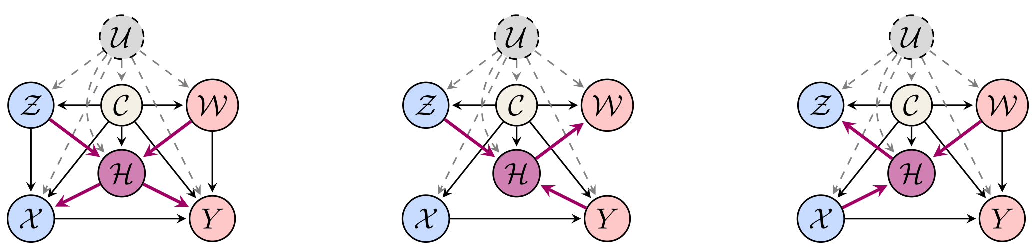


Figure 1: Three DAGs with hinges. Purple, **bold** arrows portray the two forbidden types of paths on which the **hinge** is a collider and a mediator. The blue nodes represent treatment side, and the red the outcome side.

Contributions. We develop hinge-adjusted proximal identification in general linear SCMs[1] with continuous, possibly multivariate $(\mathcal{X}, \mathcal{Z}, \mathcal{W}, \mathcal{C}, \mathcal{U})$, analyzing both outcome- and treatment-side bridges and making the required identification conditions explicit.

1. **Linear PV picture beyond scalar/binary treatments.** We formalize PV in linear SCMs and develop two complementary regression viewpoints: (i) an *outcome bridge* that uses $(\mathcal{W}, \mathcal{C})$ to absorb the $\mathcal{U}$-driven component of Y, and (ii) a *treatment bridge* that uses $(\mathcal{Z}, \mathcal{C})$ to isolate the $\mathcal{U}$-driven component of $\mathcal{X}$. In the linear multivariate regime, identification and uniqueness reduce to transparent rank/linear-independence conditions on conditional covariance operators. This characterization enables direct comparison between PV moment equations and IV regression. (Section 2)

2. **Hinges as a structural failure for conditional PV.** We formalize *hinges* as collider-mediator structures linking the treatment and outcome sides and invalidating cross-side exclusions. Treat hinges as control variable or disregard them bias will arise. We characterize the resulting bias in the moment equations. The bias arises from two sources: (i) endogeneity of noises when conditioning on children, and (ii) direct effect of the hinge. (Section 3)

3. **Nuisance PV.** We propose *Nuisance PV*, a hinge-aware correction that augments the proximal moment system with a nuisance block and identifies the target effect after projecting out the induced nuisance span. We provide hinge-aware identification results for both outcome bridge and treatment bridge constructions under analogous rank/span conditions. (Section 4)

2. Proximal Variables Regression

In the following section, we introduce the proximal regression setting based on linear structural equations, in contrast to frequently used potential outcomes. Next, we introduce the outcome and treatment bridge in the setting and discuss the estimation. We relegate all the proofs and more detailed discussions to the Appendix G and B respectively.

Setup Let $Y \in \mathbb{R}$ be an outcome, $\mathcal{X} \in \mathbb{R}^{d_x}$ a treatment vector, $\mathcal{W} \in \mathbb{R}^{d_w}$ an outcome side proxy, $\mathcal{Z} \in \mathbb{R}^{d_z}$ a treatment side proxy, $\mathcal{C} \in \mathbb{R}^{d_c}$ observed covariates, and $\mathcal{U} \in \mathbb{R}^{d_u}$ an unobserved confounder. All variables are centered and have second moments. We posit the linear structural system (we discuss extension to nonlinear settings in Appendix F.1):

$$Y = \beta^\top \mathcal{X} + \alpha_{YW}^\top \mathcal{W} + \alpha_{YC}^\top \mathcal{C} + \alpha_{YU}^\top \mathcal{U} + \varepsilon_Y, \tag{1}$$

$$\mathcal{X} = \alpha_{XC}\mathcal{C} + \alpha_{XU}\mathcal{U} + \alpha_{XZ}\mathcal{Z} + \varepsilon_X, \tag{2}$$

$$\mathcal{W} = \alpha_{WU}\mathcal{U} + \alpha_{WC}\mathcal{C} + \varepsilon_W, \tag{3}$$

$$\mathcal{Z} = \alpha_{ZC}\mathcal{C} + \alpha_{ZU}\mathcal{U} + \varepsilon_Z, \tag{4}$$

where the parameters are of respective sizes $\alpha_{ab} \in \mathbb{R}^{d_a \times d_b}$. Our object of interest is the coefficient vector β in (1), which captures the direct linear effect of $\mathcal{X}$ on Y in the structural equation. The coefficients $\alpha_{iC}, \alpha_{XZ}, \alpha_{YW}$ can be zero for any i, we require all the rest of the coefficients to be non-zero. We assume that the zero mean errors ε_i, for all i, satisfy joint independence. The system (1)-(4) encodes the cross-side exclusion and relevance of both proxies $\mathcal{Z}$ and $\mathcal{W}$.

1. We allow for some degree of nonlinearity, but the outcome and bridge equations need to be linear.

Intuition The difficulty in identifying β in (1) is that the latent variable $\mathcal{U}$ loads both into treatment and outcome, so the term $\alpha_{YU}^\top \mathcal{U}$ is correlated with $\mathcal{X}$ through $\alpha_{XU}\mathcal{U}$. The key idea behind both outcome and treatment bridges is to use observed proxies to construct a function that reproduces, up to conditional mean, the component of a target equation that is driven by $\mathcal{U}$. If such a bridge is available, then after subtracting the bridge from one side of the system, the remaining composite error no longer carries the part of $\mathcal{U}$ that creates confounding for β. Graphically, we want to erase either of the arrows from $\mathcal{U}$ to the treatment outcome pair. The informativeness assumptions formalize the requirement that the proxy delivers "rich enough" information about $\mathcal{U}$: after conditioning on controls $\mathcal{C}$, the way $\mathcal{U}$ enters the proxy spans the way $\mathcal{U}$ enters the target equation (either outcome or treatment). This guarantees the existence of a bridge in the linear class, but at this stage the bridge coefficients are only defined implicitly and are not directly observable. The second proxy is then necessary to learn these coefficients from observable moments. Rank conditions ensure that these cross-moment relations are not degenerate, i.e., that the relevant covariance matrices have full rank, so that the resulting linear system has a unique solution. In other words, the second proxy shares variation induced by $\mathcal{U}$ with the first proxy, yet it is linearly independent enough, such that after conditioning the variables of interest (other proxy and treatment) do not co-move, and reveal enough information. In this sense, the two proxies play complementary roles: one proxy is used to represent the latent confounding component (the bridge), while the other provides independent variation that makes the bridge coefficients and β identifiable from data.

2.1. Outcome bridge

The outcome equation (1) is confounded through the latent term $\alpha_{YU}^\top \mathcal{U}$. The outcome bridge strategy is to construct a function of $(\mathcal{W}, \mathcal{C})$ whose conditional expectation matches this latent component, thereby yielding observable moment restrictions involving $\mathcal{Z}$.

Definition 1 (Outcome bridge) *A function $b_Y(\mathcal{W}, \mathcal{C})$ is called an* outcome bridge *if*

$$\mathbb{E}\Big[\alpha_{YU}^\top \mathcal{U} - b_Y(\mathcal{W}, \mathcal{C}) \mid \mathcal{Z}, \mathcal{X}, \mathcal{C}\Big] = 0. \tag{5}$$

Note that our naming of the bridge and that of standard literature differs slightly, although the essence of the role of the function is the same. Once our linear bridge exists, the bridge in usual notation follows trivially. We now give a sufficient condition under which an outcome bridge exists.

Assumption 2 (Outcome informativeness) *There exists $\theta \in \mathbb{R}^{d_w}$ such that $\alpha_{YU} = \alpha_{WU}^\top \theta$.*

The outcome proxy has to be able to express the part of confounding that influences outcome (and not necessarily reconstruct the confounder). In triage example from Figure 1a, it will plausibly hold if laboratory measurement move strongly with the frailty severity. It will not hold, if the measurement is unrelated to certain comorbidity.

Lemma 3 (Existence of a linear outcome bridge) *If Assumption 2 holds, then there exists a linear outcome bridge of the form*

$$b_Y(\mathcal{W}, \mathcal{C}) = \theta^\top \mathcal{W} - \theta^\top \alpha_{WC}\mathcal{C}.$$

Given an outcome bridge, the latent term can be absorbed into a composite error that is orthogonal to $\mathcal{Z}$ (conditionally on $\mathcal{C}$), yielding linear covariance restrictions that identify β.

Assumption 4 (Outcome rank condition (joint uniqueness)) *The matrix* $\mathbb{E}[\mathrm{cov}(\mathcal{Z},(\mathcal{X},\mathcal{W}) \mid \mathcal{C})] \in \mathbb{R}^{d_z \times (d_x + d_w)}$ *has full column rank:* $\mathrm{rank}(\mathbb{E}[\mathrm{cov}(\mathcal{Z},(\mathcal{X},\mathcal{W}) \mid \mathcal{C})]) = d_x + d_w$.

The treatment proxy must provide enough independent variation to separate the causal effect on treatment from the nuisance loading on outcome proxy. In triage, different pieces of prior utilization must move treatment intensity and physiology in sufficiently different patterns. The assumption plausibly holds if one part of treatment proxy predicts intervention intensity, another part predicts physiology measurements, and they are not the same direction. It will fail if every component of treatment proxy affects treatment and outcome proxy in exactly the same way, and in turn, the data can only identify a mixture of parameters of the outcome equation.

Theorem 5 (Identification via outcome bridge) *Suppose Assumptions 2, and 4 hold. Define the nuisance vector* $\vartheta := \alpha_{YW} + \theta \in \mathbb{R}^{d_w}$. *Then* (β, ϑ) *is the unique solution to the linear system*

$$\mathbb{E}[\mathrm{cov}(\mathcal{Z}, Y \mid \mathcal{C})] = \mathbb{E}[\mathrm{cov}(\mathcal{Z}, \mathcal{X} \mid \mathcal{C})]\,\beta + \mathbb{E}[\mathrm{cov}(\mathcal{Z}, \mathcal{W} \mid \mathcal{C})]\,\vartheta. \tag{6}$$

In the Appendix B.2 we discuss how to relax slightly the rank condition. Even when ϑ is not uniquely identified, β may still be uniquely identified after projecting out the nuisance component. We can estimate the above equations with a standard Generalized Method of Moments (GMM) (Hall, 2005). We discuss the estimation procedure in detail in the Appendix B.1.2.

2.2. Treatment bridge

The treatment equation (2) is confounded through $\alpha_{XU}\mathcal{U}$. The treatment bridge strategy is to represent this confounding component using the proxy $\mathcal{Z}$ (and $\mathcal{C}$), and then form a residual $\mathcal{V}$ that can be used for identification of β. The assumptions found in this treatment bridge section follow the same logic as those from the outcome side section.

Definition 6 (Treatment bridge) *A function* $b_X(\mathcal{Z}, \mathcal{C})$ *is called a* treatment bridge *if*

$$\mathbb{E}[\alpha_{XU}\mathcal{U} - b_X(\mathcal{Z}, \mathcal{C}) \mid \mathcal{W}, \mathcal{C}] = 0. \tag{7}$$

Assumption 7 (Treatment informativeness) *There exists a matrix* $\Psi \in \mathbb{R}^{d_x \times d_z}$ *such that* $\alpha_{XU} = \Psi\,\alpha_{ZU}$.

Lemma 8 (Existence of a linear treatment bridge) *Assume Assumption 7. Then there exists a linear treatment bridge of the form*

$$b_X(\mathcal{Z}, \mathcal{C}) = \Psi \mathcal{Z} - \Psi \alpha_{ZC} \mathcal{C}.$$

Lemma 8 establishes existence of a bridge coefficient Ψ that maps $\mathcal{Z}$ to the $\mathcal{U}$-component of $\mathcal{X}$. However, the observable moment equation used for identification involves $\mathcal{X}$, which may also contain a direct $\mathcal{Z}$-term $\alpha_{XZ}\mathcal{Z}$. Consequently Ψ is not separately identified from second moments; the data identify only the composite $\Pi := \alpha_{XZ} + \Psi$. Importantly, Π is sufficient to construct a deconfounded residual and identify β. Introduce the deconfounded residual $\mathcal{V} := \mathcal{X} - \Pi(\mathcal{Z} - \alpha_{ZC}\mathcal{C})$.

Assumption 9 (Treatment rank conditions)

(i) The matrix $\mathbb{E}[\operatorname{cov}(\mathcal{W}, \mathcal{Z} \mid \mathcal{C})] \in \mathbb{R}^{d_w \times d_z}$ *has full column rank:* $\operatorname{rank}(\mathbb{E}[\operatorname{cov}(\mathcal{W}, \mathcal{Z} \mid \mathcal{C})]) = d_z$.

(ii) The matrix $\mathbb{E}[\operatorname{cov}(\mathcal{V}, \mathcal{X} \mid \mathcal{C})] \in \mathbb{R}^{d_x \times d_x}$ *is nonsingular:* $\operatorname{rank}(\mathbb{E}[\operatorname{cov}(\mathcal{V}, \mathcal{X} \mid \mathcal{C})]) = d_x$.

Theorem 10 (Identification via treatment bridge) *Suppose Assumptions 7, and 9 hold. Then:*

1. Π *the unique solution to*

$$\mathbb{E}[\operatorname{cov}(\mathcal{W}, \mathcal{X} \mid \mathcal{C})] = \mathbb{E}[\operatorname{cov}(\mathcal{W}, \mathcal{Z} \mid \mathcal{C})]\, \Pi^\top. \tag{8}$$

2. *With* $\mathcal{V}$ *defined as above,* β *the unique solution to*

$$\mathbb{E}[\operatorname{cov}(\mathcal{V}, Y \mid \mathcal{C})] = \mathbb{E}[\operatorname{cov}(\mathcal{V}, \mathcal{X} \mid \mathcal{C})]\, \beta. \tag{9}$$

We discuss how to relax the rank condition and compare to outcome bridge relaxation in the Appendix B.2.

Estimation. Again estimation proceeds with GMM. We give a detailed description of two steps estimation in the Appendix B.1.3.

2.3. Remarks

Treatment vs outcome bridges The outcome bridge and treatment bridge routes provide two complementary linear identification strategies. They can be used separately, or jointly. We compare them in detail in the Appendix B.3.

Proximal Variables vs Instrumental Variables In conditional IV (e.g., Pearl, 2009; Henckel, 2021; Thams et al., 2024), identification with instrument variable $\mathcal{I}$ is based on the moment equation

$$\mathbb{E}\Big[\operatorname{cov}\big(Y - \beta^\top \mathcal{X},\ \mathcal{I} \mid \mathcal{C}\big)\Big] = 0,$$

together with a full rank requirement on $\mathbb{E}[\operatorname{cov}(\mathcal{X}, \mathcal{I} \mid \mathcal{C})]$ (where all the other variables play the same roles: outcome, treatment, controls). Our outcome bridge moment equation has the same form but with an additional proxy term:

$$\mathbb{E}\Big[\operatorname{cov}\big(Y - \beta^\top \mathcal{X} - \vartheta^\top \mathcal{W},\ \mathcal{Z} \mid \mathcal{C}\big)\Big] = 0.$$

In the perspective of IV, PV regression can be interpreted as allowing $\mathcal{Z}$ to be an *invalid instrument* (correlated with the latent confounder), while cancelling the induced bias using the outcome proxy block $\mathcal{W}$ through a bridge (as noted in Tchetgen et al., 2020). We go into a detailed comparison between IV and PV, as well as discussion on how to exploit the similarities for diagnostic purposes in the Appendix B.4.

3. Hinges

To formalize the hinges we expand the system of equation (1)-(4) to include term $\mathcal{H} \in \mathbb{R}^{d_h}$ – an observed variable that can enter either of the equations and can be direct child of any of the other variables in the system (up to preservation of acyclicity). Again, ε_i, the noise terms, satisfy the previous joint independence.

Intuition Suppose an observed variable $\mathcal{H}$ is present and acts as a *hinge* between the treatment- and outcome-side proxy blocks. If the modeling choice does not explicitly adjust for the dual role, there are two natural defaults: (i) *ignore* $\mathcal{H}$ and proceed as if the proximal setup held with conditioning set $\mathcal{S} = \mathcal{C}$ only, or (ii) *condition on* $\mathcal{H}$ and treat it as an additional control, i.e. set $\mathcal{S} = (\mathcal{C}, \mathcal{H})$. To unify the discussion, let $\mathcal{S} \in \{\mathcal{C},\ (\mathcal{C}, \mathcal{H})\}$ denote the chosen conditioning set.

In both outcome bridge and treatment bridge PV, identification hinges on a cross-side orthogonality step for an appropriate bridge residual, expressed via conditional-covariance moments of the form $\mathbb{E}[\mathrm{cov}(\cdot, \cdot \mid \mathcal{S})]$. A hinge can break these moments through two distinct mechanisms:
(1) *Direct impact.* If $\mathcal{H}$ enters the relevant bridge residual and is *not conditioned on* (i.e. $\mathcal{S} = \mathcal{C}$), then the moment equations inherit additional terms proportional to cross-moments such as $\mathbb{E}[\mathrm{cov}(\mathcal{Z}, \mathcal{H} \mid \mathcal{C})]$ or $\mathbb{E}[\mathrm{cov}(\mathcal{W}, \mathcal{H} \mid \mathcal{C})]$, multiplied by the loading of $\mathcal{H}$ in the residual. In contrast, if $\mathcal{S} = (\mathcal{C}, \mathcal{H})$ then these leakage terms vanish mechanically because $\mathrm{cov}(\mathcal{A}, f(\mathcal{H}, \mathcal{C}) \mid \mathcal{C}, \mathcal{H}) = 0$ for any measurable $f(\mathcal{H}, \mathcal{C})$.
(2) *Noise endogeneity (failure of cross-side noise orthogonality).* The bridge proofs also rely on conditional mean-zero statements such as $\mathbb{E}[\varepsilon_W \mid \mathcal{Z}, \mathcal{X}, \mathcal{S}] = 0$ or $\mathbb{E}[\varepsilon_Z \mid \mathcal{W}, \mathcal{S}] = 0$. With hinges, these can fail even if structural noises are jointly independent across equations. There are two common ways this happens: (a) *Collider activation:* if $\mathcal{H}$ is a collider on a $\mathcal{Z}$–$\mathcal{W}$ path (e.g. $\mathcal{Z} \to \mathcal{H} \leftarrow \mathcal{W}$), then conditioning on $\mathcal{H}$ can induce cross-side dependence and yield $\mathrm{cov}(\mathcal{Z}, \varepsilon_W \mid \mathcal{C}, \mathcal{H}) \neq 0$ and/or $\mathrm{cov}(\mathcal{W}, \varepsilon_Z \mid \mathcal{C}, \mathcal{H}) \neq 0$. (b) *Descendant conditioning via mediation:* if $\mathcal{H}$ mediates across sides, e.g. $\mathcal{W} \to \mathcal{H} \to \mathcal{Z}$ or $\mathcal{W} \to \mathcal{H} \to \mathcal{X}$, then $\mathcal{Z}$ or $\mathcal{X}$ becomes a *descendant* of $\mathcal{W}$ and therefore carries information about ε_W; consequently, one can have non-zero noise covariances even without conditioning on $\mathcal{H}$. Analogous statements hold for ε_Z when $\mathcal{Z} \to \mathcal{H} \to \mathcal{W}$.

In both bridges, the estimator solves a linear moment system. Hence the population bias enters as a deviation term premultiplied by an inverse of an identifying conditional-covariance operator. Therefore, hinge bias is small when the deviation is small (weak leakage and/or weak noise endogeneity), and it can be amplified when the identifying operator is ill-conditioned.

3.1. Outcome bridge hinge bias

Outcome bridge identification targets the population restriction

$$\mathbb{E}[\mathrm{cov}(\mathcal{Z},\ R_Y(\beta, \vartheta) \mid \mathcal{S})] = 0, \qquad R_Y(\beta, \vartheta) := Y - \beta^\top \mathcal{X} - \vartheta^\top \mathcal{W}. \tag{10}$$

When a hinge is present, the same algebra yields the *biased* moment equation

$$\mathbb{E}[\mathrm{cov}(\mathcal{Z}, Y \mid \mathcal{S})] = \mathbb{E}[\mathrm{cov}(\mathcal{Z}, \mathcal{X} \mid \mathcal{S})]\, \beta + \mathbb{E}[\mathrm{cov}(\mathcal{Z}, \mathcal{W} \mid \mathcal{S})]\, \vartheta + \Delta_{OB}^{\mathcal{S}}, \tag{11}$$

where the hinge-induced deviation is

$$\Delta_{OB}^{\mathcal{S}} := \mathbb{E}[\mathrm{cov}(\mathcal{Z},\ R_Y(\beta, \vartheta) \mid \mathcal{S})]\,. \tag{12}$$

Thus, outcome bridge "fails" exactly when the hinge makes $\mathcal{Z}$ correlate with the bridge residual R_Y under the chosen S. Introduce $\alpha_{bH} \ := \ \alpha_{YH} - \alpha_{WH}^\top \theta$. Then

$$\Delta_{OB}^{\mathcal{S}} = \underbrace{\mathbb{E}[\mathrm{cov}(\mathcal{Z}, \mathcal{H} \mid \mathcal{S})]\, \alpha_{bH}}_{\text{direct impact}} + \underbrace{\mathbb{E}\Big[\mathrm{cov}\Big(\mathcal{Z}, \varepsilon_Y - \theta^\top \varepsilon_W \mid \mathcal{S}\Big)\Big]}_{\text{noise endogeneity}}. \tag{13}$$

When do the two components vanish? (i) If $\mathcal{S} = (\mathcal{C}, \mathcal{H})$, then $\mathrm{cov}(\mathcal{Z}, \mathcal{H} \mid \mathcal{C}, \mathcal{H}) = 0$ (since $\mathcal{H}$ is S-measurable), so the *leakage* term in (13) vanishes automatically; (ii) The *noise endogeneity* term vanishes under the hinge-free cross-side exclusion logic, but can become nonzero (a) when conditioning on a collider (e.g. $\mathcal{Z} \to \mathcal{H} \leftarrow \mathcal{W}$ and $\mathcal{S} = (\mathcal{C}, \mathcal{H})$), and/or (b) when $\mathcal{Z}$ or $\mathcal{X}$ becomes a descendant of $\mathcal{W}$ (e.g. $\mathcal{W} \to \mathcal{H} \to \mathcal{Z}$ or $\mathcal{W} \to \mathcal{H} \to \mathcal{X}$) even if $\mathcal{S} = \mathcal{C}$.

Let the (population) outcome bridge estimator under conditioning choice $\mathcal{S}$ be the solution of the *unbiased* system (i.e. it sets $\Delta_{OB}^{S} = 0$ in (11)). Whenever the inverse (pseudoinverse) exists,

$$\begin{pmatrix} \beta_{OB}^{\mathcal{S}} - \beta \\ \vartheta_{OB}^{\mathcal{S}} - \vartheta \end{pmatrix} = \Big(\mathbb{E}[\mathrm{cov}(\mathcal{Z}, (\mathcal{X}, \mathcal{W}) \mid \mathcal{S})]\Big)^{-1} \Delta_{OB}^{\mathcal{S}}. \tag{14}$$

Hence hinge bias is small when Δ_{OB}^{S} is small (weak leakage and/or weak noise endogeneity), and it can be large when $\mathbb{E}[\mathrm{cov}(\mathcal{Z}, (\mathcal{X}, \mathcal{W}) \mid \mathcal{S})]$ is ill-conditioned.

3.2. Treatment bridge hinge bias

Treatment bridge relies on two orthogonality conditions: a first-stage operator equation (for the coefficient used to construct V) and a second-stage IV-like equation identifying β using V. Let Ξ denote the coefficient used to construct the deconfounded residual V. We write the target system generically as

$$\mathbb{E}[\mathrm{cov}(\mathcal{W},\, \mathcal{X} - \Xi\mathcal{Z} \mid \mathcal{S})] = 0, \qquad \mathbb{E}\Big[\mathrm{cov}\Big(\mathcal{V},\, Y - \beta^{\top}\mathcal{X} \mid \mathcal{S}\Big)\Big] = 0, \qquad \mathcal{V} := \mathcal{X} - \Xi\mathcal{Z}. \tag{15}$$

First-stage deviation. With hinges, the true first-stage relation becomes

$$\mathbb{E}[\mathrm{cov}(\mathcal{W}, \mathcal{X} \mid \mathcal{S})] = \mathbb{E}[\mathrm{cov}(\mathcal{W}, \mathcal{Z} \mid \mathcal{S})]\, \Xi^{\top} + \Delta_{TB,1}^{\mathcal{S}}, \qquad \Delta_{TB,1}^{\mathcal{S}} := \mathbb{E}[\mathrm{cov}(\mathcal{W},\, \mathcal{X} - \Xi\mathcal{Z} \mid \mathcal{S})]\,. \tag{16}$$

Introduce $\kappa_H := \alpha_{XH} - \Xi\alpha_{ZH}$ so that

$$\Delta_{TB,1}^{\mathcal{S}} = \underbrace{\mathbb{E}[\mathrm{cov}(\mathcal{W}, \mathcal{H} \mid \mathcal{S})]\, \kappa_H^{\top}}_{\text{direct impact}} + \underbrace{\mathbb{E}[\mathrm{cov}(\mathcal{W}, \varepsilon_X - \Xi\varepsilon_Z \mid \mathcal{S})]}_{\text{noise endogeneity}}. \tag{17}$$

As in the outcome bridge, conditioning on $\mathcal{H}$ kills the leakage term but can create noise endogeneity via collider activation.

Second-stage deviation and bias in β. The second-stage moment becomes

$$\mathbb{E}[\mathrm{cov}(\mathcal{V}, Y \mid \mathcal{S})] = \mathbb{E}[\mathrm{cov}(\mathcal{V}, \mathcal{X} \mid \mathcal{S})]\, \beta + \Delta_{TB,2}^{\mathcal{S}}, \qquad \Delta_{TB,2}^{\mathcal{S}} := \mathbb{E}\Big[\mathrm{cov}\Big(\mathcal{V},\, Y - \beta^{\top}\mathcal{X} \mid \mathcal{S}\Big)\Big]\,. \tag{18}$$

Expanding $Y - \beta^{\top}\mathcal{X} = \alpha_{YW}^{\top}\mathcal{W} + \alpha_{YH}^{\top}\mathcal{H} + \alpha_{YU}^{\top}\mathcal{U} + \varepsilon_Y$ gives the decomposition

$$\Delta_{TB,2}^{\mathcal{S}} = \mathbb{E}[\mathrm{cov}(\mathcal{V}, \mathcal{W} \mid \mathcal{S})]\, \alpha_{YW} + \mathbb{E}[\mathrm{cov}(\mathcal{V}, \mathcal{H} \mid \mathcal{S})]\, \alpha_{YH} + \mathbb{E}[\mathrm{cov}(\mathcal{V}, \mathcal{U} \mid \mathcal{S})]\, \alpha_{YU} + \mathbb{E}[\mathrm{cov}(\mathcal{V}, \varepsilon_Y \mid \mathcal{S})]\,. \tag{19}$$

If one solves the *unbiased* second-stage equation (setting $\Delta_{TB,2}^{\mathcal{S}} = 0$), then whenever $\mathbb{E}[\mathrm{cov}(\mathcal{V}, \mathcal{X} \mid \mathcal{S})]$ is invertible,

$$\beta_{TB}^{\mathcal{S}} - \beta = \Big(\mathbb{E}[\mathrm{cov}(\mathcal{V}, \mathcal{X} \mid \mathcal{S})]\Big)^{-1} \Delta_{TB,2}^{\mathcal{S}}. \tag{20}$$

A key difference from outcome bridge is that $\Delta^{\mathcal{S}}_{TB,2}$ is *endogenous to the first stage*: if the hinge contaminates Ξ (i.e. $\Delta^{\mathcal{S}}_{TB,1} \neq 0$), then $\mathcal{V} = \mathcal{X} - \Xi\mathcal{Z}$ may retain U-dependence, making the $\mathbb{E}[\mathrm{cov}(\mathcal{V},\mathcal{U} \mid \mathcal{S})]\alpha_{YU}$ term in (19) potentially dominant. This is why hinge-robustification for treatment bridge typically needs to address first-stage deviations explicitly, whereas outcome bridge hinge bias appears as an additive deviation in a single moment system.

4. Nuisance Proximal Variables

Our goal is to identify the causal coefficient β in the presence of an observed hinge $\mathcal{H}$ that couples the treatment- and outcome-side proxy blocks. Hinge can be confounded by the unobserved variable $\mathcal{U}$. We relegate the proofs and discussion to the Appendix G and D respectively. We discuss extension to nonlinear settings in Appendix F.1. Essentially, under the minimal assumptions only the outcome model and bridges have to be linear and the process $(\mathcal{X}, \mathcal{Z}, \mathcal{W}, \mathcal{H}, \mathcal{C})$ may follow arbitrary (possibly nonlinear) mechanisms. We gather all the algorithmic steps, post-estimation diagnostic and other suggestions for practitioners in Apendix F.2.

Intuition A hinge can break proximal moment conditions through (i) *direct impact* (mediator leakage) and (ii) *noise endogeneity* (cross-side dependence of proxy noises). Our approach keeps the conditioning set fixed at $\mathcal{C}$ (so we avoid opening collider paths by conditioning on $\mathcal{H}$), and instead augments the moment system with a *nuisance block* $\mathcal{N} := (\mathcal{W}^\top, \mathcal{H}^\top)^\top$. This makes hinge leakage an explicit term involving observable cross-moments with $(\mathcal{W}, \mathcal{H})$. This adapts nuisance IV logic (Thams et al., 2024) to the PV setting. The remaining obstacle is remaining hinge-induced noise endogeneity. The same nuisance procedure can adjust for it as long as the induced violation lies in the *nuisance span* generated by cross-moments with $\mathcal{N}$. Under this span restriction, β is identified from the augmented moment system. The same proximal *informativeness* assumptions as in hinge-free PV remain the driver of bridge existence. The price paid for hinge-robustness is instead in rank conditions: adding/removing nuisance spans consumes identifying variation, so projected covariance operators must remain well-conditioned. We compare the nuisance PV and nuisance IV in detail in the Appendix D.1.

4.1. Hinge-robust outcome bridge (NPV-OB)

Under Assumption 2, define $\vartheta := \alpha_{YW} + \theta \in \mathbb{R}^{d_w}$, $\alpha_{bH} := \alpha_{YH} - \alpha_{WH}^\top\theta \in \mathbb{R}^{d_h}$, $\delta := \alpha_{YC} - \alpha_{WC}^\top\theta \in \mathbb{R}^{d_c}$, and the composite noise $\tilde{\varepsilon}_Y := \varepsilon_Y - \theta^\top\varepsilon_W$. Then the outcome equation can be rewritten as

$$Y = \beta^\top\mathcal{X} + \vartheta^\top\mathcal{W} + \alpha_{bH}^\top\mathcal{H} + \delta^\top\mathcal{C} + \tilde{\varepsilon}_Y. \tag{21}$$

Taking $\mathbb{E}[\mathrm{cov}(\mathcal{Z}, \cdot \mid \mathcal{C})]$ of (21) gives

$$\begin{aligned}\mathbb{E}[\mathrm{cov}(\mathcal{Z}, Y \mid \mathcal{C})] = \mathbb{E}[\mathrm{cov}(\mathcal{Z}, \mathcal{X} \mid \mathcal{C})]\,\beta + \mathbb{E}[\mathrm{cov}(\mathcal{Z}, \mathcal{W} \mid \mathcal{C})]\,\vartheta \\ + \mathbb{E}[\mathrm{cov}(\mathcal{Z}, \mathcal{H} \mid \mathcal{C})]\,\alpha_{bH} + \mathbb{E}[\mathrm{cov}(\mathcal{Z}, \tilde{\varepsilon}_Y \mid \mathcal{C})].\end{aligned} \tag{22}$$

where the last term we note as Δ_{noise}. The terms involving (ϑ, α_{bH}) are proportional to observable cross-moments with $(\mathcal{W}, \mathcal{H})$. The remaining obstacle is Δ_{noise}, which can be nonzero under hinge-induced noise endogeneity.

Augmented nuisance moment system. Let $\mathcal{N} := (\mathcal{W}^\top, \mathcal{H}^\top)^\top \in \mathbb{R}^{d_n}$ where $d_n = d_w + d_h$, and define the cross-moment operator

$$\mathbb{E}[\mathrm{cov}(\mathcal{Z}, \mathcal{N} \mid \mathcal{C})] = \Big(\mathbb{E}[\mathrm{cov}(\mathcal{Z}, \mathcal{W} \mid \mathcal{C})] \quad \mathbb{E}[\mathrm{cov}(\mathcal{Z}, \mathcal{H} \mid \mathcal{C})]\Big) \in \mathbb{R}^{d_z \times d_n}.$$

Stack $\gamma := (\vartheta^\top, \alpha_{bH}^\top)^\top \in \mathbb{R}^{d_n}$ so that (22) becomes

$$\mathbb{E}[\mathrm{cov}(\mathcal{Z}, Y \mid \mathcal{C})] \;=\; \mathbb{E}[\mathrm{cov}(\mathcal{Z}, \mathcal{X} \mid \mathcal{C})]\beta \;+\; \mathbb{E}[\mathrm{cov}(\mathcal{Z}, \mathcal{N} \mid \mathcal{C})]\gamma \;+\; \Delta_{\mathrm{noise}}.$$

Assumption 11 (Nuisance-span condition for hinge-induced noise) *There exists $\lambda \in \mathbb{R}^{d_n}$ such that $\Delta_{\mathrm{noise}} = \mathbb{E}[\mathrm{cov}(\mathcal{Z}, \tilde{\varepsilon}_Y \mid \mathcal{C})] = \mathbb{E}[\mathrm{cov}(\mathcal{Z}, \mathcal{N} \mid \mathcal{C})]\,\lambda$.*

Assumption 11 says that hinge-induced cross-side noise correlation in the $\mathcal{Z}$-moment equation is *linearly explainable by the same $\mathcal{Z}$-variation that $\mathcal{Z}$ shares with the observed nuisance block $\mathcal{N} = (\mathcal{W}, \mathcal{H})$*. Equivalently, there exists λ such that $\mathbb{E}[\mathrm{cov}(\mathcal{Z}, \tilde{\varepsilon}_Y - \lambda^\top \mathcal{N} \mid \mathcal{C})] = 0$. This is the algebraic statement that the hinge transmits the induced dependence through *observed* $(\mathcal{W}, \mathcal{H})$. Under Assumption 11, there exists $\gamma^\star := \gamma + \lambda$ such that the moment equation closes as

$$\mathbb{E}[\mathrm{cov}(\mathcal{Z}, Y \mid \mathcal{C})] \;=\; \mathbb{E}[\mathrm{cov}(\mathcal{Z}, \mathcal{X} \mid \mathcal{C})]\beta \;+\; \mathbb{E}[\mathrm{cov}(\mathcal{Z}, \mathcal{N} \mid \mathcal{C})]\gamma^\star. \tag{23}$$

Assumption 12 (Rank condition for β modulo the nuisance span) *The columns of $\mathbb{E}[\mathrm{cov}(\mathcal{Z}, \mathcal{X} \mid \mathcal{C})]$ are linearly independent modulo $\mathrm{col}(\mathbb{E}[\mathrm{cov}(\mathcal{Z}, \mathcal{N} \mid \mathcal{C})])$, i.e.*
$\mathbb{E}[\mathrm{cov}(\mathcal{Z}, \mathcal{X} \mid \mathcal{C})]b \in \mathrm{col}(\mathbb{E}[\mathrm{cov}(\mathcal{Z}, \mathcal{N} \mid \mathcal{C})]) \implies b = 0 \qquad$ *for all $b \in \mathbb{R}^{d_x}$.*

Theorem 13 (β-identification for hinge-robust outcome bridge (NPV-OB)) *Suppose Assumption 2, 11 and 12 hold. Then β is uniquely identified as the unique value for which there exists $\gamma^\star \in \mathbb{R}^{d_n}$ satisfying the closed moment equation (23).*

In the linear (unweighted) GMM / least-squares implementation used in our experiments, we estimate (β, γ) from the sample analogue of (23) and report the coefficient $\hat{\beta}$ on $\mathcal{X}$.

4.2. Hinge-robust treatment bridge (NPV-TB)

Under Assumption 7, define $\Pi_Z := \alpha_{XZ} + \Psi_Z$, $\kappa_H := \alpha_{XH} - \Psi_Z \alpha_{ZH}$, $\Gamma := \alpha_{XC} - \Psi_Z \alpha_{ZC}$, and $\tilde{\varepsilon}_X := \varepsilon_X - \Psi_Z \varepsilon_Z$. Then the treatment equation can be rewritten as

$$\mathcal{X} \;=\; \Pi_Z \mathcal{Z} \;+\; \kappa_H \mathcal{H} \;+\; \Gamma \mathcal{C} \;+\; \tilde{\varepsilon}_X. \tag{24}$$

Stage 1: identifying Π_Z via an augmented moment system. Taking $\mathbb{E}[\mathrm{cov}(\mathcal{W}, \cdot \mid \mathcal{C})]$ of (24) gives

$$\mathbb{E}[\mathrm{cov}(\mathcal{W}, \mathcal{X} \mid \mathcal{C})] = \mathbb{E}[\mathrm{cov}(\mathcal{W}, \mathcal{Z} \mid \mathcal{C})]\,\Pi_Z^\top + \mathbb{E}[\mathrm{cov}(\mathcal{W}, \mathcal{H} \mid \mathcal{C})]\,\kappa_H^\top + \mathbb{E}[\mathrm{cov}(\mathcal{W}, \tilde{\varepsilon}_X \mid \mathcal{C})], \tag{25}$$

where the last term we note as $\Delta_{TB,1}$. The two non-ideal terms are hinge leakage $\mathbb{E}[\mathrm{cov}(\mathcal{W}, \mathcal{H} \mid \mathcal{C})]\kappa_H^\top$ and hinge-induced noise endogeneity $\Delta_{TB,1}$.

Assumption 14 (Hinge-span condition for TB Stage 1 noise) *There exists a matrix $\Lambda_1 \in \mathbb{R}^{d_h \times d_x}$ such that $\Delta_{TB,1} = \mathbb{E}[\mathrm{cov}(\mathcal{W}, \tilde{\varepsilon}_X \mid \mathcal{C})] = M_{WH}\Lambda_1$.*

Under Assumption 14, (25) closes as

$$\mathbb{E}[\mathrm{cov}(\mathcal{W}, \mathcal{X} \mid \mathcal{C})] = M_{WZ}\,\Pi_Z^\top + M_{WH}\,\Gamma_1 \qquad \text{for some } \Gamma_1 \in \mathbb{R}^{d_h \times d_x}. \tag{26}$$

Assumption 15 (Rank condition for Stage 1, modulo the hinge span) $\mathbb{E}[\mathrm{cov}(\mathcal{W}, \mathcal{Z} \mid \mathcal{C})]$ *is injective modulo* $\mathrm{col}(\mathbb{E}[\mathrm{cov}(\mathcal{W}, \mathcal{H} \mid \mathcal{C})])$*, i.e.*
$\mathbb{E}[\mathrm{cov}(\mathcal{W}, \mathcal{Z} \mid \mathcal{C})]b \in \mathrm{col}(\mathbb{E}[\mathrm{cov}(\mathcal{W}, \mathcal{H} \mid \mathcal{C})]) \implies b = 0 \qquad$ *for all* $b \in \mathbb{R}^{d_z}$.

Proposition 16 (Identification of Π_Z (hinge-robust TB Stage 1)) *Assume Assumptions 7, 14, and 15. Then Π_Z is uniquely identified as the unique value for which there exists Γ_1 satisfying (26).*

Stage 2: identifying β from an augmented moment system. Define the generated residual $\mathcal{V} := \mathcal{X} - \Pi_Z\mathcal{Z}$ (with $\mathcal{C}$-residualization understood). From (24), $\mathcal{V} = \kappa_H\mathcal{H} + \Gamma\mathcal{C} + \tilde{\varepsilon}_X$. Under Assumption 7, the composite noise $\tilde{\varepsilon}_X = \varepsilon_X - \Psi_Z\varepsilon_Z$ has no *direct* U component. However, $\mathcal{V}$ can still be correlated with U through the hinge term $\kappa_H\mathcal{H}$ whenever $\mathcal{H}$ is confounded by U. Accordingly, we do not assume $\mathbb{E}[\mathrm{cov}(\mathcal{V}, U \mid \mathcal{C})] = 0$; any remaining U-dependence enters the second-stage deviation term below. Taking $\mathbb{E}[\mathrm{cov}(\mathcal{V}, \cdot \mid \mathcal{C})]$ of the outcome equation yields

$$\mathbb{E}[\mathrm{cov}(\mathcal{V}, Y \mid \mathcal{C})] = \mathbb{E}[\mathrm{cov}(\mathcal{V}, \mathcal{X} \mid \mathcal{C})]\,\beta + \mathbb{E}[\mathrm{cov}(\mathcal{V}, \mathcal{N} \mid \mathcal{C})]\,\gamma + \Delta_{TB,2}, \tag{27}$$

where $\mathcal{N} := (\mathcal{W}^\top, \mathcal{H}^\top)^\top$ and $\gamma := (\alpha_{YW}^\top, \alpha_{YH}^\top)^\top$, and the second-stage deviation is

$$\Delta_{TB,2} := \mathbb{E}\big[\mathrm{cov}(\mathcal{V}, \alpha_{YU}^\top U + \varepsilon_Y \mid \mathcal{C})\big] = \mathbb{E}[\mathrm{cov}(\mathcal{V}, U \mid \mathcal{C})]\,\alpha_{YU} + \mathbb{E}[\mathrm{cov}(\mathcal{V}, \varepsilon_Y \mid \mathcal{C})].$$

Assumption 17 (Nuisance-span condition for TB Stage 2 deviation) *Let* $\mathbb{E}[\mathrm{cov}(\mathcal{V}, \mathcal{N} \mid \mathcal{C})] := \mathbb{E}[\mathrm{cov}(\mathcal{V}, \mathcal{N} \mid \mathcal{C})]$. *There exists* $\lambda_2 \in \mathbb{R}^{d_w + d_h}$ *such that* $\Delta_{TB,2} = \mathbb{E}\big[\mathrm{cov}(\mathcal{V}, \alpha_{YU}^\top U + \varepsilon_Y \mid \mathcal{C})\big] = \mathbb{E}[\mathrm{cov}(\mathcal{V}, \mathcal{N} \mid \mathcal{C})]\lambda_2$.

Under Assumption 17, the second-stage moment closes as

$$\mathbb{E}[\mathrm{cov}(\mathcal{V}, Y \mid \mathcal{C})] = \mathbb{E}[\mathrm{cov}(\mathcal{V}, \mathcal{X} \mid \mathcal{C})]\,\beta + \mathbb{E}[\mathrm{cov}(\mathcal{V}, \mathcal{N} \mid \mathcal{C})]\,\gamma^\star \qquad \text{for some } \gamma^\star. \tag{28}$$

Assumption 18 (Rank condition for β modulo the nuisance span) $\mathbb{E}[\mathrm{cov}(\mathcal{V}, \mathcal{X} \mid \mathcal{C})]$ *is injective modulo*
$\mathrm{col}(\mathbb{E}[\mathrm{cov}(\mathcal{V}, \mathcal{N} \mid \mathcal{C})])$*, i.e.* $\mathbb{E}[\mathrm{cov}(\mathcal{V}, \mathcal{X} \mid \mathcal{C})]b \in \mathrm{col}(\mathbb{E}[\mathrm{cov}(\mathcal{V}, \mathcal{N} \mid \mathcal{C})]) \implies b = 0 \qquad$ *for all* $b \in \mathbb{R}^{d_x}$.

Theorem 19 (β-identification by hinge-robust treatment bridge (NPV-TB)) *Assume Assumption 7. Let Π_Z be identified from Stage 1 under Assumptions 14 and 15, and define $\mathcal{V} := \mathcal{X} - \Pi_Z\mathcal{Z}$. If Assumptions 17 and 18 hold, then β is uniquely identified as the unique value for which there exists $\gamma^\star$ satisfying the closed moment equation (28).*

4.3. Remarks

Nuisance PV is not the same as conditioning on $\mathcal{H}$. Treating $\mathcal{H}$ as a control corresponds to replacing $\mathbb{E}[\mathrm{cov}(\cdot, \cdot \mid \mathcal{C})]$ by $\mathbb{E}[\mathrm{cov}(\cdot, \cdot \mid \mathcal{C}, \mathcal{H})]$. Both terms of the covariance get residualized by the conditioning variable. This can eliminate mediator leakage (since $\mathrm{cov}(\mathcal{A}, f(\mathcal{H}, \mathcal{C}) \mid \mathcal{C}, \mathcal{H}) = 0$), but it can also *create* noise endogeneity by activating collider paths. In contrast, NPV keeps the conditioning set fixed (typically $\mathcal{C}$ only) and uses a *linear-algebraic* correction: it deletes the component of the moment equation that lies in an empirically identified nuisance span, the residualization is asymmetric.

How strong are the nuisance-span and projected-rank assumptions? Assumptions 11 and 14 are *span* restrictions: they do not claim that hinge-induced noise endogeneity vanishes, but that the induced violation lives in the linear subspace generated by observable cross-moments with the *observed nuisance block*. In linear/Gaussian SCMs this is natural when cross-side dependence is transmitted through observed hinge/proxy channels—e.g., directed pathways that pass through $(\mathcal{W}, \mathcal{H})$—and there is no additional "bypass" mechanism (extra unobserved confounders not obeying informativeness, omitted observed mediators not included in $\mathcal{N}$) that would create $\mathcal{Z}$–noise dependence in directions orthogonal to $\mathrm{col}(\mathbb{E}[\mathrm{cov}(\mathcal{Z}, \mathcal{N} \mid \mathcal{C})])$ (or $\mathcal{W}$–noise dependence orthogonal to $\mathrm{col}(\mathbb{E}[\mathrm{cov}(\mathcal{W}, \mathcal{H} \mid \mathcal{C})])$). Assumptions 12 and 15 are the complementary *relevance* conditions: after removing the nuisance span, there must remain enough identifying variation to recover the target parameters. These are direct analogues of instrument relevance and weak-identification conditions in IV literature; near failures manifest as ill-conditioning (smallest singular value close to zero) and lead to finite-sample variance amplification. Importantly, some restriction of this kind is essentially necessary for point identification from our moment information: if the hinge creates violations outside the nuisance span, or if all $\mathcal{Z}$–$\mathcal{X}$ (resp. $\mathcal{W}$–$\mathcal{Z}$) variation is absorbed by the nuisance span, then no estimator that only uses $(\mathcal{Z}, \mathcal{W}, \mathcal{H}, \mathcal{C})$ can generically isolate β (resp. Π_Z in first stage treatment bridge) without further assumptions. To build further intuition we give an example about sufficient graphical structure for the assumptions to hold in the Appendix D, and a more concrete discussion in the form of an Figure 1a example in Appendix F.3.

5. Experiments

We complement the theory with synthetic (as we focus here to validate theoretical claims) linear-Gaussian simulations in which an observed hinge $\mathcal{H}$ induces the two failure modes studied in the paper: *direct hinge impact* ($\mathcal{H}$ enters structural equations) and *hinge-induced noise endogeneity* (collider-driven dependence between proxy noises across sides). We compare three practical estimators throughout: *ignore* $\mathcal{H}$ (hinge-free proximal estimator), *condition on* $\mathcal{H}$ (treat $\mathcal{H}$ as an additional control/conditioning variable), and *nuisance PV* (augment the proximal moment system with $\mathcal{H}$ as a nuisance regressor).

We start with the setup Figure 1 (a) and modify the parametrization. We perform four experiments. In (a) we start with random parameters and increase sample size. In (b) we break the nuisance-span condition by injecting an additional latent factor that creates cross-side dependence bypassing the observed nuisance block $(\mathcal{W}, \mathcal{H})$. In (c) we scale the magnitude of the direct hinge effect. In (d) we scale the collider mechanism that produces cross-side noise dependence.

We portray the results in Figure 2. All plots report the estimation error $\hat{\beta} - \beta$ over repeated 200 draws.

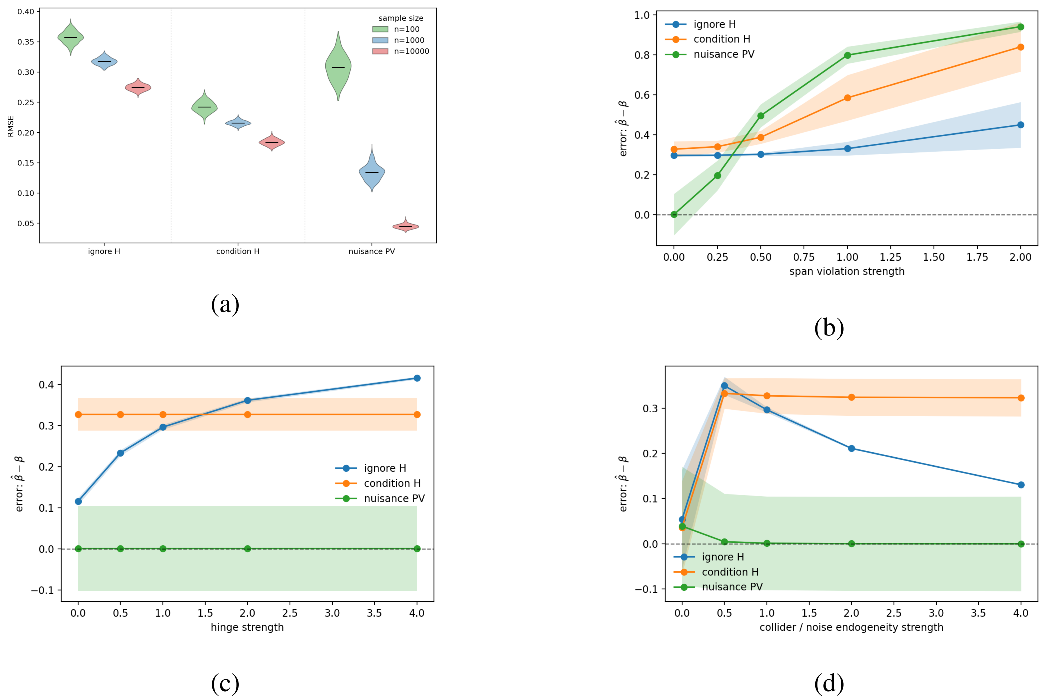


Figure 2: Synthetic experiments. Top row: (a) finite sample RMSE; (b) span violation. Bottom row: (c) mediator paths strength; (d) collider paths strength.

Results. (a) As sample size increases, nuisance PV concentrates around zero error, whereas the two baselines remain biased, although nuisance PV can exhibit higher variance in small samples. (b) As this violation grows, nuisance PV deteriorates, matching the theory that its robustness requires the induced violation to lie in the nuisance span. (c) Ignoring $\mathcal{H}$ becomes increasingly biased as the hinge effect strengthens, while nuisance PV remains stable. (d) Conditioning on $\mathcal{H}$ is sensitive to this collider strength, whereas nuisance PV remains stable as long as the nuisance-span condition holds. We provide more details in Appendix E.

6. Discussion

In this work we extend previous proximal causal inference discussions to multivariate continuous variables and more complex graphical structure. In particular, we work with multivariate, continuous treatment and propose new, regression based treatment bridge, as well as characterize, and adjust for, hinges. Working with linear regression models allows us to gain intuition about the inner workings of PV, directly, but also through a comparison to extensively studied IV. This allows us to get more intuition on the essence of why PV works, for example, as the cross-side exclusion proves not to be strictly necessary. A future direction might pick up from here and generalize to more complex graphical settings or to include more complex functional relations in bridges such as generalized linear models, for a start. An adjacent problem worth studying is partial identification in settings with proxies. Last but not least, it remains to explore which estimates based on real world data are biased due to hinges.

Acknowledgments

This paper was written within the scope of Novo Nordisk Foundation grant Project Grants in the Natural and Technical Sciences 2024 (#0096126), whose support is gratefully acknowledged.

References

Yifan Cui, Hongming Pu, Xu Shi, Wang Miao, and Eric Tchetgen Tchetgen. Semiparametric proximal causal inference. *Journal of the American Statistical Association*, pages 1–12, 2023.

Ben Deaner. Proxy controls and panel data. *arXiv preprint arXiv:1810.00283*, 2023.

AmirEmad Ghassami, Andrew Ying, Ilya Shpitser, and Eric Tchetgen Tchetgen. Minimax kernel machine learning for a class of doubly robust functionals with application to proximal causal inference. In *International conference on artificial intelligence and statistics*, pages 7210–7239. PMLR, 2022.

AmirEmad Ghassami, Ilya Shpitser, and Eric Tchetgen Tchetgen. Partial identification of causal effects using proxy variables. *arXiv preprint arXiv:2304.04374*, 2023.

Helen Guo, Elizabeth L Ogburn, and Ilya Shpitser. Comparing two proxy methods for causal identification. *arXiv preprint arXiv:2512.00175*, 2025.

Alastair R Hall. *Generalized method of moments*. Oxford University Press, 2005.

Leonard Henckel. *Graphical Tools for Efficient Causal Effect Estimation*. PhD thesis, ETH Zurich, 2021.

Nathan Kallus, Xiaojie Mao, and Masatoshi Uehara. Causal inference under unmeasured confounding with negative controls: A minimax learning approach. *arXiv preprint arXiv:2103.14029*, 2021.

Benjamin Kompa, David Bellamy, Tom Kolokotrones, Andrew Beam, et al. Deep learning methods for proximal inference via maximum moment restriction. *Advances in Neural Information Processing Systems*, 35:11189–11201, 2022.

Jiewen Liu, Chan Park, Kendrick Li, and Eric J Tchetgen Tchetgen. Regression-based proximal causal inference. *American journal of epidemiology*, page kwae370, 2024.

Afsaneh Mastouri, Yuchen Zhu, Limor Gultchin, Anna Korba, Ricardo Silva, Matt Kusner, Arthur Gretton, and Krikamol Muandet. Proximal causal learning with kernels: Two-stage estimation and moment restriction. In *International conference on machine learning*, pages 7512–7523. PMLR, 2021.

Wang Miao, Xu Shi, Yilin Li, and Eric J. Tchetgen Tchetgen. A confounding bridge approach for double negative control inference on causal effects. *Statistical Theory and Related Fields*, 8(4): 262–273, 2024.

Judea Pearl. Causal inference in statistics: An overview. 2009.

Eric J Tchetgen Tchetgen, Andrew Ying, Yifan Cui, Xu Shi, and Wang Miao. An introduction to proximal causal learning. *arXiv:2009.10982*, 2020.

Nikolaj Thams, Rikke Søndergaard, Sebastian Weichwald, and Jonas Peters. Identifying causal effects using instrumental time series: Nuisance iv and correcting for the past. *Journal of Machine Learning Research*, 25(302):1–51, 2024.

Yong Wu, Yanwei Fu, Shouyan Wang, and Xinwei Sun. Doubly robust proximal causal learning for continuous treatments. In *The Twelfth International Conference on Learning Representations*, 2024.

Liyuan Xu, Heishiro Kanagawa, and Arthur Gretton. Deep proxy causal learning and its application to confounded bandit policy evaluation. *NeurIPS*, 34:26264–26275, 2021.

Zhiheng Zhang and Xinyan Su. Partial identification with proxy of latent confoundings via sum-of-ratios fractional programming. In *Proceedings of the Fortieth Conference on Uncertainty in Artificial Intelligence*, volume 244 of *Proceedings of Machine Learning Research*, pages 4140–4172. PMLR, 2024.

Appendix Overview

We introduce a shorthand notation throughout the Appendix $M_{AB} := \mathbb{E}[\text{cov}(A, B \mid C)]$.

Appendix A. Related work

Our setting falls within proximal/negative-control identification: two proxy blocks (treatment-side and outcome-side) together with a bridge function yield moment restrictions that identify causal effects under latent confounding (e.g. Miao et al., 2024; Tchetgen et al., 2020; Cui et al., 2023). Much of this literature assumes there exists a conditioning set that enforces the required *cross-side exclusion* (conditional separation of the two proxy blocks), and then focuses on *estimation*—learning bridge functions with flexible classes, adversarial/minimax objectives, double robustness, and finite-sample analyses (e.g. Tchetgen et al., 2020; Kallus et al., 2021; Ghassami et al., 2022; Mastouri et al., 2021; Xu et al., 2021; Kompa et al., 2022; Wu et al., 2024; Deaner, 2023; Cui et al., 2023). Relative to this line, we take a step back: we study a concrete *structural* reason why cross-side exclusion can fail even in linear models. In particular, an *observable hinge* that is simultaneously a mediator and a collider can make *no single conditioning choice* satisfy the cross-side exclusions needed for the standard proximal moments, so the issue is not merely that the bridge is difficult to learn but that the baseline moment conditions are structurally misspecified.

Regression-based proximal implementations. Bridge estimation is often presented as solving an inverse (potentially ill-posed) operator equation, in contrast to the direct regression form of IV. Several works advocate regression-style implementations of proximal identification, including two-stage least squares-type procedures (Tchetgen et al., 2020) and two-stage generalized linear models regression (Liu et al., 2024). This line is directly relevant because we also emphasize regression moments. The key difference is that regression-based proximal methods operate under the standard cross-side exclusions and focus primarily on the outcome bridge perspective. In contrast, we (i) generalize to allow all, except of the outcome, variables to be multivariate and continuous (ii) develop a complementary treatment bridge regression viewpoint in a linear multivariate SCM, (iii) make identification/uniqueness conditions explicit via rank/linear-independence requirements on conditional covariance operators, and (iv) address hinge-induced structural violations by modifying the moment system through Nuisance PV under explicit rank/span conditions.

Informativeness/completeness versus rank conditions. Proxy identification ultimately requires proxies to be "informative enough" about latent confounding. Guo et al. (2025) clarify that, across proxy identification paradigms, completeness plays a role analogous to a linear-independence condition and compare bridge-equation methods to array-decomposition approaches. We remain in the bridge/moment-equation paradigm but make the connection explicit in our setting: in linear proximal regression, informativeness/completeness-type requirements reduce to transparent rank conditions on conditional cross-covariance operators. This viewpoint also clarifies how hinge-robustification consumes identifying variation: rank requirements after removing the nuisance span induced by the hinge.

Robustness when proximal assumptions fail. When proximal assumptions are doubtful, an alternative is partial identification and sensitivity analysis. Ghassami et al. (2023) and Zhang and Su (2024) develop proxy-based bounds that remain valid without strong point-identification conditions and compute bounds via optimization. This is related motivation-wise—proximal assumptions can fail—but it targets a different regime. Our hinge setting is a structural cross-side exclusion failure induced by an observed collider-mediator, and our goal is to restore point identification *within linear PV* by appropriately modifying the identifying moments (projection-based Nuisance PV), rather than switching to bounds. Moreover, the two aforementioned works assume univariate and binary treatment in contrast to our multivariate and continuous variable setting. Nevertheless, a very specific setting lies on the intersection of our paper and Ghassami et al. (2023). They provide fully observable bounds under $\mathcal{W} \perp\!\!\!\perp \mathcal{Z} \mid \mathcal{X},\mathcal{C},\mathcal{U}$. If we restrict to binary treatment and univariate proxies, and if the observed hinge $\mathcal{H}$ can be incorporated into the covariate set $\mathcal{C}$ in such a way that it blocks the proxy–proxy dependence such that the above restriction holds, then the "two independent invalid proxies" bounds apply directly ($\mathcal{H}$ has to be mediator between $\mathcal{W}$ and $\mathcal{Z}$). This is a very particular subcase of our general formulation of the problem.

Collider–mediator structures and nuisance adjustment in IV. Thams et al. (2024) study collider-mediator structures that invalidate standard IV estimation and introduce nuisance-IV constructions that augment identifying equations with nuisance regressors. This is conceptually adjacent to our approach—we also treat the hinge as a nuisance component—but the proximal setting differs in what breaks and what is available to fix it. Proximal identification relies on two cross-side exclusions and uses proxies that are deliberately correlated with latent confounding, so the relevant moment failures can arise through endogeneity of bridge residual noise in addition to direct mediated pathways. Nuisance PV adapts the nuisance-adjustment idea to this two-sided proximal moment structure, yielding hinge-aware identification results under explicit span/rank conditions.

Appendix B. Discussion for Section 2

B.1. Estimation of the bridges

B.1.1. RESIDUALIZATION AND EMPIRICAL CONDITIONAL COVARIANCES

Throughout, we work with $\mathcal{C}$-residualized variables. For any random vector A, define

$$\tilde{A} := A - \mathbb{E}[A \mid \mathcal{C}].$$

In practice, we replace $\mathbb{E}[A \mid \mathcal{C}]$ by an estimator $\widehat{\mathbb{E}}[A \mid \mathcal{C}]$ (e.g. linear regression, series regression, or cross-fitted machine learning) and form $\hat{\tilde{A}}_i := A_i - \widehat{\mathbb{E}}[A \mid \mathcal{C}_i]$.

Using $\tilde{A}$ and $\tilde{B}$, the population object $\mathbb{E}[\mathrm{cov}(A, B \mid \mathcal{C})]$ equals $\mathbb{E}[\tilde{A}\tilde{B}^\top]$, and its empirical analog is

$$\mathbb{E}[\widehat{\mathrm{cov}(A, B} \mid \mathcal{C})] := \frac{1}{n}\sum_{i=1}^{n} \hat{\tilde{A}}_i\, \hat{\tilde{B}}_i^\top.$$

Let $\{(Y_i, \mathcal{X}_i, \mathcal{W}_i, \mathcal{Z}_i, \mathcal{C}_i)\}_{i=1}^n$ be i.i.d.

B.1.2. Estimation via outcome bridge: GMM and 2SLS forms

GMM For the joint identification system (6), define the respective sample moment

$$\hat{g}_O(\beta, \vartheta) := \frac{1}{n}\sum_{i=1}^{n} \tilde{\mathcal{Z}}_i\Big(\tilde{Y}_i - \beta^\top \tilde{\mathcal{X}}_i - \vartheta^\top \tilde{\mathcal{W}}_i\Big) \in \mathbb{R}^{d_z},$$

and estimate (β, ϑ) by GMM:

$$(\hat{\beta}, \hat{\vartheta}) \in \arg\min_{\beta,\vartheta}\ \hat{g}_O(\beta, \vartheta)^\top\, \hat{W}_n\, \hat{g}_O(\beta, \vartheta),$$

with a positive definite weight matrix $\hat{W}_n$ (e.g. the inverse of the estimated moment covariance).

The outcome bridge population moments can be written as

$$\mathbb{E}\Big[\tilde{\mathcal{Z}}\left(\tilde{Y} - \beta^\top \tilde{\mathcal{X}} - \vartheta^\top \tilde{\mathcal{W}}\right)\Big] = 0.$$

2SLS form. Define the stacked regressor block

$$\tilde{\mathcal{R}} := \begin{bmatrix}\tilde{\mathcal{X}} & \tilde{\mathcal{W}}\end{bmatrix} \in \mathbb{R}^{d_X + d_W}.$$

With $W = (\frac{1}{n}\tilde{\mathcal{Z}}^\top \tilde{\mathcal{Z}})^{-1}$, the estimator equals 2SLS on residualized variables:

$$\begin{bmatrix}\hat{\beta}\\ \hat{\vartheta}\end{bmatrix} = (\tilde{\mathcal{R}}^\top P_{\tilde{\mathcal{Z}}} \tilde{\mathcal{R}})^{-1} \tilde{\mathcal{R}}^\top P_{\tilde{\mathcal{Z}}} \tilde{Y}, \qquad P_{\tilde{\mathcal{Z}}} := \tilde{\mathcal{Z}}(\tilde{\mathcal{Z}}^\top \tilde{\mathcal{Z}})^{-1}\tilde{\mathcal{Z}}^\top.$$

Inference. Standard heteroskedastic-robust (or cluster-robust) GMM/2SLS variance estimators apply directly; in overidentified settings, Hansen's J test can be used as a specification test of the bridge moment restrictions.

B.1.3. Estimation via treatment bridge

The treatment bridge is naturally estimated in two steps.

Step 1 (estimate Π). Using $\mathcal{C}$-residualized variables, the population equation

$$\mathbb{E}[\mathrm{cov}(\mathcal{W}, \mathcal{X} \mid \mathcal{C})] = \mathbb{E}[\mathrm{cov}(\mathcal{W}, \mathcal{Z} \mid \mathcal{C})]\, \Pi^\top$$

becomes $\mathbb{E}[\tilde{\mathcal{W}}\tilde{\mathcal{X}}^\top] = \mathbb{E}[\tilde{\mathcal{W}}\tilde{\mathcal{Z}}^\top]\Pi^\top$. Estimate Π by the sample analog, e.g. by solving

$$\widehat{\mathbb{E}[\tilde{\mathcal{W}}\tilde{\mathcal{X}}^\top]} = \widehat{\mathbb{E}[\tilde{\mathcal{W}}\tilde{\mathcal{Z}}^\top]}\, \hat{\Pi}^\top,$$

or (in overidentified settings) by minimizing the Frobenius norm of the corresponding residual.

Step 2 (construct $\hat{\mathcal{V}}$ and estimate β). Form the deconfounded residual

$$\hat{\tilde{\mathcal{V}}}_i := \hat{\tilde{\mathcal{X}}}_i - \hat{\Pi}\,\hat{\tilde{\mathcal{Z}}}_i.$$

Then estimate β from the IV-like moment equation

$$\mathbb{E}[\tilde{\mathcal{V}}(\tilde{Y} - \beta^\top \tilde{\mathcal{X}})] = 0 \quad \Rightarrow \quad \widehat{\mathbb{E}[\hat{\tilde{\mathcal{V}}}\hat{\tilde{Y}}]} = \widehat{\mathbb{E}[\hat{\tilde{\mathcal{V}}}\hat{\tilde{\mathcal{X}}}^\top]}\,\hat{\beta}.$$

B.2. Identification of β only by projection.

B.2.1. OUTCOME BRIDGE

Even when ϑ is not uniquely identified, β may still be uniquely identified after projecting out the nuisance component. Let

$$M_{WZ} \;:=\; \mathbb{E}[\mathrm{cov}(\mathcal{Z}, \mathcal{W} \mid \mathcal{C})] \in \mathbb{R}^{d_z \times d_w},$$

and define the orthogonal projector onto the orthogonal complement of $\mathrm{col}(M_{WZ}) \subseteq \mathbb{R}^{d_z}$ by

$$P_W^\perp \;:=\; I_{d_z} - M_{WZ}(M_{WZ}^\top M_{WZ})^\dagger M_{WZ}^\top, \tag{29}$$

where $(\cdot)^\dagger$ denotes the Moore–Penrose pseudoinverse.

Assumption 20 (Projected rank condition (β-uniqueness))

$$\mathrm{rank}\Big(P_W^\perp\,\mathbb{E}[\mathrm{cov}(\mathcal{Z}, \mathcal{X} \mid \mathcal{C})]\Big) = d_x.$$

Proposition 21 (β-identification by projection) *Suppose Assumptions 2 hold. If Assumption 20 holds, then β the unique solution to*

$$P_W^\perp\,\mathbb{E}[\mathrm{cov}(\mathcal{Z}, Y \mid \mathcal{C})] = P_W^\perp\,\mathbb{E}[\mathrm{cov}(\mathcal{Z}, \mathcal{X} \mid \mathcal{C})]\,\beta. \tag{30}$$

For the projected estimator (30), set

$$\hat{M}_{WZ} := \mathbb{E}[\widehat{\mathrm{cov}(\mathcal{Z}, \mathcal{W}} \mid \mathcal{C})], \qquad \hat{P}_W^\perp := I_{d_z} - \hat{M}_{WZ}(\hat{M}_{WZ}^\top \hat{M}_{WZ})^\dagger \hat{M}_{WZ}^\top,$$

and solve the projected sample system

$$\hat{P}_W^\perp\,\mathbb{E}[\widehat{\mathrm{cov}(\mathcal{Z}, Y} \mid \mathcal{C})] = \hat{P}_W^\perp\,\mathbb{E}[\widehat{\mathrm{cov}(\mathcal{Z}, \mathcal{X}} \mid \mathcal{C})]\,\hat{\beta}.$$

B.2.2. TREATMENT BRIDGE

For the treatment bridge, β is identified from moments involving the residual $\mathcal{V} = \mathcal{X} - b_X(\mathcal{Z}, \mathcal{C})$, where the bridge coefficients enter *inside* $\mathcal{V}$. Consequently, there is no direct analogue of "projecting out an additive nuisance" as in the outcome bridge system. Instead, when the first-stage operator $\mathbb{E}[\mathrm{cov}(\mathcal{W}, \mathcal{Z} \mid \mathcal{C})]$ is rank-deficient, one can still recover β by restricting attention to the *identifiable subspace* of $\mathcal{Z}$ – the component of $\mathcal{Z}$ that is "seen" through $\mathcal{W}$.

In this spirit, let

$$M_{WZ} \;:=\; \mathbb{E}[\mathrm{cov}(\mathcal{W}, \mathcal{Z} \mid \mathcal{C})] \in \mathbb{R}^{d_w \times d_z}, \qquad r \;:=\; \mathrm{rank}(M_{WZ}).$$

Let $R \in \mathbb{R}^{d_z \times r}$ be any matrix with orthonormal columns spanning the row space of M_{WZ}, i.e., $\mathrm{col}(R) = \mathrm{col}(M_{WZ}^\top)$. Define the reduced treatment-side proxy

$$\mathcal{Z}_R \;:=\; R^\top \mathcal{Z} \in \mathbb{R}^r.$$

By construction, the reduced cross-moment operator

$$M_{WZ_R} \;:=\; \mathbb{E}[\mathrm{cov}(\mathcal{W}, \mathcal{Z}_R \mid \mathcal{C})] = M_{WZ} R \in \mathbb{R}^{d_w \times r}$$

has full column rank r.

Assumption 22 (Projected treatment informativeness) *There exists a matrix $\Psi_R \in \mathbb{R}^{d_x \times r}$ such that the confounding component of $\mathcal{X}$ can be represented using $\mathcal{Z}_R$:*

$$\alpha_{XU} \;=\; \Psi_R \, \alpha_{Z_R U}, \qquad \textit{where } \alpha_{Z_R U} := R^\top \alpha_{ZU}.$$

Lemma 23 (Existence of a linear projected treatment bridge) *Introduce $\Pi_R = \alpha_{XZ} R + \Psi_R$. If Assumption 22 holds, then there exists a linear treatment bridge based on $\mathcal{Z}_R$ of the form*

$$b_X(\mathcal{Z}, \mathcal{C}) \;=\; \Pi_R \, \mathcal{Z}_R + \Gamma \, \mathcal{C}$$

for some $\Gamma \in \mathbb{R}^{d_x \times d_c}$.

Define the corresponding residual

$$\mathcal{V}_R \;:=\; \mathcal{X} - \Pi_R \mathcal{Z}_R - \Gamma \mathcal{C}. \tag{31}$$

Assumption 24 (Projected treatment rank for β) *The matrix*

$$\mathbb{E}[\mathrm{cov}(\mathcal{V}_R, \mathcal{X} \mid \mathcal{C})] \in \mathbb{R}^{d_x \times d_x}$$

is nonsingular:

$$\mathrm{rank}(\mathbb{E}[\mathrm{cov}(\mathcal{V}_R, \mathcal{X} \mid \mathcal{C})]) = d_x.$$

Theorem 25 (β-identification via projected treatment bridge) *Suppose Assumption 22 holds. If Assumption 24 holds, then β the unique solution to*

$$\mathbb{E}[\mathrm{cov}(\mathcal{V}_R, Y \mid \mathcal{C})] = \mathbb{E}[\mathrm{cov}(\mathcal{V}_R, \mathcal{X} \mid \mathcal{C})] \, \beta. \tag{32}$$

Estimation (two-step with subspace reduction). Use the same residualization construction as in the main text: for any $A \in \{Y, \mathcal{X}, \mathcal{W}, \mathcal{Z}\}$, let $\tilde{A}_i := A_i - \hat{m}_A(\mathcal{C}_i)$ denote residuals after regressing on $\mathcal{C}$. Form the sample analogs

$$\hat{M}_{WZ} := \frac{1}{n} \sum_{i=1}^n \tilde{\mathcal{W}}_i \, \tilde{\mathcal{Z}}_i^\top, \qquad \hat{M}_{WX} := \frac{1}{n} \sum_{i=1}^n \tilde{\mathcal{W}}_i \, \tilde{\mathcal{X}}_i^\top.$$

Compute a thin SVD $\hat{M}_{WZ} = \hat{U} \hat{\Sigma} \hat{V}^\top$ and let $\hat{R} \in \mathbb{R}^{d_z \times \hat{r}}$ collect the right singular vectors corresponding to the non-negligible singular values (a plug-in estimate of the row space). Define the reduced proxy residuals

$$\tilde{\mathcal{Z}}_{R,i} := \hat{R}^\top \tilde{\mathcal{Z}}_i, \qquad \hat{M}_{WZ_R} := \frac{1}{n} \sum_{i=1}^n \tilde{\mathcal{W}}_i \, \tilde{\mathcal{Z}}_{R,i}^\top = \hat{M}_{WZ} \hat{R}.$$

Estimate Π_R by the (unique) least-squares solution

$$\hat{\Pi}_R^\top := (\hat{M}_{WZ_R})^\dagger \hat{M}_{WX} \quad \text{(equivalently, solve } \hat{M}_{WX} = \hat{M}_{WZ_R}\hat{\Pi}_R^\top\text{)}.$$

Then form residuals

$$\hat{\mathcal{V}}_{R,i} := \tilde{\mathcal{X}}_i - \hat{\Pi}_R\, \tilde{\mathcal{Z}}_{R,i}.$$

Finally, estimate β from the sample analogue of (32), e.g. by solving

$$\Big(\frac{1}{n}\sum_{i=1}^{n} \hat{\mathcal{V}}_{R,i}\tilde{\mathcal{X}}_i^\top\Big)\hat{\beta} = \frac{1}{n}\sum_{i=1}^{n} \hat{\mathcal{V}}_{R,i}\tilde{Y}_i,$$

or, equivalently, by GMM with moment

$$\hat{g}_{T,R}(\beta) := \frac{1}{n}\sum_{i=1}^{n} \hat{\mathcal{V}}_{R,i}\big(\tilde{Y}_i - \beta^\top \tilde{\mathcal{X}}_i\big).$$

B.2.3. REMARKS

Outcome vs. treatment projections. In the outcome bridge β-only result, projection is used to *eliminate an additive nuisance*: the outcome moment equation contains β and a nuisance coefficient ϑ entering additively through the span $\mathrm{col}(M_W)$, where $M_W = \mathbb{E}[\mathrm{cov}(\mathcal{Z}, \mathcal{W} \mid \mathcal{C})]$. Left-multiplying by $P_W^\perp$ removes that span and isolates β.

In the treatment bridge procedure, there is no additive nuisance in the β-moment. Instead, the unknown bridge coefficient Π is needed to *construct* the residual $\mathcal{V}$. When M_{WZ} is rank-deficient, Π is not fully identified; the appropriate analogue of projection is therefore to *retain only the identifiable subspace of* $\mathcal{Z}$, namely $\mathrm{col}(M_{WZ}^\top)$, by replacing $\mathcal{Z}$ with $\mathcal{Z}_R = R^\top \mathcal{Z}$. Notably, $\mathrm{col}(M_{WZ}^\top) = \mathrm{col}(M_W)$, so the outcome-side projection removes $\mathrm{col}(M_W)$ while the treatment-side reduction keeps it.

Why the two sides are structurally asymmetric. The asymmetry stems from the fact that $\mathcal{W}$ appears directly in the outcome equation through $\alpha_{YW}^\top \mathcal{W}$, which induces the nuisance term ϑ in the $\mathcal{Z}$–Y cross-moment equation. In contrast, $\mathcal{W}$ does not enter the treatment equation (2), so the first-stage equation for Π is a pure linear operator equation rather than an additive decomposition with a removable nuisance component.

Relation to the full-rank treatment case. If M_{WZ} has full column rank d_z, then $r = d_z$, one may take $R = I_{d_z}$, and the projected treatment bridge construction reduces to the standard treatment bridge identification and estimation in Section 2.2.

B.3. Further comparison between treatment and outcome bridges

Both Theorem 5 and Theorem 10 show that each of the bridges identify the unique structural coefficient in (1). In that sense, they agree on the population target: they are two routes to the same parameter. However, they generally differ in nuisance quantities (outcome bridge involves ϑ, treatment bridge involves Π and $\mathcal{V}$) and therefore yield different finite-sample estimators and different overidentifying restrictions.

Table 1: Comparison of analogous elements of outcome and treatment bridges identification procedures

	Outcome bridge	Treatment bridge
Bridge moment	$\mathbb{E}[\alpha_{YU}^\top \mathcal{U} - b_Y(\mathcal{W},\mathcal{C}) \mid \mathcal{Z},\mathcal{X},\mathcal{C}] = 0$	$\mathbb{E}[\alpha_{XU}\mathcal{U} - b_X(\mathcal{Z},\mathcal{C}) \mid \mathcal{W},\mathcal{C}] = 0$
Informativeness	$\exists\, \theta$ s.t. $\theta^\top \alpha_{WU} = \alpha_{YU}^\top$	$\exists\, \Psi$ s.t. $\Psi\, \alpha_{ZU} = \alpha_{XU}$
Bridge	$b_Y(\mathcal{W},\mathcal{C}) := \theta^\top \mathcal{W} - \theta^\top \alpha_{WC}\mathcal{C}$	$\Pi := \alpha_{XZ} + \Psi$ $b_X(\mathcal{Z},\mathcal{C}) := \Pi\,\mathcal{Z} - \Pi\,\alpha_{ZC}\mathcal{C}$
Auxiliary quantity	$\vartheta := \alpha_{YW} + \theta$	$\mathcal{V} := \mathcal{X} - b_X(\mathcal{Z},\mathcal{C})$
Rank condition(s)	$\operatorname{rank}\Big(\mathbb{E}[\operatorname{cov}\left((\mathcal{X},\mathcal{W}),\mathcal{Z} \mid \mathcal{C})\right]\Big) = \dim(\mathcal{X}) + \dim(\mathcal{W})$	$\operatorname{rank}\Big(\mathbb{E}[\operatorname{cov}(\mathcal{X},\mathcal{V} \mid \mathcal{C})]\Big) = \dim(\mathcal{X})$ $\operatorname{rank}\Big(\mathbb{E}[\operatorname{cov}(\mathcal{W},\mathcal{Z} \mid \mathcal{C})]\Big) = \dim(\mathcal{Z})$
Moment equation(s)	$\mathbb{E}[\operatorname{cov}(Y,\mathcal{Z} \mid \mathcal{C})] = \mathbb{E}[\operatorname{cov}(\mathcal{X},\mathcal{Z} \mid \mathcal{C})]\,\beta + \mathbb{E}[\operatorname{cov}(\mathcal{W},\mathcal{Z} \mid \mathcal{C})]\,\vartheta$	$\mathbb{E}[\operatorname{cov}(\mathcal{W},\mathcal{X} \mid \mathcal{C})] = \mathbb{E}[\operatorname{cov}(\mathcal{W},\mathcal{Z} \mid \mathcal{C})]\,\Pi^\top$ $\mathbb{E}[\operatorname{cov}(Y,\mathcal{V} \mid \mathcal{C})] = \mathbb{E}[\operatorname{cov}(\mathcal{X},\mathcal{V} \mid \mathcal{C})]\,\beta$

This appendix contrasts the outcome bridge and treatment bridge identification strategies introduced in Sections 2.1–2.2, and discusses when one may be preferable in practice. Observe a summary portrayed in Figure 1. We also describe how both strategies can be combined, either as a diagnostic device or via stacked GMM.

Both bridges aim to identify the same structural parameter β in (1), but they remove confounding in different places. Outcome bridges "repair" the *outcome equation* by augmenting the residual; treatment bridges "repair" the *treatment equation* by constructing a deconfounded treatment residual.

The outcome bridge moment system is a single IV/GMM system for the augmented block $(\mathcal{X},\mathcal{W})$ instrumented by $\mathcal{Z}$ (after residualizing on $\mathcal{C}$). The treatment bridge strategy is naturally sequential: it first uses $(\mathcal{W},\mathcal{Z})$ moments to learn the bridge parameter Π (hence to construct $\mathcal{V}$), and then uses $\mathcal{V}$ to identify β through an IV-like covariance equation.

How they relate and when one may be preferable Neither bridge dominates the other: they rely on different informativeness and rank conditions, and one may be plausible when the other is not.

When to prefer either outcome or treatment bridge? We have three options.

1. Prefer outcome bridge when $\mathcal{Z}$ is "rich" for $(\mathcal{X},\mathcal{W})$. outcome bridge estimation is most attractive when $\mathcal{Z}$ provides enough independent variation to span the augmented block $(\mathcal{X},\mathcal{W})$ after conditioning on $\mathcal{C}$ (heuristically, $d_z \gtrsim d_x + d_w$ and the matrix $\mathbb{E}[\operatorname{cov}(\mathcal{Z},(\mathcal{X},\mathcal{W}) \mid \mathcal{C})]$ is well-conditioned). Outcome bridge is then a single-step GMM/2SLS estimator for (β,ϑ).

2. Prefer treatment bridge when $(\mathcal{W},\mathcal{Z})$ is well-conditioned but $\mathcal{Z}$ is not large enough to span $(\mathcal{X},\mathcal{W})$. treatment bridge estimation can be preferable when $\mathbb{E}[\operatorname{cov}(\mathcal{W},\mathcal{Z} \mid \mathcal{C})]$ is well-conditioned so that Π can be learned reliably, and when $\mathbb{E}[\operatorname{cov}(\mathcal{V},\mathcal{X} \mid \mathcal{C})]$ is nonsingular so that β is identified from the deconfounded residual $\mathcal{V}$.

3. If both are feasible, use them as a stability check. When both sets of assumptions appear plausible, comparing estimators for both bridges is a useful diagnostic: large discrepancies suggest failure of at least one set of bridge assumptions, misspecification of the linear bridge class, or mishandling of controls.

Interpretation of overidentification under stacking. Rejection of the stacked Hansen J test indicates that at least one set of moment restrictions is violated. Unlike classical IV, this should be interpreted as evidence against the joint validity of: (i) proxy exclusion structure, (ii) linear bridge specification, and/or (iii) correct handling of controls $\mathcal{C}$. Acceptance does not prove assumptions, but provides a useful internal consistency check when multiple proxy blocks are available.

Relation to β-only identification. Both strategies admit relaxations in which nuisance parameters are not uniquely identified but β is. For the outcome bridge, this can be achieved by projecting out the nuisance span associated with $\mathbb{E}[\mathrm{cov}(\mathcal{Z}, \mathcal{W} \mid \mathcal{C})]$ (Appendix B.2). On the treatment bridge side, relaxations typically involve identifying only the components of Π that are learnable from $\mathbb{E}[\mathrm{cov}(\mathcal{W}, \mathcal{Z} \mid \mathcal{C})]$ and constructing $\mathcal{V}$ from the corresponding projected proxy variation (also discussed in Appendix B.2).

B.4. Relation to IV and IV-style diagnostics

Conditional IV. Assume treatment $\mathcal{X}$ and outcome Y are confounded by a hidden variable U, and there exists control variable $\mathcal{C}$ and an exogenous instrument $\mathcal{I}$. Under IV conditions ($\mathcal{I}$ is related to Y only through $\mathcal{X}$, and observables conditionally independent of $\mathcal{U}$) (e.g. Henckel (2021); Thams et al. (2024)), the target coefficient β satisfies

$$\mathbb{E}\Big[\mathrm{cov}\big(Y - \beta^\top \mathcal{X},\ \mathcal{I} \mid \mathcal{C}\big)\Big] = 0, \tag{33}$$

and is uniquely identified if $\mathrm{rank}\big(\mathbb{E}[\mathrm{cov}(\mathcal{X}, \mathcal{I} \mid \mathcal{C})]\big) = d_X$.

Outcome bridge as IV on an augmented regressor block. Outcome bridge PVR yields the population moment equation

$$\mathbb{E}\Big[\mathrm{cov}\big(Y - \beta^\top \mathcal{X} - \vartheta^\top \mathcal{W},\ \mathcal{Z} \mid \mathcal{C}\big)\Big] = 0, \tag{34}$$

equivalently,

$$\mathbb{E}[\mathrm{cov}(\mathcal{Z}, Y \mid \mathcal{C})] = \mathbb{E}[\mathrm{cov}(\mathcal{Z}, \mathcal{X} \mid \mathcal{C})]\,\beta + \mathbb{E}[\mathrm{cov}(\mathcal{Z}, \mathcal{W} \mid \mathcal{C})]\,\vartheta.$$

This is precisely the IV/GMM moment system for the augmented endogenous regressor block $\mathcal{R} := (\mathcal{X}, \mathcal{W})$ with instrument block $\mathcal{Z}$, after residualizing on $\mathcal{C}$.

Exogeneity vs. bridge cancellation. In IV, $\mathcal{I}$ is required to be orthogonal to the structural residual so that (33) holds directly. In PVR, $\mathcal{Z}$ may correlate with latent confounding; instead, the proxy block $\mathcal{W}$ is used to form an augmented residual $Y - \beta^\top \mathcal{X} - \vartheta^\top \mathcal{W}$ that is orthogonal to $\mathcal{Z}$ in the conditional-covariance sense (34). Thus PVR replaces instrument exogeneity by proxy validity and bridge-existence conditions.

Rank conditions and dimensionality. In IV, uniqueness typically requires $d_I \geq d_X$ and full row rank of $\mathbb{E}[\mathrm{cov}(\mathcal{X}, \mathcal{I} \mid \mathcal{C})]$. In OB-PVR, joint uniqueness of (β, ϑ) requires that $\mathcal{Z}$ spans the augmented block $(\mathcal{X}, \mathcal{W})$, i.e. full rank of $\mathbb{E}[\mathrm{cov}(\mathcal{Z}, (\mathcal{X}, \mathcal{W}) \mid \mathcal{C})]$. If one is only interested in β, weaker conditions may suffice by projecting out the nuisance ϑ component (see Theorem 21).

Treatment bridge as a generated instrument. treatment bridge PVR constructs $\mathcal{V} := \mathcal{X} - b_X(\mathcal{Z}, \mathcal{C})$ such that

$$\mathbb{E}\Big[\mathrm{cov}\big(\mathcal{V},\ Y - \beta^\top \mathcal{X} \mid \mathcal{C}\big)\Big] = 0.$$

This is formally an IV moment condition with instrument $\mathcal{V}$, but $\mathcal{V}$ is not observed and must be estimated through the bridge. As a result, inference must account for first-stage uncertainty (e.g. via stacked GMM).

IV-style diagnostics for OB-PVR (interpretation changes). Because (34) is a standard linear GMM/2SLS moment system for $(\mathcal{X}, \mathcal{W})$ instrumented by $\mathcal{Z}$, many IV diagnostics apply directly to theoutcome bridgeestimating equations. For example, Hansen's J statistic tests the overidentifying restrictions, but rejection should be interpreted as evidence against the joint validity of proxy/bridge/control assumptions rather than against instrument exogeneity alone. Similarly, rank and weak-identification diagnostics (e.g. Kleibergen–Paap type tests) assess whether $\mathcal{Z}$ provides sufficiently informative variation for the augmented block $(\mathcal{X}, \mathcal{W})$ after conditioning on $\mathcal{C}$. In the same manner, Hausman test could be applied, among others.

B.5. Relation to nuisance instrumental variables (NIV)

Recent work on conditional instrumental variables introduces *nuisance instrumental variables* (NIV) as an extension of IV in which identification is achieved by augmenting the outcome regression with additional nuisance regressors (see, e.g., Thams et al., 2024). At the level of *moment equations*, NIV is closely related to our outcome bridge PVR moment system. At the level of *graphs and identifying assumptions*, however, the two approaches differ substantially.

NIV moment equation. Let $\mathcal{I}$ denote an instrument block, $\mathcal{X}$ the treatment block of interest, $\mathcal{Z}_N$ a block of nuisance regressors ("N" for nuisance), $\mathcal{C}$ a conditioning set, and Y the outcome. NIV posits that β satisfies the moment equation

$$\exists\, \alpha \text{ s.t.} \quad \mathbb{E}\Big[\operatorname{cov}\big(Y - \beta^\top \mathcal{X} - \alpha^\top \mathcal{Z}_N,\ \mathcal{I} \mid \mathcal{C}\big)\Big] = 0, \tag{35}$$

and β is *identified by NIV* if it is the unique β for which (35) can hold (equivalently, if IV conditions hold for the augmented regressor block $(\mathcal{X}, \mathcal{Z}_N)$).

Outcome bridge PVR has the same algebraic form. Our outcome bridge PVR moment equation can be written as

$$\mathbb{E}\Big[\operatorname{cov}\big(Y - \beta^\top \mathcal{X} - \vartheta^\top \mathcal{W},\ \mathcal{Z} \mid \mathcal{C}\big)\Big] = 0, \tag{36}$$

which is identical in form to (35) under the correspondence $\mathcal{I} \equiv \mathcal{Z}, \quad \mathcal{Z}_N \equiv \mathcal{W}, \quad \mathcal{C} \equiv \mathcal{C}, \quad \alpha \equiv \vartheta$. Thus, *from the perspective of linear GMM/2SLS estimation*, outcome bridge PV is a nuisance-IV problem for the augmented regressor block $(\mathcal{X}, \mathcal{W})$ instrumented by $\mathcal{Z}$ after residualizing on $\mathcal{C}$.

Key difference: why does the nuisance term exist? Despite the formal similarity of (35) and (36), the underlying identifying logic is different.

- **In NIV**, the nuisance regressor $\mathcal{Z}_N$ is introduced to absorb a violation of IV validity (e.g., an instrument-to-outcome path through $\mathcal{Z}_N$). The coefficient α has the interpretation of a regression/structural coefficient on the observed nuisance regressor in the outcome equation, and the instrument $\mathcal{I}$ is required to satisfy an exclusion-type condition *for the augmented regression*.

- **In outcome bridge-PV**, the additional term $\vartheta^\top \mathcal{W}$ is not introduced to block a direct path $\mathcal{Z} \to Y$ (indeed, $\mathcal{Z}$ is excluded from the structural equation for Y in (1)). Instead, $\mathcal{Z}$ is allowed to be *confounded* (correlated with $\mathcal{U}$), and $\mathcal{W}$ is an outcome-side proxy used to *cancel the confounding channel* in the moment equation via the bridge condition. Consequently, ϑ generally mixes the direct effect of $\mathcal{W}$ on Y (through α_{YW}) and the confounding-correction component (through θ), and should not be interpreted as a causal effect of $\mathcal{W}$ on Y.

Graphs and assumptions: complementary rather than nested. NIV is formulated in terms of observed-variable graphical conditions (IV-style d-separation) for an *augmented* regressor set. In contrast, PVR is designed for settings with an *unobserved confounder* $\mathcal{U}$ and two proxy blocks $(\mathcal{W}, \mathcal{Z})$ satisfying cross-side exclusion and informativeness conditions. Accordingly, there are settings where NIV applies but PVR does not (e.g., credible instruments exist and nuisance regressors are observed), and vice versa (e.g., no credible instrument exists, but proxy blocks for $\mathcal{U}$ exist).

Rank/dimension conditions: the same linear-algebra requirement on an augmented block. Both NIV and outcome bridge-PV require that the "instrument" block has enough independent variation for an *augmented* regressor block. For NIV, this is a rank condition on $\mathbb{E}[\mathrm{cov}((\mathcal{X}, \mathcal{Z}_N), \mathcal{I} \mid \mathcal{C})]$. For outcome bridge-PV, it is Assumption 4, i.e. a rank condition on $\mathbb{E}[\mathrm{cov}((\mathcal{X}, \mathcal{W}), \mathcal{Z} \mid \mathcal{C})]$. In both cases, the moment equation is a linear system in the augmented coefficients.

Appendix C. Discussion for Section 3

C.1. Bank of hinge examples

There are seven canonical cases of possible arrows orientations that create collider-mediator between the sides. We portray them in Figure 3.

Appendix D. Discussion for Section 4

D.1. Nuisance IV versus Nuisance PV

This paper's nuisance proximal variables (NPV) are conceptually related to *nuisance instrumental variables* (NIV) in linear IV settings (Thams et al., 2024). We briefly compare the two to clarify (i) what is inherited from the IV logic and (ii) what is genuinely proximal-specific.

Nuisance IV (NIV): moment closure by adding nuisance regressors. Consider a linear SCM with outcome Y, target regressors $\mathcal{X}$, instruments $\mathcal{I}$, and a conditioning set $\mathcal{C}$. In *conditional IV* (IV), one requires that (a) $\mathcal{I}$ is conditionally exogenous for the $X \to Y$ effect after removing the direct $\mathcal{X} \to Y$ edges (graphically: $\mathcal{I}$ and Y are d-separated given $\mathcal{C}$ in the edge-deleted graph), (b) $\mathcal{C}$ contains no descendants of $\mathcal{X}$ or Y, and (c) a rank/relevance condition such as $\mathrm{rank}\{\mathbb{E}[\mathrm{cov}(\mathcal{X}, \mathcal{I} \mid \mathcal{C})]\} = d_X$. Under these conditions, β is the unique solution to the IV moment equation

$$\mathbb{E}[\mathrm{cov}(Y - \beta^\top \mathcal{X}, \mathcal{I} \mid \mathcal{C})] = 0.$$

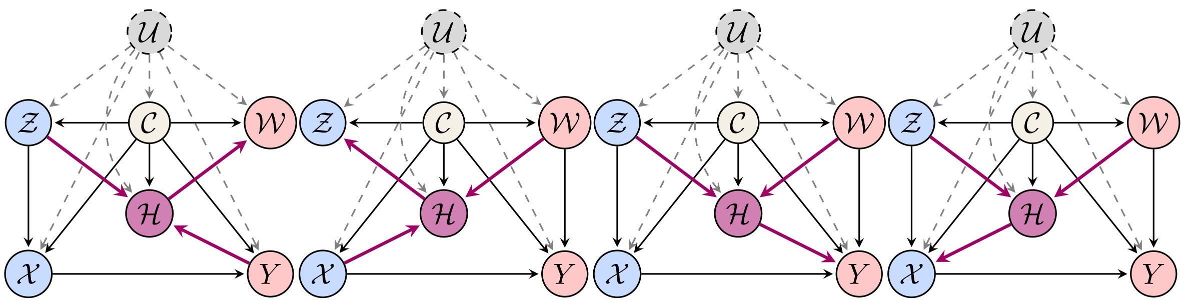
NIV extends this by allowing a *nuisance regressor* block $\mathcal{Z}$ and solving

$$\exists\, \alpha \;\; \text{s.t.} \;\; \mathbb{E}[\mathrm{cov}(Y - \beta^\top \mathcal{X} - \alpha^\top \mathcal{Z}, \mathcal{I} \mid \mathcal{C})] = 0, \tag{37}$$

with β identified if it is the unique value for which some α makes (37) hold. Operationally, NIV corresponds to applying IV to the enlarged regressor block $\tilde{N} := (\mathcal{X}^\top, \mathcal{Z}^\top)^\top$ and then extracting the component β associated with $\mathcal{X}$.

Nuisance PV (NPV-OB): the analogous closed moment system. In our outcome bridge setting, the role of the IV instrument is played by the opposite-side proxy block $\mathcal{Z}$ through moment restrictions of the form $\mathbb{E}[\mathrm{cov}(\mathcal{Z}, \cdot \mid \mathcal{C})]$ (here $\mathcal{C}$ is the proximal conditioning set). With an observed hinge $\mathcal{H}$, we form the nuisance block $\mathcal{N} := (\mathcal{W}^\top, \mathcal{H}^\top)^\top$ and obtain (Section 4.1)

$$M_{ZY} \;=\; M_{ZX}\beta \;+\; M_{ZN}\gamma^\star, \tag{38}$$

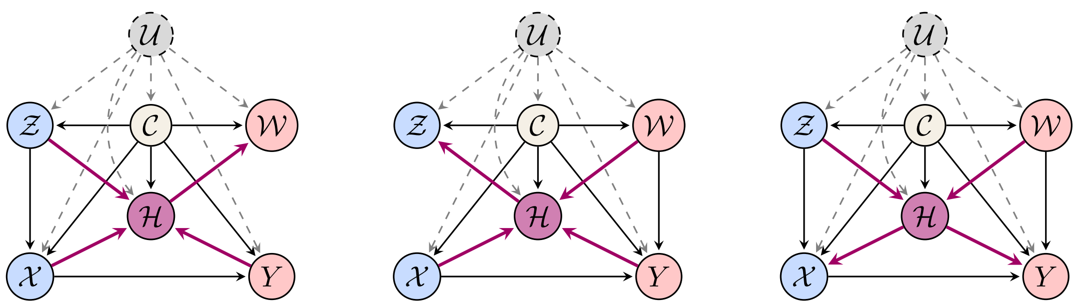


(a)
(Z, W) channel closed [open]
(Y, Z) channel open [closed]
conditioned by $\mathcal{H}$ [$\varnothing$].

(b)
(Z, W) channel closed [open]
(W, X) channel open [closed]
conditioned by $\mathcal{H}$ [$\varnothing$].

(c)
(Z, W) channel open [closed]
(Y, Z) channel closed [open]
conditioned by $\mathcal{H}$ [$\varnothing$].

(d)
(Z, W) channel open [closed]
(W, X) channel closed [open]
conditioned by $\mathcal{H}$ [$\varnothing$].



(e)
(Z, W) and (X, W) channel open [closed] (Y, Z) channel closed [open] conditioned by $\mathcal{H}$ [$\varnothing$].

(f)
(Z, W) and (Y, Z) channel open [closed] (X, W) channel closed [open] conditioned by $\mathcal{H}$ [$\varnothing$].

(g)
(Z, W) channel open [closed] (X, W) and (Y, Z) channel closed [open] conditioned by $\mathcal{H}$ [$\varnothing$].

Figure 3: Characterization of all canonical hinge cases. Purple, **bold** arrows denote hinge edges incident to $\mathcal{H}$. Dashed gray arrows indicate edges from the unobserved $\mathcal{U}$. The blue nodes represent treatment side, and the red the outcome side. Some of the black edges could have different orientation, as long as they do not create any cycles.

where $M_{ZY} := \mathbb{E}[\mathrm{cov}(\mathcal{Z}, Y \mid \mathcal{C})]$, $M_{ZX} := \mathbb{E}[\mathrm{cov}(\mathcal{Z}, \mathcal{X} \mid \mathcal{C})]$, $M_{ZN} := \mathbb{E}[\mathrm{cov}(\mathcal{Z}, \mathcal{N} \mid \mathcal{C})]$. This is the direct analogue of (37): $\mathcal{Z}$ (proxy) plays the role of $\mathcal{I}$ (instrument), and $\mathcal{N} = (\mathcal{W}, \mathcal{H})$ plays the role of the nuisance regressor block. The key subtlety is that, unlike NIV, (38) does *not* follow from a graphical exclusion alone: the proximal bridge residual contains proxy noises, so hinge mechanisms can induce $\mathbb{E}[\mathrm{cov}(\mathcal{Z}, \tilde{\varepsilon}_Y \mid \mathcal{C})] \neq 0$ even when structural errors are independent. Our Assumption 11 is exactly the condition under which the moment equation *closes* in the nuisance span: the hinge-induced violation lies in $\mathrm{col}(M_{ZN})$, so it can be absorbed into $\gamma^\star$ as in (38).

Rank/relevance: IV relevance vs. projected relevance. In NIV, identification of β is obtained from a relevance condition on the instrument–regressor covariance operator (e.g. $\mathrm{rank}\{\mathbb{E}[\mathrm{cov}(\mathcal{X}, \mathcal{I} \mid \mathcal{C})]\} = d_X$). In NPV-OB, the analogous requirement is that M_{ZX} remains injective *modulo* the nuisance span $\mathrm{col}(M_{ZN})$ (Assumption 12): after accounting for nuisance directions, there must remain enough $\mathcal{Z}$–variation that moves $\mathcal{X}$ to identify β. This is the proximal analogue of a weak-instrument/weak-identification condition: near failure corresponds to ill-conditioning and variance amplification.

What is genuinely different in PV relative to IV? The nuisance-regressor idea is shared, but PV has an additional failure mode: *noise endogeneity in bridge residuals.* In IV, the moment $\mathbb{E}[\mathrm{cov}(\mathcal{I}, \varepsilon_Y \mid \mathcal{C})] = 0$ is ensured by exclusion (d-separation + no descendant conditioning). In PV, bridge construction introduces proxy noises (e.g. $\tilde{\varepsilon}_Y = \varepsilon_Y - \theta^\top \varepsilon_W$), and hinges can make these proxy noises depend on opposite-side variables even without conditioning on $\mathcal{H}$. The nuisance-span assumptions in NPV (Assumption 11 for OB, and analogues for TB stages) should therefore be read as the proximal counterpart of "no bypass endogeneity": any hinge-induced cross-side dependence must be routed through observed nuisance variables so that it can be absorbed (as in NIV) rather than generating violations in directions orthogonal to the nuisance span.

D.2. Example of sufficient graphical condition/shared noises for Section 4.3

Recall that in NPV-OB we use the nuisance block $N := (W^\top, H^\top)^\top$ and the hinge-free (outcome bridge) rewriting $Y = \beta^\top \mathcal{X} + \gamma^\top \mathcal{N} + \delta^\top \mathcal{C} + \tilde{\varepsilon}_Y$, so that the $\mathcal{Z}$-moment equation is

$$M_{ZY} = M_{ZX}\beta + M_{ZN}\gamma + \Delta_{\text{noise}}, \qquad \Delta_{\text{noise}} := \mathbb{E}[\mathrm{cov}(\mathcal{Z}, \tilde{\varepsilon}_Y \mid \mathcal{C})].$$

Assumption 11 requires $\Delta_{\text{noise}} \in \mathrm{col}(M_{ZN})$.

Expanded-noise representation. Let $e = (e_1, \dots, e_m)^\top$ collect mutually independent noises. In a linear acyclic SCM, every C-residualized observed variable (denoted with a tilde over the variable residualized) is a linear function of e. In particular, we may write

$$\tilde{\mathcal{Z}} = B_Z e, \qquad \tilde{\mathcal{N}} = B_N e, \qquad \tilde{\varepsilon}_Y = b^\top e, \tag{39}$$

for some matrices $B_Z \in \mathbb{R}^{d_Z \times m}$, $B_N \in \mathbb{R}^{d_N \times m}$ and vector $b \in \mathbb{R}^m$. Let $\Sigma := \mathbb{E}[ee^\top]$; under independence of noises, Σ is diagonal.

Define the set of indices of shared information between $\tilde{Z}$ and $\tilde{\varepsilon}_Y$:

$$S := \{j \in [m] : (B_Z)_{:,j} \neq 0 \text{ and } b_j \neq 0\}.$$

Graphically, $j \in S$ means that the exogenous node e_j is an ancestor of both $\mathcal{Z}$ and the bridge residual noise $\tilde{\varepsilon}_Y$ in the expanded noise graph.

In the following proposition, for an index set $S \subseteq [m] := \{1, \dots, m\}$ and a vector $v \in \mathbb{R}^m$, write $v_S \in \mathbb{R}^{|S|}$ for the subvector formed by the coordinates $(v_j)_{j \in S}$ (in increasing index order). For a matrix $A \in \mathbb{R}^{p \times m}$, write $A_{:,S} \in \mathbb{R}^{p \times |S|}$ for the submatrix formed by the columns of A indexed by S. We abbreviate $A_{:,S}$ as A_S when there is no ambiguity. In particular, for $B_N \in \mathbb{R}^{d_N \times m}$ we write $B_{N,S} := (B_N)_{:,S} \in \mathbb{R}^{d_N \times |S|}$, and for $b \in \mathbb{R}^m$ we write $b_S \in \mathbb{R}^{|S|}$.

Proposition 26 (A sufficient "shared noise" condition for Assumption 11) *Assume (39) holds and Σ is diagonal with strictly positive diagonal entries on S. If there exists $\lambda \in \mathbb{R}^{d_N}$ such that*

$$b_S = B_{N,S}^\top \lambda, \tag{40}$$

then the nuisance-span condition holds:

$$\Delta_{\text{noise}} = \mathbb{E}[\text{cov}(Z, \tilde{\varepsilon}_Y \mid C)] \in \text{col}(M_{ZN}) = \text{col}\big(\mathbb{E}[\text{cov}(Z, N \mid C)]\big).$$

Proof Since $\tilde{Z} = B_Z e$, $\tilde{N} = B_N e$, and $\tilde{\varepsilon}_Y = b^\top e$,

$$M_{ZN} = \mathbb{E}[\tilde{Z}\tilde{N}^\top] = B_Z \Sigma B_N^\top, \qquad \Delta_{\text{noise}} = \mathbb{E}[\tilde{Z}\tilde{\varepsilon}_Y] = B_Z \Sigma b.$$

Because Σ is diagonal and positive on S, the identity $B_Z \Sigma b = B_Z \Sigma B_N^\top \lambda$ holds whenever $b_S = B_{N,S}^\top \lambda$ (components outside S do not contribute to $B_Z \Sigma b$ because $(B_Z)_{:,j} = 0$ there). Thus,

$$\Delta_{\text{noise}} = B_Z \Sigma b = B_Z \Sigma B_N^\top \lambda = M_{ZN} \lambda \in \text{col}(M_{ZN}),$$

which is exactly Assumption 11. ■

Corollary 27 (A rank-based sufficient condition) *Under the conditions of Proposition 26, if $B_{N,S}$ has full row rank (i.e., $\text{rank}(B_{N,S}) = |S|$), then (40) holds for* every *b_S, and hence Assumption 11 holds.*

Interpretation. Proposition 26 makes the "no bypass mechanism" intuition precise. In linear SCMs with independent noises, Δ_{noise} can only arise from *shared exogenous noises* that reach both $\mathcal{Z}$ and $\tilde{\varepsilon}_Y$ (i.e., S). Condition (40) requires that the *mixture* of these shared noises appearing in $\tilde{\varepsilon}_Y$ is linearly representable from the nuisance block $\mathcal{N} = (\mathcal{W}, \mathcal{H})$. Equivalently, every exogenous common cause of $\mathcal{Z}$ and $\tilde{\varepsilon}_Y$ must also be "visible" through $\mathcal{N}$ in a way rich enough to reproduce the shared-noise contribution to $\tilde{\varepsilon}_Y$. If there exists an exogenous noise that affects $\mathcal{Z}$ and $\tilde{\varepsilon}_Y$ but does *not* load into $\mathcal{N}$ (or loads into $\mathcal{N}$ in a way that cannot represent b_S), then Assumption 11 can fail.

A concrete hinge example where Assumption 11 is automatic (single-noise case). A particularly simple and common special case is when the only shared noise that can create $\mathcal{Z}$–$\tilde{\varepsilon}_Y$ dependence is the outcome-proxy noise ε_W (e.g., because hinge pathways make $\mathcal{Z}$ a descendant of $\mathcal{W}$-noise in reduced form, while $\tilde{\varepsilon}_Y$ contains ε_W through $\tilde{\varepsilon}_Y = \varepsilon_Y - \theta^\top \varepsilon_W$). Then S is one-dimensional. Since $\mathcal{W}$ is a component of $\mathcal{N}$, ε_W necessarily loads into $\mathcal{N}$ (through $\mathcal{W}$), so $B_{N,S}$ has rank 1 and Corollary 27 implies Assumption 11 holds. Intuitively: in the one-noise case, any $\mathcal{Z}$–$\tilde{\varepsilon}_Y$ correlation induced by ε_W must already appear in the $\mathcal{Z}$–$\mathcal{W}$ (hence $\mathcal{Z}$–$\mathcal{N}$) cross-moment, so it lies in the nuisance span.

Analogues for Assumptions 14 and 17. The same argument applies verbatim to the treatment bridge nuisance-span assumptions: replace $(\mathcal{Z}, \mathcal{N}, \tilde{\varepsilon}_Y)$ by $(\mathcal{W}, \mathcal{H}, \tilde{\varepsilon}_X)$ for Assumption 14 (Stage 1), and by $(\mathcal{V}, \mathcal{N}, \alpha_{YU}^\top U + \varepsilon_Y)$ for Assumption 17 (Stage 2), with the corresponding shared-noise sets.

Appendix E. Details of the experiments for Section 5

The simulation is parameterized to allow smooth sweeps that separately control:

1. **Projected identifying signal / weak identification:** reduces the component of $\mathcal{Z}$ that moves $\mathcal{X}$ *outside* the nuisance span generated by $(\mathcal{W}, \mathcal{H})$.
2. **Span violation (bypass dependence):** introduces an additional latent factor that affects $(\mathcal{Z}, Y)$ but is *not* transmitted through $(\mathcal{W}, \mathcal{H})$, so the hinge-induced deviation no longer lies in the nuisance span.
3. **Hinge strength (direct impact):** scales how strongly $\mathcal{H}$ enters $(\mathcal{X}, Y)$.
4. **Collider/noise-endogeneity strength:** scales how strongly the proxy blocks are coupled through $\mathcal{H}$, amplifying collider-induced cross-side noise dependence (especially when conditioning on $\mathcal{H}$).

Latent and proxy equations. Let $\mathcal{U} \sim \mathcal{N}(0, 1)$ and let all exogenous noises be jointly independent and Gaussian. We generate

$$\mathcal{W} = \mathbf{1}\mathcal{U} + \varepsilon_W, \qquad Z_{\text{par}} = \mathbf{1}\mathcal{U} + \varepsilon_{Z_{\text{par}}}.$$

The hinge is a collider between the (aggregated) proxy sides:

$$\mathcal{H} \;=\; a_W \,\overline{W} \;+\; a_Z \,\overline{Z}_{\text{par}} \;+\; \varepsilon_H,$$

where $\overline{W}$ and $\overline{Z}_{\text{par}}$ denote within-block averages (a scalar summary). We then create a hinge-descendant proxy component

$$Z_{\text{ch}} \;=\; \mathbf{1}\mathcal{U} \;+\; a_{HZ}\,\mathcal{H} \;+\; \varepsilon_{Z_{\text{ch}}},$$

and an exogenous "good" component $Z_{\text{good}} = \varepsilon_{Z_{\text{good}}}$.

Treatment and outcome. The treatment depends on U, on all $\mathcal{Z}$ components, and directly on the hinge:

$$\mathcal{X} \;=\; b_U\mathcal{U} \;+\; b_{\text{par}}\overline{Z}_{\text{par}} \;+\; b_{\text{ch}}\overline{Z}_{\text{ch}} \;+\; b_{\text{good}}\overline{Z}_{\text{good}} \;+\; b_H\mathcal{H} \;+\; \varepsilon_X.$$

The outcome is

$$Y \;=\; \beta\mathcal{X} \;+\; \gamma_U\mathcal{U} \;+\; \gamma_H\mathcal{H} \;+\; \varepsilon_Y,$$

with target β fixed (e.g. $\beta = 1$) across all experiments. We simulate each setup 200 times.

Experiment 1 (sanity check; finite-sample distributions). We fix a baseline hinge SCM (OB) in which $\mathcal{H}$ has both a direct effect and a collider role. We estimate β using all three methods and report violin plots of RMSE of $\hat{\beta} - \beta$ for $n \in \{100, 1000, 10000\}$. This experiment checks: (i) bias and variance under direct hinge impact (ignore-$\mathcal{H}$ fails), (ii) potential collider-induced noise endogeneity under conditioning (condition-$\mathcal{H}$ can fail), and (iii) consistency of nuisance PV under its assumptions (error concentrates near zero as n grows).

Experiment 2 (robustness to nuisance-span violation). Using the OB hinge SCM, we introduce the latent factor S that affects (Z_{good}, Y) but bypasses (W, H) and sweep the span-violation strength s from 0 upward. We plot mean error with ± 1 SD bands at a fixed sample size. This directly tests the necessity of the nuisance-span condition: nuisance PV should be accurate near $s = 0$ and degrade as s increases.

To test the necessity of the nuisance-span condition, we inject an additional latent factor S that induces cross-side dependence not transmitted through the observed nuisance block $(\mathcal{W}, \mathcal{H})$. Concretely, we draw $S \sim \mathcal{N}(0, 1)$ independent of all other exogenous noises and modify

$$Z_{\text{good}} \leftarrow Z_{\text{good}} + s \cdot \mathbf{1} S, \qquad Y \leftarrow Y + s \cdot \gamma_S S,$$

where $s \geq 0$ is a scalar "span-violation strength" parameter. For $s = 0$, the nuisance-span condition is satisfied in this construction; increasing s introduces a component of Δ_{noise} aligned with Z_{good} but *orthogonal* to the nuisance span generated by $(\mathcal{W}, \mathcal{H})$, so nuisance PV is expected to fail progressively.

Experiment 3 (direct hinge-strength). We scale the direct hinge coefficients $(b_H, \gamma_H) \leftarrow t \cdot (b_H, \gamma_H)$ and sweep t. This isolates the *direct-impact* component: ignore-$\mathcal{H}$ becomes more biased with t, while nuisance PV should remain stable (provided the nuisance-span condition continues to hold).

Experiment 4 (collider/noise-endogeneity). We scale the hinge collider mechanism $(a_W, a_Z, a_{HZ}) \leftarrow c \cdot (a_W, a_Z, a_{HZ})$ and sweep c. This isolates the *noise-endogeneity* component created by the hinge. Conditioning on $\mathcal{H}$ is expected to be sensitive to c because it activates collider paths, while nuisance PV remains stable as long as the induced violation is confined to the nuisance span.

Appendix F. Further discussion

F.1. Extension to non-linear settings

Relaxation of full linear-SCM assumptions. Although our exposition uses purely linear SCMs, the identification proofs for NPV are ultimately moment-based (which have no such restriction). In particular, to focus attention, NPV-OB identifies β whenever the closed conditional-covariance system $M_{ZY} = M_{ZX}\beta + M_{ZN}\gamma^\star$ holds (for some $\gamma^\star$) and M_{ZX} is injective modulo $\text{col}(M_{ZN})$; see Theorem 13. Our results extend to settings where the equation containing the target parameter is linear in $(\mathcal{X}, \mathcal{N})$ with $\mathcal{N} = (\mathcal{W}, \mathcal{H})$, while the remaining subprocesses $(\mathcal{X}, \mathcal{Z}, \mathcal{W}, \mathcal{H})$ may follow arbitrary (possibly nonlinear) mechanisms, provided the relevant second moments exist. In this "semilinear" regime the nuisance span remains the finite-dimensional subspace $\text{col}(M_{ZN})$ with $M_{ZN} = \mathbb{E}[\text{cov}(\mathcal{Z}, \mathcal{N} \mid \mathcal{C})]$. Precisely, we summarise it in Proposition 28. A fully nonparametric extension (nonlinear bridges) would likely replace $\text{col}(M_{ZN})$ by an operator range in function space and requires additional completeness/range conditions and regularization, which we leave for future work. Informally, the completeness/range condition would be analogues of span, as they would expresses that the variation of unobserved confounders is expressed by variation in proxy variables.

Proposition 28 (NPV-OB identification beyond linear SCMs (semilinear moment formulation)) *Let $(Y, \mathcal{X}, \mathcal{Z}, \mathcal{W}, \mathcal{H}, \mathcal{C})$ be i.i.d. with finite second moments and define $\mathcal{N} := (\mathcal{W}^\top, \mathcal{H}^\top)^\top$ and $M_{AB} := \mathbb{E}[\text{cov}(A, B \mid C)]$. Assume:*

*(i) (*Semilinear outcome equation*) There exist $\beta \in \mathbb{R}^{d_x}$, $\gamma \in \mathbb{R}^{d_n}$ and a measurable $r(\mathcal{C})$ such that*

$$Y = \beta^\top \mathcal{X} + \gamma^\top \mathcal{N} + r(\mathcal{C}) + \varepsilon,$$

with $\mathbb{E}[\varepsilon \mid \mathcal{C}] = 0$ (no restriction on dependence between ε and Z beyond (ii)).

*(ii) (*Nuisance-span closure*) The induced endogeneity vector $\Delta := \mathbb{E}[\mathrm{cov}(\mathcal{Z}, \varepsilon \mid \mathcal{C})]$ satisfies $\Delta \in \mathrm{col}(M_{ZN})$.*

*(iii) (*Projected relevance*) M_{ZX} is injective modulo $\mathrm{col}(M_{ZN})$, i.e., $M_{ZX} b \in \mathrm{col}(M_{ZN}) \Rightarrow b = 0$.*

Then β is uniquely identified as the unique value for which there exists $\gamma^\star$ satisfying $M_{ZY} = M_{ZX}\beta + M_{ZN}\gamma^\star$.

We discuss more explicitly two versions of setups/assumptions that lead to identification in the following. They differ in the point of treating the assumption about informativeness or bridge existence as a primitive object. Many of the proximal causal inference papers assume the existence of the bridge as the primitive assumption (see Kallus et al., 2021; Guo et al., 2025, for a discussion about assumptions used in proximal causal inference).

Semilinear SCM extension with informativeness. The linear SCM in Section 2 is a sufficient (but not necessary) structure for our moment-based identification arguments. To separate what is essential from what is merely convenient, consider the following semilinear SCM in which only the equation containing the target parameter is required to be linear.

Let $(Y, \mathcal{X}, \mathcal{Z}, \mathcal{W}, \mathcal{H}, \mathcal{C})$ be i.i.d. with finite second moments and let $\mathcal{U}$ be an unobserved confounder. Assume:

$$\mathcal{W} = m_W(\mathcal{C}) + A_{WU}\mathcal{U} + \varepsilon_W, \tag{41}$$

$$\mathcal{Z} = m_Z(\mathcal{C}) + A_{ZU}\mathcal{U} + \varepsilon_Z, \tag{42}$$

$$\mathcal{H} = f_H(\mathcal{Z}, \mathcal{W}, \mathcal{C}, \mathcal{U}, \varepsilon_H), \tag{43}$$

$$\mathcal{X} = f_X(\mathcal{Z}, \mathcal{W}, \mathcal{H}, \mathcal{C}, \mathcal{U}, \varepsilon_X), \tag{44}$$

$$Y = \beta^\top \mathcal{X} + \alpha_{YW}^\top \mathcal{W} + \alpha_{YH}^\top \mathcal{H} + m_Y(\mathcal{C}) + \alpha_{YU}^\top \mathcal{U} + \varepsilon_Y, \tag{45}$$

where m_W, m_Z, m_Y are arbitrary measurable functions and f_H, f_X are arbitrary measurable functions (subject only to acyclicity / well-definedness). Thus, the joint law of $(\mathcal{Z}, \mathcal{W}, \mathcal{H}, \mathcal{X})$ given $(\mathcal{C}, \mathcal{U})$ may be highly nonlinear.

We retain the *outcome informativeness* requirement in the same linear-algebraic form as before:

$$\exists\, \theta \in \mathbb{R}^{d_w} \text{ s.t. } \alpha_{YU} = A_{WU}^\top \theta. \tag{46}$$

Under (46), the $\mathcal{U}$-term in (45) can be rewritten using (41):

$$\alpha_{YU}^\top \mathcal{U} = \theta^\top A_{WU}\mathcal{U} = \theta^\top(\mathcal{W} - m_W(\mathcal{C}) - \varepsilon_W).$$

Substituting into (45) yields the reduced-form partially linear outcome equation

$$Y = \beta^\top \mathcal{X} + \underbrace{(\alpha_{YW} + \theta)^\top \mathcal{W} + \alpha_{YH}^\top \mathcal{H}}_{=:\gamma^\top \mathcal{N}} + \underbrace{(m_Y(\mathcal{C}) - \theta^\top m_W(\mathcal{C}))}_{=:r(\mathcal{C})} + \underbrace{(\varepsilon_Y - \theta^\top \varepsilon_W)}_{=:\varepsilon}, \tag{47}$$

where $\mathcal{N} := (\mathcal{W}^\top, \mathcal{H}^\top)^\top$ and $\mathbb{E}[\varepsilon \mid \mathcal{C}] = 0$.

Equation (47) shows that, for NPV-OB, the key requirement is partial linearity of Y in $(\mathcal{X}, \mathcal{N})$; the remaining mechanisms generating $(\mathcal{Z}, \mathcal{W}, \mathcal{H}, \mathcal{X})$ can be arbitrary. Identification then proceeds from conditional-covariance moments together with the nuisance-span and projected-relevance conditions of Section 4.1.

Semilinear extension without informativeness. An alternative (strictly weaker in terms of structural interpretation, but more agnostic about data-generating mechanisms) is to posit the semilinear representation directly, without introducing $\mathcal{U}$.

Let $(Y, \mathcal{X}, \mathcal{Z}, \mathcal{W}, \mathcal{H}, \mathcal{C})$ be i.i.d. with finite second moments and define $\mathcal{N} := (\mathcal{W}^\top, \mathcal{H}^\top)^\top$. Assume there exist (β, γ) and a measurable function $r(\mathcal{C})$ such that

$$Y = \beta^\top \mathcal{X} + \gamma^\top \mathcal{N} + r(\mathcal{C}) + \varepsilon, \qquad \mathbb{E}[\varepsilon \mid \mathcal{C}] = 0. \tag{48}$$

No structural restrictions are imposed on how $(\mathcal{Z}, \mathcal{W}, \mathcal{H}, \mathcal{X})$ are generated beyond existence of the moments used below.

Taking $\mathbb{E}[\mathrm{cov}(\mathcal{Z}, \cdot \mid \mathcal{C})]$ of (48) gives

$$M_{ZY} = M_{ZX}\beta + M_{ZN}\gamma + \Delta, \qquad \Delta := \mathbb{E}[\mathrm{cov}(\mathcal{Z}, \varepsilon \mid \mathcal{C})].$$

The NPV-OB nuisance-span condition requires $\Delta \in \mathrm{col}(M_{ZN})$, so that the system closes as $M_{ZY} = M_{ZX}\beta + M_{ZN}\gamma^\star$ for some $\gamma^\star$. Under projected relevance (M_{ZX} injective modulo $\mathrm{col}(M_{ZN})$), β is uniquely identified (Proposition 28).

Nonlinear outcome and bridges The extension to nonlinear outcome model and bridges is non-trivial and most likley requires assumption about range of a conditional expectation operator mapping the nuisance, completeness/injectivity conditions to ensure existence/uniqueness of the nonlinear bridge and estimation would require regularization to solve the ill-posedness of the intergal equation.

F.2. Practical NPV workflow and diagnostics

This section is meant to describe a practical protocol for estimation and diagnostics when fitting proximal variables regression in possible presence of hinges. First, we summarize when an observed variable $\mathcal{H}$ is expected to break standard proximal moment conditions. Second, we provide an explicit step-by-step procedure for estimating β with Nuisance Proximal Variables (NPV). Both of the algorithms at core require only linear algebra software packages (optionally any standard non-linear regression model software). Third, we discuss how to choose between the outcome and treatment bridges. Fourth we list post-estimation diagnostic tools. As these are analogues of instrumental variables diagnostic tools (but with a different interpretation), the software that is used for instrumental variables will suffice.

F.2.1. When is a hinge a problem?

Domain knowledge hints at a hinge. One could use domain knowledge to understand if the hinge is present. Hinges are most plausible when: (i) $\mathcal{H}$ is downstream of *both* proxy blocks (or their summaries), e.g. operational decisions that depend on multiple streams of information, and simultaneously (ii) $\mathcal{H}$ is upstream of (or structurally enters) $\mathcal{X}$ and/or Y, e.g. it changes allocation, measurement timing, access, or downstream processes. This is common for variables like triage/assignment rules, timing/administration variables, routing decisions, and process-control settings. That is because the proximal cross-side assumption is violated precisely when $\mathcal{H}$ has a dual role: *mediator*-like (it transmits information across proxy sides and/or into $(\mathcal{X}, Y)$) and *collider*-like (it aggregates information from both sides so that conditioning can activate cross-side dependence).

Data-driven hints at a hinge. We need to stress that because $\mathcal{U}$ is unobserved, hinges cannot be detected conclusively from observational moments alone. However, the following checks often flag hinge-like behavior (but are only heuristics at most):

1. $\mathcal{H}$ **looks collider-like:** both proxy sides help predict $\mathcal{H}$ beyond $\mathcal{C}$. Concretely, regress $\mathcal{H}$ on $(\mathcal{Z}, \mathcal{C})$ and on $(\mathcal{W}, \mathcal{C})$ (or jointly on $(\mathcal{Z}, \mathcal{W}, \mathcal{C})$). If $\mathcal{H}$ is strongly predicted by *both* sides, it is consistent with $\mathcal{H}$ being a collider/aggregator.

2. $\mathcal{H}$ **looks mediator-like:** $\mathcal{H}$ helps predict $\mathcal{X}$ and/or Y beyond $\mathcal{C}$ (and often beyond one proxy side). Regress $\mathcal{X}$ on $(\mathcal{C}, \mathcal{H})$ and Y on $(\mathcal{C}, \mathcal{H})$ (and optionally add one proxy block) and test whether $\mathcal{H}$ has non-negligible predictive power.

3. **Instability under "ignore vs condition".** Compute a baseline proximal estimate two ways: (i) *ignore* $\mathcal{H}$ (use $\mathcal{S} = \mathcal{C}$), and (ii) *condition* on $\mathcal{H}$ (use $\mathcal{S} = (\mathcal{C}, \mathcal{H})$). Large changes in $\hat{\beta}$ across (i) and (ii) are consistent with the two hinge failure modes: mediator leakage (i) and collider-induced noise endogeneity (ii).

4. **Cross-side dependence changes when adding $\mathcal{H}$.** Compute a summary dependence measure between the residualized proxy blocks, e.g. correlations between $\tilde{\mathcal{Z}}$ and $\tilde{\mathcal{W}}$ (after residualizing on $\mathcal{C}$), and compare to residualizing on $(\mathcal{C}, \mathcal{H})$. A marked increase after conditioning on $\mathcal{H}$ is consistent with collider activation; a marked decrease is consistent with mediation being blocked.

F.2.2. ALGORITHMIC STEPS: HINGE-ROBUST OUTCOME BRIDGE (NPV-OB)

NPV-OB targets the closed nuisance-augmented moment equation (Section 4.1):

$$m_{ZY} = M_{ZX}\beta + M_{ZN}\gamma^{\star}, \qquad \mathcal{N} := (\mathcal{W}^{\top}, \mathcal{H}^{\top})^{\top},$$

under the nuisance-span and projected-rank conditions (Assumptions 11 and 12).

F.2.3. ALGORITHMIC STEPS: HINGE-ROBUST TREATMENT BRIDGE (NPV-TB)

NPV-TB is two-stage (Section 4.2): Stage 1 identifies Π_Z (the coefficient used to deconfound $\mathcal{X}$) *modulo the hinge span*, and Stage 2 identifies β *modulo the nuisance span* using the generated residual $\mathcal{V}$.

Inference note. Because $\tilde{\mathcal{V}}$ is a generated regressor, standard errors should account for Stage 1 estimation. A practical approach is *stacked GMM* using the Stage 1 and Stage 2 moments jointly (analogous to IV with generated instruments), with a heteroskedastic-robust sandwich variance. Alternatively, a nonparametric bootstrap over observations is straightforward in linear implementations.

F.2.4. WHEN TO PREFER NPV-OB VS NPV-TB?

NPV-OB and NPV-TB identify the same structural target β, but they impose different *informativeness* assumptions (Assumptions 2 vs 7) and consume identifying variation in different places.

Algorithm 1 NPV-OB (hinge-robust outcome bridge)

Input: i.i.d. data $\{(Y_i, \mathcal{X}_i, \mathcal{Z}_i, \mathcal{W}_i, \mathcal{H}_i, \mathcal{C}_i)\}_{i=1}^{n}$.
Output: Estimate $\hat{\beta}$ of the causal effect β.
Step 0 (residualize on $\mathcal{C}$). Estimate $m_A(\mathcal{C}) := \mathbb{E}[A \mid \mathcal{C}]$ for each $A \in \{Y, \mathcal{X}, \mathcal{Z}, \mathcal{W}, \mathcal{H}\}$ (linear or non-linear regression), and set $\tilde{A}_i := A_i - \hat{m}_A(\mathcal{C}_i)$.
Step 1 (form the nuisance block). Set $\tilde{\mathcal{N}}_i := (\tilde{\mathcal{W}}_i^\top, \tilde{\mathcal{H}}_i^\top)^\top$.
Step 2 (estimate moment matrices). Compute sample analogs

$$\hat{m}_{ZY} := \frac{1}{n}\sum_{i=1}^{n} \tilde{\mathcal{Z}}_i \tilde{Y}_i, \quad \hat{M}_{ZX} := \frac{1}{n}\sum_{i=1}^{n} \tilde{\mathcal{Z}}_i \tilde{\mathcal{X}}_i^\top, \quad \hat{M}_{ZN} := \frac{1}{n}\sum_{i=1}^{n} \tilde{\mathcal{Z}}_i \tilde{\mathcal{N}}_i^\top.$$

Step 3 (project out the nuisance span). Form the projector onto $\mathrm{col}(\hat{M}_{ZN})^\perp$:

$$\hat{P}_N^\perp \;:=\; I_{d_z} - \hat{M}_{ZN}(\hat{M}_{ZN}^\top \hat{M}_{ZN})^\dagger \hat{M}_{ZN}^\top.$$

Step 4 (solve for β). Solve the projected system

$$\hat{P}_N^\perp \hat{m}_{ZY} \;=\; \hat{P}_N^\perp \hat{M}_{ZX}\,\hat{\beta}, \qquad \text{e.g.} \quad \hat{\beta} := (\hat{P}_N^\perp \hat{M}_{ZX})^\dagger (\hat{P}_N^\perp \hat{m}_{ZY}).$$

Prefer NPV-OB when $\mathcal{Z}$ is "large/strong" for $(\mathcal{X}, \mathcal{N})$. NPV-OB treats $\mathcal{Z}$ as the moment (instrument) block and uses $\mathcal{N} = (\mathcal{W}, \mathcal{H})$ as nuisance regressors. It tends to be favorable when: (i) d_z is large relative to d_x and the nuisance span induced by $\mathcal{N}$, (ii) the projected operator $\hat{P}_N^\perp \hat{M}_{ZX}$ is well-conditioned ("not weak" in econometrics terms), and (iii) the outcome-side informativeness assumption (existence of θ with $\alpha_{YU} = \alpha_{WU}^\top \theta$) is substantively plausible. It is also a one-shot procedure (no generated $\mathcal{V}$), so it is somewhat typically simpler to implement and diagnose.

Prefer NPV-TB when $(\mathcal{W}, \mathcal{Z})$ is "strong" but $\mathcal{Z}$ is not rich enough to span $(\mathcal{X}, \mathcal{N})$. NPV-TB uses $\mathcal{W}$ to learn Π_Z (after removing the hinge span), then uses $\mathcal{V}$ as a deconfounded treatment variation to identify β modulo nuisance. It tends to be favorable when: (i) $\hat{P}_H^\perp \hat{M}_{WZ}$ is well-conditioned (so Π_Z is learnable modulo the hinge span), (ii) $\hat{P}_{VN}^\perp \hat{M}_{VX}$ is well-conditioned (so β is learnable modulo the nuisance span), and (iii) treatment-side informativeness (existence of Ψ with $\alpha_{XU} = \Psi \alpha_{ZU}$) is more plausible than outcome-side informativeness. NPV-TB is also attractive when d_z is moderate but d_w is sufficiently rich for the Stage 1 regression.

Recommendation in practice. When both routes are feasible, report both NPV-OB and NPV-TB estimates as a stability check: substantial disagreement is a strong indication of lack of some assumptions being met (e.g. span violations, weak projected relevance, or bridge misspecification). When only one route has stable projected operators (see diagnostics below), prefer that route.

F.2.5. POST-ESTIMATION DIAGNOSTICS FOR ESTIMATION IN SECTION 4

Because NPV-OB and NPV-TB are linear (projected) moment estimators, many familiar IV diagnostics carry over, with an important interpretational change: rejections/instability indicate failure

Algorithm 2 NPV-TB (hinge-robust treatment bridge)

Input: i.i.d. data $\{(Y_i, \mathcal{X}_i, \mathcal{Z}_i, \mathcal{W}_i, \mathcal{H}_i, \mathcal{C}_i)\}_{i=1}^n$.
Output: Estimate $\hat{\beta}$ of the causal effect β.
Step 0 (residualize on $\mathcal{C}$). As in Algorithm 1, compute $\tilde{A}_i := A_i - \hat{m}_A(\mathcal{C}_i)$ for $A \in \{Y, \mathcal{X}, \mathcal{Z}, \mathcal{W}, \mathcal{H}\}$.

Stage 1: identify Π_Z.
Step 1 (estimate Stage-1 moments). Compute

$$\hat{M}_{WX} := \frac{1}{n}\sum_{i=1}^n \tilde{\mathcal{W}}_i\,\tilde{\mathcal{X}}_i^\top, \quad \hat{M}_{WZ} := \frac{1}{n}\sum_{i=1}^n \tilde{\mathcal{W}}_i\,\tilde{\mathcal{Z}}_i^\top, \quad \hat{M}_{WH} := \frac{1}{n}\sum_{i=1}^n \tilde{\mathcal{W}}_i\,\tilde{\mathcal{H}}_i^\top.$$

Step 2 (project out the hinge span in Stage 1). Form

$$\hat{P}_H^\perp \;:=\; I_{d_w} - \hat{M}_{WH}(\hat{M}_{WH}^\top \hat{M}_{WH})^\dagger \hat{M}_{WH}^\top.$$

Step 3 (solve for Π_Z). Solve the projected equation

$$\hat{P}_H^\perp \hat{M}_{WX} \;=\; \hat{P}_H^\perp \hat{M}_{WZ}\,\Pi_Z^\top, \qquad \text{e.g.} \quad \hat{\Pi}_Z^\top := (\hat{P}_H^\perp \hat{M}_{WZ})^\dagger (\hat{P}_H^\perp \hat{M}_{WX}).$$

Step 4 (construct the generated residual). Set $\tilde{\mathcal{V}}_i := \tilde{\mathcal{X}}_i - \hat{\Pi}_Z\,\tilde{\mathcal{Z}}_i$.

Stage 2: identify β from $\mathcal{V}$ modulo nuisance.
Step 5 (form the nuisance block). Set $\tilde{\mathcal{N}}_i := (\tilde{\mathcal{W}}_i^\top, \tilde{\mathcal{H}}_i^\top)^\top$.
Step 6 (estimate Stage-2 moments). Compute

$$\hat{M}_{VY} := \frac{1}{n}\sum_{i=1}^n \tilde{\mathcal{V}}_i\,\tilde{Y}_i, \quad \hat{M}_{VX} := \frac{1}{n}\sum_{i=1}^n \tilde{\mathcal{V}}_i\,\tilde{\mathcal{X}}_i^\top, \quad \hat{M}_{VN} := \frac{1}{n}\sum_{i=1}^n \tilde{\mathcal{V}}_i\,\tilde{\mathcal{N}}_i^\top.$$

Step 7 (project out the nuisance span in Stage 2). Form

$$\hat{P}_{VN}^\perp \;:=\; I_{d_x} - \hat{M}_{VN}(\hat{M}_{VN}^\top \hat{M}_{VN})^\dagger \hat{M}_{VN}^\top.$$

Step 8 (solve for β). Solve

$$\hat{P}_{VN}^\perp \hat{M}_{VY} \;=\; \hat{P}_{VN}^\perp \hat{M}_{VX}\,\hat{\beta}, \qquad \text{e.g.} \quad \hat{\beta} := (\hat{P}_{VN}^\perp \hat{M}_{VX})^\dagger (\hat{P}_{VN}^\perp \hat{M}_{VY}).$$

---

of either of (i) bridge existence (informativeness), (ii) nuisance-span closure of hinge-induced deviations, and (iii) projected relevance (rank/injectivity after removing nuisance spans). The diagnostics above are therefore best read as checks of these three ingredients rather than classical exclusion restriction violation.

1. **Overidentification tests (Hansen's J).** If the moment system is overidentified (e.g. d_z exceeds the dimension of the identified component in NPV-OB), compute the usual GMM J statistic for the sample moments $\hat{g}(\hat{\beta}, \hat{\gamma}) = \frac{1}{n}\sum_i \tilde{\mathcal{Z}}_i(\tilde{Y}_i - \hat{\beta}^\top \tilde{\mathcal{X}}_i - \hat{\gamma}^\top \tilde{\mathcal{N}}_i)$. A large J suggests that either: (i) the nuisance-span condition does not hold (the violation is not absorbed by $\mathcal{N}$), (ii) the linear bridge class is misspecified, or (iii) the residualization on $\mathcal{C}$ is inadequate.

2. **Weak-identification / conditioning diagnostics (singular values)** NPV identification is controlled by the *projected* operators:

   $$\text{NPV-OB:} \quad \hat{P}_N^{\perp}\hat{M}_{ZX}, \qquad \text{NPV-TB Stage 1:} \quad \hat{P}_H^{\perp}\hat{M}_{WZ}, \qquad \text{NPV-TB Stage 2:} \quad \hat{P}_{VN}^{\perp}\hat{M}_{VX}.$$

   Compute smallest singular values (or condition numbers) of these matrices. Near-zero singular values indicate weak identification *after* removing nuisance spans, implying inflated variance and sensitivity (direct analogue of weak instruments in IV).

3. **"First-stage" strength checks.** For NPV-TB, Stage 1 plays the role of a first stage. Report: (i) the singular values of $\hat{P}_H^{\perp}\hat{M}_{WZ}$, (ii) the stability of $\hat{\Pi}_Z$ across resamples (bootstrap), and (iii) (optionally) out-of-sample predictive performance of $\tilde{\mathcal{X}}$ using $\tilde{\mathcal{Z}}$ after controlling for $\tilde{\mathcal{H}}$ as a descriptive strength measure (not as identification proof).

4. **Moment residuals and held-out validation.** Especially when $\hat{m}_A(\cdot)$ is learned flexibly, evaluate moment residuals on held-out folds: compute $\hat{g}(\hat{\beta})$ using $\hat{\beta}$ estimated on other folds. Systematic patterns in residuals indicate misspecification or span violations.

5. **Comparative diagnostics: OB vs TB, and ignore vs condition.** Report (at least) four estimates: (i) baseline PV ignoring $\mathcal{H}$ ($\mathcal{S} = \mathcal{C}$), (ii) baseline PV conditioning on $\mathcal{H}$ ($\mathcal{S} = (\mathcal{C}, \mathcal{H})$), (iii) NPV-OB, (iv) NPV-TB. Large dispersion across these estimates is a strong sign of hinge sensitivity and/or weak projected relevance. Agreement between NPV-OB and NPV-TB is an internal consistency check.

6. **Sensitivity to proxy/nuisance block choices.** Because NPV hinges on the nuisance span generated by $\mathcal{N} = (\mathcal{W}, \mathcal{H})$, it is useful to check robustness to reasonable alternative nuisance specifications: e.g. adding observed candidates that plausibly transmit hinge-induced dependence (extra measurements or timing/process variables), or testing whether dropping/adding subsets substantially changes $\hat{\beta}$. Large sensitivity suggests that the true violation may not be captured by the current nuisance span.

F.3. Nuisance span assumptions from Section 4

We note that the nuisance-span assumption is a structural restriction and is not fully testable from observational data, in the same sense that IV exogeneity or nonparametric completeness assumptions are not fully testable.

In linear SCMs with independent exogenous shocks, Δ_{noise} can only arise from exogenous shocks that are ancestors of both Z and $\tilde{\varepsilon}_Y$. A sufficient condition for the nuisance-span assumption is that the contribution of these shared shocks to $\tilde{\varepsilon}_Y$ is representable as a linear combination of the way the same shocks load into the nuisance block $N = (W, H)$ (see Appendix D, Proposition 26 for a formal statement). Intuitively: the nuisance-span assumption is likely to hold when the only sources of Z–bridge-residual dependence are measurement/processing/operational shocks that also affect observed W or the observed hinge H (or are recorded in hinge-related metadata included in H).

Recall the emergency-care triage (Figure 1a). Let Z be prior utilization, W be admission physiology, and H be an observed triage score. If the main hinge-induced Z–bridge-residual dependence is driven by recorded aspects of the triage/measurement process (e.g., noisy vitals measurement that enters W, or recorded triage protocol features contained in H), then the shared information that induce Δ_{noise} also load into $N = (W, H)$, making the nuisance-span assumption plausible. By contrast, the assumption is likely to fail under a bypass mechanism: for example, an unrecorded system-level factor (e.g., crowding/staffing pressure or undocumented policy changes) that affects utilization proxies Z and outcomes (hence the bridge residual) but is not reflected in (W, H) can induce a component of Δ_{noise} orthogonal to $\text{col}(M_{ZN})$, in which case NPV-OB can be biased. This is exactly the failure mode we simulate in Experiment (b) (Section 5), where we inject an additional latent factor that bypasses (W, H) and observe progressive deterioration of NPV as the violation grows.

Lastly, our identification results are exact under a linear SCM and linear bridge class. If the true data-generating process is nonlinear, then the remainder term $\tilde{\varepsilon}_Y$ in Equation (21) may contain approximation error in addition to exogenous noise, and the induced deviation $\Delta_{\text{noise}} = \mathbb{E}[\text{cov}(Z, \tilde{\varepsilon}_Y \mid C)]$ may reflect both hinge-induced dependence and bridge misspecification. Such issues are common in any regression tasks. We discuss a feasible extension of our model to include non-linearities in Subsection F.1. Moreover, another practical remedy is to interpret NPV as operating in a user-chosen feature space: by augmenting W, H, Z, X with nonlinear basis expansions (or learned features of them) and applying the same projected-moment logic to these features, one can substantially reduce misspecification while retaining a regression/GMM workflow. Diagnostics based on overidentification and held-out moment residuals then serve as checks of the adequacy of the chosen feature class (see Subsection F.2.5).

Appendix G. Proofs

G.1. Proofs for Section 2

Notation and preliminaries. For random vectors A, B and covariates $\mathcal{C}$, define the conditional covariance $\text{cov}(A, B \mid \mathcal{C}) := \mathbb{E}\big[(A - \mathbb{E}[A \mid \mathcal{C}])(B - \mathbb{E}[B \mid \mathcal{C}])^\top \mid \mathcal{C}\big]$. We will repeatedly use two elementary facts:

(F1) If $f(\mathcal{C})$ is $\sigma(\mathcal{C})$-measurable, then $\text{cov}(A, f(\mathcal{C}) \mid \mathcal{C}) = 0$ almost surely.

(F2) *Tower property for orthogonality* If $\mathbb{E}[R \mid A, \mathcal{C}] = 0$, then $\mathbb{E}[(A - \mathbb{E}[A \mid \mathcal{C}])R \mid \mathcal{C}] = 0$, and hence $\mathbb{E}[\text{cov}(A, R \mid \mathcal{C})] = 0$. Indeed,

$$\mathbb{E}[(A - \mathbb{E}[A \mid \mathcal{C}])R \mid \mathcal{C}] = \mathbb{E}\big[\mathbb{E}[(A - \mathbb{E}[A \mid \mathcal{C}])R \mid A, \mathcal{C}] \mid \mathcal{C}\big] = \mathbb{E}\big[(A - \mathbb{E}[A \mid \mathcal{C}])\,\mathbb{E}[R \mid A, \mathcal{C}] \mid \mathcal{C}\big] = 0.$$

We restate error assumptions used in the proofs. We use the standard SCM mean-zero restrictions: for each structural equation, $\mathbb{E}[\varepsilon_i \mid \text{pa}(i)] = 0$ where $\text{pa}(i)$ are the parents of node i, and additionally we assume the usual independence of noise terms across equations. These imply in particular $\mathbb{E}[\varepsilon_W \mid \mathcal{Z}, \mathcal{X}, \mathcal{C}] = 0$ and $\mathbb{E}[\varepsilon_Z \mid \mathcal{W}, \mathcal{C}] = 0$ whenever the corresponding arrows are absent.

G.1.1. OUTCOME BRIDGE

Proof [Proof of Lemma 3] Assume Assumption 2, i.e., there exists $\theta \in \mathbb{R}^{d_w}$ such that $\alpha_{YU} = \alpha_{WU}^\top \theta$. Define

$$\eta := -\alpha_{WC}^\top \theta \in \mathbb{R}^{d_c}, \qquad b_Y(\mathcal{W}, \mathcal{C}) := \theta^\top \mathcal{W} + \eta^\top \mathcal{C}.$$

Using the structural equation $\mathcal{W} = \alpha_{WU}\mathcal{U} + \alpha_{WC}\mathcal{C} + \varepsilon_W$ we obtain

$$\begin{aligned}
\alpha_{YU}^\top \mathcal{U} - b_Y(\mathcal{W}, \mathcal{C}) &= (\alpha_{WU}^\top \theta)^\top \mathcal{U} - \theta^\top \mathcal{W} - \eta^\top \mathcal{C} \\
&= \theta^\top \alpha_{WU}\mathcal{U} - \theta^\top(\alpha_{WU}\mathcal{U} + \alpha_{WC}\mathcal{C} + \varepsilon_W) - \eta^\top \mathcal{C} \\
&= -\theta^\top \alpha_{WC}\mathcal{C} - \theta^\top \varepsilon_W - \eta^\top \mathcal{C} \\
&= -\theta^\top \varepsilon_W,
\end{aligned}$$

where the last equality uses $\eta = -\alpha_{WC}^\top \theta$. Therefore,

$$\mathbb{E}\Big[\alpha_{YU}^\top \mathcal{U} - b_Y(\mathcal{W}, \mathcal{C}) \mid \mathcal{Z}, \mathcal{X}, \mathcal{C}\Big] = -\theta^\top \mathbb{E}[\varepsilon_W \mid \mathcal{Z}, \mathcal{X}, \mathcal{C}] = 0,$$

by the mean-zero property of ε_W (and its independence from $(\mathcal{Z}, \mathcal{X})$ given $\mathcal{C}$). This is exactly Definition 1, so b_Y is an outcome bridge. ■

Proof [Proof of Theorem 5] By Lemma 3, an outcome bridge of the form $b_Y(\mathcal{W}, \mathcal{C}) = \theta^\top \mathcal{W} + \eta^\top \mathcal{C}$ exists. Rewrite the outcome equation (1) by adding and subtracting the bridge:

$$\begin{aligned}
Y &= \beta^\top \mathcal{X} + \alpha_{YW}^\top \mathcal{W} + \alpha_{YC}^\top \mathcal{C} + \alpha_{YU}^\top \mathcal{U} + \varepsilon_Y \\
&= \beta^\top \mathcal{X} + \alpha_{YW}^\top \mathcal{W} + \alpha_{YC}^\top \mathcal{C} + b_Y(\mathcal{W}, \mathcal{C}) + \underbrace{(\alpha_{YU}^\top \mathcal{U} - b_Y(\mathcal{W}, \mathcal{C})) + \varepsilon_Y}_{=:R_Y}.
\end{aligned}$$

Define the nuisance coefficient and the (irrelevant) covariate coefficient

$$\vartheta := \alpha_{YW} + \theta \in \mathbb{R}^{d_w}, \qquad \delta := \alpha_{YC} + \eta \in \mathbb{R}^{d_c}.$$

Then the above becomes

$$Y = \beta^\top \mathcal{X} + \vartheta^\top \mathcal{W} + \delta^\top \mathcal{C} + R_Y. \tag{49}$$

Step 1: R_Y is orthogonal to $\mathcal{Z}$ given $\mathcal{C}$. By the bridge definition (5),

$$\mathbb{E}[\alpha_{YU}^\top \mathcal{U} - b_Y(\mathcal{W}, \mathcal{C}) \mid \mathcal{Z}, \mathcal{X}, \mathcal{C}] = 0.$$

Also, by the SCM mean-zero assumption, $\mathbb{E}[\varepsilon_Y \mid \mathcal{Z}, \mathcal{X}, \mathcal{C}] = 0$. Hence $\mathbb{E}[R_Y \mid \mathcal{Z}, \mathcal{X}, \mathcal{C}] = 0$, and by the tower property, $\mathbb{E}[R_Y \mid \mathcal{Z}, \mathcal{C}] = 0$. Applying (F2) with $A = \mathcal{Z}$ and $R = R_Y$ yields

$$\mathbb{E}[\mathrm{cov}(\mathcal{Z}, R_Y \mid \mathcal{C})] = 0.$$

Step 2: derive the observable moment equation. Take conditional covariance of (49) with $\mathcal{Z}$ given $\mathcal{C}$:

$$\mathrm{cov}(\mathcal{Z}, Y \mid \mathcal{C}) = \mathrm{cov}(\mathcal{Z}, \mathcal{X} \mid \mathcal{C})\,\beta + \mathrm{cov}(\mathcal{Z}, \mathcal{W} \mid \mathcal{C})\,\vartheta + \underbrace{\mathrm{cov}(\mathcal{Z}, \delta^\top \mathcal{C} \mid \mathcal{C})}_{=0 \text{ by (F1)}} + \mathrm{cov}(\mathcal{Z}, R_Y \mid \mathcal{C}).$$

Taking expectations and using Step 1 gives

$$\mathbb{E}[\mathrm{cov}(\mathcal{Z}, Y \mid \mathcal{C})] = \mathbb{E}[\mathrm{cov}(\mathcal{Z}, \mathcal{X} \mid \mathcal{C})]\,\beta + \mathbb{E}[\mathrm{cov}(\mathcal{Z}, \mathcal{W} \mid \mathcal{C})]\,\vartheta,$$

which is (6).

Step 3: uniqueness under the rank condition. Let

$$G := \mathbb{E}[\mathrm{cov}(\mathcal{Z}, Y \mid \mathcal{C})] \in \mathbb{R}^{d_z}, \quad A := \mathbb{E}[\mathrm{cov}(\mathcal{Z}, \mathcal{X} \mid \mathcal{C})] \in \mathbb{R}^{d_z \times d_x}, \quad B := \mathbb{E}[\mathrm{cov}(\mathcal{Z}, \mathcal{W} \mid \mathcal{C})] \in \mathbb{R}^{d_z \times d_w}.$$

Then the moment system is $G = A\beta + B\vartheta$, i.e.,

$$G = [A \;\; B] \begin{bmatrix} \beta \\ \vartheta \end{bmatrix}.$$

Assumption 4 states that the matrix $[A \;\; B]$ has full column rank $d_x + d_w$. Hence its null space is trivial: if $[A \;\; B]v = 0$ then $v = 0$. Therefore, the solution (β, ϑ) to the above linear system is unique. In particular, β is uniquely identified. ■

Outcome bridge: β-only identification by projection (Appendix B.2) **Proof** [Proof of Theorem 21] From Theorem 5 (Step 2 of its proof), the population moment equation holds:

$$G = A\beta + M_W\vartheta, \qquad M_W := \mathbb{E}[\mathrm{cov}(\mathcal{Z}, \mathcal{W} \mid \mathcal{C})].$$

By construction, $P_W^\perp$ is the orthogonal projector onto $\mathrm{col}(M_W)^\perp$, hence $P_W^\perp M_W = 0$. Left-multiplying the moment equation by $P_W^\perp$ gives

$$P_W^\perp G = P_W^\perp A\beta.$$

Assumption 20 states that $P_W^\perp A$ has full column rank d_x, so the linear system above has a unique solution for β. ■

G.1.2. TREATMENT BRIDGE

Proof [Proof of Lemma 8] Assume Assumption 7, i.e., there exists $\Psi \in \mathbb{R}^{d_x \times d_z}$ such that $\alpha_{XU} = \Psi\alpha_{ZU}$. The proxy equation is $\mathcal{Z} = \alpha_{ZC}\mathcal{C} + \alpha_{ZU}\mathcal{U} + \varepsilon_Z$. Introduce $\Pi := \alpha_{XZ} + \Psi$. Define

$$\Gamma := -\Psi\alpha_{ZC} \in \mathbb{R}^{d_x \times d_c}, \qquad b_X(\mathcal{Z}, \mathcal{C}) := \Psi\mathcal{Z} + \Gamma\mathcal{C} = \Psi\mathcal{Z} - \Psi\alpha_{ZC}\mathcal{C}.$$

Then

$$\begin{aligned} \alpha_{XU}\mathcal{U} - b_X(\mathcal{Z}, \mathcal{C}) &= \alpha_{XU}\mathcal{U} - \Psi(\alpha_{ZC}\mathcal{C} + \alpha_{ZU}\mathcal{U} + \varepsilon_Z) + \Psi\alpha_{ZC}\mathcal{C} \\ &= (\alpha_{XU} - \Psi\alpha_{ZU})\mathcal{U} - \Psi\varepsilon_Z \\ &= -\Psi\varepsilon_Z, \end{aligned}$$

where the last equality uses $\alpha_{XU} = \Psi\alpha_{ZU}$. Therefore,

$$\mathbb{E}[\alpha_{XU}\mathcal{U} - b_X(\mathcal{Z}, \mathcal{C}) \mid \mathcal{W}, \mathcal{C}] = -\Psi\,\mathbb{E}[\varepsilon_Z \mid \mathcal{W}, \mathcal{C}] = 0,$$

by the SCM mean-zero assumption for ε_Z (and its independence from $(\mathcal{W}, \mathcal{C})$). This is exactly (7), so b_X is a treatment bridge. ■

Proof [Proof of Theorem 10] Assume Assumptions 7 and 9.

Part 1: identification of Π. Using $\mathcal{W} = \alpha_{WU}\mathcal{U} + \alpha_{WC}\mathcal{C} + \varepsilon_W$ and $\mathcal{X} = \alpha_{XC}\mathcal{C} + \alpha_{XU}\mathcal{U} + \alpha_{XZ}\mathcal{Z} + \varepsilon_X$, we compute conditional covariances given $\mathcal{C}$. By bilinearity of conditional covariance and because $\mathcal{C}$-measurable terms drop out,

$$\begin{aligned}\operatorname{cov}(\mathcal{W},\mathcal{X} \mid \mathcal{C}) &= \operatorname{cov}\big(\mathcal{W}, \alpha_{XU}\mathcal{U} + \alpha_{XZ}\mathcal{Z} + \varepsilon_X \mid \mathcal{C}\big) \\ &= \operatorname{cov}(\mathcal{W},\mathcal{U} \mid \mathcal{C})\, \alpha_{XU}^\top + \operatorname{cov}(\mathcal{W},\mathcal{Z} \mid \mathcal{C})\, \alpha_{XZ}^\top + \underbrace{\operatorname{cov}(\mathcal{W},\varepsilon_X \mid \mathcal{C})}_{=0}.\end{aligned}$$

Similarly, since $\mathcal{Z} = \alpha_{ZC}\mathcal{C} + \alpha_{ZU}\mathcal{U} + \varepsilon_Z$,

$$\operatorname{cov}(\mathcal{W},\mathcal{Z} \mid \mathcal{C}) = \operatorname{cov}(\mathcal{W},\mathcal{U} \mid \mathcal{C})\, \alpha_{ZU}^\top,$$

because the $\mathcal{C}$ term vanishes and the noise term vanishes by independence/mean-zero. Using treatment informativeness $\alpha_{XU} = \Psi\alpha_{ZU}$, i.e. $\alpha_{XU}^\top = \alpha_{ZU}^\top\Psi^\top$, we obtain

$$\operatorname{cov}(\mathcal{W},\mathcal{U} \mid \mathcal{C})\, \alpha_{XU}^\top = \operatorname{cov}(\mathcal{W},\mathcal{U} \mid \mathcal{C})\, \alpha_{ZU}^\top\Psi^\top = \operatorname{cov}(\mathcal{W},\mathcal{Z} \mid \mathcal{C})\, \Psi^\top.$$

Substituting into the first display yields

$$\operatorname{cov}(\mathcal{W},\mathcal{X} \mid \mathcal{C}) = \operatorname{cov}(\mathcal{W},\mathcal{Z} \mid \mathcal{C})\, (\Psi^\top + \alpha_{XZ}^\top) = \operatorname{cov}(\mathcal{W},\mathcal{Z} \mid \mathcal{C})\, \Pi^\top,$$

where $\Pi := \alpha_{XZ} + \Psi$. Taking expectations over $\mathcal{C}$ gives the population moment equation

$$\mathbb{E}[\operatorname{cov}(\mathcal{W},\mathcal{X} \mid \mathcal{C})] = \mathbb{E}[\operatorname{cov}(\mathcal{W},\mathcal{Z} \mid \mathcal{C})]\, \Pi^\top. \tag{50}$$

If $\mathbb{E}[\operatorname{cov}(\mathcal{W},\mathcal{Z} \mid \mathcal{C})]$ has full column rank (Assumption 9(i)), then Π is uniquely identified from (50).

Part 2: identification of β via $\mathcal{V}$. Define $b_X(\mathcal{Z},\mathcal{C}) := \Pi\mathcal{Z} - \Pi\alpha_{ZC}\mathcal{C}$ and $\mathcal{V} := \mathcal{X} - b_X(\mathcal{Z},\mathcal{C})$. Next $\mathcal{Z} = \alpha_{ZC}\mathcal{C} + \alpha_{ZU}\mathcal{U} + \varepsilon_Z$, hence $b_X(\mathcal{Z},\mathcal{C}) = \Pi\alpha_{ZU}\mathcal{U} + \Pi\varepsilon_Z$. Using also $\mathcal{X} = \alpha_{XC}\mathcal{C} + \alpha_{XU}\mathcal{U} + \alpha_{XZ}\mathcal{Z} + \varepsilon_X$ and substituting $\mathcal{Z}$ gives

$$\mathcal{V} = (\alpha_{XC} + \alpha_{XZ}\alpha_{ZC})\mathcal{C} + \big(\alpha_{XU} + \alpha_{XZ}\alpha_{ZU} - \Pi\alpha_{ZU}\big)\mathcal{U} + \varepsilon_X + (\alpha_{XZ} - \Pi)\varepsilon_Z.$$

Since $\Pi = \alpha_{XZ} + \Psi$ and $\alpha_{XU} = \Psi\alpha_{ZU}$, the $\mathcal{U}$-term cancels: $\alpha_{XU} + \alpha_{XZ}\alpha_{ZU} - \Pi\alpha_{ZU} = 0$. Moreover, $\alpha_{XZ} - \Pi = -\Psi$, so

$$\mathcal{V} = (\alpha_{XC} + \alpha_{XZ}\alpha_{ZC})\mathcal{C} + \varepsilon_X - \Psi\varepsilon_Z.$$

In particular, conditional on $\mathcal{C}$, $\mathcal{V}$ is independent of $\mathcal{U}$, $\mathcal{W}$, and ε_Y by noise independence. Therefore, $\operatorname{cov}(\mathcal{V},\mathcal{U} \mid \mathcal{C}) = 0$, $\operatorname{cov}(\mathcal{V},\mathcal{W} \mid \mathcal{C}) = 0$, and $\operatorname{cov}(\mathcal{V},\varepsilon_Y \mid \mathcal{C}) = 0$ almost surely.

Taking conditional covariance of the outcome equation (1) with $\mathcal{V}$ given $\mathcal{C}$ yields

$$\operatorname{cov}(\mathcal{V}, Y \mid \mathcal{C}) = \operatorname{cov}(\mathcal{V},\mathcal{X} \mid \mathcal{C})\, \beta.$$

Taking expectations gives the population moment equation

$$\mathbb{E}[\operatorname{cov}(\mathcal{V}, Y \mid \mathcal{C})] = \mathbb{E}[\operatorname{cov}(\mathcal{V},\mathcal{X} \mid \mathcal{C})]\, \beta.$$

If $\mathbb{E}[\operatorname{cov}(\mathcal{V},\mathcal{X} \mid \mathcal{C})]$ is nonsingular (Assumption 9(ii)), then β is uniquely identified. This concludes. ■

G.2. Proofs for Section 4

G.2.1. OUTCOME BRIDGE

Lemma 29 (Projected rank vs. injectivity modulo a span) *Let $A \in \mathbb{R}^{m\times d}$ and $B \in \mathbb{R}^{m\times k}$, and let $P_B^\perp := I_m - B(B^\top B)^\dagger B^\top$ be the orthogonal projector onto $\mathrm{col}(B)^\perp$. Then the following are equivalent:*

1. $\mathrm{rank}(P_B^\perp A) = d$ *(full column rank);*
2. $Ab \in \mathrm{col}(B) \;\Rightarrow\; b = 0$ *for all $b \in \mathbb{R}^d$ (injectivity of A modulo $\mathrm{col}(B)$);*
3. $\mathrm{col}(A) \cap \mathrm{col}(B) = \{0\}$ *after mapping through A (equivalently, no nontrivial linear combination of columns of A lies in $\mathrm{col}(B)$).*

Proof For any $v \in \mathbb{R}^m$, $P_B^\perp v = 0$ if and only if $v \in \mathrm{col}(B)$, since $P_B^\perp$ projects onto $\mathrm{col}(B)^\perp$. Thus $P_B^\perp Ab = 0$ iff $Ab \in \mathrm{col}(B)$. Therefore $\ker(P_B^\perp A) = \{0\}$ iff $Ab \in \mathrm{col}(B) \Rightarrow b = 0$. Finally, $\ker(P_B^\perp A) = \{0\}$ is equivalent to $\mathrm{rank}(P_B^\perp A) = d$. ■

Proof [Proof of Theorem 13] Start from (22) and define

$$m_{ZY} := \mathbb{E}[\mathrm{cov}(\mathcal{Z}, Y \mid \mathcal{C})], \quad M_{ZX} := \mathbb{E}[\mathrm{cov}(\mathcal{Z}, \mathcal{X} \mid \mathcal{C})], \quad \mathcal{N} := (\mathcal{W}^\top, \mathcal{H}^\top)^\top, \quad M_{ZN} := \mathbb{E}[\mathrm{cov}(\mathcal{Z}, \mathcal{N} \mid \mathcal{C})].$$

Stack $\gamma := (\vartheta^\top, \alpha_{bH}^\top)^\top$ and write the moment as

$$m_{ZY} = M_{ZX}\beta + M_{ZN}\gamma + \Delta_{\text{noise}}, \qquad \Delta_{\text{noise}} := \mathbb{E}[\mathrm{cov}(\mathcal{Z}, \tilde{\varepsilon}_Y \mid \mathcal{C})].$$

By Assumption 11 there exists λ such that $\Delta_{\text{noise}} = M_{ZN}\lambda$, so the equation closes as

$$m_{ZY} = M_{ZX}\beta + M_{ZN}\gamma^\star, \qquad \gamma^\star := \gamma + \lambda.$$

We now prove uniqueness of β. Suppose $(\beta_1, \gamma_1^\star)$ and $(\beta_2, \gamma_2^\star)$ both satisfy the closed system. Subtracting gives

$$M_{ZX}(\beta_1 - \beta_2) = M_{ZN}(\gamma_2^\star - \gamma_1^\star) \in \mathrm{col}(M_{ZN}).$$

Let $b := \beta_1 - \beta_2$. Then $M_{ZX}b \in \mathrm{col}(M_{ZN})$. Assumption 12 (equivalently, by Lemma 29, $\mathrm{rank}(P_N^\perp M_{ZX}) = d_x$) implies $b = 0$, hence $\beta_1 = \beta_2$.

Therefore β is uniquely identified. Equivalently, left-multiplying the closed system by $P_N^\perp$ eliminates $M_{ZN}\gamma^\star$ and yields the projected equation $P_N^\perp m_{ZY} = P_N^\perp M_{ZX}\beta$, which has a unique solution under Assumption 12. ■

G.2.2. TREATMENT BRIDGE

Proof [Proof of Proposition 16] From (25), define

$$M_{WX} := \mathbb{E}[\mathrm{cov}(\mathcal{W}, \mathcal{X} \mid \mathcal{C})], \quad M_{WZ} := \mathbb{E}[\mathrm{cov}(\mathcal{W}, \mathcal{Z} \mid \mathcal{C})], \quad M_{WH} := \mathbb{E}[\mathrm{cov}(\mathcal{W}, \mathcal{H} \mid \mathcal{C})],$$

so that

$$M_{WX} = M_{WZ}\Pi_Z^\top + M_{WH}\kappa_H^\top + \Delta_{TB,1}, \qquad \Delta_{TB,1} := \mathbb{E}[\mathrm{cov}(W, \tilde{\varepsilon}_X \mid \mathcal{C})].$$

By Assumption 14, there exists Λ_1 with $\Delta_{TB,1} = M_{WH}\Lambda_1$, hence the first stage closes as

$$M_{WX} = M_{WZ}\Pi_Z^\top + M_{WH}\Gamma_1, \qquad \Gamma_1 := \kappa_H^\top + \Lambda_1.$$

To show uniqueness of Π_Z, suppose $(\Pi_Z^{(1)}, \Gamma_1^{(1)})$ and $(\Pi_Z^{(2)}, \Gamma_1^{(2)})$ both satisfy the closed system. Subtracting gives

$$M_{WZ}\big(\Pi_Z^{(1)} - \Pi_Z^{(2)}\big)^\top = M_{WH}\big(\Gamma_1^{(2)} - \Gamma_1^{(1)}\big) \in \mathrm{col}(M_{WH}).$$

Thus for each column b of $\big(\Pi_Z^{(1)} - \Pi_Z^{(2)}\big)^\top$ we have $M_{WZ}b \in \mathrm{col}(M_{WH})$. Assumption 15 (equivalently, $\mathrm{rank}(P_H^\perp M_{WZ}) = d_z$ by Lemma 29) implies $b = 0$ for all columns, hence $\Pi_Z^{(1)} = \Pi_Z^{(2)}$.

Therefore Π_Z is uniquely identified. ■

Proof [Proof of Theorem 19] *Step 1 (identify Π_Z).* By Proposition 16, under Assumptions 14 and 15 the closed Stage-1 system identifies Π_Z uniquely.

Step 2 (construct $\mathcal{V}$). Define $\mathcal{V} := \mathcal{X} - \Pi_Z\mathcal{Z}$ (with $\mathcal{C}$-residualization understood). From (24),

$$\mathcal{V} = \kappa_H\mathcal{H} + \Gamma\mathcal{C} + \tilde{\varepsilon}_X.$$

Under Assumption 7, $\tilde{\varepsilon}_X = \varepsilon_X - \Psi_Z\varepsilon_Z$ contains no *direct* U component. We do not assume $\mathbb{E}[\mathrm{cov}(\mathcal{V}, U \mid \mathcal{C})] = 0$ in the hinge setting; any remaining U-dependence of $\mathcal{V}$ is absorbed into the second-stage deviation term below.

Step 3 (second-stage closed moment system). Taking $\mathbb{E}[\mathrm{cov}(\mathcal{V}, \cdot \mid \mathcal{C})]$ of the outcome equation yields

$$M_{VY} = M_{VX}\beta + M_{VN}\gamma + \Delta_{TB,2},$$

where $M_{VY} := \mathbb{E}[\mathrm{cov}(\mathcal{V}, Y \mid \mathcal{C})]$, $M_{VX} := \mathbb{E}[\mathrm{cov}(\mathcal{V}, \mathcal{X} \mid \mathcal{C})]$, $M_{VN} := \mathbb{E}[\mathrm{cov}(\mathcal{V}, \mathcal{N} \mid \mathcal{C})]$, $\mathcal{N} := (\mathcal{W}^\top, \mathcal{H}^\top)^\top$, $\gamma := (\alpha_{YW}^\top, \alpha_{YH}^\top)^\top$, and

$$\Delta_{TB,2} := \mathbb{E}\big[\mathrm{cov}(\mathcal{V}, \alpha_{YU}^\top U + \varepsilon_Y \mid \mathcal{C})\big].$$

By Assumption 17, $\Delta_{TB,2} \in \mathrm{col}(M_{VN})$, so there exists λ_2 such that $\Delta_{TB,2} = M_{VN}\lambda_2$ and hence the equation closes as

$$M_{VY} = M_{VX}\beta + M_{VN}\gamma^\star, \qquad \gamma^\star := \gamma + \lambda_2.$$

Step 4 (uniqueness of β). Suppose $(\beta_1, \gamma_1^\star)$ and $(\beta_2, \gamma_2^\star)$ satisfy the closed second-stage system. Subtracting gives

$$M_{VX}(\beta_1 - \beta_2) = M_{VN}(\gamma_2^\star - \gamma_1^\star) \in \mathrm{col}(M_{VN}).$$

Let $b := \beta_1 - \beta_2$. Then $M_{VX}b \in \mathrm{col}(M_{VN})$. Assumption 18 (equivalently, $\mathrm{rank}(P_{N,V}^\perp M_{VX}) = d_x$) implies $b = 0$, hence $\beta_1 = \beta_2$. Therefore β is uniquely identified. ■

G.3. Proofs for Section F

Proof [Proof of Proposition 28] Let $\tilde{A} := A - \mathbb{E}[A \mid C]$ denote C-residuals. Then for any random vectors A, B with finite second moments,

$$M_{AB} := \mathbb{E}[\mathrm{cov}(A, B \mid C)] = \mathbb{E}[\tilde{A}\,\tilde{B}^\top].$$

Step 1: Derive the implied Z-moment equation. By Assumption (i), there exist (β, γ, r) such that

$$Y = \beta^\top X + \gamma^\top N + r(C) + \varepsilon, \qquad \mathbb{E}[\varepsilon \mid C] = 0.$$

Residualizing on C gives

$$\tilde{Y} = \beta^\top \tilde{X} + \gamma^\top \tilde{N} + \tilde{\varepsilon}, \qquad \tilde{\varepsilon} := \varepsilon - \mathbb{E}[\varepsilon \mid C] = \varepsilon,$$

where the last equality uses $\mathbb{E}[\varepsilon \mid C] = 0$.

Multiply by $\tilde{Z}$ and take expectations:

$$\mathbb{E}[\tilde{Z}\tilde{Y}] = \mathbb{E}[\tilde{Z}\tilde{X}^\top]\,\beta + \mathbb{E}[\tilde{Z}\tilde{N}^\top]\,\gamma + \mathbb{E}[\tilde{Z}\,\tilde{\varepsilon}].$$

Recognizing the definitions,

$$M_{ZY} = M_{ZX}\beta + M_{ZN}\gamma + \Delta, \qquad \Delta := \mathbb{E}[\tilde{Z}\,\tilde{\varepsilon}] = \mathbb{E}[\mathrm{cov}(Z, \varepsilon \mid C)].$$

Step 2: Use nuisance-span closure to obtain a closed system. By Assumption (ii), $\Delta \in \mathrm{col}(M_{ZN})$. Hence there exists λ such that $\Delta = M_{ZN}\lambda$. Define $\gamma^\star := \gamma + \lambda$. Then

$$M_{ZY} = M_{ZX}\beta + M_{ZN}\gamma^\star,$$

so β is feasible for the closed moment system.

Step 3: Prove uniqueness of β. Suppose β_1, β_2 are two values such that there exist $\gamma_1^\star, \gamma_2^\star$ with

$$M_{ZY} = M_{ZX}\beta_1 + M_{ZN}\gamma_1^\star \quad \text{and} \quad M_{ZY} = M_{ZX}\beta_2 + M_{ZN}\gamma_2^\star.$$

Subtracting yields

$$M_{ZX}(\beta_1 - \beta_2) = M_{ZN}(\gamma_2^\star - \gamma_1^\star) \in \mathrm{col}(M_{ZN}).$$

By Assumption (iii) (injectivity of M_{ZX} modulo $\mathrm{col}(M_{ZN})$), this implies $\beta_1 - \beta_2 = 0$, hence $\beta_1 = \beta_2$.

Therefore β is uniquely identified as the unique value for which there exists $\gamma^\star$ satisfying $M_{ZY} = M_{ZX}\beta + M_{ZN}\gamma^\star$. ∎