Semiparametric Inference for Half-Trek Estimators in Linear Structural Equation Models

Leopold Mareis LEOPOLD.MAREIS@TUM.DE
Technical University of Munich
Munich, Germany

Nils Sturma NILS.STURMA@EPFL.CH
École Polytechnique Fédérale de Lausanne
Lausanne, Switzerland

Mathias Drton MATHIAS.DRTON@TUM.DE
Munich Center for Machine Learning
Technical University of Munich
Munich, Germany

Abstract

Linear structural equation models on directed mixed graphs encode causal relationships among variables subject to latent confounding. The half-trek criterion (HTC) provides a graphical sufficient condition for the structural coefficients to be rationally identifiable from the observable covariance matrix, and yields a corresponding closed-form rational estimator. Despite this, the asymptotic distribution of the HTC estimator, and hence valid standard errors and confidence regions, have not been derived. We derive the semiparametric influence function of this estimator for all HTC-identified directed mixed graphs, including cyclic ones. The influence function combines the structural residual at the target node with the identification instruments, recursively corrected for uncertainty from earlier estimation stages. The HTC estimator is asymptotically normal with variance computable in closed form, yielding confidence regions, marginal intervals, and Wald tests for individual structural coefficients. Applied to the Fulton Fish Market dataset, our theory delivers a complete inferential summary for the causal effect of supply on demand.

Keywords: Asymptotic inference, Instrumental variables, Half-trek criterion, Linear SEM, Semiparametric theory

1. Introduction

Linear structural equation models (SEMs) on directed mixed graphs are a central tool in causal inference, valued for the direct interpretability of their structural coefficients and their ability to encode latent confounding through bidirected edges (Bollen, 1989; Spirtes et al., 2000; Maathuis et al., 2019). In such a model, a random vector $X = (X_v)_{v \in V}$ satisfies

$$X_v = \sum_{w \in \mathrm{pa}(v)} \beta_{wv} X_w + \varepsilon_v, \quad v \in V,$$

according to a directed mixed graph $G = (V, D, B)$, where errors ε_v and ε_w may be dependent whenever $v \leftrightarrow w \in B$. The fundamental question is whether β is identifiable from the observed covariance matrix Σ; once it is, a natural second question arises: what is the asymptotic distribution of the resulting estimator, and how can it be used for inference? This paper answers the second question for the half-trek criterion (HTC) estimator.

The half-trek criterion. The classical approach to identification under confounding is the instrumental variable (IV) estimator of Wright (1928) and Bowden and Turkington (1985). In the graph $X_1 \to X_2 \to X_3$, $X_2 \leftrightarrow X_3$, the coefficient β_{23} is unidentified by regression but recovered by $\beta_{23} = \mathrm{E}[X_1 X_3] / \mathrm{E}[X_1 X_2]$, because X_1 is correlated with X_2 yet uncorrelated with ε_3. A sequence of graphical criteria for linear SEMs extended this idea to progressively richer confounding structures (Brito and Pearl, 2006; Tian, 2009; Foygel et al., 2012; Chen et al., 2016, 2017; Kumor et al., 2019; Barber et al., 2022), each expressing identified parameters as rational functions of the covariance matrix Σ. Among these, the HTC of Foygel et al. (2012) is the first polynomial-time constructive criterion applicable to arbitrary directed mixed graphs, including cyclic ones, without restrictions on the error-covariance structure: for each node v, it locates a *witness set* Y_v and identifies $\beta_v = (\beta_{wv})_{w \in \mathrm{pa}(v)}$, the structural coefficients of edges into v, by solving the IV system $A_v \beta_v = b_v$, where both A_v and b_v are computable from Σ. In particular, $A_v = \mathrm{E}[Z_{Y_v} X_{\mathrm{pa}(v)}^\top]$ is an instrument-relevance matrix, where, for each $y \in Y_v$, the instrument Z_y is either equal to the raw variable X_y, or equal to its structural residual $\varepsilon_y = X_y - \sum_{q \in \mathrm{pa}(y)} \beta_{qy} X_q$ when β_y has been identified at an earlier stage. The role of y is determined by a graphical reachability condition on G formalised in Section 2. This iterative construction gives an *HTC ordering* $\prec$. While the HTC is not necessary for identifiability, complete criteria via Gröbner basis computations exist (García-Puente et al., 2010; Hollering et al., 2026).

The inferential gap. Despite yielding a closed-form rational estimator, the HTC provides no asymptotic distribution, no standard errors, no confidence regions, and no such theory has been available. Answering the inferential question requires two advances absent from classical single-stage IV: a joint matrix delta-method expansion handling all parents of v simultaneously, and recursive correction terms carried through $\prec$ that account for estimation uncertainty in earlier stages $\hat{\beta}_y$. We therefore address the following question: given HTC identification of β_v, what is the asymptotic distribution of $\hat{\beta}_v$?

Semiparametric inference for structural coefficients. Semiparametric efficiency for IV estimation was characterized by Chamberlain (1987) and Newey (1990), and efficient estimators for general nonlinear causal effects are studied in Jung et al. (2021). For linear confounded graphs, Mareis and Drton (2026) derive the efficient influence function of the front-door estimator. In the confounding-free setting, Witte et al. (2020) and Henckel et al. (2022) characterize variance-optimal adjustment sets for total effect estimation, and Henckel et al. (2024) extend this to acyclic directed mixed models. For which directed mixed graphs the semiparametric efficiency bound is attainable within the HTC framework remains open: in the graph $1 \to 3 \to 4$, $2 \to 3 \leftrightarrow 4$, the bound for β_{34} is achieved by an estimator using both instruments simultaneously, a bound no single-witness HTC estimator attains. We therefore establish asymptotic normality of the HTC estimator as a foundation for inference, working within the semiparametric framework of van der Vaart (1998) and Tsiatis (2006) and imposing only finite fourth moments and differentiability in quadratic mean at the truth.

Contributions. We establish three results.

(i) **Influence function.** For any HTC-identified directed mixed graph, including cyclic ones, the semiparametric influence function of $\hat{\beta}_v$ is $\phi_{\beta_v} = A_v^{-1} R_v$, where the vector R_v

collects the row residuals of the identification system with recursive corrections from earlier estimation stages (Theorem 7).

(ii) **Asymptotic normality.** The asymptotic variance $\mathcal{V}_v = A_v^{-1}\mathrm{E}\big[R_v R_v^\top\big] A_v^{-\top}$ is expressed via finitely many cross-moments computable by descent through $\prec$, and reduces to the standard 2SLS variance for the classical IV graph (Proposition 9).

(iii) **Inference.** Confidence regions, marginal intervals, and Wald tests are provided, implemented in `R` with output mirroring `summary.lm()` (R Core Team, 2024). Applied to the Fulton Fish Market (Graddy, 1995; Angrist et al., 2000), the theory delivers a complete inferential summary for the causal effect of supply on demand in a simultaneous-equations model with correlated errors.

Organisation. Section 2 introduces the statistical model and the HTC estimator. Section 3 derives ϕ_{β_v}, establishes asymptotic normality, and provides the closed-form asymptotic variance $\mathcal{V}_v$. Section 4 reports simulation evidence for calibration of the estimator $\hat{\mathcal{V}}_v$ across Gaussian and non-Gaussian errors, and examines the effect of witness set choice on estimation variance. Section 5 translates the limit theory into valid inference, presents the `R` implementation, and illustrates the complete workflow on the Fulton Fish Market dataset. Regularity conditions are stated in Appendix A and are in force throughout.

2. HTC Identification and Estimation

2.1. Directed Mixed Graphs and Half-Treks

Let $G = (V, D, B)$ be a directed mixed graph on the finite vertex set V with directed edge set $D \subseteq V^2$ and bidirected edge set $B \subseteq \binom{V}{2}$. We also denote the directed edges $(v, w) \in D$ by $v \to w$ and the bidirected edges $\{v, w\} \in B$ by $v \leftrightarrow w$. Moreover, we assume that neither the directed part nor the bidirected part contain self-loops, that is, $v \to v \notin D$ and $v \leftrightarrow v \notin B$ for all $v \in V$. The *bidirected neighbours* of $W \subseteq V$ are $N_B(W) := \{k \in V \setminus W : k \leftrightarrow w \in B,\, w \in W\}$. Define the *parents* of node v as $\mathrm{pa}(v) = \{w \in V : w \to v \in D\}$ and its *siblings* as $\mathrm{sib}(v) = \{w \in V : w \leftrightarrow v \in B\}$. A *half-trek* is a path of the form

$$v_0 \to v_1 \to \cdots \to v_\ell \qquad \text{or} \qquad v_0 \leftrightarrow v_1 \to \cdots \to v_\ell,$$

with $\mathrm{Left}(\pi) = \{v_0\}$ and $\mathrm{Right}(\pi) = \{v_1, \dots, v_\ell\}$. The *half-trek reachable set* of v is

$$\mathrm{htr}(v) := \big\{w \in V \setminus (\{v\} \cup \mathrm{sib}(v)) : \text{there exists a half-trek from } v \text{ to } w\big\}.$$

A system of half-treks has *no sided intersection* if $\mathrm{Left}(\pi_i) \cap \mathrm{Left}(\pi_j) = \emptyset = \mathrm{Right}(\pi_i) \cap \mathrm{Right}(\pi_j)$ for all $i \neq j$.

2.2. The Semiparametric Model

The parameter matrix β is the finite-dimensional parameter of interest, while the error distribution ε is infinite-dimensional nuisance, making the model semiparametric. We write L_0^p for the space of mean-zero $\mathbb{R}^{|V|}$-valued random vectors with finite pth moment, so $h \in L_0^p$ if $\mathrm{E}[h] = 0$ and $\mathrm{E}[\|h\|^p] < \infty$.

Definition 1 (Admissible parameters and nuisance) *Let $G = (V, D, B)$ be a directed mixed graph. The set of admissible parameter matrices is*

$$\mathcal{B} = \{\beta \in \mathbb{R}^{|V|\times|V|} : \beta_{wv} = 0 \text{ whenever } w \to v \notin D, \text{ and } \det(I - \beta) \neq 0\}.$$

For acyclic graphs, $\det(I - \beta) \neq 0$ *holds automatically; for cyclic graphs it is a non-trivial constraint. The nuisance space $\mathcal{E}$ collects random vectors satisfying the* connected set Markov property*, also called the* marginal independence model*:*

$$\mathcal{E} = \left\{\varepsilon \in L_0^4 \,\middle|\, \varepsilon_W \perp\!\!\!\perp \varepsilon_{V\setminus(W\cup N_B(W))} \text{ for every } W \subseteq V \text{ connected in } (V, B)\right\}.$$

Definition 2 (Linear model) *The* linear model *of a directed mixed graph G is the set of probability distributions $\mathcal{M}_G = \left\{P_{(\beta,\varepsilon)} : \beta \in \mathcal{B},\ \varepsilon \in \mathcal{E}\right\}$, where $P_{(\beta,\varepsilon)}$ denotes the distribution of a random vector X solving the equation system $X = \beta^\top X + \varepsilon$, that is, $X = (I - \beta)^{-\top}\varepsilon$.*

The estimation target is the parameter $\beta \in \mathcal{B}$. Based on i.i.d. samples $(X^{(i)})_{i\in[n]}$ of an unknown distribution $P_0 \in \mathcal{M}_G$, a *regular asymptotically linear* (RAL) estimator $\hat{\beta}_n$ satisfies

$$\sqrt{n}(\hat{\beta}_n - \beta) = \frac{1}{\sqrt{n}}\sum_{i=1}^{n} \phi_\beta(X^{(i)}) + o_{P_0}(1) \xrightarrow{d} \mathcal{N}\left(0, \mathrm{E}\left[\phi_\beta\phi_\beta^\top\right]\right)$$

for some mean-zero, square-integrable *influence function* ϕ_β. For any parametric path $\gamma \mapsto P_\gamma \in \mathcal{M}_G$ through P_0 that is differentiable in quadratic mean at $\gamma = 0$ (Assumption A.(i)), with score $S \in L_0^2$ as in A.(i) and induced parameter curve $\beta_\gamma \in \mathcal{B}$, the influence function satisfies $\mathbb{E}[\phi_\beta S] = \frac{d}{d\gamma}\big|_{\gamma=0}\beta_\gamma$ for every such submodel (van der Vaart, 1998, §25.3). Differentiating any population identifying equation for β along $\gamma \mapsto P_\gamma$ recovers ϕ_β. In Sections 2.3–3.2 we derive the influence function of the HTC estimator of β_v.

2.3. Identification and Instruments

Definition 3 (Half-trek criterion) *A set $Y \subset V$ satisfies the* half-trek criterion *(HTC) for $v \in V$ if*

(i) $|Y| = |\mathrm{pa}(v)|$,

(ii) $Y \cap (\{v\} \cup \mathrm{sib}(v)) = \emptyset$, *and*

(iii) there exists a system of half-treks from Y to $\mathrm{pa}(v)$ *with no sided intersections.*

A set $Y \subset V$ satisfying the HTC for v is called a witness set *for v; an element $y \in Y$ is an* internal witness *if $y \in \mathrm{htr}(v)$ and an* external witness *if $y \notin \mathrm{htr}(v)$.*

Foygel et al. (2012, Theorem 1) showed that the parameter matrix β in the graph G is *HTC-identifiable* if a family $(Y_v : v \in V)$ and a total order $\prec$ on V exist such that Y_v satisfies the HTC for every v and $w \prec v$ whenever $w \in Y_v \cap \mathrm{htr}(v)$. We call $\prec$ the *HTC ordering*; nodes with $Y_v \cap \mathrm{htr}(v) = \emptyset$ are minimal in $\prec$ and serve as the base of the induction in Section 3. To identify the parameter vector $\beta_v := (\beta_{p_1v}, \ldots, \beta_{p_kv})^\top$ of directed edges pointing into v, a raw HTC instrument vector $Z_{Y_v} := (Z_{y_1}, \ldots, Z_{y_k})^\top$ is component-wise constructed through

$$Z_y := \begin{cases} \varepsilon_y = X_y - \beta_y^\top X_{\mathrm{pa}(y)}, & y \in \mathrm{htr}(v), \\ X_y, & y \notin \mathrm{htr}(v). \end{cases}$$

Each β_y was identified before since $y \in Y_v \cap \mathrm{htr}(v)$. The *HTC relevance matrix* $A_v := \mathrm{E}\Big[Z_{Y_v} X_{\mathrm{pa}(v)}^\top\Big]$ coincides with the identification matrix studied by Foygel et al. (2012):

Lemma 4 (HTC relevance matrix) *The HTC relevance matrix A_v depends on (β, ε) only through (β, Ω) where $\Omega := \mathrm{Cov}(\varepsilon)$ is in $\mathrm{PD}(V)$, the cone of positive definite matrices. It is generically invertible: the set of $(\beta, \Omega) \in \mathcal{B} \times \mathrm{PD}(V)$ for which A_v is not invertible has Lebesgue measure zero.*

All proofs are presented in Appendix B. Lemma 4 shows invertibility holds generically; we assume it holds at the true parameter throughout (Assumption A.(ii)). Setting $b_v := \mathrm{E}[Z_{Y_v} X_v]$, the rotated instruments $A_v^{-1} Z_{Y_v}$ satisfy the standard IV orthogonality conditions:

$$\mathrm{E}\Big[A_v^{-1} Z_{Y_v} X_{\mathrm{pa}(v)}^\top\Big] = I_{|\mathrm{pa}(v)|}, \qquad \mathrm{E}\big[A_v^{-1} Z_{Y_v} X_v\big] = A_v^{-1} b_v.$$

The first identity shows that the jth row $e_j^\top A_v^{-1} Z_{Y_v}$ is uncorrelated with all parents except p_j; the second recovers $\beta_v = A_v^{-1} b_v$ as a population moment via the HTC orthogonality $\mathrm{E}[Z_y \varepsilon_v] = 0$, the central equation Section 3 differentiates to yield the influence function ϕ_{β_v}.

3. Influence Function and Asymptotic Variance

Write $\hat{\beta}_{\prec v} := (\hat{\beta}_y)_{y \prec v}$ for the estimates of nodes preceding v. The empirical HTC estimator $\hat{\beta}_v := \hat{A}_v^{-1} \hat{b}_v$ is a Z-estimator: at each v in HTC order it solves

$$n^{-1} \sum_{i \in [n]} \hat{Z}_{Y_v}(X^{(i)}, \hat{\beta}_{\prec v}) \left(X_v^{(i)} - \hat{\beta}_v^\top X_{\mathrm{pa}(v)}^{(i)} \right) = 0,$$

with $\hat{Z}_y(X, \hat{\beta}_{\prec v}) = X_y - \hat{\beta}_y^\top X_{\mathrm{pa}(y)}$ for $y \in \mathrm{htr}(v)$ and $\hat{Z}_y = X_y$ otherwise. The identification equation $A_v \beta_v = b_v$ has rows indexed by witnesses $y \in Y_v$; the influence function $\phi_{\beta_v} = \phi_{A_v^{-1} b_v}$ follows by the delta method, which requires the pathwise derivatives $\phi_{M_y(t)}$ of the individual moments $M_y(t) := \mathrm{E}[Z_y X_t]$. For external witnesses, the instrument $Z_y = X_y$ is fixed and $\phi_{M_y(t)}$ is the standard covariance influence function. For internal witnesses, the residual instrument $Z_y = X_y - \beta_y^\top X_{\mathrm{pa}(y)}$ depends on β_y, so the pathwise derivative carries a correction from the earlier influence functions $\phi_{\beta_{qy}}$. Lemma 5 collects both cases; Lemma 6 combines them into the row contributions $R_{y,v}$ entering Theorem 7 for $\phi_{\beta_v} = A_v^{-1} R_v$.

Lemma 5 (Row moment derivative) *For $M_y(t) := \mathrm{E}[Z_y X_t]$ with $y \in Y_v$, $t \in V$, the influence function along any parametric submodel is*

$$\phi_{M_y(t)} = \begin{cases} X_y X_t - \Sigma_{yt}, & y \notin \mathrm{htr}(v), \\ \varepsilon_y X_t - \mathrm{E}[\varepsilon_y X_t] - \displaystyle\sum_{q \in \mathrm{pa}(y)} \Sigma_{qt}\, \phi_{\beta_{qy}}, & y \in \mathrm{htr}(v). \end{cases}$$

The row contribution $R_{y,v}$ is the influence function of the y-th identification equation residual $e_y^\top (b_v - A_v \beta_v) = M_y(v) - \sum_{p \in \mathrm{pa}(v)} \beta_{pv} M_y(p)$. It is obtained by applying Lemma 5 to $M_y(v)$ and each $M_y(p)$ and combining linearly.

Lemma 6 (Row contribution) *For witness $y \in Y_v$ and target v, the influence function of the y-th row of the identification equation*

$$R_{y,v} := \phi_{M_y(v)} - \sum_{p \in \mathrm{pa}(v)} \beta_{pv}\, \phi_{M_y(p)},$$

satisfies

$$R_{y,v} = \begin{cases} X_y \varepsilon_v, & y \notin \mathrm{htr}(v), \\ \varepsilon_y \varepsilon_v - \sum_{q \in \mathrm{pa}(y)} \mathrm{E}[X_q \varepsilon_v]\ \phi_{\beta_{qy}}, & y \in \mathrm{htr}(v). \end{cases}$$

Write $R_v := (R_{y_1,v}, \ldots, R_{y_k,v})^\top$.

3.1. The Influence Function ϕ_{β_v}

The influence function for $\beta_v = A_v^{-1} b_v$ follows by the delta method applied to the identification equation, with Lemmas 5 and 6 supplying the row-wise derivatives.

Theorem 7 (HTC influence function) *The functional $\phi_{\beta_v} = A_v^{-1} R_v$ is an influence function for β_v in $\mathcal{M}_G$. For the edge $p_j \to v$, this yields $\phi_{\beta_{p_j v}} = e_j^\top A_v^{-1} R_v$.*

Example 1 *In the classical IV case with $|\mathrm{pa}(v)| = 1$ and one external witness $y \notin \mathrm{htr}(v)$, Lemma 6 gives $R_{y,v} = X_y \varepsilon_v$ and $A_v = \mathrm{E}[X_y X_p]$, so the HTC influence function $\phi_{\beta_{pv}}$ in Theorem 7 evaluates to $X_y \varepsilon_v / \mathrm{E}[X_y X_p]$, the standard 2SLS influence function.*

Remark 8 (Semiparametric efficiency of HTC estimators) *The influence function $\phi_{\beta_v} = A_v^{-1} R_v$ of Theorem 7 is the influence function of the HTC estimator for the chosen witness set Y_v, not necessarily the efficient influence function of β_v in $\mathcal{M}_G$. In the graph $1 \to 3 \to 4$, $2 \to 3 \leftrightarrow 4$, the efficient influence function for β_{34} uses both instruments simultaneously (Chamberlain, 1987; Newey, 1990), a bound no single-witness HTC estimator achieves.*

3.2. The Asymptotic Variance $\mathcal{V}_v$

The asymptotic normality of the HTC estimator, written $\hat{\beta}_{v,n}$ to make the sample-size dependence explicit, follows from Theorem 7 by the central limit theorem: since $\phi_{\beta_v} = A_v^{-1} R_v$ is mean-zero and square-integrable, the sandwich formula below is its asymptotic variance.

Proposition 9 (Recursive variance formula) *The HTC estimator $\hat{\beta}_{v,n}$ satisfies*

$$\sqrt{n}(\hat{\beta}_{v,n} - \beta_v) \xrightarrow{d} \mathcal{N}(0, \mathcal{V}_v) \quad \text{with} \quad \mathcal{V}_v = A_v^{-1}\, \mathrm{E}\Big[R_v R_v^\top\Big]\ A_v^{-\top}.$$

For the edge $p_j \to v$, $\mathcal{V}_v[j,j] = e_j^\top \mathcal{V}_v e_j$. The covariance $\mathrm{E}\big[R_v R_v^\top\big]$ expands for $a, b \in Y_v$ to:
(i) *$a, b \notin \mathrm{htr}(v)$:*

$$\mathrm{E}[R_{a,v} R_{b,v}] = \mathrm{E}\big[X_a X_b \varepsilon_v^2\big]\,.$$

Both witnesses are external, so $R_{a,v} = X_a\varepsilon_v$ and $R_{b,v} = X_b\varepsilon_v$ with no correction.
(ii) *$a \in \mathrm{htr}(v)$, $b \notin \mathrm{htr}(v)$:*

$$\mathrm{E}[R_{a,v}R_{b,v}] = \mathrm{E}\big[\varepsilon_a X_b \varepsilon_v^2\big] - \sum_{q\in\mathrm{pa}(a)} \mathrm{E}[X_q\varepsilon_v]\;\mathrm{E}\big[\phi_{\beta_{qa}} X_b\varepsilon_v\big]\,.$$

The subtracted term corrects for the uncertainty of β_{qa}: estimating the residual instrument $Z_a = \varepsilon_a$ requires $\hat\beta_{qa}$, which introduces a correction proportional to $\phi_{\beta_{qa}}$ into $R_{a,v}$.
(iii) *$a, b \in \mathrm{htr}(v)$:*

$$\mathrm{E}[R_{a,v}R_{b,v}] = \mathrm{E}\big[\varepsilon_a\varepsilon_b\varepsilon_v^2\big] - \sum_{r\in\mathrm{pa}(b)} \mathrm{E}[X_r\varepsilon_v]\;\mathrm{E}[\varepsilon_a\varepsilon_v\phi_{\beta_{rb}}] - \sum_{q\in\mathrm{pa}(a)} \mathrm{E}[X_q\varepsilon_v]\;\mathrm{E}\big[\phi_{\beta_{qa}}\varepsilon_b\varepsilon_v\big] + \sum_{q\in\mathrm{pa}(a)}\sum_{r\in\mathrm{pa}(b)} \mathrm{E}[X_q\varepsilon_v]\;\mathrm{E}[X_r\varepsilon_v]\;\mathrm{E}\big[\phi_{\beta_{qa}}\phi_{\beta_{rb}}\big]\,.$$

With both witnesses internal, first-stage corrections from $\phi_{\beta_{qa}}$ and $\phi_{\beta_{rb}}$ enter both factors; the double sum captures the covariance between these two correction terms.

The correction terms $\phi_{\beta_{qy}}$ in R_v in Lemma 6 account for the full variation from estimating $\hat\beta_{\prec v}$, ensuring that the stated asymptotic distribution of $\hat\beta_v$ is valid at each estimation stage.

Proposition 9 involves cross-variances $\mathrm{E}\big[\phi_{\beta_{qa}} X_b\varepsilon_v\big]$ of earlier influence functions, giving $\mathcal{V}_v$ an apparently recursive definition. In Figure 1, however, the recursion for $v = 5$ bottoms out at node 2, which has $Y_2 \cap \mathrm{htr}(2) = \emptyset$; there $\phi_{\beta_{12}} = X_1\varepsilon_2/\mathrm{E}\big[X_1^2\big]$ and $\mathrm{E}[\phi_{\beta_{12}}X_b\varepsilon_v] = \mathrm{E}[X_1\varepsilon_2X_b\varepsilon_v]\,/\mathrm{E}\big[X_1^2\big]$ is a raw 2SLS fourth-order moment.

Lemma 10 (Termination and computability) *The covariance $\mathrm{E}\big[R_vR_v^\top\big]$ reduces to finitely many fourth-order moments of observables. Each cross-variance $\mathrm{E}\big[\phi_{\beta_{qa}}X_b\varepsilon_v\big]$ equals*

$$\mathrm{E}\big[\phi_{\beta_{qa}}X_b\varepsilon_v\big] = e_j^\top A_a^{-1}\,\mathrm{E}[R_aX_b\varepsilon_v]\,,$$

and for each $y \in Y_a$,

$$\mathrm{E}[R_{y,a}X_b\varepsilon_v] = \begin{cases} \mathrm{E}[X_y\varepsilon_aX_b\varepsilon_v] & y \notin \mathrm{htr}(a), \\ \mathrm{E}[\varepsilon_y\varepsilon_aX_b\varepsilon_v] - \sum_{q'\in\mathrm{pa}(y)} \mathrm{E}\big[X_{q'}\varepsilon_a\big]\;\mathrm{E}\Big[\phi_{\beta_{q'y}}X_b\varepsilon_v\Big] & y \in \mathrm{htr}(a). \end{cases}$$

Each application of the $y \in \mathrm{htr}(a)$ branch strictly descends in the HTC order $\prec$; the recursion bottoms out at nodes u with $Y_u \cap \mathrm{htr}(u) = \emptyset$, where every entry equals a pure fourth-order moment of observables.

Under Gaussian errors, Isserlis' theorem expresses these fourth-order moments as polynomials in (Σ, Ω, β), making $\mathcal{V}_v$ an explicit rational function of the model parameters.

The estimator $\hat{\mathcal{V}}_v$ is obtained by replacing every expectation in Proposition 9 and Lemma 10 with the corresponding empirical moment, evaluated on the observed sample.

4. Numerical Experiments

The simulation data aligned with the graph in Figure 1 are generated by Gaussian and modified Gamma error distributions. Full analysis details are provided in Appendix C.1.

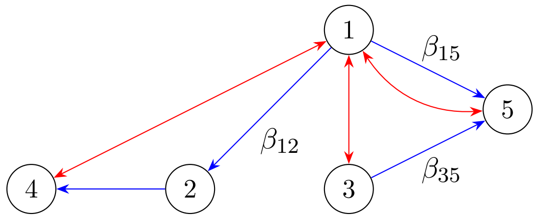


Figure 1: Five-node directed mixed graph (cf. Example 6(b) in Foygel et al. (2012)) with HTC ordering $2 \prec 4 \prec 5$.

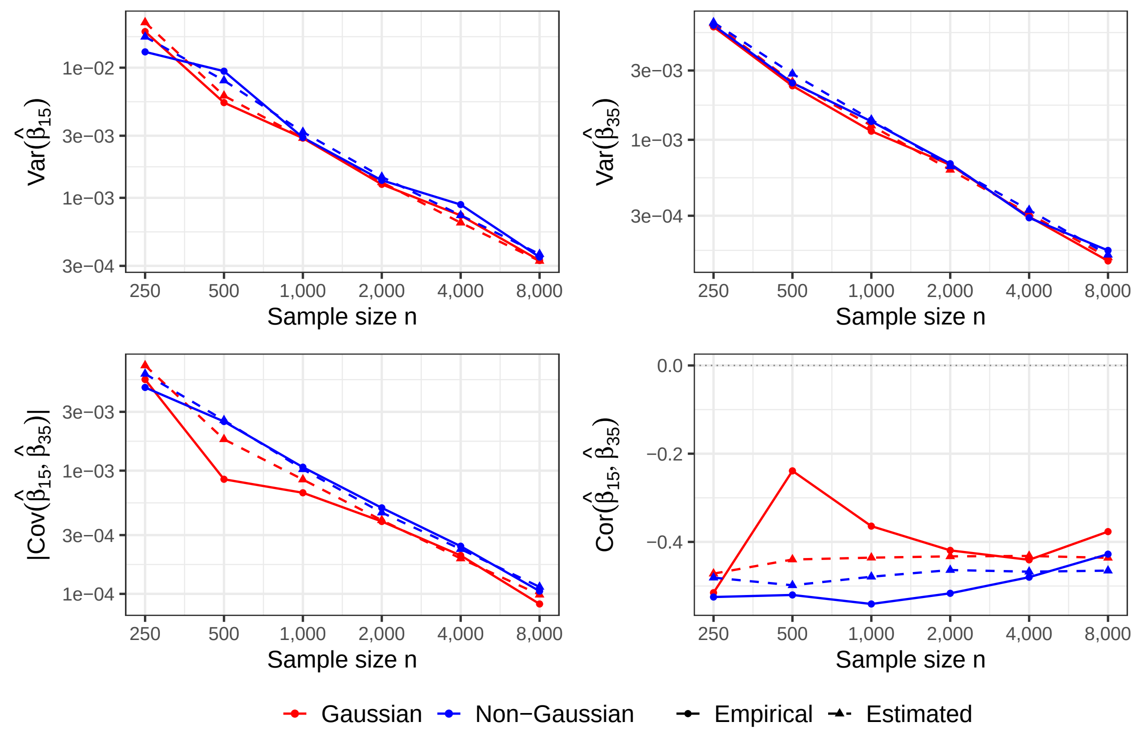


Figure 2: Calibration of $\hat{\mathcal{V}}_5$ for $(\hat{\beta}_{15}, \hat{\beta}_{35})$ in Figure 1 ($Y_5 = \{3, 4\}$, internal witness $4 \in \mathrm{htr}(5)$, external witness $3 \notin \mathrm{htr}(5)$): empirical variances, covariances, and correlations over 100 replications (solid) versus mean $\hat{\mathcal{V}}_5/n$ (dashed), across sample sizes n under Gaussian (red) and non-Gaussian (blue) errors.

4.1. Calibration

Figure 2 shows that the asymptotic variance estimator $\hat{\mathcal{V}}_v$ is well calibrated under both Gaussian and non-Gaussian errors. Since $\hat{\beta}_{15}$ and $\hat{\beta}_{35}$ are negatively correlated, the bottom-left panel displays the absolute covariance.

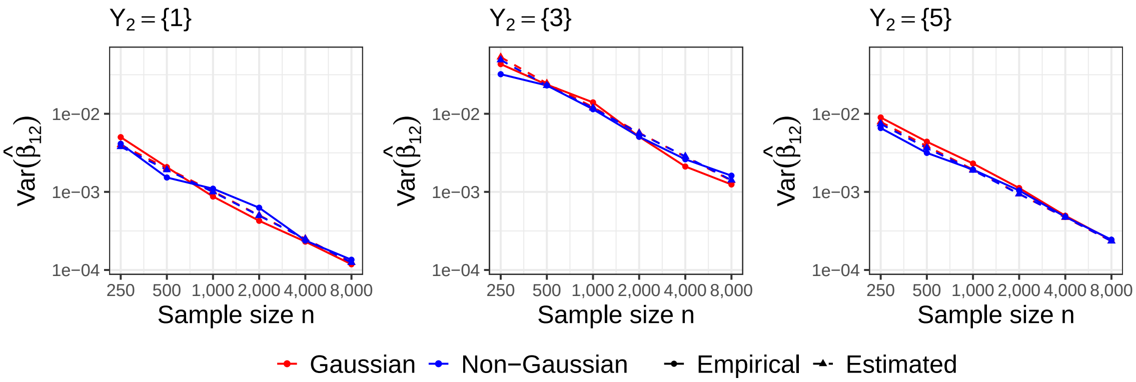


Figure 3: Estimation variance of $\hat{\beta}_{12}$ in Figure 1 for three witness choices $Y_2 \in \{\{1\}, \{3\}, \{5\}\}$: empirical variance over 100 replications (solid) versus mean $\hat{\mathcal{V}}_2/n$ (dashed), across sample sizes n under Gaussian (red) and non-Gaussian (blue) errors.

4.2. Witness Choice and Estimation Variance

The witness set choice has an effect on estimation efficiency: Figure 3 shows that the asymptotic variance of $\hat{\beta}_{12}$ varies substantially across the three valid witness choices $Y_2 \in \{\{1\}, \{3\}, \{5\}\}$. In practice, the analyst must select a witness set, and a lower-variance choice is preferable.

Remark 11 (Witness selection) *The asymptotic variance estimator $\hat{\mathcal{V}}_v$ of Proposition 9 provides, for any fixed Y_v, a consistent estimate of $\mathcal{V}_v(Y_v)$, enabling post-hoc comparison of competing witness sets. Selecting a variance-minimizing set in advance is an open problem: while valid HTC witness sets are combinatorially characterized, their complete enumeration requires searching over exponentially many candidate subsets of V, so no algorithmic guidance towards a lower-variance choice without exhaustive search is currently available.*

5. Statistical Inference

Confidence regions and Wald tests for β_v follow from the $\mathcal{N}(0, \mathcal{V}_v)$ limit of Proposition 9 via the asymptotic variance estimator $\hat{\mathcal{V}}_v$.

Proposition 12 (Confidence regions and Wald tests) *For $\alpha \in (0, 1)$, write $z_{\alpha/2}$ for the $\alpha/2$ quantile of the standard normal distribution and $\chi^2_{k,1-\alpha}$ for the $(1-\alpha)$ quantile of the chi-squared distribution with k degrees of freedom χ^2_k. The following hold.*

(i) (Confidence ellipsoid.) An asymptotic $(1-\alpha)$ confidence region for $\beta_v \in \mathbb{R}^{|\mathrm{pa}(v)|}$ is

$$\mathcal{C}_n = \left\{\beta \in \mathbb{R}^{|\mathrm{pa}(v)|} : n(\hat{\beta}_v - \beta)^\top \hat{\mathcal{V}}_v^{-1}(\hat{\beta}_v - \beta) \leq \chi^2_{|pa(v)|,1-\alpha}\right\}.$$

(ii) (Marginal intervals.) An asymptotic $(1-\alpha)$ confidence interval for the edge $\beta_{p_j v}$ is

$$\hat{\beta}_{p_j v} \pm z_{\alpha/2}\sqrt{\hat{\mathcal{V}}_v[j,j]/n}.$$

(iii) (Wald test.) For $H_0\colon C\beta_v = c$ with full row rank $C \in \mathbb{R}^{r\times|\mathrm{pa}(v)|}$ and $c \in \mathbb{R}^r$, the Wald statistic satisfies

$$W_n = n(C\hat{\beta}_v - c)^\top (C\hat{\mathcal{V}}_v C^\top)^{-1}(C\hat{\beta}_v - c) \xrightarrow{d} \chi^2_r \quad \textit{under } H_0.$$

The matrix $C\hat{\mathcal{V}}_v C^\top$ is generically invertible because C has full row rank.

For the marginal test $H_0\colon \beta_{p_j v} = 0$, the statistic reduces to $W_n = n\hat{\beta}^2_{p_j v}/\hat{\mathcal{V}}_v[j,j] \xrightarrow{d} \chi^2_1$; equivalently, $\hat{\beta}_{p_j v}/\sqrt{\hat{\mathcal{V}}_v[j,j]/n}$ is asymptotically standard normal.

5.1. R Implementation

We implement the HTC estimator, asymptotic variance estimator, and inference in the `R` function `htcfit()`, with inferential output produced by `summary(htcfit())`.[1] The output contains multiple blocks corresponding to each non-root node v in the HTC order $\prec$. Each block is structured as follows.

Block header. Identifies v, its parents $\mathrm{pa}(v)$, and the witness set Y_v with type labels `ext` (external witness, instrument $Z_y = X_y$) and `int` (internal witness, instrument $Z_y = \varepsilon_y$).

Coefficient table. For each edge $p_j \to v$: Reports the estimate $\hat{\beta}_{p_j v}$, the standard error $\sqrt{\hat{\mathcal{V}}_v[j,j]/n}$ from Proposition 12(ii), the z-statistic $\hat{\beta}_{p_j v}/\sqrt{\hat{\mathcal{V}}_v[j,j]/n}$ which is asymptotically $N(0,1)$ distributed under $H_0\colon \beta_{p_j v} = 0$, and a corresponding two-sided p-value.

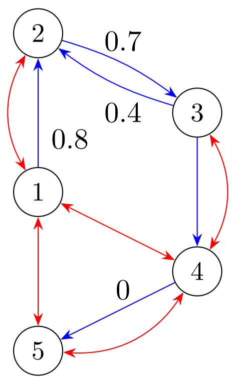
Footer. Presents the empirical residual standard deviation $\hat{\Omega}^{1/2}_{vv}$, the structural $R^2 := 1 - \hat{\Omega}_{vv}/\mathrm{Var}(X_v)$, the fraction of $\mathrm{Var}(X_v)$ explained by $X_{\mathrm{pa}(v)}$, and, when $|\mathrm{pa}(v)| > 1$, the joint Wald statistic $n\hat{\beta}_v^\top \hat{\mathcal{V}}_v^{-1}\hat{\beta}_v \sim \chi^2_{|\mathrm{pa}(v)|}$ under $H_0\colon \beta_v = 0$ from Proposition 12(iii).

Example 2 *Figure 4 shows `summary(htcfit())` applied to the cyclic five-node graph on the left, with $n = 1000$ Gaussian observations at the true parameters listed in Appendix C.2. The figure displays three of the four blocks in HTC order ($3 \prec 5 \prec 2 \prec 4$, node 4 truncated); nodes 3 and 5 each have a single parent, so their footers carry only the structural R^2, while node 2 adds the joint Wald χ^2. The block for node 2, which has two parents $\mathrm{pa}(2) = \{1,3\}$ identified by internal witnesses $Y_2 = \{3,5\} \subset \mathrm{htr}(2)$, reports a joint Wald test that strongly rejects $H_0 : \beta_{12} = \beta_{32} = 0$. The block for node 5 does not reject $H_0 : \beta_{45} = 0$, consistent with the true value $\beta_{45} = 0$.*

5.2. The Fulton Fish Market

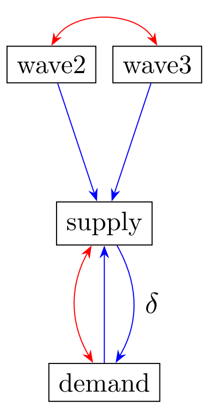
The Fulton Fish Market in New York City offers a textbook illustration of simultaneous-equations identification. Angrist et al. (2000) analyze $n = 97$ daily records of wholesale whiting transactions collected by Graddy (1995): log price (*supply*), log quantity sold (*demand*), as well as two- and three-day moving average wave height (*wave2*, *wave3*) off the

1. Source code at `https://anonymous.4open.science/r/HTC_Inference`; HTC order $\prec$ and witness sets are either determined via the `SEMID` package (Foygel Barber et al., 2025) or user-specified.




```
HTC Estimator Summary
=======================================================
Node 3  (pa: 2)  [witnesses: 1 (ext)]

          Estimate  Std. Error   z value  Pr(>|z|)
2 -> 3      0.7288      0.0163   44.6505     < 2e-16 ***
---
Signif. codes: 0 ’***’ 0.001 ’**’ 0.01 ’*’ 0.05 ’.’ 0.1 ’ ’ 1

Residual std. dev: 0.9433   Structural R-sq: 0.7946   (n = 1000)
-------------------------------------------------------
Node 5  (pa: 4)  [witnesses: 3 (int)]

          Estimate  Std. Error   z value  Pr(>|z|)
4 -> 5     -0.0156      0.0137   -1.1380       0.255
---
Signif. codes: 0 ’***’ 0.001 ’**’ 0.01 ’*’ 0.05 ’.’ 0.1 ’ ’ 1

Residual std. dev: 1.0820   Structural R-sq: -0.0268   (n = 1000)
-------------------------------------------------------
Node 2  (pa: 1, 3)  [witnesses: 3 (int), 5 (int)]

          Estimate  Std. Error   z value  Pr(>|z|)
1 -> 2      0.7946      0.0411   19.3338     < 2e-16 ***
3 -> 2      0.3893      0.0254   15.3408     < 2e-16 ***
---
Signif. codes: 0 ’***’ 0.001 ’**’ 0.01 ’*’ 0.05 ’.’ 0.1 ’ ’ 1

Residual std. dev: 1.0223   Structural R-sq: 0.7971   (n = 1000)
Joint Wald test (H0: beta_2 = 0):  chi-sq = 729.088 on 2 df,
p-value < 2e-16
-------------------------------------------------------
...
```

Figure 4: **summary(htcfit())** output for the cyclic five-node directed mixed graph (left, $n = 1000$, true $\beta_{45} = 0$); see Appendix C.2 for graph and parameter details.




```
HTC Estimator Summary
=======================================================
Node demand  (pa: supply)  [witnesses: wave2 (ext)]

                    Estimate  Std. Error   z value  Pr(>|z|)
supply -> demand     -0.8410      0.3827   -2.1976       0.028 *
---
Signif. codes: 0 ’***’ 0.001 ’**’ 0.01 ’*’ 0.05 ’.’ 0.1 ’ ’ 1

Residual std. dev: 0.6850   Structural R-sq: 0.0566   (n = 97)
-------------------------------------------------------

HTC Estimator Summary
=======================================================
Node demand  (pa: supply)  [witnesses: wave3 (ext)]

                    Estimate  Std. Error   z value  Pr(>|z|)
supply -> demand     -0.7611      0.4246   -1.7926       0.073 .
---
Signif. codes: 0 ’***’ 0.001 ’**’ 0.01 ’*’ 0.05 ’.’ 0.1 ’ ’ 1

Residual std. dev: 0.6798   Structural R-sq: 0.0708   (n = 97)
-------------------------------------------------------
```

Figure 5: **summary(htcfit())** applied to the *Fulton Fish Market* dataset ($n = 97$): cyclic directed mixed graph with correlated (*supply*, *demand*) errors (left), estimated separately with *wave2* and *wave3* as instruments for the demand elasticity δ.

Long Island coast. Supply and demand are jointly determined, and the two structural equations share correlated errors from omitted common factors such as seasonal stock variation and port conditions. The directed mixed graph in Figure 5 (left) captures this structure: directed edges $supply \to demand$ and $demand \to supply$ form the simultaneous-equations cycle, and a bidirected edge $supply \leftrightarrow demand$ represents the correlated errors. Rough seas reduce landings and raise prices without directly affecting consumer demand, making wave height a natural external instrument for the supply equation. The target parameter is $\delta := \beta_{supply,\,demand}$, the demand elasticity. Naive OLS yields $\hat{\delta}_{\mathrm{OLS}} = -0.525$.

The HTC identifies δ using *wave2* as the single external witness $Y_{demand} = \{wave2\}$. Since *wave2* is an external witness, no residual plug-in is required and the instrument is the raw variable X_{wave2}. Figure 5 (right) shows $\hat{\delta} = -0.8410$ (SE 0.3827, $z = -2.1976$, $p = 0.028$), a statistically significant negative demand elasticity directly comparable to the estimate of -1.01 (SE 0.42) from Angrist et al. (2000). The estimates agree within one standard error, and the HTC asymptotic standard error is distribution-free, requiring only finite fourth moments. The additional day of averaging in *wave3* dilutes the sharp day-before supply shock and reduces the first-stage correlation with *supply* from 0.4931 to 0.3798, making it a slightly weaker instrument for δ: Figure 5 shows $\hat{\delta} = -0.7611$ (SE 0.4246, $p = 0.073$), within one standard error of the *wave2* result. Further details on data preprocessing and identification can be found in Appendix D.

6. Conclusion

We have derived the semiparametric influence function $\phi_{\beta_v} = A_v^{-1} R_v$ of the HTC estimator for all HTC-identified directed mixed graphs (Theorem 7). The theory covers causal structures beyond the classical IV setting: multiple parents, external and internal (residual-based) instruments, latent confounding, and cyclic graphs. Under standard regularity conditions and finite fourth moments, the asymptotic variance estimator $\hat{\mathcal{V}}_v$ is computable in closed form by descent through the HTC order $\prec$ and is well calibrated across Gaussian and non-Gaussian error distributions. Applied to the Fulton Fish Market, the theory delivers a complete inferential summary for the demand elasticity in a simultaneous-equations model with correlated errors.

Two open problems emerge from this work. First, the HTC algorithm as implemented in the `SEMID` package outputs one valid witness set and HTC ordering $\prec$ without variance guarantees, while Figure 3 shows this choice has a first-order effect on variance; a polynomial-time selection rule for a lower-variance Y_v or $\prec$ has yet to be found. Second, the HTC estimator is semiparametrically efficient in the classical single-instrument IV case, where $\phi_{\beta_{pv}} = X_y \varepsilon_v / \mathrm{E}[X_y X_p]$ coincides with the efficient influence function; for which directed mixed graphs efficiency holds within the HTC class has yet to be characterized. Resolving these two problems would connect the graphical identification theory of Foygel et al. (2012) to a complete theory of variance-optimal HTC estimation.

Acknowledgments

Leopold Mareis received funding from the German Research Foundation (DFG) under the Mathematical Research Data Initiative (project No. 460135501). Mathias Drton received funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (grant agreement No. 883818). We thank Benjamin Hollering for valuable discussions concerning marginal independence models.

References

Joshua D. Angrist, Kathryn Graddy, and Guido W. Imbens. The Interpretation of Instrumental Variables Estimators in Simultaneous Equations Models with an Application to the Demand for Fish. *The Review of Economic Studies*, 67(3):499–527, 2000.

Rina Foygel Barber, Mathias Drton, Nils Sturma, and Luca Weihs. Half-trek criterion for identifiability of latent variable models. *The Annals of Statistics*, 50(6):3174–3196, 2022.

Kenneth A. Bollen. *Structural Equations with Latent Variables*. Wiley Series in Probability and Mathematical Statistics: Applied Probability and Statistics. John Wiley & Sons, Inc., New York, 1989. A Wiley-Interscience Publication.

Roger J. Bowden and Darrell A. Turkington. *Instrumental Variables*. Econometric Society Monographs. Cambridge University Press, 1985.

Carlos Brito and Judea Pearl. Graphical condition for identification in recursive SEM. In *Proceedings of the Twenty-Second Conference on Uncertainty in Artificial Intelligence*, UAI'06, page 47–54, Arlington, Virginia, USA, 2006. AUAI Press.

Gary Chamberlain. Asymptotic efficiency in estimation with conditional moment restrictions. *Journal of Econometrics*, 34(3):305–334, 1987.

Bryant Chen, Judea Pearl, and Elias Bareinboim. Incorporating knowledge into structural equation models using auxiliary variables. In *Proceedings of the Twenty-Fifth International Joint Conference on Artificial Intelligence*, IJCAI'16, page 3577–3583. AAAI Press, 2016.

Bryant Chen, Daniel Kumor, and Elias Bareinboim. Identification and Model Testing in Linear Structural Equation Models using Auxiliary Variables. In *Proceedings of the 34th International Conference on Machine Learning*, volume 70 of *Proceedings of Machine Learning Research*, pages 757–766. PMLR, 2017.

Rina Foygel, Jan Draisma, and Mathias Drton. Half-trek criterion for generic identifiability of linear structural equation models. *The Annals of Statistics*, 40(3):1682–1713, 2012.

Rina Foygel Barber, Mathias Drton, Luca Weihs, and Nils Sturma. *SEMID: Identifiability of Linear Structural Equation Models*, 2025. R package version 0.4.2.

Luis D. García-Puente, Sarah Spielvogel, and Seth Sullivant. Identifying causal effects with computer algebra. In *Proceedings of the Twenty-Sixth Conference on Uncertainty in Artificial Intelligence*, UAI'10, page 193–200, Arlington, Virginia, USA, 2010. AUAI Press.

Kathryn Graddy. Testing for Imperfect Competition at the Fulton Fish Market. *The RAND Journal of Economics*, 26(1):75–92, 1995.

Leonard Henckel, Emilija Perković, and Marloes H. Maathuis. Graphical criteria for efficient total effect estimation via adjustment in causal linear models. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 84(2):579–599, 2022.

Leonard Henckel, Martin Buttenschoen, and Marloes H. Maathuis. Graphical tools for selecting conditional instrumental sets. *Biometrika*, 111(3):771–788, 2024.

Benjamin Hollering, Pratik Misra, and Nils Sturma. Efficient Symbolic Computations for Identifying Causal Effects, 2026. arXiv:2604.20516.

Yonghan Jung, Jin Tian, and Elias Bareinboim. Estimating identifiable causal effects through double machine learning. *Proceedings of the AAAI Conference on Artificial Intelligence*, 35(13):12113–12122, May 2021.

Daniel Kumor, Bryant Chen, and Elias Bareinboim. Efficient Identification in Linear Structural Causal Models with Instrumental Cutsets. In *Advances in Neural Information Processing Systems*, volume 32. Curran Associates, Inc., 2019.

Marloes Maathuis, Mathias Drton, Steffen Lauritzen, and Martin Wainwright, editors. *Handbook of Graphical Models*. Chapman & Hall/CRC Handbooks of Modern Statistical Methods. CRC Press, Boca Raton, FL, 2019.

Leopold Mareis and Mathias Drton. Cost-Aware Optimized Front-Door Experimental Design, 2026. arXiv:2603.22024.

Whitney K. Newey. Efficient Instrumental Variables Estimation of Nonlinear Models. *Econometrica*, 58(4):809–837, 1990.

R Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria, 2024.

Peter Spirtes, Clark Glymour, and Richard Scheines. *Causation, Prediction, and Search*. Adaptive Computation and Machine Learning. MIT Press, Cambridge, MA, second edition, 2000. With additional material by David Heckerman, Christopher Meek, Gregory F. Cooper and Thomas Richardson, A Bradford Book.

Jin Tian. Parameter Identification in a Class of Linear Structural Equation Models. In *Proceedings of the 21st International Joint Conference on Artificial Intelligence*, IJCAI'09, page 1970–1975, San Francisco, CA, USA, 2009. Morgan Kaufmann Publishers Inc.

Anastasios A. Tsiatis. *Semiparametric Theory and Missing Data*. Springer Series in Statistics. Springer, New York, 2006.

Aad W. van der Vaart. *Asymptotic Statistics*, volume 3 of *Cambridge Series in Statistical and Probabilistic Mathematics*. Cambridge University Press, Cambridge, 1998.

Janine Witte, Leonard Henckel, Marloes H. Maathuis, and Vanessa Didelez. On Efficient Adjustment in Causal Graphs. *Journal of Machine Learning Research*, 21(246):1–45, 2020.

Philip Green Wright. *The Tariff on Animal and Vegetable Oils*. Macmillan, 1928.

Appendix A. Assumptions

The following conditions hold at the true parameter $(\beta_0, \varepsilon_0) \in \mathcal{B} \times \mathcal{E}$.

(i) *(Hellinger differentiability.)* Every parametric path $\gamma \mapsto P_\gamma$ in $\mathcal{M}_G$ through P_0 is differentiable in quadratic mean at $\gamma = 0$ (van der Vaart, 1998, Definition 25.14): there exists a score $S \in L_0^2$ satisfying

$$\int \left(\frac{dP_\gamma^{1/2} - dP_0^{1/2}}{\gamma} - \frac{1}{2} S \, dP_0^{1/2} \right)^2 \to 0 \quad \text{as } \gamma \to 0.$$

This is the standard regularity condition for RAL theory.

(ii) *(Non-degeneracy.)* The HTC matrix A_v is invertible for every $v \in V$. By (Foygel et al., 2012, Lemma 2), $\det(A_v)$ is a nonzero polynomial in (β, Ω), so invertibility holds for all (β, Ω) outside a proper algebraic subset of the parameter space.

Appendix B. Proofs

Lemma 4: The HTC relevance matrix A_v depends on (β, ε) only through (β, Ω) where $\Omega := \mathrm{Cov}(\varepsilon)$ is in $\mathrm{PD}(V)$, the cone of positive definite matrices. It is generically invertible: the set of $(\beta, \Omega) \in \mathcal{B} \times \mathrm{PD}(V)$ for which A_v is not invertible has Lebesgue measure zero.

Proof Write $\Sigma := \mathrm{E}\big[XX^\top\big] = (I-\beta)^{-\top}\Omega(I-\beta)^{-1}$.

Case $y_i \in \mathrm{htr}(v)$**.** From $X = \beta^\top X + \varepsilon$, we have $Z_{y_i} = \varepsilon_{y_i} = (I-\beta^\top)_{y_i \cdot}X$, so $(A_v)_{ij} = \mathrm{E}\big[Z_{y_i}X_{p_j}\big] = [(I-\beta)^\top\Sigma]_{y_i p_j}$, a rational function of (β, Ω).

Case $y_i \notin \mathrm{htr}(v)$**.** $Z_{y_i} = X_{y_i}$, so $(A_v)_{ij} = \Sigma_{y_i p_j}$, a rational function of (β, Ω).

Hence $\det(A_v)$ is rational in (β, Ω), not identically zero by (Foygel et al., 2012, Lemma 2), so A_v is invertible outside the proper algebraic subset $\{\det(A_v) = 0\} \subset \mathcal{B} \times \mathrm{PD}(V)$. ■

Lemma 5: For $M_y(t) := \mathrm{E}[Z_y X_t]$ with $y \in Y_v$, $t \in V$, the influence function along any parametric submodel is

$$\phi_{M_y(t)} = \begin{cases} X_y X_t - \Sigma_{yt}, & y \notin \mathrm{htr}(v), \\ \varepsilon_y X_t - \mathrm{E}[\varepsilon_y X_t] - \displaystyle\sum_{q \in \mathrm{pa}(y)} \Sigma_{qt}\, \phi_{\beta_{qy}}, & y \in \mathrm{htr}(v). \end{cases}$$

Proof We prove the result by induction on the HTC ordering $\prec$. The base nodes satisfy $Y_v \cap \mathrm{htr}(v) = \emptyset$, so all witnesses are external and the internal case does not arise; the base case therefore reduces to the external case below. For any $a, b \in V$, differentiating $\Sigma_{ab} = \mathrm{E}[X_a X_b]$ along a parametric submodel with score S gives $\frac{d}{d\gamma}\big|_0 \Sigma_{ab,\gamma} = \mathrm{E}[(X_a X_b - \Sigma_{ab})S]$, so $\phi_{\Sigma_{ab}} = X_a X_b - \Sigma_{ab}$ is a valid mean-zero influence function for Σ_{ab} in $\mathcal{M}_G$.

Case $y \notin \mathrm{htr}(v)$**.** $M_y(t) = \mathrm{E}[X_y X_t] = \Sigma_{yt}$, so $\phi_{M_y(t)} = \phi_{\Sigma_{yt}} = X_y X_t - \Sigma_{yt}$. This is mean-zero and satisfies $\mathrm{E}\big[\phi_{M_y(t)}S\big] = \frac{d}{d\gamma}\big|_0 M_{y,\gamma}(t)$ for all submodel scores S, hence it is a valid influence function for $M_y(t)$.

Case $y \in \mathrm{htr}(v)$**.** Since $y \prec v$ in HTC order, the influence functions $\phi_{\beta_{qy}}$ for $q \in \mathrm{pa}(y)$ are available as valid mean-zero influence functions by the induction hypothesis. Substituting $Z_y = X_y - \sum_{q \in \mathrm{pa}(y)} \beta_{qy} X_q$ into $M_y(t) = \mathrm{E}[Z_y X_t]$ gives

$$M_y(t) = \mathrm{E}[\varepsilon_y X_t] = \Sigma_{yt} - \sum_{q \in \mathrm{pa}(y)} \beta_{qy}\Sigma_{qt}.$$

Differentiating $M_{y,\gamma}(t) = \Sigma_{yt,\gamma} - \sum_q \beta_{qy,\gamma}\Sigma_{qt,\gamma}$ along the submodel using the product rule for pathwise derivatives ($\phi_{fg} = \phi_f g_0 + f_0 \phi_g$, a direct consequence of the bilinearity of differentiation; see van der Vaart (1998, §25.7)) and substituting $\phi_{\Sigma_{ab}} = X_a X_b - \Sigma_{ab}$:

$$\phi_{M_y(t)} = (X_y X_t - \Sigma_{yt}) - \sum_{q \in \mathrm{pa}(y)} \phi_{\beta_{qy}}\, \Sigma_{qt} - \sum_{q \in \mathrm{pa}(y)} \beta_{qy}(X_q X_t - \Sigma_{qt}).$$

Collecting terms: the leading random parts combine as $X_y X_t - \sum_q \beta_{qy} X_q X_t = \varepsilon_y X_t$, the constant parts reduce to $-\Sigma_{yt} + \sum_q \beta_{qy}\Sigma_{qt} = -\mathrm{E}[\varepsilon_y X_t]$, and the remaining terms yield

$-\sum_{q\in\mathrm{pa}(y)} \Sigma_{qt}\, \phi_{\beta_{qy}}$, giving

$$\phi_{M_y(t)} = \varepsilon_y X_t - \mathrm{E}[\varepsilon_y X_t] - \sum_{q\in\mathrm{pa}(y)} \Sigma_{qt}\, \phi_{\beta_{qy}}.$$

This formula is mean-zero: the first two terms are centered by construction, and each $\Sigma_{qt}\phi_{\beta_{qy}}$ is mean-zero because $\phi_{\beta_{qy}}$ is mean-zero by the induction hypothesis. Since the formula was derived by differentiating $M_{y,\gamma}(t)$ along an arbitrary parametric submodel, it satisfies $\mathrm{E}\big[\phi_{M_y(t)} S\big] = \frac{d}{d\gamma}\big|_0 M_{y,\gamma}(t)$ for all submodel scores S; hence it is a valid influence function for $M_y(t)$. ∎

Lemma 6: For witness $y \in Y_v$ and target v, the influence function of the y-th row of the identification equation

$$R_{y,v} := \phi_{M_y(v)} - \sum_{p\in\mathrm{pa}(v)} \beta_{pv}\, \phi_{M_y(p)},$$

satisfies

$$R_{y,v} = \begin{cases} X_y\varepsilon_v, & y \notin \mathrm{htr}(v), \\ \varepsilon_y\varepsilon_v - \sum_{q\in\mathrm{pa}(y)} \mathrm{E}[X_q\varepsilon_v]\ \phi_{\beta_{qy}}, & y \in \mathrm{htr}(v). \end{cases}$$

Write $R_v := (R_{y_1,v}, \ldots, R_{y_k,v})^\top$.

Proof We substitute the formulas from Lemma 5 into $R_{y,v} := \phi_{M_y(v)} - \sum_{p\in\mathrm{pa}(v)} \beta_{pv}\phi_{M_y(p)}$ and simplify case by case. Since each $\phi_{M_y(t)}$ is a valid mean-zero influence function by Lemma 5, and $R_{y,v}$ is a linear combination of these with fixed scalar coefficients β_{pv}, the row contribution $R_{y,v}$ is itself a valid mean-zero influence function for the row residual $M_y(v) - \sum_p \beta_{pv} M_y(p)$.

Case $y \notin \mathrm{htr}(v)$. Lemma 5 gives $\phi_{M_y(t)} = X_y X_t - \Sigma_{yt}$, so

$$\begin{aligned} R_{y,v} &= X_y X_v - \Sigma_{yv} - \sum_{p\in\mathrm{pa}(v)} \beta_{pv}(X_y X_p - \Sigma_{yp}) \\ &= X_y\Big(X_v - \sum_p \beta_{pv} X_p\Big) - \mathrm{E}\Big[X_y\Big(X_v - \sum_p \beta_{pv} X_p\Big)\Big] = X_y\varepsilon_v - \mathrm{E}[X_y\varepsilon_v]\,. \end{aligned}$$

The centering term vanishes: since $X = (I-\beta)^{-\top}\varepsilon$, we have $\mathrm{E}[X_y\varepsilon_v] = [(I-\beta)^{-\top}\Omega]_{yv} = \sum_k [(I-\beta)^{-1}]_{ky}\,\Omega_{kv}$. Each summand is nonzero only when a directed path runs from k to y in G (Foygel et al., 2012, Equation 2.2) and $\Omega_{kv} \neq 0$, i.e., $k \in \{v\} \cup \mathrm{sib}(v)$; such a pair constitutes a half-trek from v to y, so every summand vanishes when $y \notin \mathrm{htr}(v)$. Hence $R_{y,v} = X_y\varepsilon_v$, which is mean-zero.

Case $y \in \mathrm{htr}(v)$. Lemma 5 gives $\phi_{M_y(t)} = \varepsilon_y X_t - \mathrm{E}[\varepsilon_y X_t] - \sum_q \Sigma_{qt}\phi_{\beta_{qy}}$; substituting and expanding,

$$R_{y,v} = \varepsilon_y X_v - \mathrm{E}[\varepsilon_y X_v] - \sum_{q\in \mathrm{pa}(y)} \Sigma_{qv}\phi_{\beta_{qy}} \\ - \sum_{p\in \mathrm{pa}(v)} \beta_{pv}\Big(\varepsilon_y X_p - \mathrm{E}[\varepsilon_y X_p] - \sum_{q\in \mathrm{pa}(y)} \Sigma_{qp}\phi_{\beta_{qy}}\Big).$$

The coefficient of $\phi_{\beta_{qy}}$ in the expansion is $-\Sigma_{qv} + \sum_p \beta_{pv}\Sigma_{qp}$; since

$$\mathrm{E}[X_q\varepsilon_v] = \mathrm{E}\Big[X_q(X_v - \sum_p \beta_{pv}X_p)\Big] = \Sigma_{qv} - \sum_p \beta_{pv}\Sigma_{qp},$$

this equals $-\mathrm{E}[X_q\varepsilon_v]$. The leading random terms give $\varepsilon_y(X_v - \sum_p \beta_{pv}X_p) = \varepsilon_y\varepsilon_v$; the centering terms give $-\mathrm{E}[\varepsilon_y\varepsilon_v]$, which equals zero because the HTC requires $Y_v \cap (\{v\} \cup \mathrm{sib}(v)) = \emptyset$ (Definition 3), so $y \notin \mathrm{sib}(v)$, and the marginal independence model (Definition 1) then gives $\varepsilon_y \perp\!\!\!\perp \varepsilon_v$, hence $\mathrm{E}[\varepsilon_y\varepsilon_v] = \Omega_{yv} = 0$. Assembling these three collections gives $R_{y,v} = \varepsilon_y\varepsilon_v - \sum_{q\in\mathrm{pa}(y)} \mathrm{E}[X_q\varepsilon_v]\ \phi_{\beta_{qy}}$, which is mean-zero: $\mathrm{E}[\varepsilon_y\varepsilon_v] = 0$ by the argument above, and each $\phi_{\beta_{qy}}$ is mean-zero by Lemma 5. ■

Theorem 7: The functional $\phi_{\beta_v} = A_v^{-1}R_v$ is an influence function for β_v in $\mathcal{M}_G$. For the edge $p_j \to v$, this yields $\phi_{\beta_{p_j v}} = e_j^\top A_v^{-1}R_v$.

Proof We show that $\phi_{\beta_v} = A_v^{-1}R_v$ satisfies the three properties required of an influence function: it is mean-zero, square-integrable, and satisfies $\mathrm{E}[\phi_{\beta_v}S] = \dot\beta_v$ for every parametric submodel with score S and $\dot\beta_v = \frac{d}{d\gamma}\big|_0 \beta_{v,\gamma}$.

Mean-zero and square-integrability. Since $R_{y,v}$ is mean-zero by Lemma 6 for every $y \in Y_v$, we have $\mathrm{E}[R_v] = 0$ and hence $\mathrm{E}[\phi_{\beta_v}] = 0$. Square-integrability $\phi_{\beta_v} \in L_0^2$ follows from $\varepsilon \in L_0^4$, which guarantees $R_{y,v} \in L_0^2$ for all y, v, and hence $\phi_{\beta_v} = A_v^{-1}R_v \in L_0^2$.

Formula. Since A_v is invertible at P_0 by Assumption A.(ii), the map $\beta_{v,\gamma} = A_{v,\gamma}^{-1}b_{v,\gamma}$ is differentiable in a neighbourhood of P_0, and the chain rule gives $\phi_{\beta_v} = A_v^{-1}\{\phi_{b_v} - (dA_v)\beta_v\}$. The ith component of $\phi_{b_v} - (dA_v)\beta_v$ equals $\phi_{M_{y_i}(v)} - \sum_{p\in\mathrm{pa}(v)} \beta_{pv}\phi_{M_{y_i}(p)} = R_{y_i,v}$ by Lemma 6, so stacking over i gives $\phi_{\beta_v} = A_v^{-1}R_v$.

Score equation. Let $\gamma \mapsto P_\gamma$ be any parametric submodel in $\mathcal{M}_G$ with score S and $\dot\beta_v = \frac{d}{d\gamma}\big|_0 \beta_{v,\gamma}$. We prove $\mathrm{E}[\phi_{\beta_v}S] = \dot\beta_v$ by induction on the HTC ordering $\prec$.

Base case. The base nodes satisfy $Y_v \cap \mathrm{htr}(v) = \emptyset$, so all witnesses are external, $Z_y = X_y$ is fixed, and Lemma 6 gives $R_{y,v} = X_y\varepsilon_v$ with no correction terms. The identification equation $\mathrm{E}_{P_\gamma}[X_{Y_v}\varepsilon_{v,\gamma}] = 0$ holds for all $P_\gamma \in \mathcal{M}_G$; differentiating at $\gamma = 0$ gives $\mathrm{E}[X_y\varepsilon_v S] + \mathrm{E}\Big[X_y X_{\mathrm{pa}(v)}^\top\Big]\,\dot\beta_v\cdot(-1) = 0$, hence $\mathrm{E}[X_y\varepsilon_v S] = (A_v)_{y,\cdot}\dot\beta_v$ for each y. Stacking gives $\mathrm{E}[R_v S] = A_v\dot\beta_v$ and therefore $\mathrm{E}[\phi_{\beta_v}S] = \dot\beta_v$.

Induction hypothesis. Assume $\mathrm{E}\big[\phi_{\beta_{qy}}S\big] = \dot\beta_{qy}$ for all $q \in \mathrm{pa}(y)$, $y \prec v$.

Inductive step. The identification equation $\mathrm{E}_{P_\gamma}[Z_{Y_v}(\beta_{\prec v,\gamma})\varepsilon_{v,\gamma}] = 0$ holds at every $P_\gamma \in \mathcal{M}_G$, since Y_v is a valid HTC witness set. Differentiating at $\gamma = 0$ via the product rule gives, for each witness $y \in Y_v$,

$$\mathrm{E}\Big[\dot{Z}_y\, \varepsilon_v\Big] + \mathrm{E}[Z_y\, \dot{\varepsilon}_v] + \mathrm{E}[Z_y \varepsilon_v S] = 0.$$

For $y \notin \mathrm{htr}(v)$: $Z_y = X_y$ is fixed, so $\dot{Z}_y = 0$. The structural equation $\varepsilon_v = X_v - \beta_v^\top X_{\mathrm{pa}(v)}$ gives $\dot{\varepsilon}_v = -X_{\mathrm{pa}(v)}^\top \dot{\beta}_v$, so $\mathrm{E}[Z_y \dot{\varepsilon}_v] = -(A_v)_{y,\cdot}\dot{\beta}_v$. Rearranging: $\mathrm{E}[Z_y \varepsilon_v S] = (A_v)_{y,\cdot}\dot{\beta}_v$.

For $y \in \mathrm{htr}(v)$: $Z_y = X_y - \beta_y^\top X_{\mathrm{pa}(y)}$, so $\dot{Z}_y = -X_{\mathrm{pa}(y)}^\top \dot{\beta}_y$, giving the pathwise derivative $\mathrm{E}\Big[\dot{Z}_y \varepsilon_v\Big] = -\sum_{q \in \mathrm{pa}(y)} \mathrm{E}[X_q \varepsilon_v]\, \dot{\beta}_{qy}$. As before, $\mathrm{E}[Z_y \dot{\varepsilon}_v] = -(A_v)_{y,\cdot}\dot{\beta}_v$. Rearranging:

$$\mathrm{E}[Z_y \varepsilon_v S] = (A_v)_{y,\cdot}\dot{\beta}_v + \sum_{q \in \mathrm{pa}(y)} \mathrm{E}[X_q \varepsilon_v]\, \dot{\beta}_{qy}.$$

For each internal witness, applying the induction hypothesis $\mathrm{E}\big[\phi_{\beta_{qy}} S\big] = \dot{\beta}_{qy}$ to the explicit formula $R_{y,v} = \varepsilon_y \varepsilon_v - \sum_q \mathrm{E}[X_q \varepsilon_v]\, \phi_{\beta_{qy}}$ from Lemma 6 gives $\mathrm{E}[R_{y,v} S] = \mathrm{E}[Z_y \varepsilon_v S] - \sum_q \mathrm{E}[X_q \varepsilon_v]\, \dot{\beta}_{qy}$. Substituting the differentiation result above, the correction terms cancel:

$$\mathrm{E}[R_{y,v} S] = \Big((A_v)_{y,\cdot}\dot{\beta}_v + \textstyle\sum_q \mathrm{E}[X_q \varepsilon_v]\, \dot{\beta}_{qy}\Big) - \textstyle\sum_q \mathrm{E}[X_q \varepsilon_v]\, \dot{\beta}_{qy} = (A_v)_{y,\cdot}\dot{\beta}_v.$$

For external witnesses, $\mathrm{E}[R_{y,v} S] = \mathrm{E}[X_y \varepsilon_v S] = (A_v)_{y,\cdot}\dot{\beta}_v$ directly. Stacking over all $y \in Y_v$ gives $\mathrm{E}[R_v S] = A_v \dot{\beta}_v$, and hence $\mathrm{E}[\phi_{\beta_v} S] = A_v^{-1} A_v \dot{\beta}_v = \dot{\beta}_v$.

In particular, for submodels varying only $\varepsilon \in \mathcal{E}$ with β held fixed, $\dot{\beta}_y = 0$ for all $y \preceq v$, so the conclusion above gives $\mathbb{E}[\phi_{\beta_v} S] = 0$, confirming orthogonality to the nuisance tangent space of $\mathcal{M}_G$. ■

Proposition 9: The HTC estimator $\hat{\beta}_{v,n}$ satisfies

$$\sqrt{n}(\hat{\beta}_{v,n} - \beta_v) \xrightarrow{d} \mathcal{N}(0, \mathcal{V}_v) \quad \text{with} \quad \mathcal{V}_v = A_v^{-1}\, \mathrm{E}\Big[R_v R_v^\top\Big]\, A_v^{-\top}.$$

For the edge $p_j \to v$, $\mathcal{V}_v[j,j] = e_j^\top \mathcal{V}_v e_j$. The covariance $\mathrm{E}\big[R_v R_v^\top\big]$ expands for $a, b \in Y_v$ to:

(i) $a, b \notin \mathrm{htr}(v)$:

$$\mathrm{E}[R_{a,v} R_{b,v}] = \mathrm{E}\big[X_a X_b \varepsilon_v^2\big].$$

Both witnesses are external, so $R_{a,v} = X_a \varepsilon_v$ and $R_{b,v} = X_b \varepsilon_v$ with no correction.

(ii) $a \in \mathrm{htr}(v)$, $b \notin \mathrm{htr}(v)$:

$$\mathrm{E}[R_{a,v} R_{b,v}] = \mathrm{E}\big[\varepsilon_a X_b \varepsilon_v^2\big] - \sum_{q \in \mathrm{pa}(a)} \mathrm{E}[X_q \varepsilon_v]\, \mathrm{E}\big[\phi_{\beta_{qa}} X_b \varepsilon_v\big].$$

The subtracted term corrects for the uncertainty of β_{qa}: estimating the residual instrument $Z_a = \varepsilon_a$ requires $\hat{\beta}_{qa}$, which introduces a correction proportional to $\phi_{\beta_{qa}}$ into $R_{a,v}$.

(iii) $a, b \in \mathrm{htr}(v)$:

$$\mathrm{E}[R_{a,v}R_{b,v}] = \mathrm{E}\big[\varepsilon_a\varepsilon_b\varepsilon_v^2\big] - \sum_{r\in\mathrm{pa}(b)} \mathrm{E}[X_r\varepsilon_v]\ \mathrm{E}[\varepsilon_a\varepsilon_v\phi_{\beta_{rb}}] - \sum_{q\in\mathrm{pa}(a)} \mathrm{E}[X_q\varepsilon_v]\ \mathrm{E}\big[\phi_{\beta_{qa}}\varepsilon_b\varepsilon_v\big]$$
$$+ \sum_{q\in\mathrm{pa}(a)}\sum_{r\in\mathrm{pa}(b)} \mathrm{E}[X_q\varepsilon_v]\ \mathrm{E}[X_r\varepsilon_v]\ \mathrm{E}\big[\phi_{\beta_{qa}}\phi_{\beta_{rb}}\big]\,.$$

With both witnesses internal, first-stage corrections from $\phi_{\beta_{qa}}$ and $\phi_{\beta_{rb}}$ enter both factors; the double sum captures the covariance between these two correction terms.

Proof The proof has two parts: we first establish asymptotic normality via the multi-stage Z-estimator expansion, then derive the three covariance formulas by substituting Lemma 6 into $\mathrm{E}[R_{a,v}R_{b,v}]$.

At each node v in HTC order, $\hat\beta_v$ solves $n^{-1}\sum_i Z_{Y_v}(\hat\beta_{\prec v})\varepsilon_v(X^{(i)},\hat\beta_v) = 0$, where the instruments $Z_y = X_y - \hat\beta_y^\top X_{\mathrm{pa}(y)}$ for $y \in \mathrm{htr}(v)$ depend on the earlier-stage estimates $\hat\beta_{\prec v}$. Applying Z-estimator theory stage by stage (van der Vaart, 1998, §5), the asymptotic expansion is $\sqrt{n}(\hat\beta_v - \beta_v) = n^{-1/2}\sum_i \phi_{\beta_v}(X^{(i)}) + o_P(1)$, where the correction terms $\phi_{\beta_{qy}}$ in R_v propagate the estimation uncertainty from $\hat\beta_{\prec v}$ into the instruments, matching exactly the influence function $\phi_{\beta_v} = A_v^{-1}R_v$ of Theorem 7. Since ϕ_{β_v} is mean-zero and L_0^2 by Theorem 7, the central limit theorem gives $\sqrt{n}(\hat\beta_v - \beta_v) \xrightarrow{d} \mathcal{N}(0, \mathrm{E}\big[\phi_{\beta_v}\phi_{\beta_v}^\top\big])$. Since $A_v = \mathrm{E}\Big[Z_{Y_v}X_{\mathrm{pa}(v)}^\top\Big]$ is a non-random population moment at P_0, substituting $\phi_{\beta_v} = A_v^{-1}R_v$ gives $\mathrm{E}\Big[\phi_{\beta_v}\phi_{\beta_v}^\top\Big] = A_v^{-1}\mathrm{E}\big[R_vR_v^\top\big]\,A_v^{-\top} = \mathcal{V}_v$.

It remains to derive the three formulas for $\mathrm{E}[R_{a,v}R_{b,v}]$. In each case we substitute the formulas from Lemma 6 and expand; the scalars $\mathrm{E}[X_q\varepsilon_v]$ are non-random population moments and factor out of joint expectations. For $a, b \notin \mathrm{htr}(v)$: Lemma 6 gives $R_{a,v} = X_a\varepsilon_v$ and $R_{b,v} = X_b\varepsilon_v$, so $\mathrm{E}[R_{a,v}R_{b,v}] = \mathrm{E}\big[X_aX_b\varepsilon_v^2\big]$. For $a \in \mathrm{htr}(v)$, $b \notin \mathrm{htr}(v)$: Lemma 6 gives $R_{a,v} = \varepsilon_a\varepsilon_v - \sum_{q\in\mathrm{pa}(a)}\mathrm{E}[X_q\varepsilon_v]\,\phi_{\beta_{qa}}$ and $R_{b,v} = X_b\varepsilon_v$; multiplying and factoring out the non-random scalar $\mathrm{E}[X_q\varepsilon_v]$:

$$\mathrm{E}[R_{a,v}R_{b,v}] = \mathrm{E}\big[\varepsilon_aX_b\varepsilon_v^2\big] - \sum_{q\in\mathrm{pa}(a)} \mathrm{E}[X_q\varepsilon_v]\ \mathrm{E}\big[\phi_{\beta_{qa}}X_b\varepsilon_v\big]\,.$$

For $a, b \in \mathrm{htr}(v)$: both factors carry correction terms; substituting Lemma 6 gives

$$R_{a,v}R_{b,v} = \Big(\varepsilon_a\varepsilon_v - \sum_{q\in\mathrm{pa}(a)} \mathrm{E}[X_q\varepsilon_v]\,\phi_{\beta_{qa}}\Big)\Big(\varepsilon_b\varepsilon_v - \sum_{r\in\mathrm{pa}(b)} \mathrm{E}[X_r\varepsilon_v]\,\phi_{\beta_{rb}}\Big).$$

Expanding bilinearly and factoring out the non-random scalars $\mathrm{E}[X_q\varepsilon_v]$ and $\mathrm{E}[X_r\varepsilon_v]$ from each cross-term yields the four-term formula in the proposition. ■

Lemma 10: The covariance $\mathrm{E}\big[R_vR_v^\top\big]$ reduces to finitely many fourth-order moments of observables. Each cross-variance $\mathrm{E}\big[\phi_{\beta_{qa}}X_b\varepsilon_v\big]$ equals

$$\mathrm{E}\big[\phi_{\beta_{qa}}X_b\varepsilon_v\big] = e_j^\top A_a^{-1}\,\mathrm{E}[R_aX_b\varepsilon_v]\,,$$

and for each $y \in Y_a$,

$$\mathrm{E}[R_{y,a} X_b \varepsilon_v] = \begin{cases} \mathrm{E}[X_y \varepsilon_a X_b \varepsilon_v] & y \notin \mathrm{htr}(a), \\ \mathrm{E}[\varepsilon_y \varepsilon_a X_b \varepsilon_v] - \sum_{q' \in \mathrm{pa}(y)} \mathrm{E}\left[X_{q'} \varepsilon_a\right] \, \mathrm{E}\left[\phi_{\beta_{q'y}} X_b \varepsilon_v\right] & y \in \mathrm{htr}(a). \end{cases}$$

Each application of the $y \in \mathrm{htr}(a)$ branch strictly descends in the HTC order $\prec$; the recursion bottoms out at nodes u with $Y_u \cap \mathrm{htr}(u) = \emptyset$, where every entry equals a pure fourth-order moment of observables.

Proof The proof establishes the three claims in the lemma in order: the reduction formula (Claim 1), the case expansion (Claim 2), and termination of the recursion (Claim 3), from which computability follows.

Claim 1 — Reduction formula. By Theorem 7, the vector influence function for β_a is $\phi_{\beta_a} = A_a^{-1} R_a \in \mathbb{R}^{|\mathrm{pa}(a)|}$; the scalar component for the edge $q = p_j \to a$ is its jth entry, $\phi_{\beta_{qa}} = e_j^\top A_a^{-1} R_a$. Since $A_a = \mathrm{E}\left[Z_{Y_a} X_{\mathrm{pa}(a)}^\top\right]$ is a non-random population moment at P_0, it factors outside the expectation:

$$\mathrm{E}\left[\phi_{\beta_{qa}} X_b \varepsilon_v\right] = \mathrm{E}\left[e_j^\top A_a^{-1} R_a X_b \varepsilon_v\right] = e_j^\top A_a^{-1} \mathrm{E}[R_a X_b \varepsilon_v]\,.$$

Claim 2 — Case expansion. It remains to expand $\mathrm{E}[R_{y,a} X_b \varepsilon_v]$ for each $y \in Y_a$ by substituting Lemma 6 applied at node a. For $y \notin \mathrm{htr}(a)$: Lemma 6 gives $R_{y,a} = X_y \varepsilon_a$, so $\mathrm{E}[R_{y,a} X_b \varepsilon_v] = \mathrm{E}[X_y \varepsilon_a X_b \varepsilon_v]$, a raw fourth-order moment. For $y \in \mathrm{htr}(a)$: Lemma 6 gives $R_{y,a} = \varepsilon_y \varepsilon_a - \sum_{q' \in \mathrm{pa}(y)} \mathrm{E}\left[X_{q'} \varepsilon_a\right] \phi_{\beta_{q'y}}$, where the scalars $\mathrm{E}\left[X_{q'} \varepsilon_a\right]$ are non-random population moments (note: the subscript is ε_a, arising from applying Lemma 6 at node a, not ε_v). Factoring these scalars outside the expectation gives

$$\mathrm{E}[R_{y,a} X_b \varepsilon_v] = \mathrm{E}[\varepsilon_y \varepsilon_a X_b \varepsilon_v] - \sum_{q' \in \mathrm{pa}(y)} \mathrm{E}\left[X_{q'} \varepsilon_a\right] \, \mathrm{E}\left[\phi_{\beta_{q'y}} X_b \varepsilon_v\right],$$

which introduces new cross-variances $\mathrm{E}\left[\phi_{\beta_{q'y}} X_b \varepsilon_v\right]$ at the earlier node $y \prec a$.

Claim 3 — Termination and computability. Applying Claim 1 to each new cross-variance $\mathrm{E}\left[\phi_{\beta_{q'y}} X_b \varepsilon_v\right]$ reduces it to cross-variances $\mathrm{E}\left[R_{y',y} X_b \varepsilon_v\right]$ at node $y \prec a$; the internal case of Claim 2 then introduces cross-variances at nodes strictly below y in HTC order, and so on. Since $\prec$ is a strict partial order on the finite index set V, this descent terminates; at base nodes u with $Y_u \cap \mathrm{htr}(u) = \emptyset$, all witnesses are external, and the case expansion gives $\mathrm{E}[R_{y,u} X_b \varepsilon_v] = \mathrm{E}[X_y \varepsilon_u X_b \varepsilon_v]$, a pure fourth-order moment of observables requiring no further recursion. Since $\mathrm{E}\left[R_v R_v^\top\right]$ is expressed through Proposition 9 in terms of finitely many cross-variances, and each reduces in finitely many steps to such fourth-order moments, $\mathrm{E}\left[R_v R_v^\top\right]$ is computable in finitely many steps. ■

Proposition 12: For $\alpha \in (0, 1)$, write $z_{\alpha/2}$ for the $\alpha/2$ quantile of the standard normal distribution and $\chi^2_{k,1-\alpha}$ for the $(1-\alpha)$ quantile of the chi-squared distribution with k degrees of freedom χ^2_k. The following hold.

(i) *(Confidence ellipsoid.)* An asymptotic $(1-\alpha)$ confidence region for $\beta_v \in \mathbb{R}^{|\mathrm{pa}(v)|}$ is

$$\mathcal{C}_n = \{\beta \in \mathbb{R}^{|\mathrm{pa}(v)|} : n(\hat{\beta}_v - \beta)^\top \hat{\mathcal{V}}_v^{-1}(\hat{\beta}_v - \beta) \le \chi^2_{|pa(v)|,1-\alpha}\}.$$

(ii) *(Marginal intervals.)* An asymptotic $(1-\alpha)$ confidence interval for the edge $\beta_{p_j v}$ is

$$\hat{\beta}_{p_j v} \pm z_{\alpha/2}\sqrt{\hat{\mathcal{V}}_v[j,j]/n}.$$

(iii) *(Wald test.)* For $H_0\colon C\beta_v = c$ with full row rank $C \in \mathbb{R}^{r\times|\mathrm{pa}(v)|}$ and $c \in \mathbb{R}^r$, the Wald statistic satisfies

$$W_n = n(C\hat{\beta}_v - c)^\top (C\hat{\mathcal{V}}_v C^\top)^{-1}(C\hat{\beta}_v - c) \xrightarrow{d} \chi^2_r \quad \text{under } H_0.$$

The matrix $C\hat{\mathcal{V}}_v C^\top$ is generically invertible because C has full row rank.

Proof The finite fourth moments $\varepsilon \in L_0^4$ ensure that every term entering $\hat{\mathcal{V}}_v$ is the empirical average of an L_0^2-integrable function; by the law of large numbers, $\hat{\mathcal{V}}_v \xrightarrow{p} \mathcal{V}_v$.

(i) Since $\sqrt{n}(\hat{\beta}_v - \beta_v) \xrightarrow{d} \mathcal{N}(0, \mathcal{V}_v)$ by Proposition 9, the continuous mapping theorem gives $n(\hat{\beta}_v - \beta_v)^\top \mathcal{V}_v^{-1}(\hat{\beta}_v - \beta_v) \xrightarrow{d} \chi^2_{|\mathrm{pa}(v)|}$. By Slutsky's theorem and $\hat{\mathcal{V}}_v \xrightarrow{p} \mathcal{V}_v$, replacing $\mathcal{V}_v$ by $\hat{\mathcal{V}}_v$ does not change the limit, giving the confidence ellipsoid.

(ii) By the continuous mapping theorem (projection to coordinate j), $\sqrt{n}(\hat{\beta}_{p_j v} - \beta_{p_j v}) \xrightarrow{d} \mathcal{N}(0, \mathcal{V}_v[j,j])$. Since $\hat{\mathcal{V}}_v[j,j] \xrightarrow{p} \mathcal{V}_v[j,j] > 0$, applying Slutsky's theorem gives $\sqrt{n}(\hat{\beta}_{p_j v} - \beta_{p_j v})/\sqrt{\hat{\mathcal{V}}_v[j,j]} \xrightarrow{d} \mathcal{N}(0,1)$, yielding the stated interval.

(iii) Under $H_0\colon C\beta_v = c$, the continuous mapping theorem gives $\sqrt{n}(C\hat{\beta}_v - c) \xrightarrow{d} \mathcal{N}(0, C\mathcal{V}_v C^\top)$. By Slutsky's theorem and the continuous mapping theorem (quadratic form), replacing $C\mathcal{V}_v C^\top$ by $C\hat{\mathcal{V}}_v C^\top$ gives $W_n \xrightarrow{d} \chi^2_r$. Since $\mathcal{V}_v \succ 0$ generically under Assumption A.(i) and C has full row rank r, $C\mathcal{V}_v C^\top$ is positive definite; by consistency of $\hat{\mathcal{V}}_v$, $C\hat{\mathcal{V}}_v C^\top$ is almost surely invertible for large n. ■

Appendix C. Simulation Study: Setup and Experiments

C.1. Calibration Example

The graph in Figure 1 has directed edges $1 \to 2$, $2 \to 4$, $1 \to 5$, $3 \to 5$ and bidirected edges $1 \leftrightarrow 3$, $1 \leftrightarrow 4$, $1 \leftrightarrow 5$. The HTC ordering is $2 \prec 4 \prec 5$. The true parameter matrix has non-zero entries

$$\beta_{12} = 0.8, \quad \beta_{24} = 0.7, \quad \beta_{15} = 0.8, \quad \beta_{35} = 0.6.$$

The error covariance matrix $\Omega = \mathrm{E}\big[\varepsilon\varepsilon^\top\big]$, consistent with the marginal independence constraints $\varepsilon_w \perp\!\!\!\perp \varepsilon_{V\setminus(\{w\}\cup N_B(\{w\}))}$ for each $w \in V$, has non-zero off-diagonal entries $\Omega_{13} = 0.3$, $\Omega_{14} = 0.75$, $\Omega_{15} = 0.2$ only; all other off-diagonal entries are zero. Positive definiteness holds since $\Omega_{13}^2 + \Omega_{14}^2 + \Omega_{15}^2 = 0.6525 < 1$.

Two error distributions are studied. The Gaussian distribution samples $\varepsilon \sim \mathcal{N}_5(0, \Omega)$ directly. The non-Gaussian distribution constructs errors ε from five independent centered Gamma distributions $\varepsilon'_i \sim (\mathrm{Gamma}(2,1) - 2)/\sqrt{2}$ via the Cholesky factorization of the $\{1, 3, 4, 5\}$ block of Ω (node 2 is independent of all others). Both distributions match the error covariance Ω exactly.

Experiment to Figure 2. We created 100 replications of datasets with sample sizes $n \in \{250, 500, 1000, 2000, 4000, 8000\}$ for each of the two error distributions. For each dataset, the recursive HTC estimator $\hat{\beta}_5 = (\hat{\beta}_{15}, \hat{\beta}_{35})^\top$ and the estimated asymptotic covariance matrix $\hat{\mathcal{V}}_5$ were computed as in Proposition 9. The four panels of the combined figure compare empirical and estimated quantities across replications:

(i) $\mathrm{Var}(\hat{\beta}_{15})$ against $\hat{\mathcal{V}}_5[1,1]/n$,

(ii) $\mathrm{Var}(\hat{\beta}_{35})$ against $\hat{\mathcal{V}}_5[2,2]/n$,

(iii) $|\mathrm{Cov}(\hat{\beta}_{15}, \hat{\beta}_{35})|$ against $|\hat{\mathcal{V}}_5[1,2]|/n$ (absolute value, since the covariance is negative; log scale),

(iv) $\mathrm{Cor}(\hat{\beta}_{15}, \hat{\beta}_{35})$ against $\hat{\mathcal{V}}_5[1,2]/\sqrt{\hat{\mathcal{V}}_5[1,1]\hat{\mathcal{V}}_5[2,2]}$ (linear scale).

Estimated quantities are averaged over the 100 replications. The correlation panel (iv) is included to demonstrate that the estimators $\hat{\beta}_{15}$ and $\hat{\beta}_{35}$ are asymptotically negatively correlated, a consequence of the internal witness $4 \in \mathrm{htr}(5)$ introducing a shared correction term in $R_{4,5}$.

Experiment to Figure 3. We created 100 replications of datasets with sample sizes $n \in \{250, 500, 1000, 2000, 4000, 8000\}$ for each of the two error distributions. Three valid HTC witness sets for node 2 are compared: $Y_2 \in \big\{\{1\}, \{3\}, \{5\}\big\}$. Since $\mathrm{htr}(2) = \{4\}$, every witness is external and $Z_y = X_y$. For each dataset and witness set, the recursive HTC estimator $\hat{\beta}_{12}$ and the estimated variance $\hat{\mathcal{V}}_2[1,1]/n$ were computed as in Proposition 9. Each of the three panels plots $\mathrm{Var}(\hat{\beta}_{12})$ against the mean of $\hat{\mathcal{V}}_2[1,1]/n$ over the 100 replications, on a shared log-scale y-axis.

C.2. Inference Example

The five-node graph of Figure 4 has directed edges $1 \to 2$, $2 \to 3$, $3 \to 2$, $3 \to 4$, $4 \to 5$ and bidirected edges $1 \leftrightarrow 2$, $1 \leftrightarrow 4$, $1 \leftrightarrow 5$, $3 \leftrightarrow 4$, $4 \leftrightarrow 5$. The true parameter matrix has non-zero entries

$$\beta_{12} = 0.8, \quad \beta_{23} = 0.7, \quad \beta_{32} = 0.4, \quad \beta_{34} = 0.8, \quad \beta_{45} = 0,$$

with $\beta_{45} = 0$ included to illustrate inference at a zero coefficient. The directed cycle $2 \to 3 \to 2$ is admissible since $\det(I - \beta) \neq 0$ at the true parameter. The HTC ordering is $3 \prec 5 \prec 2 \prec 4$: node 3 is identified first via the external witness $Y_3 = \{1\}$; node 5 follows with the internal witness $Y_5 = \{3\}$; node 2 is identified using the internal witnesses $Y_2 = \{3, 5\} \subset \mathrm{htr}(2)$; node 4 is identified last through the internal witness $Y_4 = \{2\}$. The error covariance matrix $\Omega = \mathrm{E}\left[\varepsilon\varepsilon^\top\right]$, consistent with the marginal independence constraints $\varepsilon_w \perp\!\!\!\perp \varepsilon_{V \setminus (\{w\} \cup N_B(\{w\}))}$ for each $w \in V$, has non-zero off-diagonal entries $\Omega_{12} = 0.5$, $\Omega_{14} = 0.25$, $\Omega_{15} = 0.75$, $\Omega_{34} = 0.5$, $\Omega_{45} = 0.4$; all remaining off-diagonal entries are zero and all diagonal entries equal one. Errors are drawn as $\varepsilon \sim \mathcal{N}_5(0, \Omega)$ and observables are computed via $X = (I - \beta)^{-\top}\varepsilon$ for a single dataset of size $n = 1000$.

Appendix D. Application: Fulton Fish Market

The dataset of Angrist et al. (2000) records 97 daily observations from December 1991 to May 1992, with three variables: log quantity-weighted average price (`lavgprc`), log total quantity sold (`ltotqty`), and wave height off the Long Island coast (`wave2` and `wave3`). Following Angrist et al. (2000), day-of-week effects (Monday–Thursday dummies, Friday as baseline) are partialed out from all three variables by regression, yielding mean-zero working residuals that remove the systematic day-of-week trading patterns and make the i.i.d. approximation underlying the asymptotic variance estimator more credible.

In the directed mixed graph, $\mathrm{pa}(\textit{demand}) = \{\textit{supply}\}$ and $\mathrm{sib}(\textit{demand}) = \{\textit{supply}\}$, so admissible HTC witnesses must lie outside $\{\textit{supply}, \textit{demand}\}$. Both *wave2* and *wave3* satisfy this condition for $\textit{supply} \to \textit{demand}$ and are valid external witnesses, so the instruments are the raw variables $Z_{\textit{wave2}} = X_{\textit{wave2}}$ and $Z_{\textit{wave3}} = X_{\textit{wave3}}$, respectively, with first-stage correlations

$$\mathrm{Cor}(\textit{wave2}, \textit{supply}) = 0.4931 \quad \text{and} \quad \mathrm{Cor}(\textit{wave3}, \textit{supply}) = 0.3798.$$

The 3-day average smooths the sharp day-before supply shock and dilutes the first-stage signal, widening the asymptotic standard error from 0.3827 to 0.4246. The estimates $\hat{\hat{\delta}} = -0.8410$ (*wave2*) and $\hat{\hat{\delta}} = -0.7611$ (*wave3*) lie within one standard error of each other, confirming that both instruments identify the same structural parameter.

The supply slope (coefficient on $\textit{demand} \to \textit{supply}$) is not identified, as no demand-side instrument is available in this dataset.