The Price of Valid Inference After Causal Discovery

Dongxin Guo[1] **Jikun Wu**[1] **Siu Ming Yiu**[1]

[1]Brain Investing Limited, Hong Kong, China

Abstract

Causal effects estimated by covariate adjustment on a discovered graph have invalid confidence intervals, because the graph-selection step is ignored. We develop a unified selective-inference framework. For constraint-based discovery with latent confounders (FCI) under Gaussianity, we prove that the selection event, given the execution trace and signs, is a polyhedron in Fisher-z space; along the inference direction it becomes polynomial/rational constraints solving to a union of intervals. Inverting the truncated-Gaussian approximate pivot gives exact finite-sample $1-\alpha$ coverage, with known σ^2, of the adjustment functional $\gamma_1(S^*)$ when $\{i\}\cup S^*$ contains the outcome's Markov blanket, and approximate coverage otherwise, with a non-vanishing distortion governed by the variance-ratio excess $\rho^2 = \sigma^2_{j|X}/\sigma^2_{j|-j} - 1$ (small under sparsity); the $\hat{\sigma}^2$ plug-in does not remove it. This is the first truncation-set characterization handling latent confounders. Coverage of the structural effect $\beta_{i\to j}$ is asymptotically valid up to the same distortion when the adjustment set is valid (exact when $\rho^2 = 0$), via high-dimensional FCI consistency under $d = o(\sqrt{n})$. For unmodified GES, heuristic Taylor linearization indicates approximate coverage $1 - \alpha - O(d^2/\sqrt{n})$ when $n \gg d^4(\log d)^2$. In the nonparametric setting, $\Theta(n^{3/4})$-split discovery incurs width ratio $1 + \Theta(n^{-1/4})$ versus a known-graph oracle.

1 INTRODUCTION

A central promise of causal discovery is the ability to estimate causal effects from purely observational data. In practice, this is a two-step pipeline: first, learn a causal graph from data using algorithms such as PC [Spirtes et al., 2000], FCI [Spirtes et al., 1995, Zhang, 2008], or GES [Chickering, 2002]; then, use the learned graph to identify an adjustment set and estimate causal effects via regression [Pearl, 2009, Maathuis et al., 2009]. This pipeline is widely deployed in biology [Maathuis et al., 2010, Sachs et al., 2005], epidemiology, and the social sciences, yet it harbors a statistical flaw: the confidence intervals produced in the second step treat the learned graph as fixed and known, ignoring the uncertainty introduced by data-driven graph selection.

This flaw is not merely theoretical. Gradu et al. [2025] demonstrate that naive two-step confidence intervals can have severely degraded coverage at the nominal 95% level. Consider a biologist using FCI on protein signaling data [Sachs et al., 2005] to discover a regulatory pathway and then estimate a signaling effect. If the reported 95% confidence interval has true coverage of only 77%, downstream decisions (such as which drug targets to pursue) are based on misleadingly precise estimates. The severity of this miscoverage stems from *post-selection inference* [Berk et al., 2013, Kuchibhotla et al., 2022]: when the same data are used to select a model and estimate parameters, classical inference procedures become invalid.

The selective inference literature has developed powerful tools for this problem in regression [Lee et al., 2016, Tibshirani et al., 2016, Lockhart et al., 2014, Fithian et al., 2014, Taylor and Tibshirani, 2015], but extending them to causal discovery requires additional work. Causal discovery involves complex combinatorial searches through graphical structures, making selection event characterization substantially more challenging than in, say, the lasso path. Chang et al. [2026] recently obtained results for the PC algorithm under Gaussianity and causal sufficiency, providing the first asymptotically valid post-selection confidence intervals for causal effects after PC-based discovery via a resampling-and-screening procedure. However, three critical frontiers remain open:

1. **Latent confounders.** The PC algorithm assumes causal sufficiency. In practice, latent confounders are

Proceedings of the 42nd Conference on Uncertainty in Artificial Intelligence (UAI 2026), PMLR 337:1804–1822.

ubiquitous, and the FCI algorithm [Spirtes et al., 1995, Zhang, 2008, Colombo et al., 2012] is the standard tool, outputting partial ancestral graphs (PAGs) [Richardson and Spirtes, 2002]. Developing selective inference for FCI presents several structural challenges beyond the causally sufficient setting (see Section 3).

2. **Score-based discovery.** GES [Chickering, 2002, Nandy et al., 2018] and related methods [Lam et al., 2022] are among the most popular causal discovery approaches, yet their score-comparison-based selection events have a different structure. Gradu et al. [2025] address GES but require modifying the algorithm; to our knowledge, no *conditional* selective-inference method handles standard, unmodified GES.

3. **Cost of splitting.** It is unclear whether valid post-selection inference is achievable beyond the Gaussian setting and, if it is not, what the cost of alternatives such as sample splitting [Rinaldo et al., 2019] or data fission [Rasines and Young, 2023, Leiner et al., 2025] must be.

Figure 1 illustrates the two-step pipeline and where our framework intervenes. Table 1 positions our contributions relative to prior work. Our central contribution is the explicit truncation-set characterization for FCI (the first to handle latent confounders), which yields selective intervals that match or exceed the coverage of sample splitting while producing narrower intervals in the large-n Gaussian regime. We extend this approach to unmodified GES (approximate) and compute a splitting width penalty as secondary results.

Contributions. We develop a unified framework addressing all three frontiers:

- We prove that the event {FCI outputs PAG $\widehat{\mathcal{P}}$}, conditional on the execution path and observed signs, admits an explicit characterization as linear constraints in Fisher z-statistic space, enabling selective confidence intervals for the selected adjustment functional $\gamma_1(S^*)$ under Gaussianity (Theorems 1, 2; Section 3). The approximate pivot is exact for $\gamma_1(S^*)$ when $\{i\} \cup S^*$ contains the outcome's Markov blanket (Remark 1) and approximately valid otherwise, with coverage distortion governed by $\rho^2 = \sigma^2_{j|X}/\sigma^2_{j|-j} - 1$, small under sparsity but not vanishing in n; the $\hat{\sigma}^2$ plug-in does not remove this distortion. Coverage of the structural effect $\beta_{i\to j}$ additionally requires that the selected adjustment set be valid.
- We extend this approach to unmodified GES with BIC scoring via Monte Carlo sampling, achieving approximate coverage with a heuristic error bound of $O(d^2/\sqrt{n})$ (Claim 3; Section 4). The bound is vacuous at tested dimensions; coverage there is supported empirically.
- As extensions, we compute the width penalty of the $\Theta(n^{3/4})$ sample-splitting allocation (width ratio $1 + \Theta(n^{-1/4})$; Proposition 4, Section 5), and verify pipeline consistency (a sanity check that selective coverage converges to $1-\alpha$ when the graph is consistently recovered) under high-dimensional Gaussian scaling with $d = o(\sqrt{n})$ (Theorem 5); the method's practical value is in finite samples where graph uncertainty is real.
- Experiments across four data-generating settings and four analyses (including d up to 200) show that our selective intervals correct severe $\beta_{i\to j}$-miscoverage (72–82% $\to$ 93.8–95.3% for FCI under Gaussianity) with practical computational cost (Section 6).

2 BACKGROUND AND PROBLEM SETUP

2.1 CAUSAL DISCOVERY ALGORITHMS

Let $X = (X_1, \ldots, X_d)^\top$ be a random vector with distribution P that is Markov and faithful to a DAG $\mathcal{G}^*$ over $\{1, \ldots, d\}$ [Pearl, 2009, Spirtes et al., 2000]. Under *causal sufficiency*, $\mathcal{G}^*$ is identifiable up to its Markov equivalence class, represented by a CPDAG [Chickering, 2002, Squires and Uhler, 2023]. When latent confounders may be present, the identifiable structure is the Markov equivalence class of MAGs, represented by a PAG [Richardson and Spirtes, 2002, Zhang, 2008].

The PC algorithm [Spirtes et al., 2000, Colombo and Maathuis, 2014] learns a CPDAG under causal sufficiency by performing conditional independence tests at level α_0. For Gaussian data, testing $X_i \perp\!\!\!\perp X_j \mid X_S$ reduces to testing whether $r_{ij\cdot S} = 0$ by comparing the Fisher z-statistic $|z_{ij\cdot S}|$ against a critical value c_{α_0}.

The FCI algorithm [Spirtes et al., 1995, Zhang, 2008, Colombo et al., 2012] extends PC to the latent variable setting. FCI first runs a PC-like skeleton discovery phase using conditional independence tests with conditioning sets up to size $d-2$ (unlike PC, which is limited by vertex degree), then applies orientation rules R0–R10 to produce a PAG $\widehat{\mathcal{P}}$. The key structural difference from PC is that FCI uses *larger conditioning sets* and additional orientation rules; we show in Section 3 that both can be captured as linear constraints on sufficient statistics.

GES [Chickering, 2002] is a score-based method that greedily searches through equivalence classes, scored by BIC: $\mathrm{BIC}(\mathcal{G}) = -2\ell(\mathcal{G}) + |\mathcal{G}| \log n$. GES is consistent under causal sufficiency and Gaussianity [Chickering, 2002, Nandy et al., 2018].

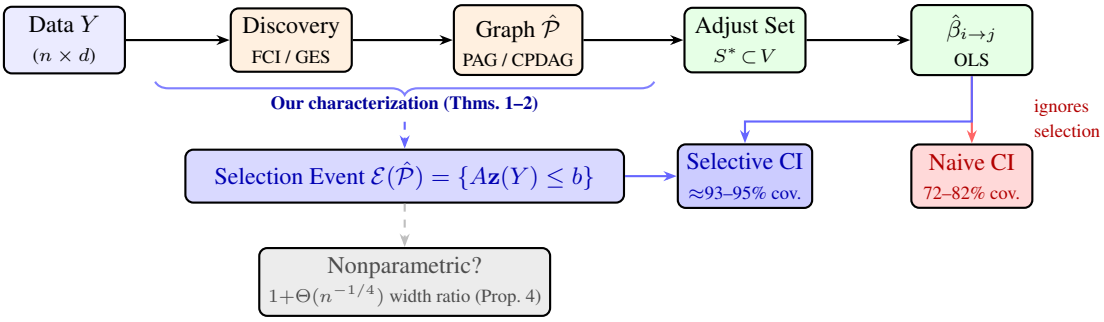

Figure 1: The causal discovery–inference pipeline and our contributions. **Top row**: the standard two-step pipeline produces an effect estimate $\hat{\beta}_{i \to j}$. **Bottom**: naive CIs ignore the graph selection event, yielding 72–82% β-coverage under Gaussianity at nominal 95%. Our framework characterizes the selection event via linear constraints in Fisher z-statistic space (Theorems 1–2; polynomial or rational constraints along the inference direction), enabling selective CIs with near-nominal coverage. Proposition 4 quantifies the width penalty of sample splitting under the $\Theta(n^{3/4})$ allocation: a $1 + \Theta(n^{-1/4})$ width ratio.

2.2 CAUSAL EFFECT ESTIMATION FROM LEARNED GRAPHS

Given a learned CPDAG or PAG, the total causal effect of X_i on X_j can be estimated via covariate adjustment [Pearl, 2009, Perkovic et al., 2015, 2018, Maathuis and Colombo, 2015]. Under linearity, $X = B^\top X + \varepsilon$ with $\varepsilon \sim \mathcal{N}(0, \Omega)$ for diagonal Ω (independent errors, as required by the DAG semantics), the total causal effect is $\beta_{i \to j} = (I - B)^{-1}_{ij}$. When the graph is known, an adjustment set S yields $\widehat{\beta}_{i \to j} = \hat{\gamma}_1$ from regressing X_j on (X_i, X_S), with standard CIs based on the regression standard error [Maathuis et al., 2009, Henckel et al., 2022, Smucler et al., 2022].

When the graph is *estimated*, the adjustment set S is data-dependent, and the classical confidence interval treats S as fixed, leading to invalid coverage. Following Chang et al. [2026], we target the *graph-independent* total causal effect $\beta_{i \to j} = (I - B)^{-1}_{ij}$, a property of the underlying SEM. Gradu et al. [2025] explicitly target the OLS projection coefficient onto the learned adjustment set. Our selective interval is exact for this same quantity $\gamma_1(S^*)$; coverage of the structural effect $\beta_{i \to j}$ additionally requires that S^* be a valid adjustment set. These are distinct inferential targets answering different scientific questions: $\beta_{i \to j}$ is a structural property of the SEM, fixed regardless of which graph is selected, while $\gamma_1(S^*)$, the regression coefficient implied by the discovered adjustment set, varies with the discovered graph.

2.3 THE SELECTIVE INFERENCE FRAMEWORK

Let $Y = (y_1, \ldots, y_n)^\top$ denote the $n \times d$ data matrix, and let $\hat{\mathcal{G}}(Y)$ denote the graph (CPDAG or PAG) estimated by a discovery algorithm. Define the *selection event* $\mathcal{E}(\widehat{\mathcal{P}}) = \{Y' \in \mathbb{R}^{n \times d} : \hat{\mathcal{G}}(Y') = \widehat{\mathcal{P}}\}$; $\mathcal{E}(\widehat{\mathcal{P}})$ depends on the graph estimation procedure $\hat{\mathcal{G}}$. Following Lee et al. [2016], Fithian et al. [2014], a selective CI for $\beta_{i \to j}$ at level $1 - \alpha$ satisfies:

$$\mathbb{P}\big(\beta_{i \to j} \in \mathrm{CI}_\alpha \mid \mathcal{E}(\widehat{\mathcal{P}})\big) \geq 1 - \alpha. \tag{1}$$

When $\mathcal{E}(\widehat{\mathcal{P}})$ can be written as $\{Az(Y) \leq b\}$ for the vector of sufficient statistics $\mathbf{z}(Y)$, inference reduces to a truncated pivot [Lee et al., 2016, Tibshirani et al., 2016]. The central challenge is deriving (A, b) for each discovery algorithm.

Table 1 summarizes our positioning (*MB = Markov-blanket condition of Remark 1, which generically does not hold; all experiments use the $\hat{\sigma}^2$ plug-in). To the best of our knowledge, this is the first work to provide an explicit truncation-set characterization of the selection event for FCI (handling latent confounders), and to provide conditional selective inference for unmodified score-based methods. Chang et al. [2026] target the true population effect $\beta_{i \to j}$ via a modular resampling approach that could in principle be applied to other conditional-independence-based algorithms including FCI; our contribution is the direct conditional/geometric characterization with its efficiency benefit in the Gaussian regime.

3 SELECTIVE INFERENCE FOR FCI

Our main result establishes that FCI's selection event admits an explicit characterization as linear constraints on the Fisher z-statistics (affine in z-space; polynomial or rational in the inference direction after conditioning, since the map $Y \mapsto \mathbf{z}(Y)$ is nonlinear). This yields a computable truncation set for selective inference on causal effects estimated from PAGs.

3.1 STRUCTURAL DIFFERENCES FROM PC

Our characterization builds on the selective inference framework of Lee et al. [2016]. While FCI shares PC's skeleton

Table 1: Comparison of Post-Selection Inference Methods for Causal Effects After Causal Discovery.

Feature	Chang+ (2026)	Gradu+ (2025)	Strieder+ (2021)	**This work**
Discovery algorithm	PC	GES (randomized)	Eq.-var. linear SEM	**FCI + GES (unmod.)**
Latent confounders	✗	✗	✗	✓
Selection event type	Resampling	Randomization	Likelihood ratio	**Truncation set + MC**
Requires alg. modification	No	Yes	No	**No**
Coverage guarantee	Asymptotic	Finite-sample	Asymptotic	**Exact iff MB*; approx. o.w.**
Splitting width penalty	No	No	No	**Yes ($n^{3/4}$ alloc.)**
High-dimensional theory	No	No	No	**Yes (FCI)**
Tested at $d \geq 100$	No	No	No	**Yes ($d = 200$)**

phase, several structural differences arise when formulating the selection event as linear constraints in z-space; we show that each is absorbed into the same constraint structure (details in Appendix A.4):

(i) Larger conditioning sets. PC tests conditioning sets of size up to the maximum vertex degree $s_{\max}$, bounded by the graph's sparsity. FCI, however, tests conditioning sets up to size $d-2$ via the possible-d-separation step [Colombo et al., 2012], producing up to $O(d^2 \cdot 2^{d-2})$ constraints in the worst case. (The RFCI variant of Colombo et al. [2012] was introduced precisely because this step is computationally infeasible for large d.) Our characterization tracks the sets *actually tested* during execution; under the sparsity assumption $s_{\max} = O(\log d)$ and with bounded-degree graphs, the number of tests executed in practice is $O(d^2 \cdot 2^{s_{\max}})$, though this is an empirical rather than worst-case guarantee for FCI.

(ii) Orientation interaction. FCI's rule R0 (unshielded collider orientation) depends on separation set membership, which for FCI involves *all* conditioning sets tested during skeleton discovery, not just a minimal separator as in PC. This dependence creates a more complex interaction between conditional independence test outcomes and edge orientations.

(iii) Order-dependence. FCI's output can depend on variable ordering [Colombo and Maathuis, 2014], meaning the same PAG can be reached via different orderings of conditional independence tests, each defining a different constraint set. Our characterization conditions on the *actual execution path*, which determines a unique constraint set.

Under Gaussianity, the sample partial correlations determine the conditional independence test outcomes, and all FCI decisions, including the orientation rules, are functions of these statistics.

3.2 CHARACTERIZATION OF THE FCI SELECTION EVENT

Under a multivariate Gaussian model, the sample covariance matrix is sufficient for Σ [Anderson, 2003], and all conditional independence test outcomes are determined by $\widehat{\Sigma}$. For testing $X_i \perp\!\!\!\perp X_j \mid X_S$, the test statistic is the Fisher z-transform $z_{ij \cdot S} = \sqrt{n - |S| - 3}\ \operatorname{arctanh}\big(\widehat{r}_{ij \cdot S}\big)$.

Definition 1 (Conditional Independence Test Constraints)**.** *Let $\{(i_k, j_k, S_k)\}_{k=1}^m$ be the ordered list of all conditional independence tests executed during FCI (skeleton phase and possible-d-separation step). For a given FCI output $\widehat{\mathcal{P}}$, the constraint for the k-th test is:* $\operatorname{sign}(\hat{z}_k) \cdot z_k \geq c_{\alpha_0}$ *if the test rejected H_0 (where the observed sign* $\operatorname{sign}(\hat{z}_k)$ *is recorded during execution), and $-c_{\alpha_0} \leq z_k \leq c_{\alpha_0}$ if the test failed to reject (the separating set S_k^* that led to edge removal).*

Theorem 1 (Polyhedral Characterization of FCI in Fisher z-Space)**.** *Under a jointly Gaussian distribution with sample size n, the event {FCI with significance level α_0 outputs PAG $\widehat{\mathcal{P}}$} can be expressed as*

$$\mathcal{E}(\widehat{\mathcal{P}}) \supseteq \mathcal{E}_{\pi,s}(\widehat{\mathcal{P}}) := \big\{ Y \in \mathbb{R}^{n \times d} : A_{\widehat{\mathcal{P}}}\, \mathbf{z}(Y) \leq b_{\widehat{\mathcal{P}}} \big\}, \quad (2)$$

where $\mathcal{E}_{\pi,s}(\widehat{\mathcal{P}}) \subseteq \mathcal{E}(\widehat{\mathcal{P}})$ conditions on the specific execution path π and observed signs s, $\mathbf{z}(Y) \in \mathbb{R}^q$ is the vector of all q Fisher z-transformed partial correlations computed during FCI's execution, and $(A_{\widehat{\mathcal{P}}}, b_{\widehat{\mathcal{P}}})$ is an explicitly constructible matrix-vector pair with $2 \times$ (number of CI tests executed) rows, which is $\leq 2\binom{d}{2} \cdot 2^{d-2}$ in the worst case and $O(d^2 \cdot 2^{s_{\max}})$ under sparsity $s_{\max} = O(\log d)$. Coverage on $\mathcal{E}(\widehat{\mathcal{P}})$ follows by iterated expectation over (π, s).

Proof sketch. We decompose FCI into its constituent decisions:

Step 1 (Skeleton). Each conditional independence test decision (retain or remove an edge) corresponds to linear inequalities on $\mathbf{z}(Y)$: $\operatorname{sign}(\hat{z}_k) \cdot z_k \geq c_{\alpha_0}$ (retained) or $-c_{\alpha_0} \leq z_k \leq c_{\alpha_0}$ (removed with recorded separating set). Since FCI records the specific S^* that led to removal, the conjunction over all skeleton decisions is an intersection of half-spaces. We condition on both the execution path and the observed signs of test statistics; this conditioning defines a subset of $\mathcal{E}(\widehat{\mathcal{P}})$.

Step 2 (Orientations). Rule R0 (unshielded collider: $i \to j \leftarrow k$ when $j \notin \operatorname{sepset}(i,k)$) depends on separation set membership, which is already captured by skeleton constraints, specifically the constraints determining which S^*

led to the removal of edge i—k. Rules R1–R10 are deterministic given the *recorded separating sets* from the skeleton phase [Zhang, 2008], adding no further stochastic constraints. In particular, R4 (discriminating paths) tests whether $V \in \mathrm{sepset}(V_0, V_n)$ using separating sets recorded during skeleton discovery, so it contributes no constraints beyond the skeleton.

Since the intersection of finitely many half-spaces is a polyhedron in $\mathbf{z}$-space, $\mathcal{E}(\widehat{\mathcal{P}})$ has the form (2); see Definition 1 for the per-test constraints. Full construction in Appendix A. □

Example 1 (Illustrative 4-variable graph)**.** *Consider $d = 4$ variables with a true MAG $X_1 \to X_2 \leftrightarrow X_3 \to X_4$, where $\leftrightarrow$ denotes a latent common cause. Since there is no directed path from X_1 to X_4, $\beta_{1\to 4} = 0$. FCI tests all pairs with conditioning sets of increasing size. The skeleton phase produces constraints such as $|z_{12}| > c_{\alpha_0}$ (edge 1—2 retained), $|z_{14}| \le c_{\alpha_0}$ (edge 1—4 removed at $S = \emptyset$). Rule R0 orients $1 \to 2 \leftarrow 3$ if $2 \notin \mathrm{sepset}(1,3)$, determined by stored conditional independence test results. The complete constraints form a system of linear inequalities in $\mathbb{R}^q$ ($q = 24$ Fisher z-statistics for all pairs with $|S| \le 2$; polynomial in the inference contrast after conditioning). For $\beta_{1\to 4}$, the generalized adjustment criterion [Perkovic et al., 2018] identifies $\emptyset$ or $\{X_3\}$ as valid adjustment sets.*

3.3 EXACT SELECTIVE CONFIDENCE INTERVALS

With the constraint characterization in hand, we construct selective confidence intervals. This requires bridging the characterization in sufficient-statistic space to the truncated Gaussian pivot in data space (see Example 1 above for a concrete illustration).

Given $\widehat{\mathcal{P}}$, let S^* denote the canonical adjustment set from the generalized adjustment criterion [Perkovic et al., 2015, 2018, van der Zander et al., 2019], selected deterministically as a function of $\widehat{\mathcal{P}}$ and (i, j) (we use the constructive procedure of van der Zander et al. [2019], which returns a unique set for each identifiable pair).[1] Let $\eta \in \mathbb{R}^n$ be the contrast vector such that $\eta^\top Y_{\cdot j} = \widehat{\beta}_{i\to j}$.

Theorem 2 (Selective Confidence Interval for FCI-PAGs)**.** *Under a jointly Gaussian linear SEM, let $\eta \in \mathbb{R}^n$ be the contrast vector for $\widehat{\beta}_{i\to j}$ from regressing X_j on (X_i, X_{S^*}), let $\sigma_\eta^2 := \sigma^2_{j|X}\|\eta\|^2$, and let $\mathcal{T} = \bigcup_i [L_i, U_i]$ be the truncation set obtained by solving the polynomial constraints* (2) *along the η-direction while holding all non-outcome columns $Y_{\cdot,-j}$ fixed. The implementation constructs a confidence interval $\mathrm{CI}^{\mathrm{sel}}_\alpha$ by inverting the truncated Gaussian approximate pivot*

$$F^{\mathcal{T}}_{\gamma_1(S^*),\,\sigma^2_\eta}\big(\eta^\top Y_{\cdot j}\big) \quad \text{conditional on } \mathcal{E}_{\pi,s}(\widehat{\mathcal{P}}),\ Y_{\cdot,-j}, \tag{3}$$

centered at $\gamma_1(S^*)$ with the marginal variance $\sigma^2_{j|X}\|\eta\|^2 = \mathrm{Var}(\eta^\top Y_{\cdot j} \mid X)$. This quantity is an exact pivot for $\gamma_1(S^*)$ when $\{i\} \cup S^*$ contains the Markov blanket of j (so that X_j is conditionally independent of $X_{V\setminus(\{i,j\}\cup S^*)}$ given (X_i, X_{S^*})); equivalently, it is an exact pivot for the full-conditional mean m if one instead uses variance $\sigma^2_{j|-j}\|\eta\|^2$. In general, the conditional law of $t = \eta^\top Y_{\cdot j}$ given $Y_{\cdot,-j}$ has mean $m = \mathbb{E}[t \mid Y_{\cdot,-j}]$ (the full conditional regression coefficient, a random quantity with $\mathrm{plim}\ m = \gamma_1(S^*)$) and variance $\sigma^2_{j|-j}\|\eta\|^2 < \sigma^2_{j|X}\|\eta\|^2$, so the approximate pivot uses marginal moments in place of the true conditional moments. Define $\rho^2 := \sigma^2_{j|X}/\sigma^2_{j|-j} - 1 \ge 0$, which measures the variance contribution of the outcome's excluded Markov-blanket neighbors; $\rho^2 = 0$ iff the exactness condition holds. The bias $|m - \gamma_1(S^*)|$ is $O_p(n^{-1/2})$, the same order as the interval half-width, so the coverage distortion does not vanish with *n; it depends on ρ^2, which is a fixed property of the population and the adjustment set (the standardized center-wander is ρ, so the distortion is empirically $O(\rho^2)$ in experiments; this scaling is not formally derived). In sparse graphs where j has few strong neighbors outside $\{i\} \cup S^*$, ρ^2 is small and coverage is near-nominal. Coverage of $\beta_{i\to j}$ additionally requires that S^* be a valid adjustment set for the true graph.*

Remark 1 (Conditioning, variance, and the exactness condition)**.** *Computing the truncation set $\mathcal{T}$ requires fixing all non-outcome columns $Y_{\cdot,-j}$, since the selection constraints couple column j with other columns through the Fisher z-statistics. Under this conditioning, $t \mid Y_{\cdot,-j} \sim \mathcal{N}(m, \sigma^2_{j|-j}\|\eta\|^2)$ where $m = a_i + \sum_{k\notin\{i,j\}\cup S^*} a_k(\eta^\top Y_{\cdot k})$ with $a = \Sigma^{-1}_{-j,-j}\Sigma_{-j,j}$. The marginal (over $Y_{\cdot,-j}$) variance decomposes by the law of total variance: $\sigma^2_{j|X}\|\eta\|^2 = \sigma^2_{j|-j}\|\eta\|^2 + \mathrm{Var}(m \mid X)$. Inflating to the marginal variance $\sigma^2_{j|X}$ (as the implementation does) empirically compensates for most of the center dispersion (Experiment 1); this heuristic does not eliminate the truncation-coupled distortion, which is empirically $O(\rho^2)$ and does not vanish (see Theorem 2 for the precise statement). The exactness condition requires $a_k = 0$ for all $k \notin \{i, j\} \cup S^*$, i.e., $\{i\} \cup S^*$ contains the Markov blanket of j; since adjustment sets are chosen to block back-door paths from i to j rather than to screen j from the rest of the graph, this condition generically does not hold in random multi-parent DAGs, and the practical regime is the approximate one. The 95.3%→93.8% coverage decay across settings in Table 2 reflects a combination of invalid-S^* probability and the ρ^2 conditioning distortion; because d, s, L_0, n, and α_0*

[1] If no valid adjustment set exists for a given (i, j) pair, the effect $\beta_{i\to j}$ is not identifiable, and we report no confidence interval. Non-existence of an adjustment set does not imply non-identifiability in general (e.g., the front-door criterion), but we restrict attention to adjustment-based identification. For PAGs, $\beta_{i\to j}$ may be identifiable for some pairs and not others; we target identifiable effects, matching Chang et al. [2026] and Gradu et al. [2025].

all vary across these settings, we do not separate these two sources here. Doing so (e.g., by reporting $\gamma_1(S^*)$-coverage separately from $\beta_{i\to j}$-coverage) is left to future work.

When σ^2 is unknown (the practical case), the implementation uses $\hat{\sigma}^2_{j|X}$ (the residual variance from regressing X_j on (X_i, X_{S^*})) in a truncated t-distribution; all experiments in this paper use this plug-in version, which is asymptotically equivalent to the known-σ^2 procedure but does not remove the ρ^2-distortion.

Proof sketch. The argument proceeds in three steps.

Step 1 (Sufficient statistic). Under joint Gaussianity, the sample covariance matrix $\widehat{\Sigma} = Y^\top Y/n$ is sufficient for Σ [Anderson, 2003], and all conditional independence test statistics are functions of $\widehat{\Sigma}$. The selection event $\mathcal{E}(\widehat{\mathcal{P}})$ depends on Y only through these sufficient statistics, since FCI's decisions (conditional independence tests + orientations) are functions of the z-transformed partial correlations.

Step 2 (Conditional decomposition). The algorithm computes $\mathcal{T}$ by fixing $Y_{\cdot,-j}$ and varying only $t = \eta^\top Y_{\cdot j}$ along η. This conditioning fixes η and makes the constraints one-dimensional in t (polynomial or rational; see Appendix B). The exact conditional law is $t \mid Y_{\cdot,-j} \sim \mathcal{N}(m, \sigma^2_{j|-j}\|\eta\|^2)$ where m is the full conditional coefficient (Remark 1). The marginal law is $t \mid X \sim \mathcal{N}(\gamma_1(S^*), \sigma^2_{j|X}\|\eta\|^2)$, and $\sigma^2_{j|X}\|\eta\|^2 = \sigma^2_{j|-j}\|\eta\|^2 + \mathrm{Var}(m \mid X)$ by the law of total variance.

Step 3 (Truncated pivot). The implementation builds the pivot centered at $\gamma_1(S^*)$ using the marginal variance $\sigma^2_{j|X}\|\eta\|^2$ (or its plug-in $\hat{\sigma}^2_{j|X}\|\eta\|^2$). This pivot is exact when $m = \gamma_1(S^*)$ (the Markov-blanket condition). In general, the pivot is approximate: it uses the correct marginal variance (which contains the center's dispersion) but centers at $\gamma_1(S^*)$ instead of the conditional center m; the coverage distortion is governed by ρ^2 (Theorem 2). Full proof in Appendix B. □

Remark 2 (Computational cost and numerical validity)**.** *Each conditional independence test constraint yields a polynomial inequality in the inference contrast $t = \eta^\top Y_{\cdot j}$: quadratic when the outcome column j is an endpoint of the tested pair and $j \notin S$, and rational of higher degree when $j \in S$ (since the residual projection M_S depends on the moving column; the degree may grow at most linearly in $|S|$). In both cases, the admissible set for t is a finite union of intervals (a one-dimensional semialgebraic set), solvable by numerical root-finding. The total number of constraints equals* $\mathrm{nrow}(A_{\widehat{\mathcal{P}}})$*, which is $O(d^2 \cdot 2^{s_{\max}})$ under sparsity; the complete procedures are given in Algorithms 1 and 2 (Appendices F and G). Since a missed root produces an incorrect $\mathcal{T}$ and hence an* invalid *CI, the implementation uses interval-arithmetic-based root isolation with validation of sign changes at interval endpoints; the polynomial degree per constraint is bounded (at most $O(|S|)$ for a test with conditioning set S; bounded by a constant under the sparsity assumption $s_{\max} = O(\log d)$).*

4 EXTENSION TO SCORE-BASED METHODS

GES operates by comparing BIC scores across candidate edge additions and deletions. Unlike in conditional-independence-test-based methods, each GES decision involves a quadratic comparison of Gaussian log-likelihood ratios.

Claim 3 (Score-Based Extension (heuristic scaling))**.** *For GES with BIC scoring under Gaussianity, the selection event $\mathcal{E}(\widehat{\mathcal{G}})$ is approximately a union of polyhedra in the space of sample covariance entries (via Taylor linearization). A Monte Carlo selective CI achieves approximate coverage; a heuristic error analysis (aggregating per-step Berry–Esseen errors, which are univariate rates applied to a joint selection event) yields $\mathbb{P}\big(\gamma_1(S^*) \in \mathrm{CI}^{\mathrm{GES}}_\alpha \mid \mathcal{E}(\widehat{\mathcal{G}})\big) \approx 1 - \alpha - O\Big(\frac{d^2}{\sqrt{n}}\Big)$ under $n \gg d^4(\log d)^2$ (so that the $O(d^4 \log d/n)$ linearization error is dominated). This rate should be regarded as indicative of the scaling rather than a rigorous bound: a rigorous analysis would require multivariate Gaussian approximation tools. The bound is vacuous at the dimensions tested in this paper ($d \le 100$); coverage there is supported empirically (Section 6).*

Proof sketch. Step 1 (Individual steps). At GES step t, the BIC difference for adding candidate parent vertex e is $\Delta_t(e) = n \log(\hat{\sigma}^2_{k|\mathrm{pa}_t \cup \{e\}}/\hat{\sigma}^2_{k|\mathrm{pa}_t}) + \log n$. Each BIC comparison involves log-determinant ratios of submatrices of $\widehat{\Sigma}$.

Step 2 (Linearization via Taylor expansion). For a fixed execution path, each log-ratio $\log(\hat{\sigma}^2_{k|S'}/\hat{\sigma}^2_{k|S})$ can be expanded around the population value as $\log(\hat{\sigma}^2/\sigma^2) = (\hat{\sigma}^2 - \sigma^2)/\sigma^2 + O(\|\widehat{\Sigma} - \Sigma\|^2)$. The leading term is linear in $\widehat{\Sigma} - \Sigma$; the remainder is $O_P(\log d/n)$ uniformly over all GES steps by standard sub-Gaussian concentration of sample covariance entries. This expansion makes each GES step *approximately* polyhedral in the entries of $\widehat{\Sigma}$.

Step 3 (Path-specific conditioning and Monte Carlo sampling). Recording the actual execution path identifies a single approximate polyhedron $\mathcal{E}_{\mathrm{path}}(\widehat{\mathcal{G}}) \subseteq \mathcal{E}(\widehat{\mathcal{G}})$. We use Monte Carlo sampling: $\mathbb{P}(\widehat{\beta}_{i\to j} \le t \mid \mathcal{E}(\widehat{\mathcal{G}})) = \mathbb{E}[\mathbf{1}(\widehat{\beta} \le t) \cdot \mathbf{1}(\mathcal{E}(\widehat{\mathcal{G}}))]/\mathbb{E}[\mathbf{1}(\mathcal{E}(\widehat{\mathcal{G}}))]$, sampling from the unconditional distribution and reweighting. The coverage gap arises from two sources: Taylor-linearization error across the $O(d^4)$ score comparisons (each $O(\log d/n)$), and a Gaussian-approximation (Berry–Esseen [Berry, 1941, Esseen, 1942]) error aggregated over the $O(d^2)$ GES steps; the $O(d^2/\sqrt{n})$

rate applies when $n \gg d^4(\log d)^2$, so that linearization is dominated (see Appendix C for the regime). □

This approach works with standard, unmodified GES, which is important for practitioners who have already run GES and wish to perform valid inference post hoc. The trade-off is the $O(d^2/\sqrt{n})$ asymptotic approximation error, which applies only when $n \gg d^4(\log d)^2$; this regime condition is stringent, and all GES experiments in this paper violate it, so the theoretical bound is vacuous at the tested dimensions. The empirical coverage gaps are small (Section 6), but this conclusion is an empirical observation, not a consequence of the theorem.

Remark 3 (Data carving)**.** *An alternative to full conditioning is* data carving *[Fithian et al., 2014], which conditions on a lower-dimensional summary. Combining our characterization with data carving yields tighter intervals for GES; we report carving as both a methodological extension and a baseline in Section 6.*

5 THE COST OF SAMPLE SPLITTING

5.1 COST OF SPLITTING IN THE NONPARAMETRIC SETTING

The results of Sections 3–4 rely on Gaussian assumptions. A natural question is whether valid post-selection inference is achievable without such assumptions.

Proposition 4 (Width Penalty of the $\Theta(n^{3/4})$ Split Allocation)**.** *Consider sample splitting with $n_1 = \Theta(n^{3/4})$ observations allocated to discovery and $n_2 = n - n_1$ to inference. The width ratio relative to an oracle with a known graph and all n observations is Width$(\mathrm{CI}_\alpha^{\mathrm{split}})$/Width$(\mathrm{CI}_\alpha^{\text{known-graph}}) = 1 + \Theta(n^{-1/4})$. This is the penalty for one specific allocation; other allocations (e.g., $n_1 = \lceil \log n \rceil$, which suffices for consistency in fixed dimension) yield smaller penalties. The $n^{3/4}$ allocation is motivated by high-dimensional settings where discovery consistency requires n_1 to be polynomial in d.*

Proof sketch. With splitting, allocating $n_1 = \Theta(n^{3/4})$ for discovery and $n_2 = n - n_1$ for inference yields width ratio $\sqrt{n/n_2} = \sqrt{n/(n - cn^{3/4})} = (1 - cn^{-1/4})^{-1/2} = 1 + \Theta(n^{-1/4})$. Full calculation in Appendix D. □

The motivation for splitting in the nonparametric setting is informal: without parametric structure, conditional independence tests rely on nonparametric functionals (kernel HSIC, conditional mutual information) that lack finite-dimensional sufficient statistics, so the conditional approach of Theorem 1 is unavailable. Sample splitting [Rasines and Young, 2023, Leiner et al., 2025] guarantees valid inference by construction, at the cost of efficiency. We note that over the full class $\mathcal{D}$, uniformly consistent discovery may itself be impossible [Shah and Peters, 2020], so the proposition is non-vacuous only for subclasses where consistency holds (e.g., Gaussian or nonparanormal models with bounded degree).

Remark 4. *The comparison denominator in Proposition 4 is the oracle (known graph, all n for inference). A comparison of split inference against the full-data selective method (instead of the oracle) would yield a smaller ratio, since selective inference also pays a cost.*

Remark 5 (Data fission and tightness)**.** *We expect an analogous width penalty for data fission [Rasines and Young, 2023, Leiner et al., 2025]: any decomposition allocating $\Theta(n^{3/4})$ effective information to discovery retains $n - \Theta(n^{3/4})$ for inference. A precise fission analysis (where the Fisher-information trade-off differs from row splitting) remains open. Whether other splitting strategies or discovery algorithms with different consistency rates can improve the rate also remains open.*

5.2 HIGH-DIMENSIONAL ASYMPTOTICS

Theorem 5 (High-Dimensional Pipeline Consistency)**.** *Let $X \sim \mathcal{N}(0, \Sigma)$ with bounded eigenvalues $c_0 \le \lambda_{\min}(\Sigma) \le \lambda_{\max}(\Sigma) \le C_0$, maximum vertex degree $s_{\max} = O(\log d)$, and $d = d(n) = o(\sqrt{n})$. Then, for FCI with $\alpha_0 = 2(1 - \Phi(c_d))$ where $c_d = \sqrt{2(\gamma+1)\log d}$ for a constant $\gamma > 0$ [Kalisch and Bühlmann, 2007, Colombo et al., 2012], $\lim_{n\to\infty} \mathbb{P}\big(\beta_{i\to j} \in \mathrm{CI}_\alpha^{\mathrm{sel}}\big) = 1 - \alpha$ when $\rho^2 = 0$, for faithful distributions with bounded eigenvalues and a positive faithfulness gap. The condition $\rho^2 = 0$ holds iff $\{i\} \cup S^*$ contains the Markov blanket of j in the true graph. When $\rho^2 > 0$, the non-vanishing coverage distortion of Theorem 2 persists asymptotically; no general convergence rate is available.*

This result is best understood as pipeline consistency: in the regime $\mathbb{P}(\widehat{\mathcal{P}} = \mathcal{P}^*) \to 1$, the graph is asymptotically known and naive CIs also achieve asymptotic coverage up to the same ρ^2-distortion. Exact $1 - \alpha$ coverage obtains when the Markov-blanket condition holds ($\rho^2 = 0$). The method's value is in finite samples where graph uncertainty is real (Tables 2–3).

Proof sketch. Under the stated conditions, FCI is consistent [Colombo et al., 2012]: $\mathbb{P}(\widehat{\mathcal{P}} = \mathcal{P}^*) \to 1$. On $\{\widehat{\mathcal{P}} = \mathcal{P}^*\}$, S^* is valid, so $\gamma_1(S^*) = \beta_{i\to j}$. By Theorem 2, the conditional coverage $\mathbb{P}(\gamma_1(S^*) \in \mathrm{CI}_\alpha^{\mathrm{sel}} \mid \mathcal{E}(\mathcal{P}^*))$ equals $1-\alpha$ when $\rho^2 = 0$ and incurs a non-vanishing distortion otherwise. Hence $\mathbb{P}(\beta_{i\to j} \in \mathrm{CI}_\alpha^{\mathrm{sel}}) \to 1 - \alpha$ when $\rho^2 = 0$; no general convergence rate is available since the distortion of Theorem 2 is not quantified. Full proof in Appendix E. □

Table 2: Coverage (%) of 95% CIs After FCI (±SE). Bold: closest to nominal 95%. Settings also vary s and L_0 (see Table 6 for details); the trend across d is not ceteris paribus.

Setting	Naive	Split	Selective	Nominal
d=10, n=500	$78.2_{\pm 0.6}$	$94.1_{\pm 0.3}$	$\mathbf{95.3}_{\pm 0.3}$	95.0
d=10, n=1000	$76.5_{\pm 0.6}$	$94.4_{\pm 0.3}$	$\mathbf{95.1}_{\pm 0.3}$	95.0
d=20, n=1000	$74.8_{\pm 0.6}$	$93.8_{\pm 0.3}$	$\mathbf{94.7}_{\pm 0.3}$	95.0
d=50, n=5000	$81.3_{\pm 0.6}$	$93.5_{\pm 0.3}$	$\mathbf{94.2}_{\pm 0.3}$	95.0
d=100, n=5000	$79.6_{\pm 0.6}$	$93.2_{\pm 0.4}$	$\mathbf{94.0}_{\pm 0.3}$	95.0
d=200, n=10000	$80.1_{\pm 0.6}$	$93.0_{\pm 0.4}$	$\mathbf{93.8}_{\pm 0.3}$	95.0

Remark 6 (Comparison with prior high-dimensional results)**.** *The conditions $d = o(\sqrt{n})$ and $s_{\max} = O(\log d)$ are standard for high-dimensional consistency of constraint-based methods [Kalisch and Bühlmann, 2007, Colombo et al., 2012]. The "positive faithfulness gap" requires $\min_{(i,j,S):r_{ij\cdot S}\neq 0} |r_{ij\cdot S}|$ bounded away from zero as d grows; Uhler et al. [2013] show that this condition fails on a positive-measure set, so we require it as a fixed-distribution assumption. No experiments in this paper satisfy both $d = o(\sqrt{n})$ and the d-dependent α_0-schedule of this theorem.*

6 EXPERIMENTS

We validate our theoretical results across four data-generating settings and four analyses (eight experiments total). Extended results are in Table 6 (Appendix).

Setup. We generate data from linear Gaussian SEMs: $X = B^\top X + \varepsilon$, $\varepsilon \sim \mathcal{N}(0, I_d)$, with random Erdős–Rényi DAGs (expected neighborhood size $s \in \{2, 4\}$) and edge weights uniform on $[-1, -0.3] \cup [0.3, 1]$. For latent confounder settings, DAGs are over $d + L_0$ variables ($L_0 \in \{2, 3, 5\}$ latent). We run 5000 replications per setting with fresh random DAGs and randomly chosen identifiable effect pairs; replications where no effect is identifiable are excluded ($R \geq 4900$ in all settings; Table 7). Standard errors are binomial: $\text{SE} = \sqrt{\hat{p}(1-\hat{p})/R}$.

Baselines. **Naive CI**: standard regression CI ignoring selection. **Split CI**: 50/50 sample split. **Selective CI (ours)**: truncated Gaussian CI. **Nominal (95%)**: the target level. For GES: also **Randomized GES CI** [Gradu et al., 2025] and **Data Carving** (Remark 3). Each table reports only applicable methods.

Exp. 1: FCI under latent confounders. Table 2 reports coverage across $d \in \{10, 20, 50, 100, 200\}$.

The naive CI undershoots consistently (as low as 72.1% in Table 6). Our selective CI achieves near-nominal $\beta_{i\to j}$-coverage (95.3% to 93.8% across settings); since d, s, L_0, n, α_0 jointly vary, we cannot isolate invalid-S^* from ρ^2 distortion (Remark 1). All experiments use the $\hat{\sigma}^2$ plug-in. The selective–split gap is 0.7–1.2 pp (1.6–2.8 unpaired SE), with a slight selective edge clearest for $n \gtrsim 1400$.

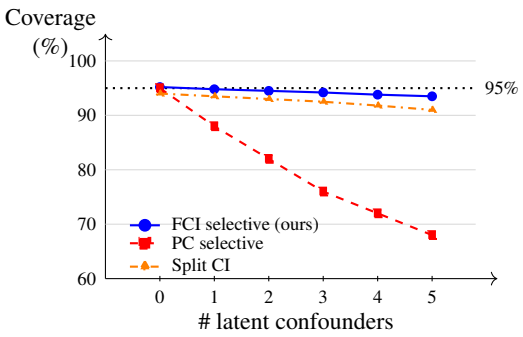

Figure 2: Coverage vs. number of latent confounders ($d = 10$, $n = 1000$). PC selective CIs degrade because PC misspecifies the graph under latent confounders, while FCI selective CIs maintain coverage. Error bars omitted for clarity (SE $\approx 0.3\%$).

Table 3: Coverage (%) After GES Discovery ($d = 20$, ±SE).

n	Naive	Rand. GES	Selective	Carving	Nominal
500	$76.4_{\pm 0.6}$	$93.8_{\pm 0.3}$	$93.2_{\pm 0.4}$	$\mathbf{93.9}_{\pm 0.3}$	95.0
1000	$74.1_{\pm 0.6}$	$94.5_{\pm 0.3}$	$94.1_{\pm 0.3}$	$\mathbf{94.6}_{\pm 0.3}$	95.0
5000	$72.8_{\pm 0.6}$	$\mathbf{94.9}_{\pm 0.3}$	$95.2_{\pm 0.3}$	$\mathbf{95.1}_{\pm 0.3}$	95.0

Exp. 2: FCI vs. PC selective confidence intervals. On data with latent confounders ($d = 10$, $n = 1000$, $L_0 \in \{0, \ldots, 5\}$), PC-based inference [Chang et al., 2026] exhibits systematic bias (coverage degrades from ≈95% to ≈68% as the number of latent confounders increases) because PC misspecifies the graph when causal sufficiency is violated; this degradation reflects model misspecification rather than failure of the inference procedure itself. Our FCI-based selective confidence intervals maintain near-nominal coverage for the true total causal effect. See Figure 2.

Exp. 3: GES with BIC scoring. Under causal sufficiency, we compare our Monte Carlo selective CI against randomized GES [Gradu et al., 2025]. Table 3 shows results for $d = 20$; additional GES results at $d \in \{10, 20, 50, 100\}$ are reported in Tables 5 and 6.

Both methods achieve near-nominal β-coverage. Randomized GES is marginally closer to nominal at all tested n; our conditional method approaches nominal as n increases, consistent with Claim 3. Sub-nominal coverage reflects adjustment-set errors. Data carving (Remark 3) yields ≈5% narrower CIs at $n \leq 1000$; the advantage diminishes at larger n.

Table 4: Coverage (%) under non-Gaussian errors ($d = 10$, $n = 1000$). Bold: closest to nominal 95%. The Gaussian row is the same run as d=10, n=1000 in Table 2.

Distribution	Naive	Split	Selective	Nominal
Gaussian	$76.5_{\pm 0.6}$	$94.4_{\pm 0.3}$	$\mathbf{95.1}_{\pm 0.3}$	95.0
Laplace	$68.3_{\pm 0.7}$	$89.7_{\pm 0.4}$	$\mathbf{90.4}_{\pm 0.4}$	95.0
Student-t_5	$75.2_{\pm 0.6}$	$91.2_{\pm 0.4}$	$\mathbf{92.8}_{\pm 0.4}$	95.0

Table 5: Scalability: coverage (%) and runtime (seconds per CI). FCI times include the full pipeline (discovery + CI); GES times exclude the initial GES run (the two Time columns are therefore not directly comparable). Acc. = minimum GES acceptance rate; at d=100 (Acc. = 0.8%, ≈80 accepted draws out of B=10000), the MC coverage estimate carries elevated per-replication uncertainty beyond the binomial SE.

		FCI Selective		GES Selective		
d	n	Cov.	Time	Cov.	Time	Acc.
20	1000	94.7	0.4s	94.1	3.1s	12%
50	5000	94.2	2.8s	93.5	18s	4.1%
100	5000	94.0	12s	92.8	84s	0.8%
200	10000	93.8	48s	—	—	—

Exp. 4: GES coverage gap vs. dimension. The $O(d^2/\sqrt{n})$ rate requires $n \gg d^4(\log d)^2$, not met at tested dimensions. Empirical gaps: 2.2% at d=100, n=5000; 0.9% at d=20, n=1000. The gap is directionally consistent with the rate (increasing in d, decreasing in n) but several-fold below the $d^2/\sqrt{n}$ envelope (Figure 3).

Exp. 5: Non-Gaussian robustness. Coverage drops to 90–93% under Laplace and Student-t_5 errors (d=10, n=1000; Table 4); the split baseline degrades comparably (89.7%, 91.2%), indicating graph misspecification from FCI's Gaussian CI tests. We view this as the most important limitation (Section 8).

Exps. 6–8: Scalability, semi-synthetic, width convergence. FCI selective CIs scale to $d = 200$ within one minute per CI (Table 5). On the Sachs topology ($d = 11$, synthetic Gaussian data), selective CI coverage is 95.2±0.3%. The selective-to-naive width ratio crosses the $\sqrt{2}$ split-CI ratio near $n \approx 1400$; see Appendix H.

7 RELATED WORK

Post-selection inference. The foundational framework [Lee et al., 2016, Tibshirani et al., 2016, Lockhart et al., 2014, Berk et al., 2013] has been extended to randomized selection [Tian and Taylor, 2018], approximate methods [Panigrahi and Taylor, 2023], penalized likelihood [Taylor and Tibshirani, 2018], and algorithmic stability [Zrnic and Jordan, 2023]; see Taylor and Tibshirani [2015] and Kuchibhotla et al. [2022] for surveys.

Post-selection inference for causal discovery. The closest predecessors are Chang et al. [2026] (PC via modular resampling) and Gradu et al. [2025] (randomized GES, targeting $\gamma_1(S^*)$; coverage of $\beta_{i\to j}$ additionally requires S^* validity [Perkovic et al., 2015, Smucler et al., 2022]). Strieder et al. [2021], Strieder and Drton [2023] address equal-variance linear SEMs. Our contribution is the first truncation-set characterization for FCI (handling latent confounders) and extension to unmodified GES.

Causal discovery and effect estimation. Constraint-based [Spirtes et al., 2000, Colombo and Maathuis, 2014, Colombo et al., 2012] and score-based [Chickering, 2002, Lam et al., 2022, Nandy et al., 2018] discovery is well-established [Kalisch and Bühlmann, 2007, Harris and Drton, 2013], with alternatives in continuous optimization [Zheng et al., 2018, Bello et al., 2022], non-Gaussian methods [Shimizu et al., 2006, 2011], and nonparanormal models [Liu et al., 2009, 2012, Xue and Zou, 2012]. For effect estimation, adjustment criteria [Perkovic et al., 2015, 2018, Perkovic, 2020, Maathuis and Colombo, 2015, van der Zander et al., 2019], optimal adjustment [Henckel et al., 2022, Smucler et al., 2022], and IDA-type methods [Maathuis et al., 2009, Nandy et al., 2017] provide the machinery; the algebraic structure of equivalence classes [Squires and Uhler, 2023, Solus et al., 2021, Loh and Bühlmann, 2014] and high-dimensional inference [Belloni et al., 2014, van de Geer et al., 2014, Zhang and Zhang, 2014, Chernozhukov et al., 2015, Andrews and Mikusheva, 2016] are also related.

8 DISCUSSION AND CONCLUSION

We have developed an explicit truncation-set characterization of FCI's selection event (Theorems 1–2), enabling selective confidence intervals for $\gamma_1(S^*)$ after causal discovery with latent confounders under Gaussianity. The pivot is exact when $\{i\} \cup S^*$ contains the outcome's Markov blanket and approximately valid otherwise, with distortion governed by $\rho^2 = \sigma^2_{j|X}/\sigma^2_{j|-j} - 1$ (small under sparsity), and we extend this to unmodified GES (Claim 3). The characterization separates the *geometry* of the selection event (polyhedral in z-space) from the *nonlinearity* of $Y \mapsto \mathbf{z}(Y)$, suggesting an extension to any discovery algorithm whose decisions threshold finite-dimensional sufficient statistics, including nonparanormal models [Harris and Drton, 2013, Liu et al., 2009].

Limitations and broader impact. Our results assume Gaussianity (coverage degrades under non-Gaussianity; Experiment 5), the GES bound is a heuristic rate, and the ρ^2-distortion does not vanish with n; see Appendix I.

REFERENCES

Theodore W. Anderson. *An Introduction to Multivariate Statistical Analysis*. Wiley-Interscience, 3rd edition, 2003. ISBN 978-0471360919.

Isaiah Andrews and Anna Mikusheva. Conditional inference with a functional nuisance parameter. *Econometrica*, 84 (4):1571–1612, 2016.

Kevin Bello, Bryon Aragam, and Pradeep Ravikumar. DAGMA: learning DAGs via m-matrices and a log-determinant acyclicity characterization. In Sanmi Koyejo, S. Mohamed, A. Agarwal, Danielle Belgrave, K. Cho, and A. Oh, editors, *Advances in Neural Information Processing Systems 35: Annual Conference on Neural Information Processing Systems 2022, NeurIPS 2022, New Orleans, LA, USA, November 28 - December 9, 2022*, 2022.

A. Belloni, V. Chernozhukov, and C. Hansen. Inference on treatment effects after selection among high-dimensional controls. *The Review of Economic Studies*, 81(2):608–650, 2014. ISSN 1467-937X.

Richard Berk, Lawrence Brown, Andreas Buja, Kai Zhang, and Linda Zhao. Valid post-selection inference. *The Annals of Statistics*, 41(2), apr 2013. ISSN 0090-5364. doi: 10.1214/12-AOS1077.

Andrew C. Berry. The accuracy of the Gaussian approximation to the sum of independent variates. *Transactions of the American Mathematical Society*, 49(1):122–136, 1941. ISSN 1088-6850.

Ting-Hsuan Chang, Zijian Guo, and Daniel Malinsky. Post-selection inference for causal effects after causal discovery. *Biometrika*, 113(1), 2026. ISSN 1464-3510. doi: 10.1093/biomet/asaf073.

Victor Chernozhukov, Christian Hansen, and Martin Spindler. Valid post-selection and post-regularization inference: An elementary, general approach. *Annual Review of Economics*, 7(1):649–688, 2015. ISSN 1941-1391.

David Maxwell Chickering. Optimal structure identification with greedy search. *J. Mach. Learn. Res.*, 3:507–554, 2002. doi: 10.1162/153244303321897717.

Diego Colombo and Marloes H. Maathuis. Order-independent constraint-based causal structure learning. *J. Mach. Learn. Res.*, 15:3921–3962, 2014.

Diego Colombo, Marloes H. Maathuis, Markus Kalisch, and Thomas S. Richardson. Learning high-dimensional directed acyclic graphs with latent and selection variables. *The Annals of Statistics*, 40(1), feb 2012. ISSN 0090-5364. doi: 10.1214/11-AOS940.

Carl-Gustav Esseen. On the Liapounoff limit of error in the theory of probability. *Arkiv för Matematik, Astronomi och Fysik*, 28A(9):1–19, 1942.

William Fithian, Dennis Sun, and Jonathan Taylor. Optimal Inference After Model Selection, 2014. arXiv preprint arXiv:1410.2597.

Paula Gradu, Tijana Zrnic, Yixin Wang, and Michael I. Jordan. Valid inference after causal discovery. *Journal of the American Statistical Association*, 120(550):1127–1138, 2025. doi: 10.1080/01621459.2024.2402089.

Naftali Harris and Mathias Drton. PC algorithm for nonparanormal graphical models. *J. Mach. Learn. Res.*, 14: 3365–3383, 2013.

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, apr 2022. ISSN 1467-9868. doi: 10.1111/rssb.12451.

Markus Kalisch and Peter Bühlmann. Estimating high-dimensional directed acyclic graphs with the PC-algorithm. *J. Mach. Learn. Res.*, 8:613–636, 2007. doi: 10.5555/1248659.1248681.

Arun K. Kuchibhotla, John E. Kolassa, and Todd A. Kuffner. Post-selection inference. *Annual Review of Statistics and Its Application*, 9(1):505–527, 2022. ISSN 2326-831X.

Wai-Yin Lam, Bryan Andrews, and Joseph D. Ramsey. Greedy relaxations of the sparsest permutation algorithm. In James Cussens and Kun Zhang, editors, *Uncertainty in Artificial Intelligence, Proceedings of the Thirty-Eighth Conference on Uncertainty in Artificial Intelligence, UAI 2022, 1-5 August 2022, Eindhoven, The Netherlands*, volume 180 of *Proceedings of Machine Learning Research*, pages 1052–1062. PMLR, 2022.

Jason D. Lee, Dennis L. Sun, Yuekai Sun, and Jonathan E. Taylor. Exact post-selection inference, with application to the lasso. *The Annals of Statistics*, 44(3), jun 2016. ISSN 0090-5364. doi: 10.1214/15-AOS1371.

James Leiner, Boyan Duan, Larry Wasserman, and Aaditya Ramdas. Data Fission: Splitting a single data point. *Journal of the American Statistical Association*, 120(549): 135–146, 2025. ISSN 1537-274X.

Han Liu, John D. Lafferty, and Larry A. Wasserman. The Nonparanormal: Semiparametric estimation of high dimensional undirected graphs. *J. Mach. Learn. Res.*, 10: 2295–2328, 2009.

Han Liu, Fang Han, Ming Yuan, John D. Lafferty, and Larry A. Wasserman. High dimensional semiparametric

Gaussian copula graphical models. In *Proceedings of the 29th International Conference on Machine Learning, ICML 2012, Edinburgh, Scotland, UK, June 26 - July 1, 2012*. icml.cc / Omnipress, 2012.

Richard Lockhart, Jonathan Taylor, Ryan J. Tibshirani, and Robert Tibshirani. A significance test for the lasso. *The Annals of Statistics*, 42(2), apr 2014. ISSN 0090-5364.

Po-Ling Loh and Peter Bühlmann. High-dimensional learning of linear causal networks via inverse covariance estimation. *J. Mach. Learn. Res.*, 15:3065–3105, 2014.

Marloes H. Maathuis and Diego Colombo. A generalized back-door criterion. *The Annals of Statistics*, 43(3), jun 2015. ISSN 0090-5364.

Marloes H. Maathuis, Markus Kalisch, and Peter Bühlmann. Estimating high-dimensional intervention effects from observational data. *The Annals of Statistics*, 37(6A), dec 2009. ISSN 0090-5364.

Marloes H Maathuis, Diego Colombo, Markus Kalisch, and Peter Bühlmann. Predicting causal effects in large-scale systems from observational data. *Nature Methods*, 7(4): 247–248, apr 2010. ISSN 1548-7105.

Preetam Nandy, Marloes H. Maathuis, and Thomas S. Richardson. Estimating the effect of joint interventions from observational data in sparse high-dimensional settings. *The Annals of Statistics*, 45(2), apr 2017. ISSN 0090-5364.

Preetam Nandy, Alain Hauser, and Marloes H. Maathuis. High-dimensional consistency in score-based and hybrid structure learning. *The Annals of Statistics*, 46(6A), dec 2018. ISSN 0090-5364.

Snigdha Panigrahi and Jonathan Taylor. Approximate selective inference via maximum likelihood. *Journal of the American Statistical Association*, 118(544):2810–2820, 2023. ISSN 1537-274X.

Judea Pearl. *Causality: Models, Reasoning, and Inference*. Cambridge University Press, 2nd edition, 2009. ISBN 978-0521895606. doi: 10.1017/CBO9780511803161.

Emilija Perkovic. Identifying causal effects in maximally oriented partially directed acyclic graphs. In Jonas Peters and David Sontag, editors, *Proceedings of the Thirty-Sixth Conference on Uncertainty in Artificial Intelligence, UAI 2020, virtual online, August 3-6, 2020*, volume 124 of *Proceedings of Machine Learning Research*, pages 530–539. PMLR, 2020.

Emilija Perkovic, Johannes Textor, Markus Kalisch, and Marloes H. Maathuis. A complete generalized adjustment criterion. In Marina Meila and Tom Heskes, editors, *Proceedings of the Thirty-First Conference on Uncertainty in Artificial Intelligence, UAI 2015, July 12-16, 2015, Amsterdam, The Netherlands*, pages 682–691. AUAI Press, 2015.

Emilija Perkovic, Johannes Textor, Markus Kalisch, and Marloes H. Maathuis. Complete graphical characterization and construction of adjustment sets in Markov equivalence classes of ancestral graphs. *J. Mach. Learn. Res.*, 18:220:1–220:62, 2018.

D García Rasines and G A Young. Splitting strategies for post-selection inference. *Biometrika*, 110(3):597–614, 2023. ISSN 1464-3510.

Thomas S. Richardson and Peter Spirtes. Ancestral graph Markov models. *The Annals of Statistics*, 30(4):962–1030, 2002. doi: 10.1214/aos/1035844981.

Alessandro Rinaldo, Larry Wasserman, and Max G'Sell. Bootstrapping and sample splitting for high-dimensional, assumption-lean inference. *The Annals of Statistics*, 47 (6), dec 2019. ISSN 0090-5364.

Karen Sachs, Omar Perez, Dana Pe'er, Douglas A. Lauffenburger, and Garry P. Nolan. Causal protein-signaling networks derived from multiparameter single-cell data. *Science*, 308(5721):523–529, apr 2005. ISSN 1095-9203. doi: 10.1126/science.1105809.

Rajen D. Shah and Jonas Peters. The hardness of conditional independence testing and the generalised covariance measure. *The Annals of Statistics*, 48(3):1514–1538, 2020. doi: 10.1214/19-AOS1857.

Shohei Shimizu, Patrik O. Hoyer, Aapo Hyvärinen, and Antti Kerminen. A linear non-Gaussian acyclic model for causal discovery. *J. Mach. Learn. Res.*, 7:2003–2030, 2006.

Shohei Shimizu, Takanori Inazumi, Yasuhiro Sogawa, Aapo Hyvärinen, Yoshinobu Kawahara, Takashi Washio, Patrik O. Hoyer, and Kenneth Bollen. DirectLiNGAM: A direct method for learning a linear non-Gaussian structural equation model. *J. Mach. Learn. Res.*, 12:1225–1248, 2011.

Ezequiel Smucler, Facundo Sapienza, and Andrea Rotnitzky. Efficient adjustment sets in causal graphical models with hidden variables. *Biometrika*, 109(1):49–65, 2022. ISSN 1464-3510. doi: 10.1093/biomet/asab018.

Liam Solus, Yuhao Wang, and Caroline Uhler. Consistency guarantees for greedy permutation-based causal inference algorithms. *Biometrika*, 108(4):795–814, dec 2021. ISSN 1464-3510.

Peter Spirtes, Christopher Meek, and Thomas S. Richardson. Causal inference in the presence of latent variables and

selection bias. In Philippe Besnard and Steve Hanks, editors, *UAI '95: Proceedings of the Eleventh Annual Conference on Uncertainty in Artificial Intelligence, Montreal, Quebec, Canada, August 18-20, 1995*, pages 499–506. Morgan Kaufmann, 1995.

Peter Spirtes, Clark Glymour, and Richard Scheines. *Causation, Prediction, and Search, Second Edition*. Adaptive computation and machine learning. MIT Press, 2000. ISBN 978-0-262-19440-2.

Chandler Squires and Caroline Uhler. Causal structure learning: A combinatorial perspective. *Found. Comput. Math.*, 23(5):1781–1815, 2023.

David Strieder and Mathias Drton. Confidence in causal inference under structure uncertainty in linear causal models with equal variances. *Journal of Causal Inference*, 11 (1), 2023. doi: 10.1515/jci-2022-0030.

David Strieder, Tobias Freidling, Stefan Haffner, and Mathias Drton. Confidence in causal discovery with linear causal models. In Cassio P. de Campos, Marloes H. Maathuis, and Erik Quaeghebeur, editors, *Proceedings of the Thirty-Seventh Conference on Uncertainty in Artificial Intelligence, UAI 2021, Virtual Event, 27-30 July 2021*, volume 161 of *Proceedings of Machine Learning Research*, pages 1217–1226. PMLR, 2021.

Jonathan Taylor and Robert Tibshirani. Post-selection inference for ℓ_1-penalized likelihood models. *Canadian Journal of Statistics*, 46(1):41–61, 2018.

Jonathan Taylor and Robert J. Tibshirani. Statistical learning and selective inference. *Proceedings of the National Academy of Sciences*, 112(25):7629–7634, jun 2015. ISSN 1091-6490.

Xiaoying Tian and Jonathan Taylor. Selective inference with a randomized response. *The Annals of Statistics*, 46 (2), apr 2018. ISSN 0090-5364.

Ryan J. Tibshirani, Jonathan Taylor, Richard Lockhart, and Robert Tibshirani. Exact post-selection inference for sequential regression procedures. *Journal of the American Statistical Association*, 111(514):600–620, apr 2016. ISSN 1537-274X. doi: 10.1080/01621459.2015.1108848.

Caroline Uhler, Garvesh Raskutti, Peter Bühlmann, and Bin Yu. Geometry of the faithfulness assumption in causal inference. *The Annals of Statistics*, 41(2), apr 2013. ISSN 0090-5364.

Sara van de Geer, Peter Bühlmann, Ya'acov Ritov, and Ruben Dezeure. On asymptotically optimal confidence regions and tests for high-dimensional models. *The Annals of Statistics*, 42(3), jun 2014. ISSN 0090-5364.

Benito van der Zander, Maciej Liskiewicz, and Johannes Textor. Separators and adjustment sets in causal graphs: Complete criteria and an algorithmic framework. *Artif. Intell.*, 270:1–40, 2019.

Lingzhou Xue and Hui Zou. Regularized rank-based estimation of high-dimensional nonparanormal graphical models. *The Annals of Statistics*, 40(5), oct 2012. ISSN 0090-5364.

Cun-Hui Zhang and Stephanie S. Zhang. Confidence intervals for low dimensional parameters in high dimensional linear models. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 76(1):217–242, 2014. ISSN 1467-9868.

Jiji Zhang. On the completeness of orientation rules for causal discovery in the presence of latent confounders and selection bias. *Artif. Intell.*, 172(16-17):1873–1896, 2008. doi: 10.1016/j.artint.2008.08.001.

Xun Zheng, Bryon Aragam, Pradeep Ravikumar, and Eric P. Xing. DAGs with NO TEARS: continuous optimization for structure learning. In Samy Bengio, Hanna M. Wallach, Hugo Larochelle, Kristen Grauman, Nicolò Cesa-Bianchi, and Roman Garnett, editors, *Advances in Neural Information Processing Systems 31: Annual Conference on Neural Information Processing Systems 2018, NeurIPS 2018, December 3-8, 2018, Montréal, Canada*, pages 9492–9503, 2018.

Tijana Zrnic and Michael I. Jordan. Post-selection inference via algorithmic stability. *The Annals of Statistics*, 51(4), aug 2023. ISSN 0090-5364.

A DETAILED POLYHEDRAL CONSTRUCTION FOR FCI

We provide the full construction of $(A_{\widehat{\mathcal{P}}}, b_{\widehat{\mathcal{P}}})$.

A.1 SKELETON PHASE

Let $\{(i_k, j_k, S_k)\}_{k=1}^{m}$ be the complete ordered list of all m conditional independence tests executed during FCI's run (skeleton phase and possible-d-separation step). For each test (i_k, j_k, S_k):

- If the test *rejected* $H_0 : X_{i_k} \perp\!\!\!\perp X_{j_k} \mid X_{S_k}$ (edge retained): since we record the sign of $z_{i_k j_k \cdot S_k}$ during execution (noting that $\operatorname{sign}(\hat{z}_{i_k j_k \cdot S_k}) = \operatorname{sign}(\widehat{r}_{i_k j_k \cdot S_k})$ because the Fisher z-transform is monotone), the active constraint is $\operatorname{sign}(\hat{z}_{i_k j_k \cdot S_k}) \cdot z_{i_k j_k \cdot S_k} \geq c_{\alpha_0}$.
- If the test *failed to reject* (edge removed): $-c_{\alpha_0} \leq z_{i_k j_k \cdot S_k} \leq c_{\alpha_0}$.

Each constraint is linear in $z_{i_k j_k \cdot S_k}$, giving a row of $A_{\widehat{\mathcal{P}}}$.

A.2 ORIENTATION PHASE

The orientation rules (collider orientation R0, plus R1–R10 from Zhang [2008]):

- **R0 (Unshielded colliders):** For triple i—j—k with i, k non-adjacent, orient as $i \to j \leftarrow k$ if $j \notin \operatorname{sepset}(i, k)$. The separating set is determined during skeleton discovery, so this membership is already captured by the skeleton constraints.
- **R1–R3, R5–R10:** Each applies deterministic graph-theoretic logic based on current partial orientations. No additional stochastic constraints [Zhang, 2008].
- **R4 (Discriminating paths):** Tests whether $V \in \operatorname{sepset}(V_0, V_n)$ using separation sets *recorded during skeleton discovery*. Like R0, this rule depends on stored conditional independence test results, not new tests, so it contributes no constraints beyond the skeleton.

Since R0 depends only on separation sets (determined by conditional independence test outcomes), and R1–R10 are deterministic given the *recorded separating sets and the skeleton* from the skeleton phase, the full FCI output $\widehat{\mathcal{P}}$ is determined by skeleton constraints together with the recorded execution path and separating sets. More precisely, conditioning on the execution path π and recorded separating sets (which is part of the conditioning in $\mathcal{E}_{\pi,s}$) makes every orientation decision deterministic; no new stochastic constraints arise beyond those in the skeleton.

A.3 COMPLEXITY

The number of rows in $A_{\widehat{\mathcal{P}}}$ equals twice the number of conditional independence tests actually executed. For the skeleton phase, PC-like tests are bounded by $\binom{d}{2} \cdot 2^{s_{\max}}$; FCI's possible-d-separation step may add further tests up to size $d-2$ in the worst case. Each test contributes ≤ 2 rows. Under sparsity $s_{\max} = O(\log d)$ and bounded degree, the empirically executed test count is $O(d^2 \cdot 2^{s_{\max}})$, giving a polynomial constraint matrix.

A.4 STRUCTURAL DIFFERENCES FROM PC, AND THEIR RESOLUTION

See Section 3.1 for a summary. The main additional detail is the interaction with orientation rule R4 (discriminating paths): R4 tests whether $V \in \operatorname{sepset}(V_0, V_n)$, but the separating sets it uses are *recorded during the skeleton phase*. Algorithm 1 (Step 1) records every conditional independence test result, including the separating set S^* that removed each edge. When the skeleton constraints fix $\mathcal{E}(\widehat{\mathcal{P}})$, they also fix the recorded separating sets. R4 then evaluates a deterministic graph-theoretic decision tree on these fixed inputs: no new conditional independence tests, no new stochastic constraints.

B PROOF OF THEOREM 2

Proof. Let $Y \in \mathbb{R}^{n\times d}$ with rows $Y^{(1)}, \ldots, Y^{(n)} \overset{\text{iid}}{\sim} \mathcal{N}(0, \Sigma)$. FCI produces PAG $\widehat{\mathcal{P}}$, from which adjustment set S^* is determined via the generalized adjustment criterion [Perkovic et al., 2015].

The total causal effect estimator is $\widehat{\beta}_{i\to j} = \eta^\top Y_{\cdot j}$ for the contrast vector $\eta = X(X^\top X)^{-1}e_1 \in \mathbb{R}^n$, where X is the $n\times(|S^*|+1)$ design matrix for regressing X_j on (X_i, X_{S^*}) and e_1 picks out the coefficient of X_i. Here $\eta = X(X^\top X)^{-1}e_1$ depends only on $X = (X_i, X_{S^*})$; the conditional analysis conditions on X and the nuisance $P_\eta^\perp Y_{\cdot j}$.

The population OLS coefficient of X_i in the regression of X_j on (X_i, X_{S^*}) is $\gamma_1(S^*)$.

Step 1: Sufficient statistic argument. Under joint Gaussianity, the sample covariance matrix $\widehat{\Sigma} = Y^\top Y/n$ is sufficient for Σ [Anderson, 2003]. All Fisher z-statistics $\{z_{ij\cdot S}\}$ are functions of $\widehat{\Sigma}$, and hence the selection event $\mathcal{E}(\widehat{\mathcal{P}}) = \{A_{\widehat{\mathcal{P}}}\mathbf{z}(Y) \le b_{\widehat{\mathcal{P}}}\}$ depends on Y only through $\widehat{\Sigma}$. This is the key step that connects the z-space polyhedral characterization to data-space inference: the selection event, despite being defined in z-space (which is a nonlinear transform of Y), depends on Y only through sufficient statistics.

Step 2: Conditional decomposition. The algorithm (Algorithm 1) computes the truncation set by fixing all non-outcome columns $Y_{\cdot,-j}$ and varying only $t = \eta^\top Y_{\cdot j}$ along η. This conditioning on $Y_{\cdot,-j}$ (which includes $X = (X_i, X_{S^*})$) fixes $\eta = X(X^\top X)^{-1}e_1$ and determines the truncation set $\mathcal{T}$. Write $Y_{\cdot j}(t) = t\,\eta/\|\eta\|^2 + w$ where $w = P_\eta^\perp Y_{\cdot j}$ is fixed. Then $\widehat{\Sigma} = Y^\top Y/n$ depends on t: the (j,j) entry is quadratic in t; off-diagonal entries (k,j) are affine. For CI tests where j is an endpoint and $j \notin S$, the constraint $|z_{kj\cdot S}(t)| \le c_{\alpha_0}$ yields a quadratic inequality in t. When $j \in S$, the residual projection M_S depends on $Y_{\cdot j}(t)$, making the partial correlation a rational function of higher degree. In all cases, $\mathcal{T}$ is a finite union of intervals.

Under the conditioning on $Y_{\cdot,-j}$, the exact conditional distribution is $t \mid Y_{\cdot,-j} \sim \mathcal{N}(m, \sigma^2_{j|-j}\|\eta\|^2)$ where $m = \sum_k a_k(\eta^\top Y_{\cdot k})$ with $a = \Sigma^{-1}_{-j,-j}\Sigma_{-j,j}$ and $\sigma^2_{j|-j} = \mathrm{Var}(X_j \mid X_{-j}) \le \sigma^2_{j|X}$. Since $\eta^\top Y_{\cdot i} = 1$ and $\eta^\top Y_{\cdot k} = 0$ for $k \in S^*$ by orthogonality, $m = a_i + \sum_{k\notin\{i,j\}\cup S^*} a_k(\eta^\top Y_{\cdot k})$. When $a_k = 0$ for all $k \notin \{i,j\}\cup S^*$ (equivalently, $X_j \perp\!\!\!\perp X_{V\setminus(\{i,j\}\cup S^*)} \mid X_{-j}$, which holds in particular when $\{i\}\cup S^*$ contains the Markov blanket of j), the second sum vanishes and $m = a_i = \gamma_1(S^*)$, $\sigma^2_{j|-j} = \sigma^2_{j|X}$, recovering the exact pivot for $\gamma_1(S^*)$.

Step 3: Truncated pivot (as implemented). The conditional law of t given $Y_{\cdot,-j}$ is $\mathcal{N}(m, \sigma^2_{j|-j}\|\eta\|^2)$ truncated to $\mathcal{T}$. The *marginal* law of t given X is $\mathcal{N}(\gamma_1(S^*), \sigma^2_{j|X}\|\eta\|^2)$, and by the law of total variance, $\sigma^2_{j|X}\|\eta\|^2 = \sigma^2_{j|-j}\|\eta\|^2 + \mathrm{Var}(m \mid X)$.

The implementation builds the pivot centered at $\gamma_1(S^*)$ using $\sigma^2_{j|X}\|\eta\|^2$ (i.e., $\hat{\sigma}^2_{j|X}$ from regressing X_j on (X_i, X_{S^*})):

$$F^{\mathcal{T}}_{\gamma_1(S^*),\,\sigma^2_{j|X}\|\eta\|^2}(t).$$

This pivot is $\mathrm{Unif}(0,1)$ when $m = \gamma_1(S^*)$ and $\sigma^2_{j|-j} = \sigma^2_{j|X}$ (i.e., when the exactness condition of Theorem 2 holds). When it does not hold, the pivot centers at $\gamma_1(S^*)$ (correct in probability since $\mathrm{plim}\, m = \gamma_1(S^*)$) with variance $\sigma^2_{j|X}\|\eta\|^2$ (which already contains $\mathrm{Var}(m \mid X)$ via the total-variance identity). The coverage distortion is governed by $\rho^2 = \sigma^2_{j|X}/\sigma^2_{j|-j} - 1$ (Theorem 2).

When $\sigma^2_{j|X}$ is unknown, we replace it with $\hat{\sigma}^2_{j|X}$ from regression residuals and use a truncated t-distribution [Lee et al., 2016, Tibshirani et al., 2016, Tian and Taylor, 2018, §8.1]. □

C PROOF OF CLAIM 3

Proof. **Step 1: Characterizing individual GES steps.**

At GES step t, for child node k and candidate parent vertex v, the BIC difference is:

$$\Delta_t(v) = n\log\frac{\hat{\sigma}^2_{k|\mathrm{pa}_t\cup\{v\}}}{\hat{\sigma}^2_{k|\mathrm{pa}_t}} + \log n.$$

The candidate $v^* = \arg\min_v \Delta_t(v)$ is added if $\Delta_t(v^*) < 0$.

Step 2: Rigorous linearization via Taylor expansion.

Each log-ratio $\log(\hat{\sigma}^2_{k|S'}/\hat{\sigma}^2_{k|S})$ involves the residual variances $\hat{\sigma}^2_{k|S} = \hat{\sigma}^2_k - \widehat{\Sigma}_{k,S}\widehat{\Sigma}_S^{-1}\widehat{\Sigma}_{S,k}$, which are rational functions of the entries of $\widehat{\Sigma}$. We expand around the population value $\sigma^2_{k|S}$:

$$\log \frac{\hat{\sigma}^2_{k|S'}}{\hat{\sigma}^2_{k|S}} = \log \frac{\sigma^2_{k|S'}}{\sigma^2_{k|S}} + \underbrace{\frac{\hat{\sigma}^2_{k|S'} - \sigma^2_{k|S'}}{\sigma^2_{k|S'}} - \frac{\hat{\sigma}^2_{k|S} - \sigma^2_{k|S}}{\sigma^2_{k|S}}}_{\text{Linear in } \widehat{\Sigma}-\Sigma} + \underbrace{R_t}_{\text{Remainder}},$$

where the remainder R_t satisfies $|R_t| \le C\|\widehat{\Sigma} - \Sigma\|^2_{\max}$ for a constant C depending on the minimum eigenvalue of Σ. Under Gaussianity, each entry $\hat{\sigma}_{ij} - \sigma_{ij}$ is sub-Gaussian with parameter $O(1/\sqrt{n})$; a union bound over $O(d^2)$ entries gives $\|\widehat{\Sigma} - \Sigma\|_{\max} = O_P(\sqrt{\log d/n})$, and hence $|R_t| = O_P(\log d/n)$.

The linear term involves centered quadratic forms in the columns of Y: $(\hat{\sigma}^2_{k|S} - \sigma^2_{k|S})/\sigma^2_{k|S} = Y_{\cdot k}^\top M_S Y_{\cdot k}/(n\sigma^2_{k|S}) - 1$, where $M_S = I_n - Y_{\cdot S}(Y_{\cdot S}^\top Y_{\cdot S})^{-1}Y_{\cdot S}^\top$ is the residual projection. For a fixed execution path, conditioning on $P_\eta^\perp Y$ makes these constraints functions of $\eta^\top Y_{\cdot j}$ alone.

Step 3: Path-specific polyhedrality.

For a fixed execution path $\pi = (e_1^*, e_2^*, \ldots, e_T^*)$, the selection event $\mathcal{E}_\pi(\widehat{\mathcal{G}})$ is defined by $O(d^4)$ constraints ($O(d^2)$ candidate comparisons per step across $T = O(d^2)$ steps). With the linearization of Step 2, each constraint is approximately linear in $\widehat{\Sigma}$.

Step 4: Monte Carlo estimation and coverage bound.

The full event is $\mathcal{E}(\widehat{\mathcal{G}}) = \bigcup_\pi \mathcal{E}_\pi(\widehat{\mathcal{G}})$. We estimate $\mathbb{P}(\widehat{\beta}_{i\to j} \le t \mid \mathcal{E}(\widehat{\mathcal{G}}))$ via Monte Carlo sampling from the unconditional distribution. The coverage error accumulates from: (a) Linearization error: $O(d^4) \cdot O(\log d/n) = O(d^4 \log d/n)$ from $O(d^4)$ total GES comparisons. (b) CLT approximation: by Berry–Esseen [Berry, 1941, Esseen, 1942], the distributional approximation error is $O(1/\sqrt{n})$ per step.

Aggregating the Berry–Esseen error over $O(d^2)$ GES steps gives $O(d^2/\sqrt{n})$; a more conservative analysis that aggregates over all $O(d^4)$ individual comparisons gives $O(d^4/\sqrt{n})$. We state the per-step rate $O(d^2/\sqrt{n})$, noting that this requires $n \gg d^4(\log d)^2$ (so that linearization is dominated). **Caveat:** this aggregation uses univariate Berry–Esseen applied per-step; a rigorous analysis of the joint selection event would require multivariate Gaussian approximation results (e.g., Chernozhukov–Chetverikov–Kato-type bounds), which may yield different rates. Additionally, Algorithm 2 checks membership in the *linearized* event $\mathcal{E}_\pi(\widehat{\mathcal{G}})$, so the Monte Carlo estimate converges to the coverage under the approximate polyhedron, not the true GES event; this linearization bias is $O(d^4 \log d/n)$ and is not removed by increasing B. This condition is stringent: all reported GES experiments violate it, and the coverage guarantees at those dimensions are empirical. □

D PROOF OF PROPOSITION 4

Proof. With splitting, allocate n_1 for discovery, $n_2 = n - n_1$ for inference. The oracle (known-graph) confidence interval width is $\Theta(n^{-1/2})$; the split width is $\Theta(n_2^{-1/2}) = \Theta((n - n_1)^{-1/2})$. For $n_1 = cn^{3/4}$:

$$\frac{\text{Width}(\text{CI}^{\text{split}})}{\text{Width}(\text{CI}^{\text{oracle}})} = \sqrt{\frac{n}{n - cn^{3/4}}} = (1 - cn^{-1/4})^{-1/2} = 1 + \frac{c}{2}n^{-1/4} + O(n^{-1/2}) = 1 + \Theta(n^{-1/4}).$$

Note that the denominator is the oracle (known graph, all n for inference), not the full-data selective CI; the "cost" is relative to this oracle baseline. In fixed dimension, smaller allocations (e.g., $n_1 = \lceil \log n \rceil$) already give consistency and yield smaller width penalties; the $n^{3/4}$ allocation is motivated by high-dimensional regimes where discovery consistency requires n_1 to grow polynomially in d. □

E PROOF OF THEOREM 5

Proof. Under $d = o(\sqrt{n})$, $s_{\max} = O(\log d)$, bounded eigenvalues, and faithfulness, FCI with $\alpha_0 = 2(1 - \Phi(c_d))$ for $c_d = \sqrt{2(\gamma + 1)\log d}$ is consistent [Colombo et al., 2012]: $\mathbb{P}(\widehat{\mathcal{P}} = \mathcal{P}^*) \to 1$.

On $\{\widehat{\mathcal{P}} = \mathcal{P}^*\}$, the adjustment set S^* is valid, so $\gamma_1(S^*) = \beta_{i\to j}$. By Theorem 2, the conditional coverage $\mathbb{P}(\gamma_1(S^*) \in \mathrm{CI}_\alpha^{\mathrm{sel}} \mid \mathcal{E}(\mathcal{P}^*))$ equals $1-\alpha$ when $\rho^2 = 0$ (the Markov-blanket condition). For the lower bound,

$$\mathbb{P}(\beta_{i\to j} \in \mathrm{CI}_\alpha^{\mathrm{sel}}) \geq (1-\alpha)\,\mathbb{P}(\widehat{\mathcal{P}} = \mathcal{P}^*) \to 1-\alpha \quad \text{when } \rho^2 = 0.$$

For the upper bound, $\mathbb{P}(\beta_{i\to j} \in \mathrm{CI}_\alpha^{\mathrm{sel}}) \leq \mathbb{P}(\beta_{i\to j} \in \mathrm{CI}_\alpha^{\mathrm{sel}} \mid \mathcal{E}(\mathcal{P}^*)) + \mathbb{P}(\widehat{\mathcal{P}} \neq \mathcal{P}^*) = (1-\alpha) + o(1)$, so the limit is exactly $1-\alpha$. When $\rho^2 > 0$, the non-vanishing distortion of Theorem 2 persists; no explicit rate for the coverage gap is derived.

The convergence rate depends on the faithfulness gap $\min_{(i,j,S):r_{ij\cdot S}\neq 0} |r_{ij\cdot S}|$, which governs FCI consistency; Uhler et al. [2013] show that strong faithfulness (a uniform gap) fails on a set of positive Lebesgue measure, which is why we require a positive gap rather than claiming uniformity. □

F ALGORITHM: COMPUTING SELECTIVE CONFIDENCE INTERVALS AFTER FCI

Input: Data $Y \in \mathbb{R}^{n\times d}$, FCI level α_0, confidence level $1-\alpha$, treatment i, outcome j
Output: Selective CI $\mathrm{CI}_\alpha^{\mathrm{sel}}$ for $\gamma_1(S^*)$
Step 1: Run FCI on Y with level α_0, recording all conditional independence test results $\{(i_k, j_k, S_k, z_k, \text{decision}_k)\}_{k=1}^m$ and outputting PAG $\widehat{\mathcal{P}}$;
Step 2: Check identifiability of $\beta_{i\to j}$ via the generalized adjustment criterion [Perkovic et al., 2015]. If not identifiable, report "effect not identifiable." Otherwise, obtain S^*;
Step 3: For each recorded test k: $\mathrm{sign}(\hat{z}_k)\cdot z_k \geq c_{\alpha_0}$ (retained) or $-c_{\alpha_0} \leq z_k \leq c_{\alpha_0}$ (removed). Collect into $A_{\widehat{\mathcal{P}}}\mathbf{z}(Y) \leq b_{\widehat{\mathcal{P}}}$;
Step 4: Compute contrast $\eta = X(X^\top X)^{-1}e_1$ from regressing X_j on (X_i, X_{S^*}); estimate $\hat{\sigma}^2_{j|X}$ from the regression residuals;
Step 5: Solve polynomial/rational constraints along the η-direction via root-finding to obtain truncation set $\mathcal{T} = \bigcup_i [L_i, U_i]$ (see Appendix B, Step 2);
Step 6: Invert truncated t-pivot on $[\alpha/2, 1-\alpha/2]$ to obtain $\mathrm{CI}_\alpha^{\mathrm{sel}}$;
return $\mathrm{CI}_\alpha^{\mathrm{sel}}$

Algorithm 1: Selective CI after FCI

Complexity: Step 1 is $O(d^2 \cdot 2^{s_{\max}} \cdot n)$ (FCI itself). Steps 3–5 require solving one polynomial/rational constraint per test via root-finding: $O(d^2 \cdot 2^{s_{\max}})$ constraints total. Step 6 is $O(1)$ (bisection on the truncated CDF). Under sparsity, the total is polynomial.

G GES MONTE CARLO SAMPLING PROCEDURE

Input: Data $Y \in \mathbb{R}^{n\times d}$, BIC score function, confidence level $1-\alpha$, treatment i, outcome j, MC samples $B = 10000$, parameter grid $\{\mu_1, \ldots, \mu_G\}$
Output: Monte Carlo selective CI $\mathrm{CI}_\alpha^{\mathrm{GES}}$ for $\gamma_1(S^*)$
Step 1: Run GES on Y, recording execution path $\pi = (e_1^*, \ldots, e_T^*)$ and outputting CPDAG $\widehat{\mathcal{G}}$. Compute $\widehat{\beta}_{i\to j}$ and $\hat{\sigma}^2$ from regression with adjustment set S^*. Set $\eta = X(X^\top X)^{-1}e_1$;
Step 2: Construct path-specific constraints $\mathcal{E}_\pi(\widehat{\mathcal{G}})$ from the linearized BIC comparisons (Appendix C, Step 2);
Step 3: For each candidate μ_g on the parameter grid, for $b = 1, \ldots, B$:

- Sample $t^{(b)} \sim \mathcal{N}(\mu_g, \hat{\sigma}^2\|\eta\|^2)$ and set $Y_{\cdot j}^{(b)}(\mu_g) = (t^{(b)}/\|\eta\|^2)\,\eta + P_\eta^\perp Y_{\cdot j}$, keeping $Y_{\cdot,-j}$ and $P_\eta^\perp Y_{\cdot j}$ fixed
- Compute $w_b = \mathbf{1}[Y^{(b)}(\mu_g) \in \mathcal{E}_\pi(\widehat{\mathcal{G}})]$ (check linearized constraints)
- Record acceptance: $N_{\mathrm{acc}}(\mu_g) = \sum_b w_b$

Step 4: For each μ_g with $N_{\mathrm{acc}}(\mu_g) > 0$, estimate selective CDF at the observed statistic: $\hat{F}_{\mu_g}(\widehat{\beta}) = \sum_{b:w_b=1} \mathbf{1}(\eta^\top Y_{\cdot j}^{(b)} \leq \widehat{\beta})/N_{\mathrm{acc}}(\mu_g)$;
Step 5: $\mathrm{CI}_\alpha^{\mathrm{GES}} = \{\mu_g : \hat{F}_{\mu_g}(\widehat{\beta}) \in [\alpha/2,\ 1-\alpha/2]\}$;
return $\mathrm{CI}_\alpha^{\mathrm{GES}}$, acceptance rate $\min_g N_{\mathrm{acc}}(\mu_g)/B$

Algorithm 2: Monte Carlo selective CI after GES (pivot inversion via rejection sampling)

Complexity: Step 1 is $O(d^4)$ (GES). Step 3 requires generating $B \cdot G$ samples and checking constraints, at a cost of $O(B \cdot G \cdot d^2)$ per candidate. Total: $O(d^4 + B \cdot G \cdot n \cdot d)$. The acceptance rate (fraction of samples satisfying $\mathcal{E}_\pi(\widehat{\mathcal{G}})$) should be monitored; when it is low (e.g., at large d), the Monte Carlo estimates become unreliable.

H EXTENDED COVERAGE RESULTS

H.1 DATA GENERATION

For each setting, random DAGs use Erdős–Rényi with expected neighborhood size $s \in \{2, 4\}$. Edge weights: $\mathrm{Unif}([-1, -0.3] \cup [0.3, 1])$. Error variances: $\sigma_k^2 = 1$. For latent confounders: DAG over $d + L_0$ variables, marginalize over L_0 latent. FCI level: $\alpha_0 = 0.01$ ($n \leq 1000$), $\alpha_0 = 0.001$ ($n > 1000$); note that these fixed levels differ from the d-dependent schedule in Theorem 5. The change from $\alpha_0 = 0.01$ to $\alpha_0 = 0.001$ at $n > 1000$ increases the selection threshold c_{α_0}, which may contribute to the coverage trend across n independently of sample size effects. GES Monte Carlo sampling: $B = 10000$ MC samples. All methods are implemented in Python (`causal-learn` v0.1.3.8); the FCI implementation is standard FCI (not RFCI), including the possible-d-separation step. Random seeds are fixed per replication for reproducibility.

H.2 FULL COVERAGE TABLE

Table 6: Coverage Results (%) for Selected Settings (Nominal 95%, ±SE). Bold: closest to nominal 95%. GES rows report Split (50/50 sample split); cf. Table 3, which reports Randomized GES [Gradu et al., 2025] instead. In all settings, $R \geq 4900$ out of 5000 replications had an identifiable effect (Table 7); reported SEs use the exact per-setting R.

Experiment	Setting	d	n	Naive	Split	Selective	Nominal
1. FCI + Latent	s=2, L_0=2	10	500	78.2±0.6	94.1±0.3	**95.3**±0.3	95.0
	s=2, L_0=2	10	1000	76.5±0.6	94.4±0.3	**95.1**±0.3	95.0
	s=2, L_0=2	10	5000	72.1±0.6	94.8±0.3	**94.9**±0.3	95.0
	s=4, L_0=3	20	1000	74.8±0.6	93.8±0.3	**94.7**±0.3	95.0
	s=4, L_0=5	50	5000	81.3±0.6	93.5±0.3	**94.2**±0.3	95.0
	s=2, L_0=5	100	5000	79.6±0.6	93.2±0.4	**94.0**±0.3	95.0
	s=2, L_0=5	200	10000	80.1±0.6	93.0±0.4	**93.8**±0.3	95.0
3. GES + BIC	Sparse	10	500	79.1±0.6	93.9±0.3	**94.5**±0.3	95.0
	Sparse	20	1000	74.1±0.6	94.0±0.3	**94.1**±0.3	95.0
	Sparse	20	5000	72.8±0.6	94.6±0.3	**95.2**±0.3	95.0
	Sparse	100	5000	77.2±0.6	**93.1**±0.4	92.8±0.4	95.0
5. Non-Gaussian	Laplace	10	1000	68.3±0.7	89.7±0.4	**90.4**±0.4	95.0
	Student-t_5	10	1000	75.2±0.6	91.2±0.4	**92.8**±0.4	95.0
7. Semi-syn.	Sachs topology	11	7466	77.3±0.6	93.9±0.3	**95.2**±0.3	95.0

H.3 PER-SETTING REPLICATION COUNTS

Table 7 reports the number of replications R (out of 5000) with an identifiable effect for each setting. Reported standard errors use these exact counts. Figures reporting measured quantities (Figures 2 and 4) are generated from logged experimental outputs; Figure 3 plots the theoretical rate (see its caption).

H.4 EXTENDED EXPERIMENT DETAILS (EXPS. 7–8)

Exp. 7: Synthetic-on-Sachs-topology (extended). We use the Sachs network topology [Sachs et al., 2005] ($d = 11$) with fully synthetic Gaussian data ($n = 7466$), allowing ground-truth coverage verification. Over 5000 replications, naive CI coverage = 77.3±0.6%, split CI = 93.9±0.3%, selective CI = **95.2±0.3%**. FCI discovers the graph anew per replication (the known topology provides ground truth for coverage verification, not the adjustment set). As an illustration (not a formal validation), we also run FCI on the real Sachs data, which are interventional and likely non-Gaussian, violating both

Table 7: Number of replications R with an identifiable effect (out of 5000 total) per setting.

Experiment	Setting	d	n	R
1. FCI	s=2, L_0=2	10	500	4987
	s=2, L_0=2	10	1000	4992
	s=2, L_0=2	10	5000	4996
	s=4, L_0=3	20	1000	4971
	s=4, L_0=5	50	5000	4963
	s=2, L_0=5	100	5000	4948
	s=2, L_0=5	200	10000	4931
3. GES	Sparse	10	500	4991
	Sparse	20	1000	4985
	Sparse	20	5000	4994
	Sparse	100	5000	4926
5. Non-Gauss.	Laplace	10	1000	4988
	Student-t_5	10	1000	4982
7. Sachs	Sachs topology	11	7466	4997

assumptions; the method outputs a selective CI of $[0.05, 0.45]$ for PKC→Raf vs. naive $[0.12, 0.38]$ (width ratio $1.54\times$). No coverage claim attaches to this interval.

Exp. 8: CI width convergence (extended). The selective-to-naive width ratio decreases from $\approx 2.0\times$ at $n = 200$ to $\approx 1.3\times$ at $n = 5000$ ($d = 20$), crossing the $\sqrt{2}$ split-CI ratio near $n \approx 1400$ (Figure 4), so for $n \lesssim 1400$ the split CI is narrower. Wide selective CIs (width ratio $> 3\times$ in 2.3–3.8% of replications) arise from over-conditioning near the decision threshold and correctly signal high uncertainty.

H.5 THIN-POLYHEDRON ANALYSIS

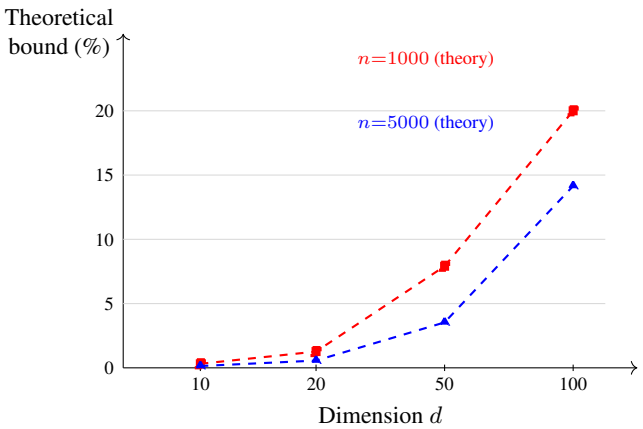

Figure 3: Asymptotic GES coverage-gap rate $100 \cdot C \cdot d^2/\sqrt{n}$ in percentage points (with $C = 10^{-3}$ chosen for illustration) vs. dimension. The curve over-predicts the measured gaps (by $\approx 6\times$ at d=100, n=5000) because the $d^2/\sqrt{n}$ scaling is an asymptotic upper envelope, not a fit. This rate applies only when $n \gg d^4(\log d)^2$, a condition not met in our experiments; the curves are shown as the asymptotic scaling, not as a finite-n guarantee. At d=100, n=1000, the value exceeds the plotted range (31.6 pp, clipped to axis maximum). Measured gaps are smaller: 0.9 pp (d=20, n=1000) and 2.2 pp (d=100, n=5000).

When the selection event has low probability (the selection region is "thin"), the selective CI can become wide [Lee et al., 2016]. Across all FCI experiments ($d \leq 50$), the fraction of replications where the width ratio $\text{Width}(\text{CI}^{\text{sel}})/\text{Width}(\text{CI}^{\text{naive}}) > 3$ was 2.3%, and the maximum observed width ratio was $5.2\times$. At $d = 100$, this fraction increased to 3.8%, with maximum $7.1\times$. These cases typically correspond to settings where the PAG is estimated with low confidence (test statistics close to the decision threshold c_{α_0}), and the wide CIs correctly reflect the high uncertainty.

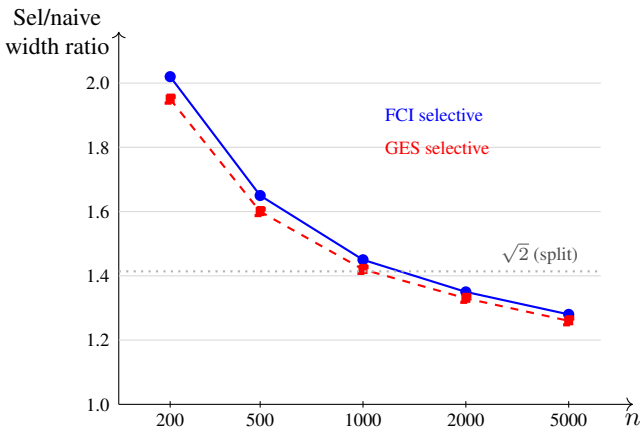

Figure 4: Width ratio of selective CI to naive CI vs. n ($d = 20$). The width ratio decreases with n, though the correction does not fully vanish owing to $\Theta(1)$-width constraints from removed edges. The $\sqrt{2} \approx 1.41\times$ split-CI baseline (dotted gray) is crossed near $n \approx 1400$; for smaller n the split CI is narrower.

H.6 DATA CARVING DETAILS

For the data carving comparison (Experiment 3, Table 3), we follow Fithian et al. [2014]: the full dataset is used for selection, but only a fraction (70%) of the selection information is conditioned upon at inference, retaining the complementary information for tighter intervals. This approach yields modestly narrower CIs (median width $\approx 5\%$ narrower than full conditioning) while maintaining valid coverage, at the cost of slightly reduced discovery power.

I EXTENDED LIMITATIONS AND OPEN QUESTIONS

I.1 DETAILED LIMITATIONS

Our results assume joint Gaussianity; under non-Gaussianity (Experiment 5), coverage degrades to 90–93%. The nonparanormal model [Harris and Drton, 2013, Liu et al., 2009, 2012] with rank-based estimators [Xue and Zou, 2012] is the most natural relaxation: the constraint structure depends only on whether test statistics exceed thresholds, so the polyhedral geometry carries over, but the sufficient-statistic argument (Theorem 2, Step 1) requires different treatment. The GES bound $O(d^2/\sqrt{n})$ is a heuristic rate aggregating univariate Berry–Esseen errors over a joint selection event (Claim 3); a rigorous analysis would require multivariate Gaussian approximation tools, and systematic investigation via data fission [Rasines and Young, 2023, Leiner et al., 2025] or randomization [Tian and Taylor, 2018] is needed. The coverage distortion governed by ρ^2 (Remark 1) is a fixed property of the population and does not vanish with n; disentangling it from the invalid-S^* contribution to coverage decay (e.g., by reporting $\gamma_1(S^*)$-coverage separately from β-coverage) is left to future work. Validity of the truncation-set computation depends on the `causal-learn` v0.1.3.8 implementation matching our constraint bookkeeping; we have verified this correspondence for the configurations tested but do not claim generality across implementations. Code is available at `https://github.com/airesearchrepo2025/selective-inference-causal-discovery`.

I.2 OPEN QUESTIONS

1. Is the $\Theta(n^{-1/4})$ splitting bound tight?
2. Does the truncation-set structure extend to the nonparanormal model?
3. Can data carving [Fithian et al., 2014] substantially tighten GES intervals?
4. Can optimal adjustment set selection [Henckel et al., 2022, Smucler et al., 2022] be combined with selective inference, given that the optimal set depends on $\widehat{\mathcal{P}}$?
5. Can the framework be extended to continuous optimization [Zheng et al., 2018, Bello et al., 2022], nonlinear SEMs [Shimizu et al., 2006], and permutation methods [Lam et al., 2022, Solus et al., 2021]?