THE BLESSING OF LINEAR NON-GAUSSIAN MODELS: POWERFUL INDEPENDENT SUBSPACE INFERENCE FOR LATENT-VARIABLE CAUSAL DISCOVERY

Anonymous authors
Paper under double-blind review

ABSTRACT

The linear non-Gaussian model is a classical parametric model for causal discovery, known for its strong identifiability. A key ingredient underlying such identifiability is independence induced by linear transformations. For example, given two random vectors $\mathbf{X}$ and $\mathbf{Z}$, one may seek a transformation $\omega^\top \mathbf{X}$ such that $\omega^\top \mathbf{X} \perp\!\!\!\perp \mathbf{Z}$, where ω may arise from linear regression or more general constructions. In the presence of latent confounders, two established tools, generalized independent noise (GIN) and transformed independent noise (TIN), exploit such constraints to characterize independent linear transformations. Despite their success in causal identification, existing implementations do not fully exploit the statistical structure of linear non-Gaussian models, which may compromise the reliability of causal identification. To address this issue, building on the characterization of independence through constant conditional means and variances, we construct a fixed linear operator whose kernel exactly characterizes all independent directions, enabling unified estimation and testing using only low-order information. We develop a paired-bootstrap GIN test with pointwise asymptotic Type I error control and a jointly calibrated TIN rank estimator. Synthetic and real-world experiments across several settings show gains in GIN testing power, TIN dimension recovery, and downstream latent-variable causal discovery.

1 INTRODUCTION

Causal discovery, which aims to uncover causal relationships from observational data, has become an important paradigm in artificial intelligence and has attracted increasing attention across various disciplines (Spirtes et al., 2010; Pearl, 2009). A fundamental question in causal discovery is identifiability (Spirtes et al., 2000), i.e., under what conditions the underlying causal structure can be uniquely determined from the observed distribution. To achieve stronger identifiability, recent studies often impose additional parametric assumptions on the data-generating process, with the linear non-Gaussian acyclic model (LiNGAM) (Shimizu et al., 2006) being a prominent example. In LiNGAM, structural identifiability is closely related to independent-noise (IN) constraints (Shimizu et al., 2011), which characterize the independence between regression residuals and explanatory variables. For instance, given two random vectors $\mathbf{X}$ and $\mathbf{Z}$, one may seek a linear transformation $\omega^\intercal \mathbf{X}$, typically obtained through linear regression, such that $\omega^\intercal \mathbf{X} \perp\!\!\!\perp \mathbf{Z}$. Such IN constraints provide useful information for identifying the causal structure among observed variables (Shimizu, 2014).

In many practical scenarios, the variables of interest may be latent (Silva et al., 2006; Kummerfeld & Ramsey, 2016). For example, in a real-data mental ability test, visual, verbal, and speed abilities may be unobserved, with only scores on individual tests available. In such cases, IN-based methods are not directly applicable for identifying the causal relations among the latent variables. One way to extend this principle to settings with latent confounders is to exploit richer independence constraints over the entire set of linear transformations for structural identification (Spirtes & Zhang, 2016). Concretely, let $\mathbf{X} \in \mathbb{R}^p$ be observed variables to be transformed and $\mathbf{Z}$ the reference variables. Under the linear non-Gaussian models considered here, these linear transformations form the independent transformation subspace (Dai et al., 2022):

$$\Omega_{\mathbf{Z};\mathbf{X}} := \{w \in \mathbb{R}^p : w^\top \mathbf{X} \perp\!\!\!\perp \mathbf{Z}\}.$$

Reviewers: please read the Reviewer Guidelines (iclr.cc/Conferences/2027/ReviewerGuidelines) and the AI Policy for Reviewers (iclr.cc/Conferences/2027/AIPolicyForReviewers). If you used AI to expand, edit, or polish your review, please provide the input text to the LLM. Better still, consider skipping the LLM and submitting your original text: we, and the authors, are much more interested in your unedited thoughts than in what an LLM has to say. AI-assisted or not, you are putting your name and reputation behind your review: LLM-generated falsehoods, hallucinations or misrepresentations are subject to disciplinary action, which may include desk-rejecting all papers you have authored.

We use *independent subspace inference* to refer broadly to statistical inference about the independent transformation subspace. Two fundamental tools pose different questions about it. The generalized independent noise (GIN) condition (Xie et al., 2020; 2024) examines whether all linear combinations of $\mathbf{X}$ canceling cross-covariances fall in $\Omega_{\mathbf{Z};\mathbf{X}}$, provided that a non-zero such direction exists. The transformed independent noise (TIN) condition (Dai et al., 2022) further studies the codimension of $\Omega_{\mathbf{Z};\mathbf{X}}$. Although these tools serve as important building blocks for latent structure identifiability (Xie et al., 2022; Jin et al., 2023; Xia et al., 2026), their statistical implementation typically relies on off-the-shelf independence tests, such as HSIC (Gretton et al., 2005), without explicitly exploiting the linear non-Gaussian structure underlying the data. Ignoring such intrinsic structure may reduce statistical efficiency and, consequently, affect the reliability of the identification procedure.

Our key insight is that parametric assumptions, such as linearity and non-Gaussianity, can not only enable causal identification but also simplify the statistical inference required for identification. Recent work provides initial evidence in independence testing. Specifically, for two linear mixtures of independent non-Gaussian sources, LiNGIC (Li et al., 2026) shows that their independence can be characterized by constant conditional first and second moments (see Theorem 2.6), thereby reducing the complexity of general kernel-based independence tests. This leads to our question:

"Can the linear non-Gaussian structure also be exploited to simplify inference of the independent transformation subspace underlying GIN and TIN conditions?"

Building on the independence characterization of LiNGIC (Li et al., 2026), one direct step is to replace the independence tests in existing procedures with LiNGIC. However, independent subspace inference requires estimating the linear transformations $\omega^\top \mathbf{X}$ from data, whose uncertainty may lead to miscalibrated independence tests. To address this issue, we combine first-order responses with mixed second-order conditional responses that are linear in the transformation coefficients, yielding a fixed operator whose null space exactly characterizes the independent transformation subspace. We theoretically show that finitely many random kernel evaluations suffice to preserve this subspace, enabling unified inference for GIN and TIN using only low-order information. We then use a paired bootstrap to calibrate a GIN test over the covariance-null space and joint multiplier calibration to estimate the rank required by TIN. The GIN test has pointwise asymptotic Type I error control under the stated conditions.

Our contributions are three-fold:

1. We characterize the entire independent subspace through a fixed linear operator of low-order conditional moments, and establish conditions under which a finite random representation preserves it almost surely.
2. We develop calibrated procedures for both GIN testing and TIN dimension inference based on this operator, and establish pointwise asymptotic Type I error control for the GIN test.
3. We evaluate the proposed procedures against existing implementations. Experiments demonstrate significant gains in GIN testing power and TIN dimension recovery in broad settings, as well as improved downstream latent-variable causal discovery performance.

2 PRELIMINARIES

Notation. Let $\mathbf{X} = (X_1, \ldots, X_p)^\top \in \mathbb{R}^p$ and $\mathbf{Z} = (Z_1, \ldots, Z_q)^\top \in \mathbb{R}^q$ denote observed random vectors, and let $w \in \mathbb{R}^p$ denote a projection direction. For a matrix $\mathbf{A}$, $\mathbf{A}_{:j}$ is its jth column, $\mathbf{A}_J$ contains the columns indexed by J, and $\ker(\mathbf{A})$ and $\mathrm{rank}(\mathbf{A})$ denote its null space and rank. Write $\Sigma_{\mathbf{XX}} := \mathrm{Cov}(\mathbf{X}, \mathbf{X})$, $\Sigma_{\mathbf{ZX}} := \mathrm{Cov}(\mathbf{Z}, \mathbf{X})$. We use $\|\cdot\|$ for the Euclidean norm of vectors and $\|\cdot\|_{\mathrm{F}}$ for the Frobenius norm of matrices. For sampling statements, $\{(\mathbf{X}^{(i)}, \mathbf{Z}^{(i)})\}_{i=1}^n$ are independent copies of $(\mathbf{X}, \mathbf{Z})$. Population variables are centered for notation. Empirical moments use $\mathbf{X}^{(i)} - \overline{\mathbf{X}}$ and $\mathbf{Z}^{(i)} - \overline{\mathbf{Z}}$, where bars denote sample means. Hats denote empirical quantities.

We work with the following linear non-Gaussian source model.

Definition 2.1 (Linear non-Gaussian model)**.** The centered random vectors $\mathbf{X} \in \mathbb{R}^p$ and $\mathbf{Z} \in \mathbb{R}^q$ satisfy $\mathbf{X} = \mathbf{ME}$ and $\mathbf{Z} = \mathbf{NE}$, where $\mathbf{M} \in \mathbb{R}^{p\times m}$ and $\mathbf{N} \in \mathbb{R}^{q\times m}$ are fixed loading matrices. The coordinates of $\mathbf{E} = (E_1, \ldots, E_m)^\top$ are mutually independent, centered, non-Gaussian sources with strictly positive, finite variances.

Independent subspace inference. Under a linear non-Gaussian model, the directions whose projections are independent of $\mathbf{Z}$ form a linear subspace of $\mathbb{R}^p$, which is formulated as follows.

Definition 2.2 (Independent transformation subspace (Dai et al., 2022))**.** The independent transformation subspace is $\Omega_{\mathbf{Z};\mathbf{X}} := \{w \in \mathbb{R}^p : w^\top \mathbf{X} \perp\!\!\!\perp \mathbf{Z}\}$.

Two fundamental tools, GIN and TIN conditions, describe two properties of this subspace.

Definition 2.3 (GIN condition (Xie et al., 2020))**.** The pair $(\mathbf{Z}, \mathbf{X})$ satisfies GIN if the covariance-null space is non-trivial and equals the independent transformation subspace: $\ker(\Sigma_{\mathbf{ZX}}) \neq \{0\}$ and $\ker(\Sigma_{\mathbf{ZX}}) = \Omega_{\mathbf{Z};\mathbf{X}}$.

Definition 2.4 (TIN condition (Dai et al., 2022))**.** The TIN condition of $\mathbf{Z}$ and $\mathbf{X}$ is defined as the codimension of the independent transformation subspace: $\mathrm{TIN}(\mathbf{Z}, \mathbf{X}) := p - \dim(\Omega_{\mathbf{Z};\mathbf{X}})$.

The following two graphs provide concrete examples to illustrate these concepts.

Example 2.5. *In panel (a), let $Z = L + E_Z$, $X_1 = L + E_1$, $X_2 = L + E_2$, and write $\mathrm{Var}(L) = \sigma^2$. Then $\Sigma_{\mathbf{ZX}} = \sigma^2(1,1), \ker(\Sigma_{\mathbf{ZX}}) = \mathrm{span}\{(1,-1)^\top\}$. For $w = (1,-1)^\top$, $w^\top \mathbf{X} = X_1 - X_2 = E_1 - E_2 \perp\!\!\!\perp \mathbf{Z}$. Every covariance-null direction is a multiple of w, and eliminating the only source shared with $\mathbf{Z}$ requires exactly $w_1 + w_2 = 0$. Hence $\Omega_{\mathbf{Z};\mathbf{X}} = \mathrm{span}\{(1,-1)^\top\}$, $\mathrm{TIN}(\mathbf{Z}, \mathbf{X}) = 2 - \dim(\Omega_{\mathbf{Z};\mathbf{X}}) = 1$, so the GIN condition is satisfied.*



(a) one shared source *(b) two shared sources*

In panel (b), let $Z = L_1 + L_2 + E_Z$, $X_1 = L_1 + E_1$, $X_2 = L_2 + E_2$, where $\mathrm{Var}(L_1) = \mathrm{Var}(L_2) = \sigma^2$. $\Sigma_{\mathbf{ZX}}$ and its null space are identical to those in panel (a). Nevertheless, for $w = (1,-1)^\top$, $w^\top \mathbf{X} = L_1 - L_2 + E_1 - E_2$ still shares both L_1 and L_2 with $\mathbf{Z}$ and is therefore dependent on $\mathbf{Z}$. Consequently, the GIN condition is violated in panel (b). More generally, independence from Z requires eliminating both shared sources, forcing $w_1 = w_2 = 0$. Thus, $\Omega_{\mathbf{Z};\mathbf{X}} = \{0\}$, $\mathrm{TIN}(\mathbf{Z}, \mathbf{X}) = 2 - 0 = 2$.

Independence tests in a linear non-Gaussian model. For arbitrary random variables, $U \perp\!\!\!\perp V$ if and only if $\mathrm{Cov}(f(U), g(V)) = 0$ for all bounded continuous functions f and g (Jacod & Protter, 2004). Under Def. 2.1, the following theorem shows that this requirement can be relaxed.

Theorem 2.6 (Independence characterization in linear non-Gaussian model (Li et al., 2026))**.** *Under Def. 2.1, two scalar variables $U = a^\top \mathbf{E}$ and $V = b^\top \mathbf{E}$ are independent if and only if $\mathbb{E}[U \mid V] = 0$ and $\mathbb{E}[U^2 \mid V] = \mathbb{E}[U^2]$ almost surely.*

Notably, without Def. 2.1, constant conditional mean and variance do not generally imply independence. To express these conditions using a rich class of functions of the reference variables, we briefly recall the kernel concepts used below.

Definition 2.7 (Kernel and reproducing kernel Hilbert space)**.** A kernel on a domain $\mathcal{T}$ is a symmetric, positive-semidefinite function $\kappa : \mathcal{T} \times \mathcal{T} \to \mathbb{R}$. Its associated reproducing kernel Hilbert space (RKHS) $\mathcal{H}$ is a Hilbert space of functions on $\mathcal{T}$ satisfying $f(t) = \langle f, \kappa(t, \cdot)\rangle_{\mathcal{H}}, f \in \mathcal{H}, t \in \mathcal{T}$. The canonical feature map is $\varphi(t) = \kappa(t, \cdot)$, so that $\kappa(t, t') = \langle \varphi(t), \varphi(t')\rangle_{\mathcal{H}}$. A bounded kernel is characteristic if the mean embedding $P \mapsto \mathbb{E}_{T\sim P}[\varphi(T)]$ is injective on probability measures.

3 Characterizing the Independent Subspace via a Linear Operator

In this section, we characterize $\Omega_{\mathbf{Z};\mathbf{X}}$ as the null space of a fixed linear operator constructed from first- and second-order conditional moments. We then show that finitely many random kernel reference locations preserve this null space almost surely, giving a matrix representation as a foundation for subsequent GIN/TIN inference.

Let $\mathbf{Z} = (Z_1, \ldots, Z_q)^\top$. For each coordinate Z_ℓ, let $\kappa_\ell : \mathbb{R} \times \mathbb{R} \to \mathbb{R}$ be a kernel with RKHS $\mathcal{H}_\ell$. Let $\mathcal{H} = \bigoplus_{\ell=1}^q \mathcal{H}_\ell$ (here $\bigoplus$ denotes the direct sum of spaces) and define the centered feature map $\Phi(\mathbf{Z}) = (\kappa_\ell(Z_\ell, \cdot) - \mathbb{E}[\kappa_\ell(Z_\ell, \cdot)])_{\ell=1}^q$.

Assumption 3.1. For every $\ell = 1, \ldots, q$, the kernel $\kappa_\ell : \mathbb{R} \times \mathbb{R} \to \mathbb{R}$ with RKHS $\mathcal{H}_\ell$ is continuous and characteristic, and $\sup_{s,t\in\mathbb{R}} |\kappa_\ell(s,t)| < \infty$.

Definition 3.2 (Independent-subspace moment operator)**.** For $\mathbf{X}$ and $\mathbf{Z}$ with finite second moments, define $\mathcal{L} : \mathbb{R}^p \to \mathbb{R}^q \oplus \mathcal{H}^{p+1}$ by $\mathcal{L}w = \big(\Sigma_{\mathbf{ZX}} w,\ \mathbb{E}[(w^\top \mathbf{X})\Phi(\mathbf{Z})],\ \big(\mathbb{E}[X_i (w^\top \mathbf{X})\Phi(\mathbf{Z})]\big)_{i=1}^p\big)$.

The expectations exist by kernel boundedness and finite second moments. By the reproducing property, $\mathcal{L}w = 0$ is equivalent to $\Sigma_{\mathbf{ZX}} w = 0$, $\mathbb{E}[\mathbf{X}h(\mathbf{Z})]^\top w = 0$ and $\mathbb{E}[\mathbf{X}\mathbf{X}^\top h(\mathbf{Z})]w = 0$ for every centered reference function $h(\mathbf{Z}) = f(Z_\ell) - \mathbb{E}[f(Z_\ell)]$, with $f \in \mathcal{H}_\ell$ and $\ell = 1, \ldots, q$.

Theorem 3.3. *Suppose that* $\mathbf{X}$ *and* $\mathbf{Z}$ *satisfy Def. 2.1 and that the kernels* $\{\kappa_\ell\}_{\ell=1}^q$ *satisfy Assumption 3.1. Then the operator in Def. 3.2 satisfies* $\ker(\mathcal{L}) = \Omega_{\mathbf{Z};\mathbf{X}}$.

Theorem 3.3 turns the search for independent projections into a linear equation system. Once the kernels are fixed, every coefficient of $\mathcal{L}$ is determined by $\mathbb{P}(\mathbf{X}, \mathbf{Z})$ and does not depend on w. Its null space characterizes the entire independent subspace, including its orientation and dimension.

Notably, the distinction from the independence characterization in Theorem 2.6 is in the use of second-order information. The scalar response $\mathbb{E}[(w^\top \mathbf{X})^2 h(\mathbf{Z})]$ is quadratic in w, whereas Def. 3.2 retains the complete response $\mathbb{E}[\mathbf{X}(w^\top \mathbf{X})h(\mathbf{Z})] = \mathbb{E}[\mathbf{X}\mathbf{X}^\top h(\mathbf{Z})]w$, which is linear in w. The linear non-Gaussian model makes this stronger constraint necessary. Thus the full second-order response can be used to construct independent directions while preserving exactly the target subspace.

However, this operator characterization is at the population level, and $\mathcal{L}$ may take values in an infinite-dimensional space. To use it for estimation, we need a finite representation. Since $\mathcal{L}$ acts on $\mathbb{R}^p$, its rank is at most p, suggesting that finitely many evaluations may suffice. We next establish conditions under which finitely many kernel evaluations preserve its null space.

Definition 3.4 (Finite-location independent-subspace moment matrix)**.** For each coordinate ℓ of $\mathbf{Z}$, choose K_ℓ locations $\{a_{\ell,1}, \ldots, a_{\ell,K_\ell}\}$ and enumerate the functions $z \mapsto \kappa_\ell(z_\ell, a_{\ell,s})$ as $F = \{f_1, \ldots, f_K\}$, where $K = \sum_{\ell=1}^q K_\ell$. Define

$$\mathbf{L}_F w = \Big(\Sigma_{\mathbf{ZX}} w,\ \big(\mathbb{E}[(w^\top \mathbf{X})h_j(\mathbf{Z})]\big)_{j=1}^K,\ \big(\mathbb{E}[\mathbf{X}(w^\top \mathbf{X})h_j(\mathbf{Z})]\big)_{j=1}^K\Big). \tag{1}$$

Here $h_j(\mathbf{Z}) = f_j(\mathbf{Z}) - \mathbb{E}[f_j(\mathbf{Z})]$. The blocks are stacked vertically, giving a $\{q + K(p+1)\} \times p$ matrix whose entries are observable moments. To ensure that random locations identify the entire subspace, we next impose an additional condition.

Assumption 3.5. For every $\ell = 1, \ldots, q$ and every Borel probability measure P on $\mathbb{R}$, the function $a \mapsto \int_{\mathbb{R}} \kappa_\ell(t, a) dP(t)$ is real analytic on $\mathbb{R}$.

Theorem 3.6 (Identification by finitely many random locations)**.** *Suppose that* $\mathbf{X}$ *and* $\mathbf{Z}$ *satisfy Def. 2.1 and that the kernels* $\{\kappa_\ell\}_{\ell=1}^q$ *satisfy Assumptions 3.1 and 3.5*[1]*. Let* $\mathbf{L}_F$ *be defined by Equ. (1). If for each coordinate* ℓ, $K_\ell \geq p - \operatorname{rank}(\Sigma_{\mathbf{ZX}})$*, and the joint distribution of the* K *locations has a Lebesgue density on* $\mathbb{R}^K$*, then*

$$\mathbb{P}\big(\ker(\mathbf{L}_F) = \Omega_{\mathbf{Z};\mathbf{X}}, \operatorname{rank}(\mathbf{L}_F) = p - \dim(\Omega_{\mathbf{Z};\mathbf{X}})\big) = 1. \tag{2}$$

More generally, $K_\ell \geq p - \dim(\Omega_{\mathbf{Z};\mathbf{X}}) - \operatorname{rank}(\Sigma_{\mathbf{ZX}})$ *for every* ℓ *suffices. If this bound is zero, covariance alone identifies the subspace.*

The proof extends analytic random-evaluation arguments (Chwialkowski et al., 2015; Jitkrittum et al., 2017) to preservation of a row space. Writing $k = p - \dim(\Omega_{\mathbf{Z};\mathbf{X}})$, a basis of the covariance row space can be extended to the full row space using $k - \operatorname{rank}(\Sigma_{\mathbf{ZX}})$ additional response rows. A corresponding nonzero minor is analytic in the locations and hence vanishes only on a Lebesgue-null set. Consequently, Equ. (2) holds for each fixed data-generating distribution, simultaneously for all directions. These results represent the independent subspace as the null space of a finite matrix of observable moments.

[1] For example, a Gaussian kernel with fixed positive bandwidths satisfies both Assumptions 3.1 and 3.5.

4 TESTING THE GIN CONDITION

Section 3 characterizes the independent transformation subspace $\Omega_{\mathbf{Z};\mathbf{X}}$ through the finite moment matrix $\mathbf{L}_F$. GIN asks whether every covariance-null direction belongs to $\Omega_{\mathbf{Z};\mathbf{X}}$. Existing implementations often select one such direction and test $w^\top \mathbf{X}$ against $\mathbf{Z}$ using an independence test such as HSIC (Gretton et al., 2005; Xie et al., 2024). When the covariance-null space has dimension greater than one, this leaves other directions unchecked. We instead test the entire space jointly and account for estimating it from the data. Throughout this section, the reference functions $F = \{f_1, \ldots, f_K\}$ are chosen independently of the testing sample and held fixed across bootstrap draws.

Let $r = \operatorname{rank}(\Sigma_{\mathbf{ZX}}) < p$. Let $\mathbf{V} \in \mathbb{R}^{p\times(p-r)}$ have orthonormal columns spanning $\ker(\Sigma_{\mathbf{ZX}}\Sigma_{\mathbf{XX}}^{-1/2})$, and define $\mathbf{W} = \Sigma_{\mathbf{XX}}^{-1/2}\mathbf{V}$. Then $\Sigma_{\mathbf{ZX}}\mathbf{W} = 0$ and $\mathbf{W}^\top\Sigma_{\mathbf{XX}}\mathbf{W} = \mathbf{I}_{p-r}$. Under the identification conditions of Theorem 3.6, the GIN null can therefore be written as

$$H_0^{\mathrm{GIN}} \quad\Longleftrightarrow\quad \ker(\Sigma_{\mathbf{ZX}}) = \Omega_{\mathbf{Z};\mathbf{X}} \quad\Longleftrightarrow\quad \mathbf{L}_F\mathbf{W} = 0. \tag{3}$$

Thus covariance determines the candidate directions, whereas the remaining first- and second-order rows of $\mathbf{L}_F$ determine whether every candidate direction is independent of $\mathbf{Z}$.

The empirical counterpart of Equ. (1) acts on $w \in \mathbb{R}^p$ as

$$\widehat{\mathbf{L}}_F w = \left(\widehat{\Sigma}_{\mathbf{ZX}}w, \left(\widehat{\mathrm{Cov}}\left(w^\top\mathbf{X}, \widehat{h}_j(\mathbf{Z})\right)\right)_{j=1}^K, \left(\widehat{\mathrm{Cov}}\left(\mathbf{X}\left(w^\top\mathbf{X}\right), \widehat{h}_j(\mathbf{Z})\right)\right)_{j=1}^K\right). \tag{4}$$

Every second-order response is a p-vector, so the operator retains all p corresponding rows rather than a single quadratic moment. Assume $\widehat{\Sigma}_{\mathbf{XX}} \succ 0$ and define the whitened matrix $\widehat{\mathbf{M}} = \widehat{\mathbf{L}}_F\widehat{\Sigma}_{\mathbf{XX}}^{-1/2}$. After selecting the covariance rank $\widehat{r} < p$, let $\widehat{\mathbf{V}}$ contain the $p - \widehat{r}$ right singular vectors of $\widehat{\Sigma}_{\mathbf{ZX}}\widehat{\Sigma}_{\mathbf{XX}}^{-1/2}$ associated with its smallest singular values. The corresponding directions are $\widehat{\mathbf{W}} = \widehat{\Sigma}_{\mathbf{XX}}^{-1/2}\widehat{\mathbf{V}}$ and satisfy $\widehat{\mathbf{W}}^\top\widehat{\Sigma}_{\mathbf{XX}}\widehat{\mathbf{W}} = \mathbf{I}_{p-\widehat{r}}$. Whitening therefore puts the candidate projections on a common variance scale without altering the exact covariance-null space.

We measure the aggregate violation of the restrictions in Equ. (3) by $T_0 = n\|\widehat{\mathbf{M}}\widehat{\mathbf{V}}\|_F^2$. This is the sum of squared empirical covariance, first-order, and complete second-order restrictions over the entire estimated covariance-null space. The statistic is unchanged if $\widehat{\mathbf{V}}$ is replaced by another orthonormal basis of the same space. The first q rows of $\widehat{\mathbf{M}}$ contain the whitened cross-covariance; retaining them in T_0 includes the small singular values discarded when selecting $\widehat{r}$.

Paired-Bootstrap Calibration. Because the covariance directions and moment restrictions are estimated from the same observations, their sampling fluctuations are dependent. We preserve this dependence by resampling $(\mathbf{X}^{(i)}, \mathbf{Z}^{(i)})$ pairs and refitting the directions in each draw. To calibrate under the null, we first remove the empirical moment response along the estimated candidate space.

Define the null-centered operator $\widehat{\mathbf{M}}_0 = \widehat{\mathbf{M}}\big(\mathbf{I}_p - \widehat{\mathbf{V}}\widehat{\mathbf{V}}^\top\big)$. For each bootstrap draw b, sample indices $I_1^{(b)}, \ldots, I_n^{(b)}$ independently and uniformly from $\{1, \ldots, n\}$, and resample the pairs $(\mathbf{X}^{(I_i^{(b)})}, \mathbf{Z}^{(I_i^{(b)})})$. For each draw, compute the sample means of $\mathbf{X}$, $\mathbf{Z}$, and $f_j(\mathbf{Z})$ from the resampled pairs. Use these means to recompute $\widehat{\mathbf{L}}_F^{*(b)}$ as in Equ. (4) and the covariance estimate $\widehat{\Sigma}_{\mathbf{XX}}^{*(b)}$. Write $\widehat{\mathbf{M}}^{*(b)} = \widehat{\mathbf{L}}_F^{*(b)}(\widehat{\Sigma}_{\mathbf{XX}}^{*(b)})^{-1/2}$ and set $\widehat{\mathbf{M}}^{0*(b)} = \widehat{\mathbf{M}}_0 + (\widehat{\mathbf{M}}^{*(b)} - \widehat{\mathbf{M}})$.

Resampling complete observation pairs and recomputing the moments preserves their joint sampling variation while imposing the fitted null at the bootstrap center. Holding $\widehat{r}$ fixed, obtain $\widehat{\mathbf{V}}^{*(b)}$ from the $p - \widehat{r}$ smallest right singular directions of the first q rows of $\widehat{\mathbf{M}}^{0*(b)}$, and calculate $T_b = n\|\widehat{\mathbf{M}}^{0*(b)}\widehat{\mathbf{V}}^{*(b)}\|_{\mathrm{F}}^2$. Recomputing the directions from these rows repeats the direction-selection step used for T_0. The Monte Carlo p-value is computed by

$$\widehat{p} = \frac{1 + \sum_{b=1}^{n_{\mathrm{boot}}} \mathbf{1}\{T_b \geq T_0\}}{n_{\mathrm{boot}} + 1}. \tag{5}$$

We reject the GIN condition when $\widehat{p} \leq \alpha$.

Proposition 4.1 (Bootstrap approximation with refitted directions)**.** *Suppose that the observations are i.i.d., the reference functions are bounded and chosen independently of the sample,* $\mathbb{E}\|\mathbf{X}\|^4 + \mathbb{E}\|\mathbf{Z}\|^4 <$

∞, $\Sigma_{\mathbf{XX}} \succ 0$. *At a fixed GIN null with* $\operatorname{rank}(\Sigma_{\mathbf{ZX}}) = r < p$, *assume that the nonzero singular values of* $\Sigma_{\mathbf{ZX}}\Sigma_{\mathbf{XX}}^{-1/2}$ *are separated from zero and that* $\Pr(\widehat{r} = r) \to 1$. *If the limiting distribution of* T_0 *is continuous, then, conditionally on the observations,*

$$\sup_t \left| \overset{*}{\Pr}(T_b \le t) - \Pr(T_0 \le t) \right| \xrightarrow{p} 0, \tag{6}$$

where $\Pr^*(T_b \le t)$ *is the probability over the bootstrap resampling indices.*

Thus bootstrap quantiles approximate the null quantiles of T_0 while accounting for estimation of the covariance-null space. The paired resampling preserves the joint fluctuation of $\widehat{\mathbf{L}}_F$ and $\widehat{\Sigma}_{\mathbf{XX}}$. Recomputing candidate directions from the first q rows of each null-centered bootstrap matrix accounts for their estimation variability.

Theorem 4.2 (Validity and consistency)**.** *Fix* n_{boot} *such that* $1/(n_{\text{boot}} + 1) \le \alpha$. *Under the conditions of Proposition 4.1,*

$$\limsup_{n\to\infty} \Pr(\widehat{p} \le \alpha) \le \frac{\lfloor \alpha(n_{\text{boot}} + 1) \rfloor}{n_{\text{boot}} + 1} \le \alpha. \tag{7}$$

At a fixed alternative, suppose the reference functions are bounded and chosen independently of the sample, $\mathbb{E}\|\mathbf{X}\|^4 + \mathbb{E}\|\mathbf{Z}\|^4 < \infty$, *and* $\Sigma_{\mathbf{XX}} \succ 0$. *Let* $r = \operatorname{rank}(\Sigma_{\mathbf{ZX}}) < p$ *and suppose* $\Pr(\widehat{r} = r) \to 1$. *If* $\mathbf{L}_F\mathbf{W} \neq 0$, *with* $\mathbf{W}$ *defined above, then* $\Pr(\widehat{p} \le \alpha) \to 1$.

The first conclusion bounds the probability of rejecting a true GIN condition. For the fixed alternatives in the second conclusion, $T_0/n \to \|\mathbf{L}_F\mathbf{W}\|_{\text{F}}^2 > 0$, whereas the null-recentered bootstrap statistics remain bounded in probability. These guarantees are pointwise in the data-generating distribution: they do not imply uniform validity as a nonzero covariance singular value approaches zero or $\Sigma_{\mathbf{XX}}$ approaches singularity. Under Theorem 3.6, almost every identifying draw of the feature locations detects a fixed violation of GIN. Algorithm 1 summarizes the procedure.

Algorithm 1 GIN testing

Input: Data $\{(\mathbf{X}^{(i)}, \mathbf{Z}^{(i)})\}_{i=1}^n$, reference functions F, significance level α, number of bootstrap draws n_{boot}.
Output: Estimated rank $\widehat{r}$, statistic T_0, p-value $\widehat{p}$, and GIN decision.

1: Center $\mathbf{X}$, $\mathbf{Z}$, and the values of each reference function empirically. Construct $\widehat{\Sigma}_{\mathbf{XX}}$, $\widehat{\Sigma}_{\mathbf{ZX}}$, and $\widehat{\mathbf{L}}_F$ as above.
2: Estimate $\widehat{r}$ from $\widehat{\Sigma}_{\mathbf{ZX}}$. If $\widehat{r} = p$, reject GIN as no non-zero candidate direction exists.
3: Set $\widehat{\mathbf{M}} = \widehat{\mathbf{L}}_F\widehat{\Sigma}_{\mathbf{XX}}^{-1/2}$ and obtain $\widehat{\mathbf{V}}$ from the $p - \widehat{r}$ smallest right singular directions of $\widehat{\Sigma}_{\mathbf{ZX}}\widehat{\Sigma}_{\mathbf{XX}}^{-1/2}$.
4: Compute $T_0 = n\|\widehat{\mathbf{M}}\widehat{\mathbf{V}}\|_F^2$ and $\widehat{\mathbf{M}}_0 = \widehat{\mathbf{M}}(\mathbf{I}_p - \widehat{\mathbf{V}}\widehat{\mathbf{V}}^\top)$.
5: For each bootstrap draw b, resample observation pairs. Then, construct $\widehat{\mathbf{M}}^{0*(b)}$ and recompute $\widehat{\mathbf{V}}^{*(b)}$ from its first q rows with rank $\widehat{r}$. Finally, compute $T_b = n\|\widehat{\mathbf{M}}^{0*(b)}\widehat{\mathbf{V}}^{*(b)}\|_{\text{F}}^2$.
6: Compute $\widehat{p}$ by Equ. (5). Reject GIN when $\widehat{p} \le \alpha$; otherwise do not reject.

5 Testing the TIN Condition

Different from the GIN condition that tests a covariance-defined candidate space, TIN condition requires the dimension of the entire independent space. If F identifies the independent subspace, Theorem 3.6 gives

$$k_F := \operatorname{rank}(\mathbf{L}_F) = \operatorname{TIN}(\mathbf{Z}, \mathbf{X}), \qquad \dim(\Omega_{\mathbf{Z};\mathbf{X}}) = p - k_F. \tag{8}$$

As in Section 4, we choose F independently of the testing sample and hold it fixed during calibration. We estimate k_F from the full empirical matrix $\widehat{\mathbf{L}}_F$ in Equ. (4), retaining every row of each second-order moment matrix. We scale the rows associated with the reference functions by $1/\sqrt{K}$ so that adding locations does not inflate their aggregate weight:

$$\widehat{\mathbf{L}}_F^{\text{sc}} = \mathbf{D}_F\widehat{\mathbf{L}}_F, \qquad \mathbf{D}_F = \operatorname{diag}\Big(I_q, 1/\sqrt{K}I_K, 1/\sqrt{K}I_{Kp}\Big). \tag{9}$$

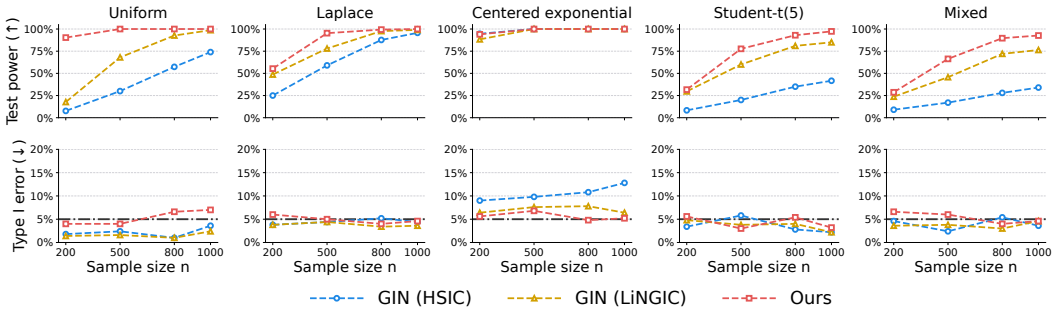

Figure 1: GIN testing in Setting I. Columns show five source distributions. The top row gives power at $\theta = 0.8$, and the bottom row gives Type I error under $\theta = 0$. The curves compare GIN (HSIC), GIN (LiNGIC), and our test across sample sizes at $\alpha = 0.05$.

Since $\mathbf{D}_F$ is invertible, this scaling preserves the rank. For a candidate rank $r < p$, we test $H_{0,r} : \mathrm{rank}(\mathbf{L}_F) = r$. Let $\widehat{\sigma}_j$ be the ordered singular values of $\widehat{\mathbf{L}}_F^{\mathrm{sc}}$ and $\widehat{\mathbf{W}}_r$ its $p - r$ trailing right singular vectors. We use two test statistics:

$$T_{0,r}^{\mathrm{full}} = n \sum_{j>r} \widehat{\sigma}_j^2, \qquad T_{0,r}^{\mathrm{quad}} = \frac{np}{p-r} \sum_{j=1}^{K} \left\| \widehat{\mathbf{W}}_r^\top \frac{\widehat{\mathbf{B}}_j}{\sqrt{K}} \widehat{\mathbf{W}}_r \right\|_{\mathrm{F}}^2. \tag{10}$$

Here $\widehat{\mathbf{B}}_j = \frac{1}{n} \sum_i \widehat{h}_j(\mathbf{Z}^{(i)}) \widehat{x}_i \widehat{x}_i^\top$ is the jth second-order moment matrix, with $\widehat{x}_i = \mathbf{X}^{(i)} - \overline{\mathbf{X}}$. The full statistic tests the rank of all constraints; the quadratic statistic adds sensitivity to variation concentrated in the fitted candidate space.

Joint calibration. The fitted singular subspace and the residuals fluctuate together. For each of N multiplier draws, apply one Rademacher sign per observation to the empirically centered influence matrices, including the effect of estimating the means. This yields $\mathbf{G}_b = n^{-1/2} \sum_i \xi_i^{(b)} \widehat{\Psi}_i$ on the scale of $\widehat{\mathbf{L}}_F^{\mathrm{sc}}$. Let $\widehat{\mathbf{U}}_r$ contain the r leading left singular vectors of $\widehat{\mathbf{L}}_F^{\mathrm{sc}}$. Using the same draw for both statistics, compute

$$T_{b,r}^{\mathrm{full}} = \left\| (I - \widehat{\mathbf{U}}_r \widehat{\mathbf{U}}_r^\top) \mathbf{G}_b \widehat{\mathbf{W}}_r \right\|_{\mathrm{F}}^2, \qquad T_{b,r}^{\mathrm{quad}} = \frac{p}{p-r} \sum_{j=1}^{K} \left\| \widehat{\mathbf{W}}_r^\top (\mathbf{G}_b)_{B_j} \widehat{\mathbf{W}}_r \right\|_{\mathrm{F}}^2. \tag{11}$$

The left projection accounts for fitting the rank-r component. Combine the two statistics by their matched upper-tail ranks, testing candidate ranks increasingly until the first joint p-value exceeds α. Report this rank as $\widehat{k}_F$ and $p - \widehat{k}_F$ as the independent-space dimension. Set $\widehat{k}_F = p$ if every candidate is rejected. With fixed reference functions that identify the subspace, a separated positive spectrum, and $2/(N+1) \le \alpha$, this procedure rejects every fixed rank below k_F with probability tending to one and controls rejection at the true rank pointwise. The centering correction, joint rank rule, and precise guarantees are given in Appendix D.

6 Experiments

We evaluate the proposed inference procedures both as statistical tests and as components of latent-variable causal discovery. The main text first examines GIN testing in synthetic settings that vary the number of covariance-null directions and reference variables, and then evaluates its effect within the latent-variable causal-discovery pipeline on simulated latent-variable models. Experiments on TIN dimension inference, component ablations, and a real-data application are reported in the Appendix.

6.1 GIN tests

Data generating process. We consider three data-generating settings based on independent, standardized non-Gaussian sources E_{j1}, E_{j2} and private sources V_k. Let $Z_j = (E_{j1} + E_{j2})/\sqrt{2}$ and

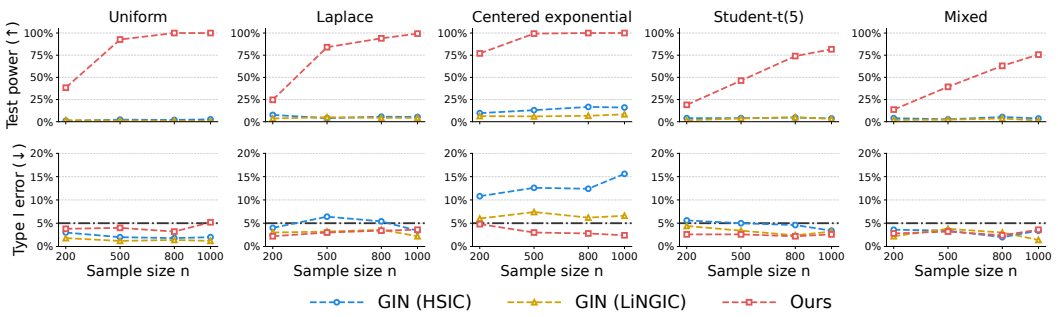

Figure 2: GIN testing in Setting II. Columns show five source distributions. The top row gives power at $\theta = 0.8$, and the bottom row gives Type I error under $\theta = 0$. The curves compare GIN (HSIC), GIN (LiNGIC), and our test across sample sizes at $\alpha = 0.05$.

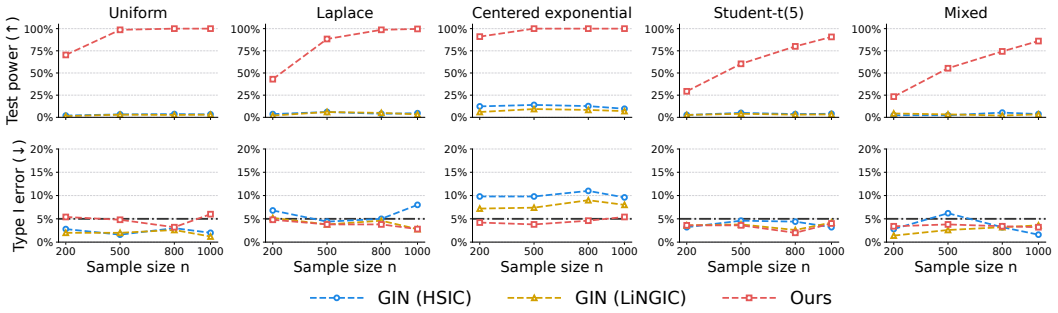

Figure 3: GIN testing in Setting III. Columns show five source distributions. The top row gives power at $\theta = 0.8$, and the bottom row gives Type I error under $\theta = 0$. The curves compare GIN (HSIC), GIN (LiNGIC), and our test across sample sizes at $\alpha = 0.05$.

$L_j = (E_{j1} - E_{j2})/\sqrt{2}$. Setting I uses $(p, q) = (2, 1)$, $X_1 = V_1 + Z_1 + L_1$, and $X_2 = V_2 + \theta L_1$. Setting II uses $(p, q) = (4, 2)$: $X_j = V_j + Z_j + L_j$ for $j = 1, 2$, $X_3 = V_3 + \theta L_1$, and $X_4 = V_4$. Setting III extends Setting I with $X_3 = V_3$ and $X_4 = V_4$, giving $(p, q) = (4, 1)$ and a three-dimensional covariance-null space. In all settings, $\theta = 0$ defines the null and $\theta = 0.8$ defines the alternatives. Since L_1 and Z_1 are uncorrelated but dependent under the specified non-Gaussian distributions, the perturbation preserves covariance rank while introducing dependence in one candidate direction. Settings II and III additionally test whether a procedure examines the full covariance-null space rather than a single selected direction. Each setting is evaluated at $n \in \{200, 500, 800, 1000\}$. The four source distributions are $\mathrm{Unif}(-\sqrt{3}, \sqrt{3})$, $\mathrm{Laplace}(0, 1/\sqrt{2})$, $\mathrm{Exp}(1) - 1$, and $t_5/\sqrt{5/3}$. In the mixed case, each noise draw is drawn independently and uniformly from these four distributions.

Compared methods and evaluation. We compare our GIN test with two baselines: the original HSIC-based GIN test and a scalar LiNGIC replacement. For multivariate $\mathbf{Z}$, the baseline coordinate-wise p-values are combined using Fisher's method (Fisher, 1930). Type I error is assessed on 500 separate null data sets and power on 300 alternative data sets. The inner bootstrap uses 199 resamples.

Performance analysis. At $n = 800$, our mean power across five source distributions is 92.1% in Setting III and 89.5% in Setting II, compared with 5.2% and 5.7% for GIN (HSIC), and 4.9% and 5.9% for GIN (LiNGIC), respectively. This gap is consistent with a single selected covariance-null direction missing the injected contrast: Fisher combination across $\mathbf{Z}$ coordinates cannot recover omitted X directions. From $n = 200$ to $n = 1000$, the power of ours rises from 53.0% to 95.9% in Setting III and from 41.3% to 93.7% in Setting II. The larger advantage in Setting III cannot be attributed to candidate-space dimension alone, since Setting II also changes the number of Z coordinates. In Setting I, where all methods use one candidate direction, the gap narrows: at $n = 800$, mean power is 96.5% for ours, 66.4% for GIN (HSIC), and 90.7% for GIN (LiNGIC). The mixed distribution remains challenging (89.7%, 35.7%, and 76.0%, respectively), whereas all three methods

reach 100% with centered-exponential sources by $n = 500$. Across the 15 setting–distribution combinations at $n = 800$, the matched-protocol mean Type I error rates are 4.6%, 4.6%, and 4.8%, respectively, although individual estimates range from 3.0% to 8.0%.

6.2 LATENT-VARIABLE CAUSAL DISCOVERY

Data generating process. We simulated linear structural equation models with four latent variables and 12 observed variables (three pure indicators per latent). The latent DAG had one of six topologies: chain, fork, collider, diamond, dense DAG, or two disconnected components. For a DAG-compatible coefficient matrix $\mathbf{B}$, we generated $\mathbf{H} = \mathbf{D}^{-1}(\mathbf{I} - \mathbf{B})^{-1}\xi$ and $X_{jk} = \lambda_{jk}H_j + 0.5|\lambda_{jk}|\eta_{jk}$ for $j \in \{1, \ldots, 4\}$ and $k \in \{1, 2, 3\}$. Here $\mathbf{D}$ standardizes latent variances, and ξ and η are mutually independent, centered, unit-variance sources. Each λ_{jk} has a random sign and magnitude drawn from $\mathrm{Unif}(0.8, 1.2)$. Each nonzero coefficient in $\mathbf{B}$ independently receives a random sign and magnitude $c\ \mathrm{Unif}(0.8, 1.2)$, with $c \in \{0.3, 0.7, 1.2\}$ denoting weak, medium, and strong connections. The measurement-noise-to-signal standard-deviation ratio is 0.5. We used standardized Uniform, Laplace, centered exponential, Logistic, and near-Gaussian source distributions ($\sqrt{0.9}G + \sqrt{0.1}U$ for independent standard Gaussian G and standardized Uniform U). In the heterogeneous setting, the five distributions were assigned to the 16 exogenous sources once per parameterized graph and held fixed across observations. We randomly permuted the observed variables to hide their latent grouping. Each topology had 14 independent parameterizations, evaluated on the same datasets across methods. We used $n \in \{200, 500, 1000\}$ at $c = 0.7$ and $c \in \{0.3, 1.2\}$ at $n = 500$.

Compared methods and metrics. We used the latent-variable causal discovery routine (Xie et al., 2022) and CCA rank queries for all methods, varying the GIN implementation between our test, original GIN (HSIC) with Fisher combination, and a LiNGIC replacement with the same combination. For scoring, inferred latent variables were aligned to the reference groups by indicator overlap. The bounded edit loss is $E/(E + T)$, where E counts latent-node, measurement-edge, and latent-edge edits and T counts true nodes and edges. Latent-edge F_1 scores directed edges, and clustering F_1 scores pairs of observed variables assigned to the same latent group.

Method	Edit loss ↓	Latent-edge F_1 ↑	Clustering F_1 ↑
Ours	**0.2088**	**0.457**	**0.859**
Original GIN (HSIC)	0.2659	0.355	0.782
Original GIN (LiNGIC)	0.2596	0.374	0.786

Table 1: End-to-end latent causal-discovery performance with original GIN baselines at $\alpha = 0.05$.

Performance analysis. Table 1 shows that our GIN implementation improves all three reported metrics on the synthetic datasets. Mean edit loss is 0.2088, compared with 0.2659 for GIN (HSIC) and 0.2596 for GIN (LiNGIC). The gains extend to both parts of the recovered structure: latent-edge F_1 increases to 0.457 (from 0.355 and 0.374), and clustering F_1 to 0.859 (from 0.782 and 0.786). The lower edit loss and higher F_1 scores suggest that improvements in GIN testing carry through to latent grouping and edge recovery.

7 CONCLUSION

We studied statistical inference of the independent transformation subspace under linear non-Gaussian models. Our main finding is that non-Gaussianity can be further exploited to improve two widely used tools for latent-variable causal discovery, the generalized independent noise (GIN) condition and transformed independent noise (TIN) condition, using only first- and second-order conditional moments. We develop a low-order conditional-moment framework for independent subspace inference, together with calibrated procedures for GIN and TIN that account for subspace estimation uncertainty and enjoy theoretical guarantees. Experiments further demonstrated improved statistical efficiency and downstream causal discovery performance. Several directions remain open, including applying our moment operator to other questions about the independent transformation subspace and developing adaptive weighting and feature-location selection with theoretical guarantees. We hope our results provide a useful foundation for these developments and, more broadly, for statistically principled causal discovery under parametric models.

AI USE STATEMENT

We used generative AI tools to assist with editing the manuscript for grammar and readability, writing and modifying parts of the experimental code, and with parts of the theoretical derivations. All AI-assisted text, code, and derivations were carefully reviewed, tested, and verified by the authors, who take full responsibility for the content of this paper.

REPRODUCIBILITY STATEMENT

Anonymized source code is provided in the supplementary material. Implementation details and evaluation metrics in different experiments are all described in the paper, respectively.

REFERENCES

Jeffrey Adams, Niels Hansen, and Kun Zhang. Identification of partially observed linear causal models: Graphical conditions for the non-gaussian and heterogeneous cases. *Advances in Neural Information Processing Systems*, 34:22822–22833, 2021.

Melisande Albert, Beatrice Laurent, Amandine Marrel, and Anouar Meynaoui. Adaptive test of independence based on HSIC measures. *Annals of Statistics*, 50(2):858–879, 2022.

Gilles Blanchard, Motoaki Kawanabe, Masashi Sugiyama, Vladimir Spokoiny, and Klaus-Robert Muller. In Search of Non-Gaussian Components of a High-Dimensional Distribution. *Journal of Machine Learning Research*, 7:247–282, 2006.

Ruichu Cai, Feng Xie, Clark Glymour, Zhifeng Hao, and Kun Zhang. Triad constraints for learning causal structure of latent variables. *Advances in Neural Information Processing Systems*, 32: 12883–12892, 2019.

Ruichu Cai, Zhiyi Huang, Wei Chen, Zhifeng Hao, and Kun Zhang. Causal discovery with latent confounders based on higher-order cumulants. In *International conference on machine learning*, pp. 3380–3407. PMLR, 2023.

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

Aiyou Chen and Peter J. Bickel. Efficient independent component analysis. *Annals of Statistics*, 34 (6):2825–2855, 2006.

Qihui Chen and Zheng Fang. Improved inference on the rank of a matrix. *Quantitative Economics*, 10(4):1787–1824, 2019.

Kacper Chwialkowski, Aaditya Ramdas, Dino Sejdinovic, and Arthur Gretton. Fast two-sample testing with analytic representations of probability measures. In *Advances in Neural Information Processing Systems*, volume 28, 2015.

R. Dennis Cook and Sanford Weisberg. Sliced Inverse Regression for Dimension Reduction: Comment. *Journal of the American Statistical Association*, 86(414):328–332, 1991.

Haoyue Dai, Peter Spirtes, and Kun Zhang. Independence testing-based approach to causal discovery under measurement error and linear non-gaussian models. *Advances in Neural Information Processing Systems*, 35:27524–27536, 2022.

Haoyue Dai, Immanuel Albrecht, Peter Spirtes, and Kun Zhang. Distributional equivalence in linear non-gaussian latent-variable cyclic causal models: Characterization and learning. In *The Fourteenth International Conference on Learning Representations*, 2026.

Xinshuai Dong, Biwei Huang, Ignavier Ng, Xiangchen Song, Yujia Zheng, Songyao Jin, Roberto Legaspi, Peter Spirtes, and Kun Zhang. A versatile causal discovery framework to allow causally-related hidden variables. In *The Twelfth International Conference on Learning Representations*, 2023.

Ronald Aylmer Fisher. *Statistical methods for research workers*. Number 5. Oliver and Boyd, 1930.

Arthur Gretton, Olivier Bousquet, Alex Smola, and Bernhard Schölkopf. Measuring statistical dependence with hilbert-schmidt norms. In *International conference on algorithmic learning theory*, pp. 63–77. Springer, 2005.

Arthur Gretton, Kenji Fukumizu, Choon Teo, Le Song, Bernhard Schölkopf, and Alex Smola. A kernel statistical test of independence. *Advances in Neural Information Processing Systems*, 20: 585–592, 2007.

Lars Peter Hansen. Large Sample Properties of Generalized Method of Moments Estimators. *Econometrica*, 50(4):1029–1054, 1982.

Patrik O. Hoyer, Shohei Shimizu, Antti Kerminen, and Markus Palviainen. Estimation of causal effects using linear non-Gaussian causal models with hidden variables. *International Journal of Approximate Reasoning*, 49(2):362–378, 2008.

Aapo Hyvarinen. Fast and Robust Fixed-Point Algorithms for Independent Component Analysis. *IEEE Transactions on Neural Networks*, 10(3):626–634, 1999. doi: 10.1109/72.761722.

Aapo Hyvarinen and Stephen M. Smith. Pairwise Likelihood Ratios for Estimation of Non-Gaussian Structural Equation Models. *Journal of Machine Learning Research*, 14(4):111–152, 2013.

Jean Jacod and Philip Protter. *Probability essentials*. Springer Science & Business Media, 2004.

Songyao Jin, Feng Xie, Guangyi Chen, Biwei Huang, Zhengming Chen, Xinshuai Dong, and Kun Zhang. Structural estimation of partially observed linear non-gaussian acyclic model: A practical approach with identifiability. In *The Twelfth International Conference on Learning Representations*, 2023.

Wittawat Jitkrittum, Zoltán Szabó, and Arthur Gretton. An adaptive test of independence with analytic kernel embeddings. In *Proceedings of the 34th International Conference on Machine Learning*, volume 70 of *Proceedings of Machine Learning Research*, pp. 1742–1751, 2017.

Erich Kummerfeld and Joseph Ramsey. Causal clustering for 1-factor measurement models. In *Proceedings of the 22nd ACM SIGKDD international conference on knowledge discovery and data mining*, pp. 1655–1664, 2016.

Bing Li and Shaoli Wang. On Directional Regression for Dimension Reduction. *Journal of the American Statistical Association*, 102(479):997–1008, 2007.

Ker-Chau Li. Sliced Inverse Regression for Dimension Reduction. *Journal of the American Statistical Association*, 86(414):316–327, 1991.

Xiu-Chuan Li, Kun Zhang, and Tongliang Liu. Causal structure recovery with latent variables under milder distributional and graphical assumptions. In *The Twelfth International Conference on Learning Representations*, 2024.

Yiqing Li, Xiaofei Wang, Boyang Sun, Yewei Xia, and Kun Zhang. Independence test for linear non-gaussian data and applications in causal discovery. In *The Fourteenth International Conference on Learning Representations*, 2026.

Krikamol Muandet, Wittawat Jitkrittum, and Jonas Kubler. Kernel Conditional Moment Test via Maximum Moment Restriction. In *Proceedings of the Conference on Uncertainty in Artificial Intelligence*, volume 124 of *Proceedings of Machine Learning Research*, pp. 41–50, 2020.

Ignavier Ng, Xinshuai Dong, Haoyue Dai, Biwei Huang, Peter Spirtes, and Kun Zhang. Score-based causal discovery of latent variable causal models. In *International Conference on Machine Learning*, pp. 37531–37552. PMLR, 2024.

Sunmin Oh, Seungsu Han, and Gunwoong Park. Optimal estimation of linear non-gaussian structural equation models. *Proceedings of Machine Learning Research*, 258:748–756, 2025.

Judea Pearl. *Causality*. Cambridge University Press, 2 edition, 2009.

Jean-Marc Robin and Richard J. Smith. Tests of rank. *Econometric Theory*, 16(2):151–175, 2000.

Antonin Schrab, Ilmun Kim, Benjamin Guedj, and Arthur Gretton. Efficient Aggregated Kernel Tests using Incomplete U-statistics. In *Advances in Neural Information Processing Systems*, volume 35, 2022.

Arnab Sen and Bodhisattva Sen. Testing independence and goodness-of-fit in linear models. *Biometrika*, 101(4):927–942, 2014.

Shohei Shimizu. Lingam: Non-gaussian methods for estimating causal structures. *Behaviormetrika*, 41(1):65–98, 2014.

Shohei Shimizu, Patrik O Hoyer, Aapo Hyvärinen, Antti Kerminen, and Michael Jordan. A linear non-gaussian acyclic model for causal discovery. *Journal of Machine Learning Research*, 7(10): 2003–2030, 2006.

Shohei Shimizu, Takanori Inazumi, Yasuhiro Sogawa, Aapo Hyvärinen, Yoshinobu Kawahara, Takashi Washio, Patrik O Hoyer, Kenneth Bollen, and Patrik Hoyer. Directlingam: A direct method for learning a linear non-gaussian structural equation model. *Journal of Machine Learning Research*, 12(Apr):1225–1248, 2011.

Ricardo Silva, Richard Scheines, Clark Glymour, and Peter Spirtes. Learning the Structure of Linear Latent Variable Models. *Journal of Machine Learning Research*, 7(8):191–246, 2006.

Alexander Sokol, Marloes H. Maathuis, and Benjamin Falkeborg. Quantifying identifiability in independent component analysis. *Electronic Journal of Statistics*, 8(1):1438 – 1459, 2014.

Peter Spirtes and Kun Zhang. Causal discovery and inference: concepts and recent methodological advances. In *Applied informatics*, volume 3, pp. 1–28. SpringerOpen, 2016.

Peter Spirtes, Christopher Meek, and Thomas Richardson. Causal inference in the presence of latent variables and selection bias. In *Proceedings of the Eleventh conference on Uncertainty in artificial intelligence*, pp. 499–506, 1995.

Peter Spirtes, Clark N Glymour, Richard Scheines, and David Heckerman. *Causation, prediction, and search*. MIT press, 2000.

Peter Spirtes, Clark Glymour, Richard Scheines, and Robert Tillman. Automated search for causal relations: Theory and practice. *Department of Philosophy technical report*, 2010.

Seth Sullivant, Kelli Talaska, and Jan Draisma. Trek separation for gaussian graphical models. *Annals of statistics*, 38(3):1665–1685, 2010.

Gabor J. Szekely, Maria L. Rizzo, and Nail K. Bakirov. Measuring and testing dependence by correlation of distances. *Annals of Statistics*, 35(6):2769–2794, 2007.

Y. Samuel Wang and Mathias Drton. Causal Discovery with Unobserved Confounding and Non-Gaussian Data. *Journal of Machine Learning Research*, 24(271):1–61, 2023.

Yewei Xia, Zhengming Chen, Haoyue Dai, Fuhong Wang, Yixin Ren, Yiqing Li, Kun Zhang, and Shuigeng Zhou. Conditional independent component analysis for estimating causal structure with latent variables. In *The Fourteenth International Conference on Learning Representations*, 2026.

Feng Xie, Ruichu Cai, Biwei Huang, Clark Glymour, Zhifeng Hao, and Kun Zhang. Generalized independent noise condition for estimating latent variable causal graphs. *Advances in Neural Information Processing Systems*, 33:14891–14902, 2020.

Feng Xie, Biwei Huang, Zhengming Chen, Yangbo He, Zhi Geng, and Kun Zhang. Identification of linear non-gaussian latent hierarchical structure. In *International Conference on Machine Learning*, pp. 24370–24387. PMLR, 2022.

Feng Xie, Biwei Huang, Zhengming Chen, Ruichu Cai, Clark Glymour, Zhi Geng, and Kun Zhang. Generalized independent noise condition for estimating causal structure with latent variables. *Journal of Machine Learning Research*, 25:1–61, 2024.

APPENDIX OVERVIEW

- Appendix A proves the independent-subspace characterization and finite-location identification results.
- Appendix B proves the GIN bootstrap approximation, level control, and consistency results.
- Appendix C presents pseudocode for the original GIN test and four original TIN estimators.
- Appendix D details the moment estimates, multiplier calibration, and rank decision used for TIN inference.
- Appendix E reports additional GIN, TIN, and latent-variable causal discovery results.
- Appendix F examines the contributions of the GIN test components.
- Appendix G reports the real-data application.
- Appendix H discusses related work.

A PROOFS OF THE INDEPENDENCE AND OPERATOR RESULTS

Proof of Theorem 3.3. We first record the source geometry behind the result. Put $J = \{j : \mathbf{N}_{:j} \neq 0\}$. If $w^\top \mathbf{X} \perp\!\!\!\perp \mathbf{Z}$, it is independent of each coordinate Z_ℓ. For every $j \in J$, some Z_ℓ contains E_j with a nonzero coefficient. The Darmois–Skitovich theorem and the non-Gaussianity of E_j then force $w^\top \mathbf{M}_{:j} = 0$. Conversely, these equalities leave $w^\top \mathbf{X}$ with sources absent from $\mathbf{Z}$. Thus

$$\Omega_{\mathbf{Z};\mathbf{X}} = \{w : w^\top \mathbf{M}_{:j} = 0 \text{ for every } j \in J\}. \tag{12}$$

Suppose that $\mathcal{L}w = 0$ and put $Y = w^\top \mathbf{X}$. For each ℓ and $f \in \mathcal{H}_\ell$, set $g(Z_\ell) = f(Z_\ell) - \mathbb{E}[f(Z_\ell)]$. The reproducing property gives $\mathbb{E}[Y g(Z_\ell)] = \mathbb{E}[Y^2 g(Z_\ell)] = 0$: the second equality follows by multiplying $\mathbb{E}[\mathbf{X}Y\Phi(\mathbf{Z})] = 0$ by $w^\top$. For $H = Y$ or Y^2, define the finite signed measure

$$\nu_{H,\ell}(A) = \mathbb{E}[(H - \mathbb{E}H)\mathbf{1}\{Z_\ell \in A\}].$$

Its kernel mean embedding is zero. This measure has total mass zero. Writing its positive and negative parts as equally weighted probability measures shows that the characteristic property of κ_ℓ forces $\nu_{H,\ell} = 0$. Hence $\mathbb{E}[Y|Z_\ell] = 0$ and $\mathbb{E}[Y^2|Z_\ell] = \mathbb{E}[Y^2]$ almost surely. Theorem 2.6 gives $Y \perp\!\!\!\perp Z_\ell$ for every ℓ. Applying the source argument in Equ. (12) to each coordinate eliminates every source in J, and therefore $Y \perp\!\!\!\perp \mathbf{Z}$.

For the reverse inclusion, let $w \in \Omega_{\mathbf{Z};\mathbf{X}}$ and split $\mathbf{X} = \mathbf{X}_J + \mathbf{X}_{J^c}$ according to the source indices. Equ. (12) gives $Y = w^\top \mathbf{X}_{J^c}$, while $(\mathbf{X}_{J^c}, Y) \perp\!\!\!\perp (\mathbf{X}_J, \mathbf{Z})$. Because $\mathbb{E}Y = 0$, conditioning on $\mathbf{Z}$ yields $\mathbb{E}[Y|\mathbf{Z}] = 0$, $\mathbb{E}[\mathbf{X}_J Y|\mathbf{Z}] = 0$, and $\mathbb{E}[\mathbf{X}_{J^c} Y|\mathbf{Z}] = \mathbb{E}[\mathbf{X}_{J^c} Y]$. Every centered feature response in $\mathcal{L}w$ therefore vanishes. Independence also gives $\Sigma_{\mathbf{Z}\mathbf{X}} w = 0$. Thus $w \in \ker(\mathcal{L})$, completing both inclusions. □

Proof of Theorem 3.6. Fix the distribution of $(\mathbf{X}, \mathbf{Z})$, and write $r_0 = \operatorname{rank}(\Sigma_{\mathbf{Z}\mathbf{X}})$ and $k = p - \dim(\Omega_{\mathbf{Z};\mathbf{X}})$. By Theorem 3.3, the row space of $\mathcal{L}$ has dimension k. Its covariance rows span an r_0-dimensional subspace. The remaining rows can be obtained from kernel evaluations: if a vector annihilates the covariance rows and every evaluation row at every location, it annihilates each RKHS component of $\mathcal{L}$, and hence belongs to $\ker(\mathcal{L})$. Since the domain is finite-dimensional, the evaluation rows therefore span the same row space as the corresponding components of $\mathcal{L}$.

Choose r_0 independent covariance rows and extend them with $d = k - r_0$ evaluation rows to a basis of the row space. Each selected evaluation row specifies a coordinate ℓ, one response entry and one location. There are at most d such rows for any coordinate, so $K_\ell \geq d$ permits us to assign them to distinct location slots. At this particular location configuration, some $k \times k$ minor of $\mathbf{L}_F$ is nonzero. The same minor, with its row and column indices fixed, is a function of all K location variables.

Every entry in an evaluation row is an integral of $\kappa_\ell(Z_\ell, a)$ against a finite signed measure: its weight is an entry of $\mathbf{X}$, an entry of $\mathbf{X}\mathbf{X}^\top$, or the centered version of one of these. Such a measure is a linear combination of finite positive measures. After normalization, Assumption 3.5 makes each integral real analytic in a. The selected minor is therefore real analytic on $\mathbb{R}^K$ and is not identically zero. Its

zero set has Lebesgue measure zero, so the joint density of the locations makes this minor nonzero almost surely. Every row of $\mathbf{L}_F$ annihilates $\Omega_{\mathbf{Z};\mathbf{X}}$, while the nonzero minor gives $\operatorname{rank}(\mathbf{L}_F) \geq k$. Rank-nullity now yields $\operatorname{rank}(\mathbf{L}_F) = k$ and $\ker(\mathbf{L}_F) = \Omega_{\mathbf{Z};\mathbf{X}}$ on the same probability-one event. Finally, $d = k - r_0 \leq p - r_0$, which proves both location-count bounds in the statement. If $d = 0$, the covariance rows already give the result. □

B BOOTSTRAP

Proof of Proposition 4.1. Condition on the fixed reference functions F and work on the event $\widehat{r} = r$, whose probability tends to one. Write $\mathbf{M} = \mathbf{L}_F \Sigma_{\mathbf{XX}}^{-1/2}$ and $\mathbf{C}_s = \Sigma_{\mathbf{ZX}} \Sigma_{\mathbf{XX}}^{-1/2}$. Their empirical versions are $\widehat{\mathbf{M}}$ and $\widehat{\mathbf{C}}_s = \widehat{\Sigma}_{\mathbf{ZX}} \widehat{\Sigma}_{\mathbf{XX}}^{-1/2}$ from Section 4. Under GIN, the covariance kernel is contained in $\ker(\mathbf{L}_F)$, so $\mathbf{MV} = 0$. The positive covariance singular values stay separated from zero. After orthogonally aligning $\widehat{\mathbf{V}}$ with $\mathbf{V}$, write the aligned basis as $\widetilde{\mathbf{V}}$. Differentiating the covariance-null frame gives its transverse displacement $-\mathbf{C}_s^{+}(\widehat{\mathbf{C}}_s - \mathbf{C}_s)\mathbf{V}$. An internal rotation contributes nothing because $\mathbf{MV} = 0$. Consequently,

$$\sqrt{n}\,\widehat{\mathbf{M}}\widetilde{\mathbf{V}} = \left[\sqrt{n}(\widehat{\mathbf{M}} - \mathbf{M}) - \mathbf{M}\mathbf{C}_s^{+}\sqrt{n}(\widehat{\mathbf{C}}_s - \mathbf{C}_s)\right]\mathbf{V} + o_p(1). \tag{13}$$

The first term is the fluctuation of the moments at fixed directions. The second term accounts for fitting the covariance-null space. Fluctuations of $\widehat{\Sigma}_{\mathbf{XX}}$ enter both terms through whitening. The stated fourth moments make the finite vector of centered covariance, first-order and complete second-order moment entries square-integrable. Its joint central limit theorem, followed by the smooth map $\Sigma_{\mathbf{XX}} \mapsto \Sigma_{\mathbf{XX}}^{-1/2}$, makes the right side of Equ. (13) converge to a centered Gaussian matrix G. The empirical centerings are included in this joint expansion. Orthogonal alignment does not change a Frobenius norm, and therefore $T_0 \Rightarrow \|G\|_{\mathrm{F}}^2$.

We check that the bootstrap uses the same local null geometry. Let $\widehat{\mathbf{C}}_{s,0}$ be the first q rows of $\widehat{\mathbf{M}}_0$. The singular-vector definition gives $\widehat{\mathbf{C}}_{s,0} = \widehat{\mathbf{C}}_s(I_p - \widehat{\mathbf{V}}\widehat{\mathbf{V}}^{\top})$, which has rank r and kernel spanned by $\widehat{\mathbf{V}}$ with probability tending to one. Also $\widehat{\mathbf{M}}_0\widehat{\mathbf{V}} = 0$, so $\widehat{\mathbf{M}}_0$ itself has rank r, and under GIN it converges to $\mathbf{M}$. In a bootstrap draw, $\widehat{\mathbf{M}}^{0*(b)} - \widehat{\mathbf{M}}_0 = \widehat{\mathbf{M}}^{*(b)} - \widehat{\mathbf{M}}$. Conditional on the data, resampling pairs and recomputing every centering and the whitener reproduces the joint limiting law of $\sqrt{n}(\widehat{\mathbf{M}} - \mathbf{M}, \widehat{\mathbf{C}}_s - \mathbf{C}_s)$. Align the bootstrap frame $\widehat{\mathbf{V}}^{*(b)}$ with $\widehat{\mathbf{V}}$ and denote the aligned frame by $\widetilde{\mathbf{V}}^{*(b)}$. Applying the same transverse expansion at the rank-r center gives

$$\sqrt{n}\widehat{\mathbf{M}}^{0*(b)}\widetilde{\mathbf{V}}^{*(b)} = \left[\sqrt{n}(\widehat{\mathbf{M}}^{*(b)} - \widehat{\mathbf{M}}) - \widehat{\mathbf{M}}_0\widehat{\mathbf{C}}_{s,0}^{+}\sqrt{n}\big(\widehat{\mathbf{M}}^{*(b)} - \widehat{\mathbf{M}}\big)_{1:q,:}\right]\widehat{\mathbf{V}} + o_{p^*}(1). \tag{14}$$

Its right side converges conditionally in probability to the same Gaussian matrix as Equ. (13), because $\widehat{\mathbf{M}}_0 \to \mathbf{M}$ and $\widehat{\mathbf{C}}_{s,0}^{+} \to \mathbf{C}_s^{+}$. Hence $\sqrt{n}\widehat{\mathbf{M}}^{0*(b)}\widetilde{\mathbf{V}}^{*(b)} \Rightarrow G$ conditionally. Since orthogonal alignment leaves the Frobenius norm unchanged, T_b converges conditionally to $\|G\|_{\mathrm{F}}^2$. The assumed continuity of this limit upgrades convergence of both distribution functions to uniform convergence, and the triangle inequality gives Equ. (6). □

Proof of Theorem 4.2. Under the null, Proposition 4.1 and independent bootstrap index draws imply joint convergence of $T_0, T_1, \ldots, T_{n_{\mathrm{boot}}}$ to independent copies of $\|G\|_{\mathrm{F}}^2$. Their common distribution is continuous by assumption, so the rank of the observed statistic among these $n_{\mathrm{boot}} + 1$ values is uniform. With the conservative upper-tail count in Equ.(5), at most $\lfloor \alpha(n_{\mathrm{boot}} + 1) \rfloor$ of these ranks lead to rejection. This proves Equ. (7), including the finite-bootstrap resolution factor. The same bound holds without conditioning on the independently chosen reference functions by integrating over them.

At a fixed alternative, the law of large numbers and consistency of $\widehat{r}$ and the separated covariance-null frame give

$$\frac{T_0}{n} \xrightarrow{p} \|\mathbf{L}_F\mathbf{W}\|_{\mathrm{F}}^2 > 0. \tag{15}$$

The null-centered matrix $\widehat{\mathbf{M}}_0$ has rank r and annihilates $\widehat{\mathbf{V}}$ even under this alternative. Its first q rows retain the fixed positive covariance singular gap. The paired bootstrap perturbation $\widehat{\mathbf{M}}^{*(b)} - \widehat{\mathbf{M}}$ is

$O_{p^*}(n^{-1/2})$ in probability. The refitted trailing frame is therefore $O_{p^*}(n^{-1/2})$ from the null frame up to a rotation, and $T_b = O_{p^*}(1)$ in probability. With a fixed number of draws, T_0 exceeds all T_b with probability tending to one, so $\widehat{p} = 1/(n_{\mathrm{boot}} + 1) \le \alpha$ with probability tending to one. This proves consistency. □

C ORIGINAL GIN AND TIN IMPLEMENTATIONS

This section presents the procedures in the original GIN and TIN papers (Xie et al., 2024; Dai et al., 2022). The input is n paired observations, and $\widehat{\Sigma}_{\mathbf{ZX}}$ is their centered cross-covariance. For a scalar R, let $\widehat{p}_{\mathrm{H}}(R, Z_j)$ be the scalar HSIC independence-test p-value. The original implementations use these coordinate-wise tests to assess independence from $\mathbf{Z}$. This is the finite-sample surrogate used below. All procedures return a non-rejection or an estimated dimension, not a proof of population independence.

C.1 THE ORIGINAL GIN TEST

The population GIN condition requires a nonzero covariance-null direction and independence for every direction in $\ker(\Sigma_{\mathbf{ZX}})$. The original finite-sample TESTGIN routine selects one such direction and combines its coordinate-wise HSIC p-values by Fisher's rule (Xie et al., 2024). Algorithm 2 spells out this routine. When the empirical covariance has no exact null space, its last right singular vector supplies the approximate null direction. A covariance-rank rule can first determine whether a non-zero candidate space is admitted. Selecting one vector does not check every direction when that space has dimension greater than one. The pseudocode below follows the Fisher-combined TESTGIN in the published paper.

Algorithm 2 Original single-direction GIN test (TESTGIN)

Input: Paired data $\{(\mathbf{X}^{(i)}, \mathbf{Z}^{(i)})\}_{i=1}^{n}$, level α, and an optional covariance-rank rule.
Output: GIN non-rejection decision and combined p-value.

1: Compute $\widehat{\Sigma}_{\mathbf{ZX}}$. If a rank rule is used and returns $\widehat{r} = p$, report that no nonzero covariance-null candidate is available.
2: Select a unit vector $\widehat{\omega}$ in the estimated right null space of $\widehat{\Sigma}_{\mathbf{ZX}}$. For the original single-direction implementation, take a right singular vector associated with its smallest singular value. Set $R_i = \widehat{\omega}^{\top}\mathbf{X}^{(i)}$.
3: For $j = 1, \ldots, q$, compute $p_j = \widehat{p}_{\mathrm{H}}(R, Z_j)$.
4: Compute $S = -2\sum_{j=1}^{q} \log p_j$ and $p_{\mathrm{F}} = 1 - F_{\chi^2_{2q}}(S)$, using the Fisher's method.
5: Report GIN as non-rejected if $p_{\mathrm{F}} > \alpha$ and rejected otherwise. Return p_{F}.

C.2 THE FOUR ORIGINAL TIN ESTIMATORS

TIN asks for $p - \dim \Omega_{\mathbf{Z};\mathbf{X}}$, rather than a decision about one covariance-null direction. The TIN paper gives four approaches: subsets, constrained independent subspace analysis (ISA), stacked-cumulant rank, and two-step equation solving followed by independence testing (Dai et al., 2022). For the three approaches using HSIC, write

$$P_{\mathrm{ind}}(\omega; \mathbf{Z}) := \min_{1 \le j \le q} \widehat{p}_{\mathrm{H}}(\omega^{\top}\mathbf{X}, Z_j), \tag{16}$$

so a tested direction is counted as independent when $P_{\mathrm{ind}}(\omega; \mathbf{Z}) > \alpha$.

The two-step and subset implementations use a fixed list of coordinate-wise reference functions

$$\mathcal{F}_{\mathrm{orig}} = \{z, z^2, z^3, |z|, e^z, \log|z|, \sin z, \cos z, (1 + e^{-z})^{-1}, \tanh z\}. \tag{17}$$

For a function f, $f(\mathbf{Z})$ means componentwise application. For a selected prefix $F = (f_1, \ldots, f_m)$, form the stacked $mq \times p$ matrix

$$A_F = \left[\widehat{\mathrm{Cov}}(f_1(\mathbf{Z}), \mathbf{X})^{\top} \quad \cdots \quad \widehat{\mathrm{Cov}}(f_m(\mathbf{Z}), \mathbf{X})^{\top}\right]^{\top}. \tag{18}$$

The symbols A_F and $\mathcal{F}_{\mathrm{orig}}$ here designate the original equations using nonlinear reference functions and are separate from our complete operator $\mathbf{L}_F$ in Section 5.

Algorithm 3 Original TIN-subsets estimator

Input: Paired data $\{(\mathbf{X}^{(i)}, \mathbf{Z}^{(i)})\}_{i=1}^{n}$, level α, and the ordered list of reference functions $\mathcal{F}_{\text{orig}}$.
Output: Estimated $\text{TIN}(\mathbf{Z}, \mathbf{X})$.

1: For each nonempty coordinate subset $S \subseteq \{1, \ldots, p\}$, run Algorithm 6 on $\mathbf{X}_S$ to obtain its right-singular-vector candidates using the specified prefixes of the list of reference functions. In Equ. (16), replace $\mathbf{X}$ by $\mathbf{X}_S$ for this call. Let $b(S)$ be the largest value of $P_{\text{ind}}(\omega; \mathbf{Z})$ among these candidates. It measures whether an independent transformation exists for S.
2: For $k = 0, \ldots, p-1$, collect $\{b(S) : |S| = k+1\}$, when $k > 0$, $\{b(S) : |S| = k\}$.
3: Return the first k for which the 10th percentile of the first collection exceeds α and, when $k > 0$, the 90th percentile of the second collection is below α.
4: If no $k < p$ satisfies both criteria, return p.

TIN-subsets. At population level, Theorem 4 of (Dai et al., 2022) characterizes a positive value k by existence of an independent projection for every $(k+1)$-subset and nonexistence for some k-subset. The percentiles in Algorithm 3 are a finite-sample relaxation of these quantifiers. Appendix G.1 of (Dai et al., 2022) describes a 90th/10th-percentile heuristic for the two conditions without fixing the orientation of its independence score. For the HSIC p-value score $b(S)$ defined here, the released solver uses the explicit 10th/90th-percentile comparisons shown above (See TIN implementation). Algorithm 4 makes the imposed block mask and subsequent independence count explicit without introducing an optimization objective absent from that paper.

Algorithm 4 Original TIN-ISA estimator

Input: Paired data $\{(\mathbf{X}^{(i)}, \mathbf{Z}^{(i)})\}_{i=1}^{n}$ and level α.
Output: Estimated $\text{TIN}(\mathbf{Z}, \mathbf{X})$.

1: Fit a constrained ISA demixing transform to $(\mathbf{Z}^\top, \mathbf{X}^\top)^\top$ with block matrix $\left[\begin{smallmatrix} I_q & 0 \\ 0 & W_{\mathbf{XX}} \end{smallmatrix}\right]$. Keep the I_q and zero blocks fixed, update only the invertible $p \times p$ block $\mathbf{W}_{\mathbf{XX}}$, and optimize separation of $\mathbf{W}_{\mathbf{XX}}\mathbf{X}$ from $\mathbf{Z}$ as in the original ISA subroutine.
2: For each row $w_\ell^\top$ of the fitted $\mathbf{W}_{\mathbf{XX}}$, compute $P_{\text{ind}}(w_\ell; \mathbf{Z})$.
3: Let d be the number of rows with $P_{\text{ind}}(w_\ell; \mathbf{Z}) > \alpha$ and return $p - d$.

TIN-rank. For the cumulant estimator, define the $q \times p$ slice $C^{(m)}_{\mathbf{ZX}}[j, \ell] = \text{cum}(Z_j, \ldots, Z_j, X_\ell)$, where Z_j appears $m-1$ times, and let $K^{(m)}_{\mathbf{Z};\mathbf{X}} = [C^{(2)\top}_{\mathbf{ZX}}, \ldots, C^{(m)\top}_{\mathbf{ZX}}]^\top$. In particular, $C^{(2)}_{\mathbf{ZX}} = \Sigma_{\mathbf{ZX}}$. This K is the stacked cumulant matrix called Ψ in (Dai et al., 2022). We use K to avoid confusion with the influence contributions $\widehat{\Psi}_i$ above.

Algorithm 5 Original TIN-rank estimator

Input: Paired data $\{(\mathbf{X}^{(i)}, \mathbf{Z}^{(i)})\}_{i=1}^{n}$, maximum cumulant order $m_{\max}$, and SVD tolerance ϵ.
Output: Estimated $\text{TIN}(\mathbf{Z}, \mathbf{X})$ and whether the rank stopped before the order cap.

1: For $m = 2, \ldots, m_{\max}$, estimate $C^{(m)}_{\mathbf{ZX}}$ from the samples and vertically stack $\widehat{K}^{(m)}_{\mathbf{Z};\mathbf{X}}$.
2: Use SVD to set r_m to the number of singular values of $\widehat{K}^{(m)}$ exceeding $\epsilon \max\{(m-1)q, p\}\, s_{\max}(\widehat{K}^{(m)})$.
3: At the first $m \geq 3$ with $r_m = r_{m-1}$, return r_m. If no such order is reached by $m_{\max}$, return $r_{m_{\max}}$ and flag that the stopping rule did not fire within the chosen cap.

The original paper reports $\epsilon = 0.005$ for its numerical SVD rank. The released solver uses 0.004 and, when no adjacent ranks agree, returns the last computed rank. Its population "rank stops increasing" conclusion relies on the genericity assumption for the stacked cumulants (Dai et al., 2022). Equality of two estimated ranks at finite n is a numerical stopping rule.

TIN-2steps. The original two-step implementation tests all p right singular vectors rather than retaining only vectors below a singular-value cutoff. The first step supplies candidate directions, and

Algorithm 6 Original TIN-2steps candidate construction and testing

Input: Paired data, level α, and an ordered prefix of $\mathcal{F}_{\mathrm{orig}}$. For a scalar output, a prespecified prefix index $m_\star$.
Output: Candidate bases and their coordinate-wise HSIC p-value tables. A rank after a stated counting rule.

1: For each prefix $F_m = (f_1, \ldots, f_m)$ of the list of reference functions, form $\widehat{A}_{F_m}$ in Equ. (18) and compute a full SVD $\widehat{A}_{F_m} = U_m D_m V_m^\top$.
2: For every right singular vector $v_{m\ell}$, $\ell = 1, \ldots, p$, compute $p_{m\ell j} = \widehat{p}_{\mathrm{H}}(v_{m\ell}^\top \mathbf{X}, Z_j)$ for $j = 1, \ldots, q$. Store the full table over m, ℓ, j.
3: For a scalar output, count $d_{m_\star} = \#\{\ell : \min_j p_{m_\star \ell j} > \alpha\}$ and return $p - d_{m_\star}$. If several prefixes are compared, report the selection rule together with the result.

the second step tests their independence. The published implementation returns the candidate and p-value tables for each prefix of the list of reference functions, so any single-number rule that selects among those prefixes is an additional finite-sample convention.

D TIN Calibration and Rank Selection

This section gives the complete procedure behind Section 5. If F identifies the independent subspace, TIN is $k_F = \mathrm{rank}(\mathbf{L}_F)$, and the independent transformation subspace has dimension $p - k_F$. Unlike GIN testing, here each candidate dimension is obtained from the singular vectors of the same full operator $\widehat{\mathbf{L}}_F^{\mathrm{sc}}$ in Equ. (9). These reference functions are chosen independently of the testing sample and held fixed throughout the rank scan and multiplier calibration.

D.1 Moment construction and centering correction

Write $\widehat{x}_i = \mathbf{X}^{(i)} - \overline{\mathbf{X}}$, $\widehat{z}_i = \mathbf{Z}^{(i)} - \overline{\mathbf{Z}}$, and $\widehat{h}_{ji} = f_j(\mathbf{Z}^{(i)}) - n^{-1}\sum_{\ell=1}^n f_j(\mathbf{Z}^{(\ell)})$. Compute the empirical cross-covariance and first- and second-order moments

$$\widehat{\Sigma}_{\mathbf{ZX}} = n^{-1}\sum_i \widehat{z}_i \widehat{x}_i^\top, \qquad \widehat{a}_j = n^{-1}\sum_i \widehat{h}_{ji}\widehat{x}_i, \qquad \widehat{\mathbf{B}}_j = n^{-1}\sum_i \widehat{h}_{ji}\widehat{x}_i\widehat{x}_i^\top. \tag{19}$$

Stack $\widehat{\Sigma}_{\mathbf{ZX}}$, $\widehat{a}_j^\top$, and every row of $\widehat{\mathbf{B}}_j$ in the order of Equ. (4), then multiply by $\mathbf{D}_F$ to obtain $\widehat{\mathbf{L}}_F^{\mathrm{sc}}$. This retains the complete second-order response used to identify the independent subspace.

The multiplier draw must reproduce the fluctuation of the centered estimates, including the sample means of $f_j(\mathbf{Z})$ and $\mathbf{X}$. Let $\widehat{\Sigma}_{\mathbf{XX}} = n^{-1}\sum_i \widehat{x}_i\widehat{x}_i^\top$. The unscaled observation-level contributions are

$$\widehat{\Psi}_{i,C}^{\mathrm{raw}} = \widehat{z}_i\widehat{x}_i^\top - \widehat{\Sigma}_{\mathbf{ZX}}, \qquad \widehat{\Psi}_{i,A_j}^{\mathrm{raw}} = \widehat{h}_{ji}\widehat{x}_i^\top - \widehat{a}_j^\top, \tag{20}$$

$$\widehat{\Psi}_{i,B_j}^{\mathrm{raw}} = \widehat{h}_{ji}(\widehat{x}_i\widehat{x}_i^\top - \widehat{\Sigma}_{\mathbf{XX}}) - \widehat{\mathbf{B}}_j - \widehat{x}_i\widehat{a}_j^\top - \widehat{a}_j\widehat{x}_i^\top. \tag{21}$$

The covariance subtraction and the last two terms in Equ. (21) are the centering correction. Stack these contributions in the same order as the operator and set $\widehat{\Psi}_i = \mathbf{D}_F\widehat{\Psi}_i^{\mathrm{raw}}$. For each multiplier draw $b = 1, \ldots, N$, draw independent Rademacher signs $\xi_i^{(b)}$ and compute

$$\mathbf{G}_b = n^{-1/2}\sum_{i=1}^n \xi_i^{(b)}\widehat{\Psi}_i. \tag{22}$$

One sign multiplies all contributions from the same observation, preserving their joint fluctuation. In particular, $(\mathbf{G}_b)_{B_j} = K^{-1/2}n^{-1/2}\sum_i \xi_i^{(b)}\widehat{\Psi}_{i,B_j}^{\mathrm{raw}}$. Thus the observed $\widehat{\mathbf{B}}_j/\sqrt{K}$ in Equ. (10) and its multiplier counterpart have the same scale.

D.2 Joint calibration and rank decision

For a candidate rank $r < p$, a singular-value decomposition of $\widehat{\mathbf{L}}_F^{\mathrm{sc}}$ supplies the r leading left singular vectors $\widehat{\mathbf{U}}_r$ and the $p - r$ trailing right singular vectors $\widehat{\mathbf{W}}_r$. The latter are the candidate independent

directions. Compute the observed full-tail and quadratic statistics in Equ. (10) and, using the same $\mathbf{G}_b$ for both statistics, their multiplier versions in Equ. (11). The full statistic measures all restrictions remaining beyond rank r. The quadratic statistic measures second-order variation within the fitted candidate space. For a symmetric projected moment matrix, the implementation may use svec, with off-diagonal entries multiplied by $\sqrt{2}$, since $\|\operatorname{svec}(M)\|_2^2 = \|M\|_{\mathrm{F}}^2$.

At a true rank-r null, the left projection $I - \widehat{\mathbf{U}}_r\widehat{\mathbf{U}}_r^\top$ in the full multiplier statistic removes first-order variation absorbed by fitting the leading rank-r component. Each population B_j is symmetric and satisfies $B_j W_r = 0$, so the first derivative of $W^\top B_j W$ with respect to the candidate frame vanishes. The quadratic multiplier statistic therefore needs no additional first-order frame correction. The factor $p/p-r$ and the division by $\sqrt{K}$ are the same for the observed and multiplier statistics.

To combine the two statistics, include the observed row $\ell = 0$ with the N multiplier rows. For $a \in \{\mathrm{full}, \mathrm{quad}\}$, set

$$\rho_{\ell,r}^a = \frac{1}{N+1}\sum_{m=0}^{N} \mathbf{1}\{T_{m,r}^a \geq T_{\ell,r}^a\}, \qquad s_{\ell,r} = \min_{a\in\{\mathrm{full},\mathrm{quad}\}} \rho_{\ell,r}^a. \tag{23}$$

The calibrated candidate p-value is

$$\widehat{p}_r = \frac{1}{N+1}\sum_{\ell=0}^{N} \mathbf{1}\{s_{\ell,r} \leq s_{0,r}\}. \tag{24}$$

The matched row index preserves dependence between the statistics. The inequalities specify how ties are counted. We test ranks in the order $r = 0, \ldots, p-1$ and stop at the first $\widehat{p}_r > \alpha$. The resulting $\widehat{k}_F = r$ estimates $\mathrm{TIN}(\mathbf{Z}, \mathbf{X})$. If every candidate is rejected, set $\widehat{k}_F = p$. No separate singular-value threshold is used to choose $\widehat{k}_F$.

Algorithm 7 collects the steps. In the TIN experiments, we use $K_\ell = \max\{p, 8\}$ Gaussian-kernel locations per coordinate of $\mathbf{Z}$, drawn once from $\mathcal{N}(0, 1.5^2)$ independently of the testing data, with kernel bandwidth 1 on the raw observed coordinates. We use $N = 199$ multiplier draws.

Algorithm 7 TIN dimension inference with joint rank calibration

Input: Paired data $\{(\mathbf{X}^{(i)}, \mathbf{Z}^{(i)})\}_{i=1}^n$, fixed reference functions F, level α, and number N of multiplier draws with $2/(N+1) \leq \alpha$.

Output: Estimated TIN dimension $\widehat{k}_F$, independent-space dimension $p - \widehat{k}_F$, estimated basis $\widehat{\mathbf{W}}_{\widehat{k}_F}$ when nonempty, and tested p-values $\widehat{p}_r$.

1: Center $\mathbf{X}$, $\mathbf{Z}$, and the values $f_j(\mathbf{Z})$. Construct $\widehat{\Sigma}_{\mathbf{ZX}}$, $\widehat{a}_j$, and $\widehat{\mathbf{B}}_j$ by Equ. (19). Stack and scale them to form $\widehat{\mathbf{L}}_F^{\mathrm{sc}}$.

2: Compute one singular-value decomposition of $\widehat{\mathbf{L}}_F^{\mathrm{sc}}$. Construct the corrected $\widehat{\Psi}_i$ by Eqs. (20)–(21), then draw $G_1, \ldots, G_N$ by Equ. (22).

3: For $r = 0, \ldots, p-1$ in order, take $\widehat{\mathbf{U}}_r$ and $\widehat{\mathbf{W}}_r$ from that decomposition. Compute both observed statistics by Equ. (10) and both multiplier statistics by Equ. (11), pairing them by the same multiplier index. Calculate $\widehat{p}_r$ by Eqs. (23)–(24). If $\widehat{p}_r > \alpha$, stop and return $(\widehat{k}_F, p - \widehat{k}_F, \widehat{\mathbf{W}}_{\widehat{k}_F}) = (r, p-r, \widehat{\mathbf{W}}_r)$ and the computed p-values.

4: If all $r < p$ are rejected, return $\widehat{k}_F = p$, dimension 0, an empty basis, and the computed p-values.

Assume F identifies the independent subspace and is chosen independently of the sample. Suppose its functions are bounded, the data have finite fourth moments, the positive singular spectrum is separated at the true rank, and the limiting null distribution is continuous. Fix N with $2/(N+1) \leq \alpha$. At the true rank, the observed and multiplier pairs of statistics converge jointly to exchangeable draws. Ranking the observed value of $s_{\ell,r}$ among these draws therefore gives pointwise asymptotic rejection probability at most $\lfloor \alpha(N+1) \rfloor/(N+1) \leq \alpha$.

At any fixed rank $r < k_F$, the observed full statistic diverges, whereas its multiplier counterparts remain bounded in probability. Hence the observed row has full-statistic upper-tail rank $1/(N+1)$ with probability tending to one, and $s_{0,r} = 1/(N+1)$. No multiplier row can attain this rank

through the full statistic. Among the quadratic-statistic upper-tail ranks, at most one row can equal $1/(N+1)$, even when ties occur. Thus at most one multiplier row can have $s_{\ell,r} \le s_{0,r}$, giving $\widehat{p}_r \le 2/(N+1) \le \alpha$ with probability tending to one. Since there are only finitely many ranks below k_F, each is rejected with probability tending to one. Consequently, when $k_F < p$,

$$\limsup_{n\to\infty} \Pr(\widehat{k}_F > k_F) \le \alpha, \qquad \Pr(\widehat{k}_F < k_F) \longrightarrow 0. \tag{25}$$

Overestimation requires rejecting the true rank, so fixed-level testing does not imply exact rank recovery with probability tending to one. If $k_F = p$, rejection of every smaller rank yields $\Pr(\widehat{k}_F = p) \to 1$.

The stronger draw condition is needed for this joint rule. To see why, take independent Rademacher variables E, η, an independent Laplace variable U, $\mathbf{Z} = E$, and $\mathbf{X} = (E + \eta, U)^\top$. For Gaussian-kernel reference functions at nonzero locations, the centered feature values are proportional to E. Thus $\Sigma_{\mathbf{ZX}} = (1, 0)$ while every population $\mathbf{B}_j$ vanishes, giving $k_F = 1$. At the false candidate rank $r = 0$, the full statistic diverges while the quadratic statistic has a continuous null limit. When $N = 19$ and $\alpha = 0.05$, the joint p-value is $1/20$ only if the observed quadratic statistic is the largest of the 20 rows; otherwise it is $2/20$. Hence $\Pr(\widehat{k}_F = 0) \to 19/20$, although $1/(N+1) \le \alpha$. The experimental choice $N = 199$ satisfies $2/(N+1) \le \alpha$ at $\alpha = 0.05$. These statements are pointwise and do not assert uniform calibration near changing ranks.

E More Results

This section extends the experiments in Section 6. It reports GIN testing at the weaker alternative $\theta = 0.4$, TIN dimension inference in four synthetic settings, and latent-variable causal discovery results by source setting, latent-DAG topology, sample size, and structural strength. The discovery tables use the same edit loss, latent-edge F_1, and clustering F_1 metrics as the main text.

E.1 GIN tests

The figures below report the weaker alternative $\theta = 0.4$ for the three GIN settings in the main text. The Type I error panels retain the null $\theta = 0$. At $n = 800$, our mean power across five source distributions is 72.2% in Setting I, 42.2% in Setting II, and 57.8% in Setting III. The corresponding GIN (HSIC) rates are 27.0%, 5.5%, and 5.4%, while GIN (LiNGIC) reaches 34.9%, 2.9%, and 4.2%. In Settings II and III, the single-direction baselines stay near their null rejection rates even as n increases, whereas our power rises to 50.7% and 64.7% at $n = 1000$. The weaker alternative also reduces our power relative to the main-text $\theta = 0.8$ results, especially at smaller samples. Across the 15 setting–distribution combinations at $n = 800$, mean Type I error is 3.7% for ours, 5.2% for GIN (HSIC), and 3.9% for GIN (LiNGIC).

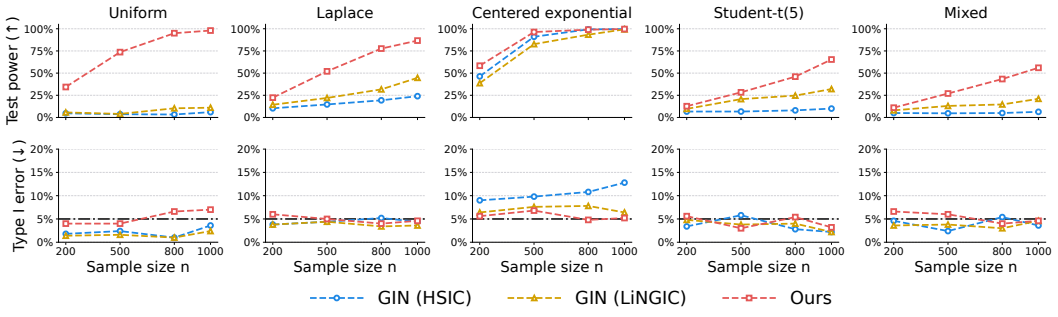

Figure 4: GIN testing in Setting I. Columns show five source distributions. The top row gives power at $\theta = 0.4$, and the bottom row gives Type I error under $\theta = 0$. The curves compare GIN (HSIC), GIN (LiNGIC), and our test across sample sizes at $\alpha = 0.05$.

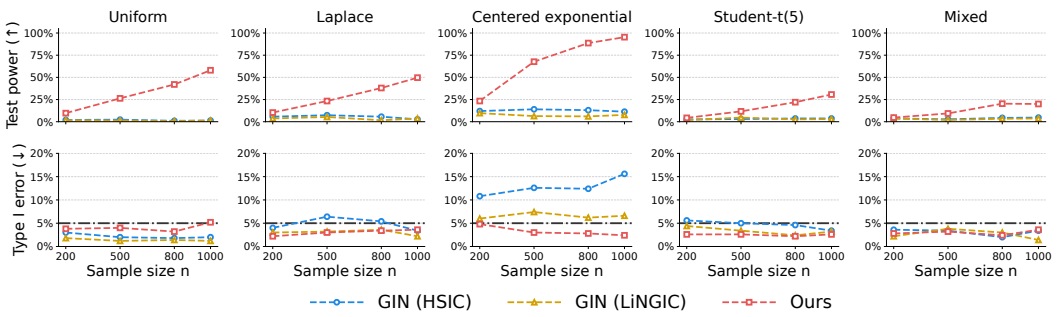

Figure 5: GIN testing in Setting II. Columns show five source distributions. The top row gives power at $\theta = 0.4$, and the bottom row gives Type I error under $\theta = 0$. The curves compare GIN (HSIC), GIN (LiNGIC), and our test across sample sizes at $\alpha = 0.05$.

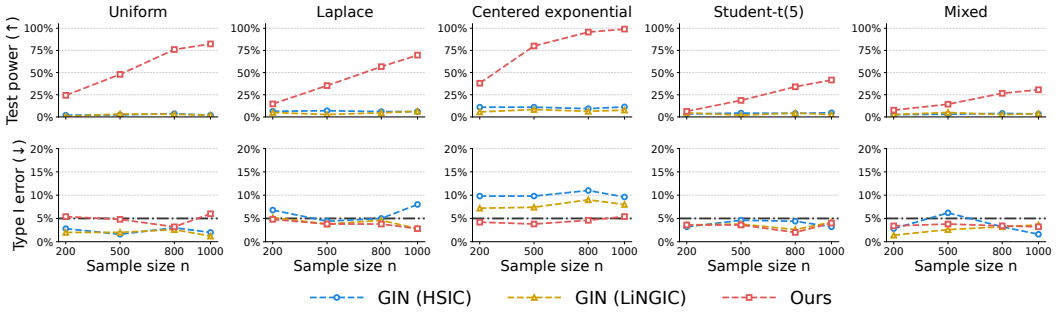

Figure 6: GIN testing in Setting III. Columns show five source distributions. The top row gives power at $\theta = 0.4$, and the bottom row gives Type I error under $\theta = 0$. The curves compare GIN (HSIC), GIN (LiNGIC), and our test across sample sizes at $\alpha = 0.05$.

E.2 TIN DIMENSION INFERENCE

Data generating process. We evaluate $k = \mathrm{TIN}(\mathbf{Z}, \mathbf{X})$ in four settings with independent, standardized non-Gaussian sources. In the single shared source setting, $(p, q, k) = (3, 1, 1)$, so one source contributes to both $\mathbf{X}$ and $\mathbf{Z}$. In multi-Z, $(p, q, k) = (4, 2, 3)$. Both generate a k-vector of shared sources $\mathbf{E}$ and individual sources $\mathbf{U}, \mathbf{V}$ according to

$$\mathbf{X} = A(L_X \mathbf{E} + 0.35\,\mathbf{U}), \qquad \mathbf{Z} = L_Z \mathbf{E} + 0.35\,\mathbf{V}, \tag{26}$$

where L_X has orthonormal columns, each entry of L_Z has a random sign and magnitude $\mathrm{Unif}(0.5, 1.5)/\sqrt{k}$, and A is drawn from random orthogonal factors with singular values evenly spaced between 0.75 and 1.5. The covariance-cancellation setting has $(p, q, k) = (3, 1, 2)$. For independent identically distributed E_1, E_2, it uses $S = (E_1 + E_2)/\sqrt{2}$ and $D = (E_1 - E_2)/\sqrt{2}$: up to invertible mixing and individual noise, $\mathbf{X}$ contains S, D, and one independent coordinate, while $\mathbf{Z}$ contains a signed copy of S. Although $\mathrm{Cov}(D, S) = 0$, the non-Gaussian contrast remains dependent on S.

The contrast-blind setting fixes $(p, q, k) = (5, 2, 2)$ at the $\theta = 0$ point, where $\mathbf{X}$ has no direct S loading. With independent identically distributed E_1, E_2, an independent source T, $S = (E_1 + E_2)/\sqrt{2}$, and $D = (E_1 - E_2)/\sqrt{2}$, it generates

$$\mathbf{X} = 1.2(\mathbf{m}_1 T + \mathbf{m}_2 D) + 0.30\,\mathbf{U}, \qquad \mathbf{Z} = (T, S)^\top + 0.30\,\mathbf{V}, \tag{27}$$

where $\mathbf{m}_1, \mathbf{m}_2$ are the first two columns of a fixed dense orthogonal matrix. Here $\mathrm{rank}(\Sigma_{\mathbf{ZX}}) = 1$, while the true TIN dimension is two. Each setting uses $n \in \{200, 500, 800, 1000\}$ and the four standardized source distributions $\mathrm{Unif}(-\sqrt{3}, \sqrt{3})$, $\mathrm{Laplace}(0, 1/\sqrt{2})$, $\mathrm{Exp}(1) - 1$, and $t_5/\sqrt{5/3}$. All exogenous coordinates in a cell follow the same distribution. We evaluate 30 independently generated paired datasets per setting–distribution–sample-size cell.

Table 2: Mean exact TIN-dimension recovery rates across the four settings at $\alpha = 0.05$. Each setting averages four source distributions and four sample sizes, with 30 paired datasets per cell; the Overall column pools all 1,920 datasets. Higher is better. Bold marks the best method in each column. Underestimation and overestimation rates are shown in Figures 7–10. TIN (two-step*) denotes TIN (two-step + LiNGIC).

Method	Single shared source	Covariance cancellation	Multi-Z	Contrast blind	Overall
Ours	94.4%	**64.0%**	**85.4%**	**92.5%**	**84.1%**
TIN (ISA)	6.7%	63.3%	18.8%	5.8%	23.6%
TIN (two-step)	95.4%	35.4%	69.2%	56.2%	64.1%
TIN (subset)	81.2%	5.6%	27.1%	14.8%	32.2%
TIN (cumulant)	53.5%	32.9%	70.0%	59.8%	54.1%
TIN (two-step*)	**95.6%**	53.5%	79.0%	58.5%	71.7%

Compared methods and evaluation. We compare our TIN dimension estimator with the four original estimators (Dai et al., 2022), TIN (ISA), TIN (two-step), TIN (subset), and TIN (cumulant), and with TIN (two-step + LiNGIC), which replaces the two-step estimator's scalar HSIC tests. All six methods receive the same $\mathbf{X}$ and $\mathbf{Z}$ in each repetition. We report the proportion of datasets with exact recovery of k, as well as the proportions that underestimate or overestimate it. Numerical failures, if any, remain in the denominator. Procedures with independence decisions use $\alpha = 0.05$. Our estimator uses 199 multiplier draws and the LiNGIC replacement uses 99 permutations. No numerical failures occurred in the displayed grid.

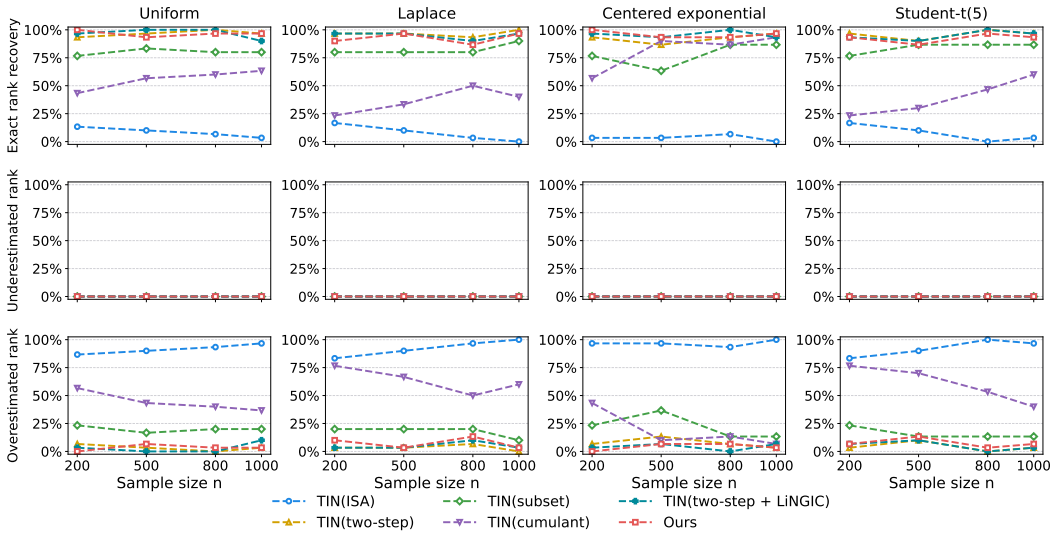

Figure 7: TIN dimension recovery in the single shared source setting, with $(p, q, k) = (3, 1, 1)$. Columns show four source distributions. Rows give exact rank recovery, underestimation, and overestimation. Curves compare our estimator, TIN (ISA), TIN (two-step), TIN (subset), TIN (cumulant), and TIN (two-step + LiNGIC) across sample sizes at $\alpha = 0.05$.

Performance analysis. Table 2 makes the variation across settings explicit. In the single shared source setting, our 94.4% exact-recovery rate is slightly below TIN (two-step) at 95.4% and TIN (two-step + LiNGIC) at 95.6%; all three methods already recover the true dimension on more than 94% of datasets. Averaged equally over all four settings, ours reaches 84.1%, compared with 71.7% for TIN (two-step + LiNGIC), the strongest baseline on this aggregate. This average summarizes the chosen balanced grid and does not imply that ours is best in every setting.

Covariance cancellation is harder: ours recovers the true rank in 64.0% of datasets, compared with 63.3% for TIN (ISA) and 53.5% for TIN (two-step + LiNGIC). At $n = 200$, TIN (ISA) is higher (68.3% versus 36.7%). By $n = 1000$, ours is higher (78.3% versus 62.5%), with most of our remaining errors due to underestimation. In multi-Z, ours reaches 85.4% mean exact recovery,

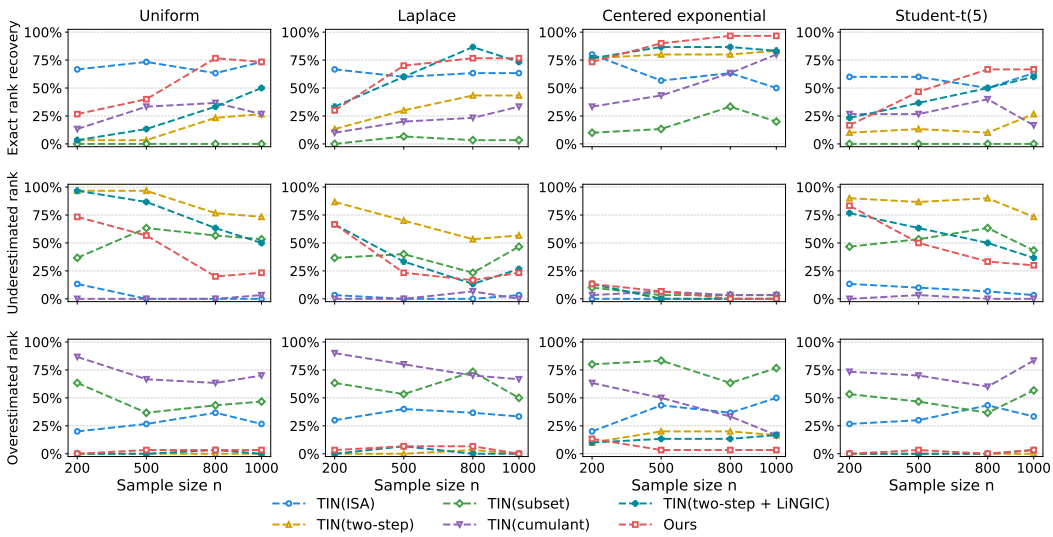

Figure 8: TIN dimension recovery in the covariance-cancellation setting, with $(p, q, k) = (3, 1, 2)$. Columns show four source distributions. Rows give exact rank recovery, underestimation, and overestimation. Curves compare our estimator, TIN (ISA), TIN (two-step), TIN (subset), TIN (cumulant), and TIN (two-step + LiNGIC) across sample sizes at $\alpha = 0.05$.

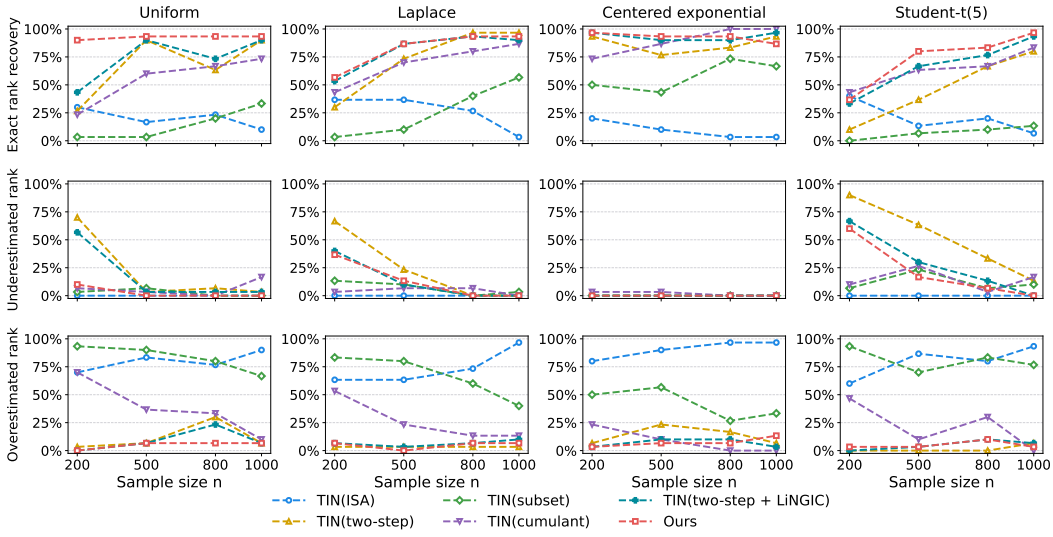

Figure 9: TIN dimension recovery in the multi-Z setting, with $(p, q, k) = (4, 2, 3)$. Columns show four source distributions. Rows give exact rank recovery, underestimation, and overestimation. Curves compare our estimator, TIN (ISA), TIN (two-step), TIN (subset), TIN (cumulant), and TIN (two-step + LiNGIC) across sample sizes at $\alpha = 0.05$.

compared with 79.0% for TIN (two-step + LiNGIC) and 70.0% for TIN (cumulant). At $n = 1000$, ours and TIN (two-step + LiNGIC) both reach 92.5% mean exact recovery.

In the contrast-blind setting, ours recovers $k = 2$ in 92.5% of datasets, compared with 59.8% for TIN (cumulant), 58.5% for TIN (two-step + LiNGIC), and 56.2% for TIN (two-step). The covariance-null contrast is invisible to cross-covariance rank, but remains detectable through its dependence on S. The baseline errors here are predominantly overestimates, whereas our underestimation and overestimation rates are 3.1% and 4.4%. These comparisons describe the four source distributions shown in the figures. Each plotted cell is based on 30 repetitions.

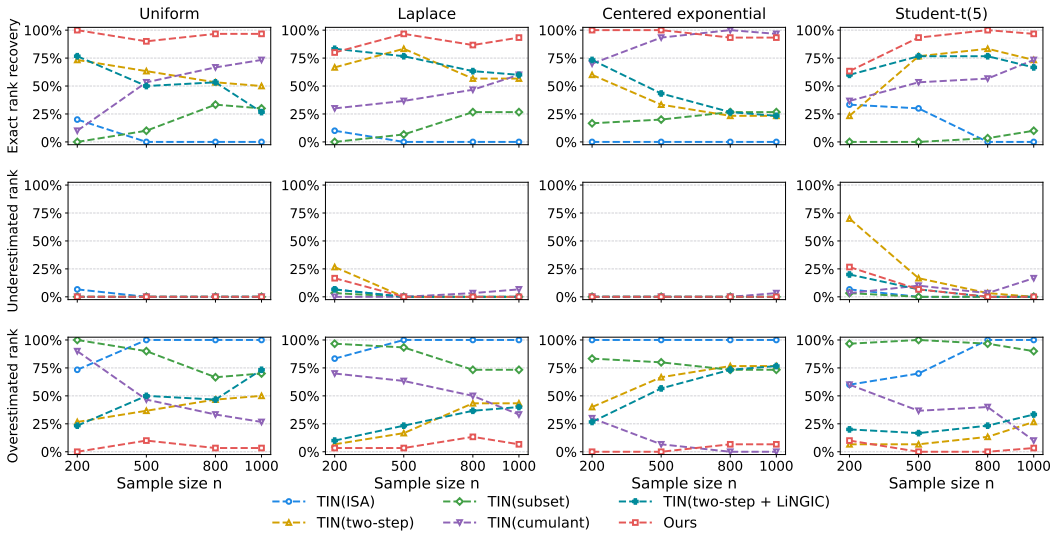

Figure 10: TIN dimension recovery in the contrast-blind setting, with $(p, q, k) = (5, 2, 2)$ and $\theta = 0$. Columns show four source distributions. Rows give exact rank recovery, underestimation, and overestimation. Curves compare our estimator, TIN (ISA), TIN (two-step), TIN (subset), TIN (cumulant), and TIN (two-step + LiNGIC) across sample sizes at $\alpha = 0.05$.

E.3 Latent-variable causal discovery

The following tables further break down the end-to-end results in the main text by source setting, latent-DAG topology, sample size, and structural strength.

Implementation details. All three GIN implementations use the same standardized observations, CCA rank queries, and pure-measurement discovery routine for latent grouping, ordering, and edge pruning. In the heterogeneous source setting, the 16 exogenous sources are assigned distributions once per parameterized graph, and these assignments are held fixed across observations. Our test selects the covariance rank from the data and uses $n_{\text{boot}} = 499$ paired-bootstrap draws per GIN query. GIN (HSIC) and GIN (LiNGIC) each test one right singular direction: the former uses a Gamma approximation and the latter uses 499 permutations to obtain coordinate-wise p-values. These values are combined using Fisher's χ^2_{2q} reference after applying a 10^{-5} floor, where $q = \dim(\mathbf{Z})$.

Table 3: Latent-variable causal discovery by source setting and latent-DAG topology. Each entry reports mean edit loss / latent-edge F_1 / clustering F_1. Lower loss and higher F_1 are better. Bold identifies the best method for each metric within a row. Each row averages 420 paired datasets.

Condition	Ours	GIN (HSIC)	GIN (LiNGIC)
Source setting			
Uniform	**0.2231 / 0.458 / 0.840**	0.2561 / 0.366 / 0.798	0.2514 / 0.380 / 0.803
Laplace	**0.2127 / 0.438 / 0.853**	0.2810 / 0.319 / 0.754	0.2659 / 0.390 / 0.758
Centered exponential	**0.1743 / 0.637 / 0.867**	0.2688 / 0.454 / 0.754	0.2726 / 0.445 / 0.739
Logistic	**0.2100 / 0.384 / 0.878**	0.2623 / 0.305 / 0.803	0.2532 / 0.339 / 0.815
Near-Gaussian	**0.2254 / 0.338 / 0.858**	0.2651 / 0.292 / 0.796	0.2567 / 0.301 / 0.809
Heterogeneous	**0.2073 / 0.486 / 0.858**	0.2621 / 0.393 / 0.786	0.2579 / 0.387 / 0.792
Latent-DAG topology			
Chain	**0.2269 / 0.394 / 0.846**	0.2728 / 0.321 / 0.778	0.2700 / 0.326 / 0.777
Fork	**0.1774 / 0.549 / 0.884**	0.2493 / 0.421 / 0.804	0.2372 / 0.442 / 0.810
Collider	**0.2069 / 0.453 / 0.864**	0.2622 / 0.354 / 0.785	0.2567 / 0.384 / 0.792
Diamond	**0.2309 / 0.388 / 0.854**	0.2757 / 0.304 / 0.785	0.2775 / 0.303 / 0.780
Dense DAG	**0.2348 / 0.423 / 0.846**	0.2861 / 0.324 / 0.772	0.2803 / 0.353 / 0.779
Disconnected	**0.1757 / 0.534 / 0.860**	0.2493 / 0.406 / 0.767	0.2359 / 0.433 / 0.778

Table 4: Latent-variable causal discovery by sample size and structural strength at $\alpha = 0.05$. Each entry reports mean edit loss / latent-edge F_1 / clustering F_1. Lower loss and higher F_1 are better; bold identifies the best method for each metric within a row. Each row averages 504 paired datasets.

Condition	Ours	GIN (HSIC)	GIN (LiNGIC)
Sample size, medium structural strength			
$n = 200$	**0.2111 / 0.420 / 0.864**	0.2807 / 0.308 / 0.778	0.2701 / 0.328 / 0.777
$n = 500$	**0.2076 / 0.476 / 0.858**	0.2640 / 0.362 / 0.787	0.2564 / 0.395 / 0.796
$n = 1000$	**0.2051 / 0.503 / 0.860**	0.2548 / 0.422 / 0.791	0.2508 / 0.443 / 0.791
Structural strength, $n = 500$			
Weak ($c = 0.3$)	**0.2024 / 0.450 / 0.858**	0.2628 / 0.345 / 0.771	0.2527 / 0.360 / 0.789
Medium ($c = 0.7$)	**0.2076 / 0.476 / 0.858**	0.2640 / 0.362 / 0.787	0.2564 / 0.395 / 0.796
Strong ($c = 1.2$)	**0.2177 / 0.436 / 0.854**	0.2672 / 0.338 / 0.781	0.2680 / 0.343 / 0.779

Performance analysis. Table 3 shows that our GIN test has lower edit loss and higher latent-edge and clustering F_1 than both baselines in every source and topology stratum. The centered-exponential source setting has the largest edit-loss gap: 0.1743 for ours, compared with 0.2688 for GIN (HSIC) and 0.2726 for GIN (LiNGIC). Its latent-edge F_1 is also higher (0.637 versus 0.454 and 0.445). Near-Gaussian sources give the lowest latent-edge F_1 for all three methods, but ours still reaches 0.338, compared with 0.292 and 0.301. This source-setting contrast is consistent with stronger non-Gaussian signals making latent-edge recovery easier. Across topologies, our edit loss is lowest for disconnected components (0.1757) and the fork (0.1774), and highest for the dense DAG (0.2348). The diamond has our lowest latent-edge F_1 (0.388), whereas clustering F_1 remains between 0.846 and 0.884 across all six topologies. The larger variation in edge F_1 suggests that topology affects edge recovery more than latent grouping in this experiment.

Table 4 separates sample-size and structural-strength effects. At medium strength, increasing n from 200 to 1000 raises our latent-edge F_1 from 0.420 to 0.503, while edit loss declines from 0.2111 to 0.2051 and clustering F_1 stays near 0.86. Thus, in this grid, the larger sample size is most visible in directed latent-edge recovery. At $n = 500$, performance does not improve monotonically with structural strength: our edit loss is 0.2024, 0.2076, and 0.2177 at weak, medium, and strong connections, respectively. Our method nevertheless has the best value for all three metrics in each strength stratum. These stratified comparisons are descriptive; the Fisher-combined baselines were evaluated after the original study design, and the table does not establish equal query-level error rates across methods.

F Ablation Study

Our GIN test combines the complete second-order response in Equ. (4), all $p - r$ covariance-null directions, equal weights in $T_0 = n\|\widehat{\mathbf{M}}\widehat{\mathbf{V}}\|_F^2$, and the paired bootstrap of Section 4. We first compare the complete procedure with the single-direction baselines used in the main experiments, then change individual components while holding the remaining construction fixed. The two studies use different simulation grids and calibration pools, so their estimates are reported separately.

Comparison with single-direction tests. The first study uses 40 regular linear alternative settings and 20 matched null settings. Original GIN (HSIC) tests one estimated covariance-null direction with scalar HSIC. Original GIN (LiNGIC) replaces that scalar test with LiNGIC. Both combine coordinate-wise p-values across $\mathbf{Z}$ by Fisher's method, as in the main experiments. The latter is a constructed LiNGIC replacement. Each method is evaluated on the same 300 alternative data sets per setting. For each null setting, 999 independent data sets determine an external cutoff and 500 further data sets audit Type I error. Alongside native $\widehat{p} \leq 0.05$ decisions, we report a protected comparison using the 26th ordered null score, with marginal expected size 2.6%, and a separate 50th-order comparison with marginal expected size 5%. These external cutoffs require the matched null distribution and are used only to compare power at similar size in the simulation.

Performance analysis. Under the protected cutoff, our test gains 40.04 percentage points of power over GIN (HSIC) and 29.52 points over GIN (LiNGIC). Paired 95% intervals are $[38.96, 41.58]$ and $[28.68, 31.08]$, respectively. They include calibration-pool variation but are pointwise across

Table 5: Power (%) from the same study, stratified by the population covariance-null dimension $p - r$ and evaluated under the protected cutoff. The final column uses the separate 5%-matched cutoff. The population rank is used only for reporting these strata.

Method	$p - r = 1$	$p - r > 1$	Matched power
Ours	**57.12**	**55.65**	**63.19**
Original GIN (HSIC)	25.54	2.92	21.04
Original GIN (LiNGIC)	42.96	3.08	31.12

Table 6: Component changes under the protected cutoff.

Change (before → after)	Old	New	Combined	95% interval
Within-space → complete second-order response	+17.69	+22.04	+19.86	$[+18.64, +21.09]$
Corrected multiplier → paired bootstrap	-0.06	+0.59	+0.26	$[-0.30, +0.82]$
Block-trace → equal weights	+20.45	+19.66	+20.06	$[+18.58, +21.54]$
One direction → full space ($p - r > 1$)	+53.91	+8.75	+31.33	$[+29.93, +32.73]$
Uncorrected → corrected reference	-0.36	-4.03	-2.20	$[-2.78, -1.62]$
Euclidean → variance-normalized directions	-1.37	-0.35	-0.86	$[-1.23, -0.48]$
Minimum-p → energy sum ($q > 1$)	-7.31	+3.30	-2.01	$[-3.69, -0.33]$
Gamma → corrected multiplier	+0.39	+0.16	+0.27	$[-0.23, +0.78]$

these exploratory comparisons. The 5%-matched gains are 42.15 and 32.07 points, with paired intervals $[40.83, 43.31]$ and $[31.02, 33.24]$. Table 5 shows that the baselines have little protected power when $p - r > 1$. In those settings they test one selected direction, whereas our test examines the entire estimated covariance-null space, so this gap includes a difference in direction coverage. The advantage remains for $p - r = 1$, where the candidate-direction target agrees across procedures. In 32 of the 40 alternative settings $q = \dim(\mathbf{Z}) = 1$, and Fisher combination leaves the scalar-test decisions unchanged. Its distinct effect is confined to the eight settings with $q = 2$.

Component changes. The second study freezes 24 procedures and compares them on the historical 40 alternatives and 36 additional alternative views. The historical and new panels each receive half of the combined weight. Transformed new views that share source samples are kept together in uncertainty estimation. Each cell uses 1999 independent matched-null data sets for external calibration and 1000 alternative data sets. The protected cutoff is the 64th ordered null score, with marginal expected size 3.2%. Table 6 reports changes in power in percentage points. The first two rows compare paired-bootstrap procedures directly or replace corrected multiplier calibration with the paired bootstrap for the same statistic. Remaining rows change components of a corrected-multiplier procedure. They do not isolate their effects under paired bootstrap. "Old" and "new" denote the two simulation panels. The interval is simultaneous for the combined effect. Positive values favor the change. Its 95% intervals account for calibration-pool variation and are simultaneous over the prespecified family of 79 component and interaction contrasts across panels. Because this study differs from the Fisher comparison in data, repetitions, and cutoff, its effect estimates cannot be pooled with those above.

Interpretation. Keeping the complete second-order response, including components outside the candidate space, improves protected power by 19.86 points within the paired-bootstrap design. With corrected multiplier calibration, equal weights improve power by 20.06 points over the tested same-sample block-trace weights. Testing all candidate directions improves power by 31.33 points when $p - r > 1$, although this comparison changes the direction coverage and varies substantially between the historical and new panels. These latter two effects have not been isolated under paired bootstrap. For the identical equal-weight statistic, replacing corrected multipliers with paired bootstrap changes protected power by only 0.26 points, and the simultaneous interval crosses zero. Thus this study does not establish an additional size-matched power gain from paired resampling, even though Section 4 uses it to account for refitting the covariance-null space. Correcting direction estimation in the multiplier reference raises native mean Type I error from 0.67% to 5.23%, addressing severe conservatism. After external size matching, its mean power is 2.20 points lower. Variance normalization, energy aggregation across $\mathbf{Z}$ coordinates, and the change from Gamma to corrected

multipliers likewise show no average protected-power gain over their respective controls. The component effects depend on their reference procedures and should not be added to explain the complete-procedure gap.

G Real-Data Experiments

Setup. We applied the latent-variable causal discovery pipeline with different GIN implementations to a real-world dataset. Our GIN test, GIN (HSIC), and GIN (LiNGIC) used the same standardized observations, CCA rank gate, search rules, and significance level. Published measurement groups were withheld from fitting and used only to score clustering F_1, as defined in the main experiments. Our test used $n_{\text{boot}} = 499$ null-recentered paired-bootstrap draws per GIN query. For GIN (HSIC) and GIN (LiNGIC), we combined coordinate-wise p-values by Fisher's χ^2_{2q} reference with the same 10^{-5} floor as in the synthetic discovery experiments, where $q = \dim(\mathbf{Z})$.

Multitasking behavior. We also analyzed the public data of `https://osf.io/tn6hp/`, used in the original GIN study (Xie et al., 2024). The 202 participants have three speed indicators S_1, S_2, S_3, three error indicators E_1, E_2, E_3, and three question indicators Q_1, Q_2, Q_3. We specified these groups as the scoring reference before fitting; the discovery methods did not use them. At the prespecified $\alpha = 0.01$, our test recovered $\{E_1, E_2, E_3\}$, whereas neither Fisher-combined baseline recovered a cluster. Clustering F_1 was 0.50 versus 0 and 0. At $\alpha = 0.001$, our test recovered the speed and error groups ($F_1 = 0.80$), while both baselines recovered only the speed group ($F_1 = 0.50$). The CCA rank gate changes the candidate clusters across levels, so these outputs are not nested in α. Moreover, 499 bootstrap draws give our direct GIN p-values a minimum of 0.002, which limits the $\alpha = 0.001$ sensitivity comparison. To examine stability at $\alpha = 0.01$, we drew 40 shared random 80% subsamples (162 participants each). Mean clustering F_1 was 0.505 for our test and 0.113 for each baseline. Our score was higher in 33 subsamples and tied in seven against either baseline. These overlapping subsamples are a within-dataset sensitivity analysis rather than independent replications. None of the methods fully recovered the published measurement grouping on the complete dataset.

H Related Work

Learning causal structure with latent variables. Constraint-based methods such as FCI infer causal features among measured variables in the presence of latent confounding (Spirtes et al., 1995). A different line aims to locate latent variables and recover relations involving them. Vanishing tetrads identify measurement groups, and trek separation connects graph structure to ranks of observed covariance submatrices (Silva et al., 2006; Sullivant et al., 2010). RLCD extends rank-based recovery to causally related latent variables (Dong et al., 2023); score-based and other identification approaches study related graph classes under different structural or distributional assumptions (Ng et al., 2024; Li et al., 2024). Non-Gaussian approaches can use information beyond covariance: overcomplete ICA studies latent sources (Adams et al., 2021; Dai et al., 2026), while higher-order cumulants have been used to recover latent confounding and causal order (Hoyer et al., 2008; Cai et al., 2023; Wang & Drton, 2023). These works address different graph-recovery targets. In particular, the numerical rank of an observed covariance matrix need not equal the rank of a source-sensitive operator. Our focus is the statistical reliability of subspace queries used inside latent-variable discovery, rather than a new graph class or a general graph-search algorithm.

Independent-noise constraints. LiNGAM identifies causal directions among observed variables through independent non-Gaussian disturbances, and DirectLiNGAM turns residual independence into a search principle (Shimizu et al., 2006; 2011). Triad constraints apply related reasoning to simple latent measurement structures (Cai et al., 2019). GIN uses covariance-canceling linear combinations to infer latent clusters and their relations (Xie et al., 2020; 2024); TIN considers the dimension of the space of combinations independent of a reference block and applies it to discovery under measurement error (Dai et al., 2022). Conditional independent components and rank-deficiency proxies provide another recent route under different assumptions (Xia et al., 2026). We retain the GIN and TIN population questions. A selected covariance-null projection need not represent every direction in that space, while TIN concerns the dimension of the entire independent space. Their sample implementations must also account for directions fitted from the data on which the test statistics are evaluated.

Non-Gaussianity and statistical efficiency. Non-Gaussian structure has long served statistical precision as well as identification. ICA contrasts and semiparametrically efficient ICA estimators adapt to source distributions, and likelihood-based LiNGAM criteria use non-Gaussian information for causal direction estimation (Hyvarinen, 1999; Chen & Bickel, 2006; Hyvarinen & Smith, 2013). Recent LiNGAM work studies structure-recovery sample complexity under specified distributional conditions (Oh et al., 2025). These results concern mixing parameters or particular graph decisions and do not provide an efficiency bound for a fitted GIN/TIN test. Moreover, near-Gaussian ICA models show that qualitative identifiability may coexist with deteriorating statistical discrimination (Sokol et al., 2014). More directly related to our response design, LiNGIC shows that constancy of conditional first and second moments characterizes independence of a scalar projection within its linear independent-source model (Li et al., 2026). We retain mixed second-order responses so that the constraints are linear in the projection coefficients and one fixed operator can identify the entire independent subspace. Additional moments are useful only when their signal survives covariance adjustment, sampling variation and fitted-direction error. Optimal moment weighting provides a statistical precedent (Hansen, 1982; Chamberlain, 1987), but increased information in a specified moment experiment does not by itself ensure greater power for an unweighted omnibus test.

Conditional-moment subspaces. SIR, SAVE and directional regression recover regression-reduction spaces from first- and second-order conditional responses, while non-Gaussian component analysis uses nonlinear functions to find directions departing from a Gaussian background (Li, 1991; Cook & Weisberg, 1991; Li & Wang, 2007; Blanchard et al., 2006). Their targets differ from projections of $\mathbf{X}$ independent of a specified $\mathbf{Z}$. Kernel conditional-moment tests show how a continuum of conditional restrictions can be represented and tested through RKHS moments (Muandet et al., 2020). Finite analytic evaluations have also been used to build adaptive independence tests (Jitkrittum et al., 2017). In our setting, the evaluations must preserve an entire operator kernel, rather than detect dependence for one fixed pair of variables. This is a population identification statement: the number and locations of finite features still determine the dependence signal and sampling noise available to a test.

Independence testing and fitted-subspace calibration. General-purpose HSIC and distance covariance tests can detect the non-Gaussian dependence relevant to latent-source cancellation (Gretton et al., 2007; Szekely et al., 2007). Finite-feature tests, adaptive HSIC and incomplete-U kernel procedures already address feature selection, power and computation (Jitkrittum et al., 2017; Albert et al., 2022; Schrab et al., 2022). Hence model-specific low-order responses do not imply uniform superiority over well-tuned kernel tests; a fair power comparison must match the causal target, actual test level and computational budget. Fitting also changes calibration: a null law for observed errors cannot generally be reused for estimated residuals (Sen & Sen, 2014), and matrix-rank inference depends on whether the null specifies an exact rank or a composite rank stratum (Robin & Smith, 2000; Chen & Fang, 2019). Our procedures therefore account for fitted candidate spaces and state pointwise guarantees under separated ranks. They do not claim uniform validity as covariance or operator singular values approach zero.