ONE TRANSITION SCORE, TWO GRAPHS: UNTANGLING DYNAMICS AND SYNCHRONOUS INNOVATION DEPENDENCE

Anonymous authors
Paper under double-blind review

ABSTRACT

Finite-interval sampling folds dynamical propagation and synchronous innovation dependence into the same transition, making the two mechanisms difficult to distinguish. We study how to jointly recover the dynamics and synchronous graphs from irregularly sampled observations. For additive-noise diffusions with correlated innovations, differentiating the conditional transition score with respect to the terminal increment and the starting state yields a curvature block and a cross block, respectively. The curvature block characterizes synchronous dependence and demixes the cross block to reveal dynamical sensitivity, allowing the two blocks to jointly identify both graphs. We construct systems with identical cross blocks but different two-layer structures, establishing that the cross block alone cannot identify the two graphs. We also derive support-recovery conditions that account for finite observation gaps and score-estimation error. CAST (Curvature and Sensitivity from Transitions) learns a terminal score conditioned on the actual observation gap directly from irregular transition pairs, without estimating trajectory derivatives or imputing onto a regular grid. Experiments on synthetic systems and external benchmarks validate the two-block readout and gap-conditioned learning. CAST outperforms published baselines for dynamical structure learning on three of four benchmark datasets.

1 INTRODUCTION

Interactions in many dynamical systems unfold on timescales far shorter than the observation gap. Between consecutive daily financial records or fMRI scans acquired seconds apart, directed interactions may propagate between variables while multiple variables are jointly perturbed at the same instant; we call the latter synchronous innovations. Downsampling folds fine-scale directed paths and synchronous innovations into the same finite-gap transition. Both mechanisms shape the finite-gap transition law, and multiple fine-scale generative graphs may be compatible with the same observations (Gong et al., 2015; Danks & Plis, 2013). This creates a fundamental temporal-resolution limit: associations over a finite gap mix distinct generative mechanisms and do not directly reveal instantaneous causation in the generating process (Granger, 1969; Fan et al., 2024). We study continuous-time systems that combine directed dynamics with synchronous innovation dependence (Figure 1) and target their two-layer structure (Peters et al., 2017; Mogensen & Hansen, 2022; Sokol & Hansen, 2014). The directed dynamics graph describes how interactions propagate across time, whereas the undirected synchronous dependence graph encodes conditional dependence among innovations at the same instant. We refer to them as the dynamics and synchronous graphs, respectively. Both graphs are properties of the generating process; the observation gap affects recovery difficulty and bias, not their definitions.

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.

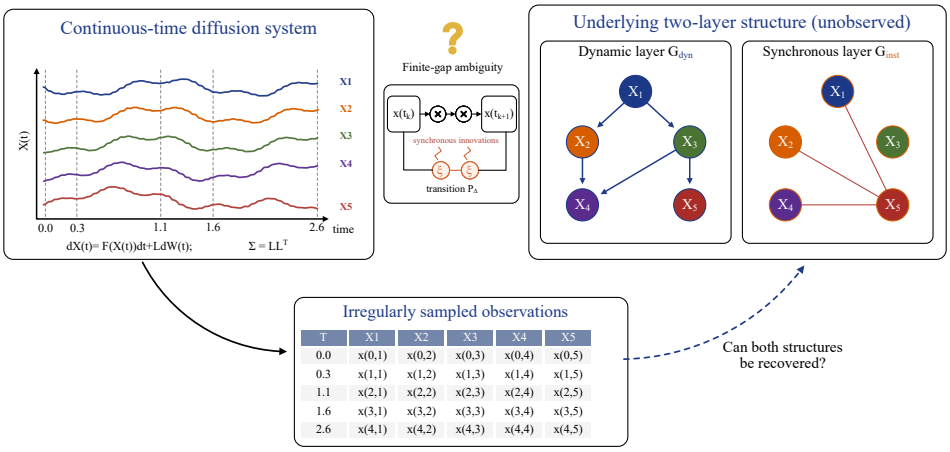


Figure 1: Problem setting. Finite-interval sampling folds temporally propagated directed influences and conditional dependence among synchronous innovations into a single transition. CAST jointly recovers the dynamics and synchronous graphs from irregular observations.

Existing continuous-time structure learners typically recover only a directed dynamics graph under diagonal diffusion (Bellot et al., 2022; Wang et al., 2024; Tagliapietra et al., 2026). Discrete-time methods can estimate contemporaneous structure (Hyvärinen et al., 2010; Runge, 2020; Gong et al., 2023), but their estimand is a directed same-time edge on a discrete grid rather than the undirected support of innovation precision. This mismatch is structural, not merely an omitted output: ordinary differential equation (ODE) models contain no innovation covariance, while the identifiability guarantees for several stochastic differential equation (SDE) methods assume diagonal diffusion. With correlated innovations, the derivative of the transition score with respect to the starting state—the cross block—mixes dynamical sensitivity with innovation precision. At a fixed observation gap, systems with different two-layer structures can even share the same cross block (Proposition 3.1). An alternative route first estimates trajectory derivatives (Brunton et al., 2016), but two-sided differencing leaks synchronous innovations into drift regression, while finite differences amplify measurement noise by $1/\Delta$ (Section 4). These obstacles motivate separating the two structures directly from finite-gap transitions.

We introduce **CAST** (Curvature and Sensitivity from Transitions), which directly learns a terminal score conditioned on the actual observation gap Δ from observed transition pairs. The curvature block yields a readout of innovation precision, while curvature-based demixing turns the cross block into dynamical sensitivity. The two blocks of a single score therefore identify both the synchronous and dynamics graphs. CAST shares parameters across observation gaps and operates on transition pairs, eliminating trajectory-derivative estimation and regular-grid imputation.

Contributions. **(i)** We formulate structure discovery from irregularly sampled continuous-time trajectories as the joint recovery of a dynamics graph and a synchronous dependence graph. From a single gap-conditioned terminal score, CAST reads out innovation precision and demixed dynamical sensitivity without derivative estimation, regular-grid imputation, or a separate graph search. **(ii)** We prove that the cross block alone cannot identify the two-layer structure, characterize synchronous leakage from derivative front ends, quantify finite-gap bias, and connect C^1 score error to exact support recovery. **(iii)** We validate joint recovery of the two layers and show how synchronous curvature improves the dynamics readout. CAST achieves the strongest results as dimension increases and outperforms published baselines for dynamical structure learning on the main synthetic system, NetSim, and DblMass.

2 PRELIMINARIES

Continuous-time model and observations. Consider the continuous-time stochastic system

$$\mathrm{d}X_t \;=\; F(X_t)\,\mathrm{d}t \;+\; L\,\mathrm{d}W_t, \qquad \Sigma \;=\; LL^\top \succ 0, \tag{1}$$

where $X_t \in \mathbb{R}^D$, F is the drift function, W_t is standard Brownian motion, and Σ is the innovation covariance per unit time, which may be non-diagonal. Our method uses all valid transition pairs $(X_t, X_{t+\Delta}, \Delta)$ from irregularly sampled trajectories. We assume that all variables share observation times and are fully observed, and that the observed coordinates form a Markov diffusion; the effects of common drivers acting at the same instant are absorbed into Σ.

Two-layer structural targets. Under standard regularity conditions, the two estimands are independent of the observation gap and are defined directly by the infinitesimal conditional mean and covariance of the process:

$$F(x) = \lim_{\Delta\to 0} \frac{\mathbb{E}[X_{t+\Delta} - X_t \mid X_t = x]}{\Delta}, \qquad \Sigma = \lim_{\Delta\to 0} \frac{\mathrm{Cov}(X_{t+\Delta} - X_t \mid X_t = x)}{\Delta}. \tag{2}$$

Given an evaluation-state distribution π, the directed dynamics graph G_{dyn} is the off-diagonal support of the drift Jacobian: an edge $i \to j$ is present if and only if $\pi(\{x : \partial F_j(x)/\partial x_i \neq 0\}) > 0$. If $F \in C^1$ and π has a full-support density, this definition coincides with nonzero support in the Lebesgue sense. The undirected synchronous dependence graph G_{inst} is the off-diagonal support of the innovation precision matrix $\Theta = \Sigma^{-1}$, which encodes conditional dependence among innovation components in the Gaussian graphical model sense.

The observation gap Δ does not enter the definitions in Equation (2). Finite-gap readouts approximate these infinitesimal targets with a gap-dependent bias; the sampling rate affects their recoverability, not the graph definitions. The graph G_{inst} encodes only undirected conditional dependence among innovations. From second-order information alone, instantaneous causation, marginalized fast latent processes, and common causes cannot be distinguished; G_{inst} therefore carries no directional interpretation. Although the two layers are defined by different infinitesimal conditional moments, the same population lagged moments can jointly encode both. For example, consider a stable stationary Ornstein–Uhlenbeck (OU) reference model and define the lagged moment $K_\Delta = \mathbb{E}[X_{t+\Delta} X_t^\top]$. If K_Δ is invertible and the chosen branch of the matrix logarithm is unique, then $(K_\Delta, K_{2\Delta})$ jointly and uniquely determine the drift and innovation covariance, and hence both graphs (Appendix Theorem B.1).

3 METHOD

CAST operates directly on observed transition pairs $(X_t, X_{t+\Delta})$ and their actual observation gaps Δ. For each pair, define the standardized increment and its terminal score conditioned on the current state and observation gap as

$$Z_\Delta = \frac{X_{t+\Delta} - X_t}{\sqrt{\Delta}}, \qquad s(z \mid x, \Delta) = \nabla_z \log p_\Delta(z \mid x). \tag{3}$$

Here, the score is the gradient of the conditional log-density with respect to the terminal increment z. After training, automatic differentiation with respect to z and the starting state x yields the terminal curvature block $J_{zz} := \partial_z s$ and the cross block $J_{zx} := \partial_x s$. The negated curvature block, $-J_{zz}$, provides the synchronous precision readout, while its inverse removes noise weighting from the cross block. A single network shares parameters across all observation gaps, and the structural readout requires neither graph search nor topological sorting.

3.1 TWO-BLOCK READOUT: FROM NOISE WEIGHTING TO DYNAMICAL SENSITIVITY

The dynamics layer cannot be read directly from the cross block because the block captures how the conditional center changes with the starting state after weighting by transition precision. In the linear Gaussian reference $F(x) = Ax$, let

$$H_\Delta = \frac{e^{A\Delta} - I}{\sqrt{\Delta}}, \qquad Q_\Delta = \int_0^\Delta e^{Au}\Sigma e^{A^\top u}\,\mathrm{d}u, \qquad S_\Delta = \frac{Q_\Delta}{\Delta}.$$

The standardized increment satisfies $Z_\Delta \mid x \sim \mathcal{N}(H_\Delta x, S_\Delta)$, so $s(z \mid x, \Delta) = -S_\Delta^{-1}(z - H_\Delta x)$ and $J_{zx} = S_\Delta^{-1} H_\Delta$. The cross block therefore expresses dynamical sensitivity on the noise scale set by the precision S_Δ^{-1}. With correlated innovations, left multiplication by this precision matrix mixes the rows of H_Δ, so a dynamics edge can appear in other rows of the cross block; in general, $\mathrm{supp}(J_{zx}) \neq \mathrm{supp}(H_\Delta)$.

Proposition 3.1 (Non-identifiability from the cross block alone at fixed Δ)**.** *For any $\Delta > 0$, there exist two stable two-dimensional OU models with identical cross blocks J_{zx} but different G_{dyn} and G_{inst}. Together, (J_{zz}, J_{zx}) identify both layers when the chosen branch of the matrix logarithm is unique. Appendix B gives the proof and an explicit construction.*

The construction makes the ambiguity explicit. Choose a symmetric positive-definite contraction M and set $Q_\Delta = \sqrt{\Delta}(I - M)$; then $S_\Delta = (I - M)/\sqrt{\Delta}$, $H_\Delta = (M - I)/\sqrt{\Delta}$, and hence $J_{zx} = -I$. Taking M_1 diagonal with distinct eigenvalues and $M_2 = U M_1 U^\top$ for a non-permutation rotation U leaves this cross block unchanged while introducing off-diagonal entries in both the drift and innovation precision. Thus the cross block alone cannot distinguish the two-layer structures, whereas the curvature block resolves the ambiguity. This is a fixed-gap construction; Remark B.2 discusses its dependence on the observation gap, and Section 5 examines the corresponding finite-sample behavior.

The same score also contains the required noise scale. In the linear reference, $J_{zz} = -S_\Delta^{-1}$, so multiplication by $-J_{zz}^{-1}$ exactly removes the precision weighting. The implicit function theorem extends this relation beyond Gaussian distributions: differentiating the mode condition $s(z^*(x) \mid x) = 0$ at a conditional mode $z^*(x)$ gives

$$\frac{\mathrm{d}z^*}{\mathrm{d}x} \;=\; -J_{zz}^{-1} J_{zx}. \tag{4}$$

Thus, the curvature block captures the local scale, while the cross block after curvature-based demixing gives the sensitivity of the conditional mode to the starting state. This identity requires only that J_{zz} be nonsingular at the mode; Appendix Theorem C.2 gives the exact linear form.

For the linear specialization of Equation (1), substituting H_Δ and S_Δ gives the two readouts

$$\begin{aligned} P_\Delta &= -J_{zz} = S_\Delta^{-1} = \Sigma^{-1} + O(\Delta), \\ B_\Delta &= \frac{\underbrace{-J_{zz}^{-1} J_{zx}}_{\text{demixing}}}{\sqrt{\Delta}} = \frac{e^{A\Delta} - I}{\Delta} = A + O(\Delta). \end{aligned} \tag{5}$$

The first line shows that the negated curvature block P_Δ is a finite-gap approximation to innovation precision; its off-diagonal entries are used to estimate synchronous edges. For $P_\Delta \succ 0$, normalizing its entries as partial correlations, $-P_{\Delta,ij}/\sqrt{P_{\Delta,ii} P_{\Delta,jj}}$, preserves the support of P_Δ. The second line uses inverse curvature only in the dynamics readout: it removes precision weighting from the cross block, and division by $\sqrt{\Delta}$ yields a finite-gap readout of the drift Jacobian. The $O(\Delta)$ remainder in the dynamics readout arises from the finite-gap expansion. For nonlinear systems, the same interpretation holds pointwise at small Δ, with the remainder controlled by Proposition C.6. The two blocks thus provide local structural readouts at each state and observation gap.

3.2 Gap-conditioned score learning from standardized increments

CAST estimates the population score needed for the two structural readouts directly from irregular transition pairs. Sample paths of the SDE in Equation (1) are not pointwise differentiable, and two-sided derivative front ends leak synchronous innovations into the dynamics layer (Section 4). We therefore work with the standardized increment Z_Δ in Equation (3). As the gap shrinks, the conditional covariance of a raw increment is $O(\Delta)$, whereas that of a finite-difference derivative is $O(1/\Delta)$. By contrast, the standardized increment has Σ as its limiting covariance and remains on an $O(1)$ scale across gaps. This normalization gives a single network a common training scale across gaps and makes $-J_{zz}$ a direct readout of innovation precision. Appendix E.5 compares alternative target scalings.

We estimate the score in Equation (3) using a denoising network conditioned on (x, Δ) (Vincent, 2011; Ho et al., 2020). For each transition pair, we sample a noise level $\bar{\alpha} \in (0, 1)$ and $\varepsilon \sim \mathcal{N}(0, I)$,

then minimize

$$U = \sqrt{\bar{\alpha}}\, Z_\Delta + \sqrt{1-\bar{\alpha}}\, \varepsilon, \qquad \mathcal{L}(\theta) = \mathbb{E}\big[\|\varepsilon - \varepsilon_\theta(U, x, \Delta, \bar{\alpha})\|_2^2\big]. \tag{6}$$

Training perturbs only the terminal increment; the starting state x and gap Δ remain conditioning inputs. At each noise level, the perturbed score is obtained as $s = -\varepsilon_\theta/\sqrt{1-\bar{\alpha}}$. Perturbing only the terminal increment also preserves a Gaussian conditional law in the linear reference and hence the relation across noise levels used below. We log-standardize Δ before embedding it in the network, allowing a single model to learn from all valid transition pairs without regular-grid imputation (Giorgini, 2026a;b).

3.3 From local readouts to stable edge scores

The trained network evaluates the score at any held-out transition, observation gap, and noise level, providing a local two-block readout. Structural recovery requires a single score for each candidate edge. For each $(\Delta, \bar{\alpha})$, the implementation averages the cross block over terminal-perturbation draws for each held-out transition $W = (x, z)$, pools curvature over transitions and draws, and uses the inverse pooled curvature to demix each transition. Let $B(W, \Delta, \bar{\alpha})$ denote the resulting dynamics sensitivity normalized by $\sqrt{\bar{\alpha}\Delta}$. It then takes an RMS over held-out transitions and noise levels within each gap and averages the gap-specific scores. For the uniform prespecified gap set $\mathcal{D}$, CAST defines

$$\widehat{g}_{ji} = \mathbb{E}_{\Delta\in\mathcal{D}}\left[\left(\mathbb{E}_{W,\bar{\alpha}|\Delta}[B_{ji}(W,\Delta,\bar{\alpha})^2]\right)^{1/2}\right]. \tag{7}$$

Here, W is drawn from the held-out transition distribution at gap Δ, while gaps and noise levels come from prespecified sets. The within-gap RMS takes magnitudes before averaging, matching the dynamics-edge definition in Section 2 and preventing nonlinear effects with opposite signs across states from canceling. The final average across gaps balances estimation variance at small gaps against spurious edges induced by multihop propagation at large gaps, an effect we call path fill-in. Averaging across noise levels reduces dependence on any single artificial noise level. In the linear reference, perturbing only the terminal increment leaves the demixed dynamical sensitivity invariant to $\bar{\alpha}$, while terminal curvature admits a closed-form correction (Appendix Corollary C.3). Averaging across levels therefore preserves the dynamics target in the reference model, stabilizes the learned score, and avoids post hoc noise-level selection.

For the synchronous layer, CAST aggregates precision-like curvature $P = -J_{zz}$ and converts its off-diagonal entries to partial correlations, which preserve support. The main experiments rank edges using the uncorrected scores; the known covariance shift induced by terminal perturbation also admits an analytic correction. Appendix Corollary C.3 gives the full expression, while the experiments and Appendix G examine state aggregation and state-dependent diffusion, respectively.

4 Theory

Beyond population-level two-block identification, support recovery is governed by finite temporal resolution and score-estimation error. We first quantify the error scales induced by temporal resolution and then propagate score-Jacobian error through the two structural readouts.

4.1 Three error scales from finite temporal resolution

Finite temporal resolution induces three distinct error scales. In a stationary OU reference, let $P = \mathrm{Cov}(X_t)$, and let w_- denote the total first-order weight that a differencing or smoothing stencil assigns to observations before the target time. When the stencil uses observations on both sides of the target time, the derivative of the OU covariance has a jump at zero lag, injecting the synchronous-innovation term $w_-\Sigma P^{-1}$ into drift regression. This $O(1)$ term persists as the step size tends to zero. Centered differencing corresponds to $w_- = 1/2$, whereas one-sided forward differencing corresponds to $w_- = 0$ (Appendix Theorem C.1).

CAST avoids this derivative leakage by operating directly on transitions, but both of its readouts retain finite-gap bias. In the linear reference, the transition identities in Section 3.1 give

$$\begin{aligned} B_\Delta &= A + \frac{\Delta}{2}A^2 + O(\Delta^2), \\ \Sigma_\Delta := \frac{Q_\Delta}{\Delta} &= \Sigma + \frac{\Delta}{2}(A\Sigma + \Sigma A^\top) + O(\Delta^2), \\ P_\Delta = \Sigma_\Delta^{-1} &= \Theta - \frac{\Delta}{2}(\Theta A + A^\top \Theta) + O(\Delta^2). \end{aligned} \tag{8}$$

Thus temporal propagation perturbs both the dynamics and synchronous readouts. For a chain $1 \to 2 \to 3$ with no direct edge $1 \to 3$, for example, $(B_\Delta)_{31} = (\Delta/2)A_{32}A_{21} + O(\Delta^2)$ creates an $O(\Delta)$ path-fill-in term. Appendix Theorem C.4 gives explicit remainder bounds. When the starting state is observed with noise, errors-in-variables (EIV) bias instead grows as $O(1/\Delta)$ (Appendix Proposition C.7). Smaller gaps suppress finite-gap bias but amplify measurement-noise bias, motivating aggregation across several prespecified gaps. Appendix C verifies the expansion numerically, and Section 5 evaluates the strong-noise regime.

4.2 From score error to support recovery

The learned score contributes additional Jacobian error to both readouts. Let C_* denote the population curvature used for demixing, let B_* and Σ_* denote the corresponding population dynamics and corrected covariance readouts, and suppose their finite-gap biases are bounded by $b_{\mathrm{dyn}}(\Delta)$ and $b_{\mathrm{inst}}(\Delta)$. Set $m = \lambda_{\min}(C_*)$, $L_B = \sup_w \|B_*(w)\|$, and $\mu = \lambda_{\min}(\Sigma)$. If the learned cross block has error η_x and the curvature blocks used by the two readouts have errors $\eta_{z,\mathrm{dyn}}$ and $\eta_{z,\mathrm{inst}}$, with both curvature errors below m, then Appendix Theorem D.2 gives

$$\begin{aligned} \delta_{\mathrm{dyn}} &= b_{\mathrm{dyn}}(\Delta) + \frac{\eta_x/\sqrt{\bar{\alpha}\Delta} + \eta_{z,\mathrm{dyn}} L_B}{m - \eta_{z,\mathrm{dyn}}}, \\ \delta_\Sigma &= b_{\mathrm{inst}}(\Delta) + \frac{\eta_{z,\mathrm{inst}}}{\bar{\alpha} m (m - \eta_{z,\mathrm{inst}})}, \qquad \delta_{\mathrm{inst}} = \frac{\delta_\Sigma}{\mu(\mu - \delta_\Sigma)}. \end{aligned} \tag{9}$$

The last bound requires $\delta_\Sigma < \mu$. Equation (9) exposes the separate effects of finite gaps, the $1/\sqrt{\bar{\alpha}\Delta}$ amplification of cross-block error, and the spectral stability of the two matrix inversions. The synchronous guarantee concerns the analytically corrected readout; the main experiments use the uncorrected curvature ranking in Section 3.3, with the correction evaluated separately in Appendix E.5.

A bound on score values alone is insufficient because both readouts differentiate the learned score: a small-amplitude, high-frequency error can still induce a large Jacobian error. Accordingly, η_x and the curvature errors combine three sources: the network's C^1 approximation error, sampling error from finitely many independent trajectory segments, and conditional Monte Carlo error from finitely many artificial-noise perturbations. Additional perturbation draws reduce only the last term; they do not increase the effective number of independent trajectory segments. Appendix D gives the corresponding concentration bounds.

Corollary 4.1 (Two-layer support recovery under margin conditions)**.** *Define the dynamics-edge strength as* $g_{ji} = \{\mathbb{E}_{X\sim\pi}[(\partial_i F_j(X))^2]\}^{1/2}$*, and let* $\gamma_{\mathrm{dyn}} = \min_{(i\to j)\in G_{\mathrm{dyn}}} g_{ji}$ *and* $\gamma_{\mathrm{inst}} = \min_{\{i,j\}\in G_{\mathrm{inst}}} |\Theta_{ij}|$*. If* $\gamma_{\mathrm{dyn}} > 2\delta_{\mathrm{dyn}}$ *and* $\gamma_{\mathrm{inst}} > 2\delta_{\mathrm{inst}}$*, then each layer admits a threshold that exactly recovers its support. Top-k selection is also exact when the true edge count of each layer is known. Appendix D gives the proof.*

5 Experiments

We first evaluate structural recovery and the readout mechanism on synthetic systems with ground truth for both layers, then examine irregular observation gaps, external systems, dimensional scaling, and measurement noise. Baselines include PCMCI+, VARLiNGAM, DYNOTEARS, NGM, CaDyT, and a matched Δ-Gaussian model with the same transition inputs and model capacity (Runge, 2020; Hyvärinen et al., 2010; Pamfil et al., 2020; Bellot et al., 2022; Tagliapietra et al., 2026). For the dynamics layer, we report AUPRC, top-k F1, and quadratic-edge recall; for the synchronous layer, we report top-k F1 against the support of innovation precision. In all top-k evaluations, k equals the number of true edges. Appendix E.1 gives the full protocol.

Table 1: Two-layer recovery on ten nonlinear ordered ER-style systems (mean $\pm$ standard deviation). The synchronous column uses the uncorrected support readout. ^{c} denotes the matched control, † averages over the eight successful runs, and N/A indicates that a method does not output the synchronous layer.

D	Method	G_{dyn} top-k F1↑	Quadratic-edge recall↑	G_{inst} top-k F1↑
5	**CAST**	$\mathbf{0.960} \pm 0.052$	1.000 ± 0.000	1.000 ± 0.000
	Δ-Gaussian NLL^{c}	0.930 ± 0.067	1.000 ± 0.000	1.000 ± 0.000
	VARLiNGAM	0.610 ± 0.110	0.708 ± 0.281	N/A
10	**CAST**	$\mathbf{0.850} \pm 0.058$	$\mathbf{0.831} \pm 0.171$	1.000 ± 0.000
	Δ-Gaussian NLL^{c}	0.705 ± 0.104	0.745 ± 0.164	1.000 ± 0.000
	VARLiNGAM	0.275 ± 0.079	0.314 ± 0.176	N/A
15	**CAST**	$\mathbf{0.733} \pm 0.059$	$\mathbf{0.890} \pm 0.074$	0.986 ± 0.045
	Δ-Gaussian NLL^{c}	0.567 ± 0.077	0.760 ± 0.097	0.986 ± 0.045
	VARLiNGAM†	0.238 ± 0.060	0.362 ± 0.141	N/A

Table 2: Structural readout analysis. For each system, all variants use the same trained score model, held-out transitions, observation gaps, noise levels, and perturbation draws. Only the dynamics readout and state aggregation rule vary.

Dynamics readout	State aggregation	Dynamics AUPRC↑	Top-k F1↑	Quadratic-edge recall↑
Full curvature demixing (CAST)	RMS	**0.967**	**0.927**	**0.890**
Raw cross block	RMS	0.793	0.717	0.760
Diagonal curvature scaling	RMS	0.854	0.777	0.800
Full curvature demixing	Absolute of signed mean	0.893	0.817	0.490

Two-layer structural recovery. The main synthetic system samples its dynamics graph from an ordered ER-style graph model and assigns each edge a linear, hyperbolic-tangent, or quadratic mechanism at random; correlated innovation pairs provide ground truth for the synchronous graph. The main protocol uses 100 training and 50 held-out trajectory segments, with readout gaps $\{0.076, 0.2\}$; Appendix E.1 gives the full generator. CAST recovers both layers (Table 1). At low dimensions, CAST and the matched Gaussian both achieve near-ceiling recovery. As dimension increases, CAST maintains the strongest dynamics-graph recovery and quadratic-edge recall, while synchronous-graph recovery remains nearly exact. On the same trajectories, CAST also outperforms PCMCI+, VARLiNGAM, and DYNOTEARS; Appendix E.6 reports the full results.

To verify the estimand of the synchronous layer, we also construct three-node chains whose covariance and precision supports differ. In a ranking audit of the saved terminal-curvature matrices, AUPRC against precision support exceeds AUPRC against covariance support by $+0.0936$, with a seed-bootstrap 95% interval of $[+0.0386, +0.1410]$ (Appendix E.4).

Structural readout analysis. The curvature block provides both the synchronous-graph readout and the information needed to remove noise weighting from the cross block. With RMS aggregation fixed, we compare the raw cross-block readout, diagonal curvature scaling, and full curvature demixing. This is a separate $D = 15$ mechanism study, not the $D = 15$ slice of Table 1. It uses 2,000 independent training and 500 held-out episodes with renewal-sampled gaps $\{0.04, 0.08, 0.10, 0.12, 0.16\}$; all comparisons are paired within each system. Across these ten nonlinear ER-style systems, dynamics-layer AUPRC increases from 0.793 to 0.854 and then to 0.967. The fraction of variable pairs connected only by a synchronous relation but incorrectly reported as a dynamics edge falls from 10% to 7% and then to 2% (Appendix E.4). The curvature information therefore improves dynamics-graph recovery in addition to defining the synchronous layer.

With full curvature demixing fixed, Table 2 compares RMS aggregation with the absolute value of the signed mean. At each gap, these rules are $\sqrt{\mathbb{E}_{W,\bar{\alpha}|\Delta}[B_{ji}^2]}$ and $|\mathbb{E}_{W,\bar{\alpha}|\Delta}[B_{ji}]|$, respectively; perturbation-draw and gap aggregation remain fixed. Signed averaging cancels quadratic-edge effects whose signs differ across states, whereas RMS preserves the signal. Appendix E.5 compares noise-level aggregation, target scaling, and analytic correction.

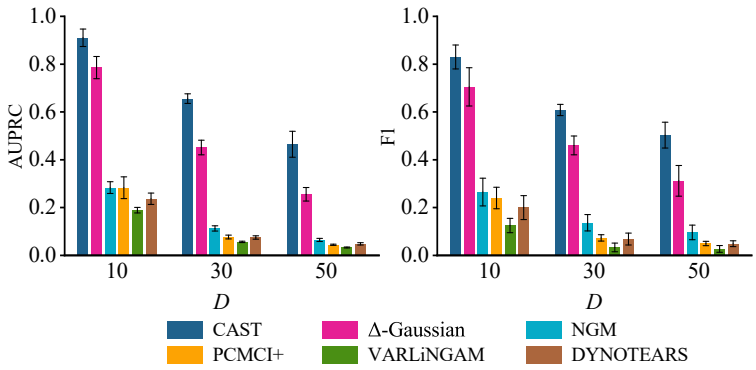


Figure 2: Dynamics-graph scaling on ER-style systems. Left: dynamics-edge AUPRC. Right: top-k F1. Bars show means, and error bars show bootstrap 95% confidence intervals.

From irregular observation gaps to external systems. We keep the latent SDE paths fixed and vary only the observation timestamps. Controls include ridge regression with explicit Δ and two impute-then-discover pipelines. The latter fill a regular grid using linear interpolation or a CSDI-style model (Tashiro et al., 2021), estimate derivatives by finite differences, and apply the prespecified structure-discovery pipeline to state–derivative pairs. CAST achieves higher dynamics-layer AUPRC than these controls at seen gaps, interpolation gaps, and shifted (extrapolation) gaps, while exactly recovering the synchronous graph in all three regimes. The matched Gaussian achieves comparable accuracy in this near-ceiling setting (Appendix E.2).

The ranking changes across external benchmarks. On NetSim (Smith et al., 2011), CAST outperforms all published baselines for dynamical structure learning considered here. On DblMass (Tagliapietra et al., 2026), CAST requires no derivative presmoothing and achieves the highest mean AUPRC and lowest nSHD among these baselines. CausalDynamics-CI (Herdeanu et al., 2025) is the only exception: NGM ranks first and CAST second. Across the main synthetic system and three external benchmarks, CAST outperforms the evaluated published baselines in three of four settings. The relative ranking of CAST and the matched Gaussian reverses between NetSim and CausalDynamics-CI, showing that gains from score parameterization are mechanism-dependent (Appendices F and E.7). On NetSim, adding an earlier observation improves prediction of the next state even after conditioning on the current BOLD state (Appendix J), suggesting that the observed process may contain memory beyond a first-order Markov model.

Dimensional scaling, finite gaps, and measurement noise. CAST remains stable on the synchronous layer as dimension increases, maintaining top-k F1 at or above 0.980 through $D = 50$, while the matched Gaussian falls from 0.980 at $D = 10$ to 0.760 at $D = 50$ (Appendix Table 12). On the dynamics layer, performance declines gradually, with quadratic edges most affected (Figure 2). CAST outperforms the other five methods on the dynamics layer at every dimension, and its advantage over the matched Gaussian baseline grows with dimension.

Training remains the dominant computational cost at the evaluated dimensions; Appendix I gives the full cost analysis. Appendix C evaluates finite-gap path fill-in.

At the highest measurement-noise level in DblMass, increasing the readout gap attenuates the $1/\Delta$ amplification of starting-state noise. Without derivative presmoothing, CAST still achieves the highest AUPRC and lowest nSHD among the evaluated published baselines (Appendix E.7, Table 8).

6 Related Work

Time-series causal discovery and continuous-time structure. Discrete-time causal discovery includes neural Granger methods, conditional-independence tests, differentiable acyclicity optimization, and non-Gaussian identification (Tank et al., 2022; Runge et al., 2019; Zheng et al., 2018; Pamfil et al., 2020; Shimizu et al., 2006; Hyvärinen et al., 2010). These methods typically assume a regular time grid; when they model contemporaneous structure, the target is a directed same-time edge in discrete time (Gong et al., 2023). Continuous-time structure learning uses neural ODE/SDE mod-

els, Gaussian processes, and impute-then-Granger pipelines; other methods estimate dynamics from state samples that are unpaired across time (Bellot et al., 2022; Wang et al., 2024; Tagliapietra et al., 2026; Cheng et al., 2024; Guan et al., 2024). APPEX jointly estimates drift and diffusion, whereas CAST defines the synchronous graph by the off-diagonal support of innovation precision (Lauritzen, 1996; Meinshausen & Buhlmann, 2006). This target differs from directed contemporaneous edges in discrete time and from mixed edges induced by marginalizing latent processes in continuous-time local-independence graphs (Mogensen & Hansen, 2022; Sokol & Hansen, 2014).

Transition scores and structural readout. Denoising score matching provides tools for conditional density estimation (Hyvärinen, 2005; Vincent, 2011; Song et al., 2021; Ho et al., 2020). SBTG uses joint-window score geometry to estimate noise-scaled, lag-specific transition Jacobians and test directed edges (Kinger et al., 2026). Conditional score-based Langevin modeling uses finite-lag score identities to calibrate drift and mobility from correlation constraints (Giorgini, 2026a;b). CAST instead targets infinitesimal dynamics support jointly with innovation-precision support: terminal curvature supplies the second target and demixes the cross block, while the analysis tracks finite-gap and score-Jacobian error in both readouts. In static settings, SCORE and DiffAN use score Jacobians to recover topological orderings in additive-noise models (Rolland et al., 2022; Sanchez et al., 2023).

7 Discussion and Conclusion

The two structural readouts are coupled rather than parallel outputs. Terminal curvature identifies synchronous dependence and supplies the inverse noise scale that demixes the cross block. Its dual role explains why using the full curvature improves dynamics recovery: synchronous information is part of the dynamics readout, not an auxiliary prediction target.

The ordering of CAST and the matched Gaussian varies across systems, showing that gains from score parameterization depend on the transition mechanism. The linear reference yields exact two-block identities, while for general conditional laws the implicit-function identity is exact at non-singular modes. Nonlinear finite-gap readouts are controlled by the short-time score-Jacobian and population-residual conditions. Finite gaps induce the $O(\Delta)$ biases above, whereas measurement noise favors larger readout gaps; aggregation across prespecified gaps balances these effects. External benchmarks primarily assess the dynamics layer, while constructed systems with known innovation precision verify the synchronous semantics. Appendix J details these assumptions, the finite-gap and measurement-noise regimes, and the diagnostic for Markov compatibility.

Under these conditions, a single gap-conditioned transition score separates the two mechanisms mixed by finite-interval sampling. Its curvature block recovers synchronous precision, and curvature-based demixing turns the cross block into dynamical sensitivity; the resulting error bounds specify when both supports remain recoverable from irregular observations.

Reproducibility Statement

Section 3 specifies CAST's standardized transition representation, observation-gap-conditioned terminal score network, denoising objective, two-block Jacobian readout, and aggregation across states, noise levels, and observation gaps. Section 5 and Appendix E.1 describe the data-generating processes, train/held-out construction, preprocessing, evaluation metrics, baseline configurations, and experimental protocols. The appendix reports the fixed hyperparameters, single-GPU computational budget, and random seeds. The supplementary material includes an anonymized minimal implementation of CAST, the frozen main configuration, and a script for reproducing the main synthetic experiment.

AI Use Statement

Generative AI tools (a large language model) were used to assist with English translation, language polishing, code and experiment debugging, and consistency checking across the paper. The authors made all final methodological and interpretive decisions, independently reviewed and verified all AI-assisted material, and take full responsibility for the paper.

REFERENCES

Alexis Bellot, Kim Branson, and Mihaela van der Schaar. Neural graphical modelling in continuous-time: consistency guarantees and algorithms, 2022. URL https://arxiv.org/abs/2105.02522.

Steven L. Brunton, Joshua L. Proctor, and J. Nathan Kutz. Discovering governing equations from data by sparse identification of nonlinear dynamical systems. *Proceedings of the National Academy of Sciences*, 113(15):3932–3937, 2016. doi: 10.1073/pnas.1517384113. URL https://www.pnas.org/doi/abs/10.1073/pnas.1517384113.

Robrecht Cannoodt, Wouter Saelens, Louise Deconinck, and Yvan Saeys. Spearheading future omics analyses using dyngen, a multi-modal simulator of single cells. *Nature Communications*, 12:3942, 2021. doi: 10.1038/s41467-021-24152-2.

Yuxiao Cheng, Lianglong Li, Tingxiong Xiao, Zongren Li, Jinli Suo, Kunlun He, and Qionghai Dai. CUTS+: High-dimensional causal discovery from irregular time-series. In *Proceedings of the AAAI Conference on Artificial Intelligence*. AAAI Press, 2024. ISBN 978-1-57735-887-9. doi: 10.1609/aaai.v38i10.29034. URL https://doi.org/10.1609/aaai.v38i10.29034.

David Danks and Sergey Plis. Learning causal structure from undersampled time series. Proceedings of the NIPS 2013 Workshop on Causality, December 2013. URL https://kilthub.cmu.edu/articles/journal_contribution/Learning_Causal_Structure_from_Undersampled_Time_Series/6492101.

Payam Dibaeinia and Saurabh Sinha. SERGIO: A single-cell expression simulator guided by gene regulatory networks. *Cell Systems*, 11(3):252–271.e11, 2020. doi: 10.1016/j.cels.2020.08.003.

Shunxing Fan, Mingming Gong, and Kun Zhang. On the recoverability of causal relations from temporally aggregated I.I.D. data. In Ruslan Salakhutdinov, Zico Kolter, Katherine Heller, Adrian Weller, Nuria Oliver, Jonathan Scarlett, and Felix Berkenkamp (eds.), *Proceedings of the 41st International Conference on Machine Learning*, volume 235 of *Proceedings of Machine Learning Research*, pp. 12806–12831. PMLR, 21–27 Jul 2024. URL https://proceedings.mlr.press/v235/fan24a.html.

Ludovico T. Giorgini. Conditional score-based modeling of effective Langevin dynamics, 2026a. URL https://arxiv.org/abs/2604.23952.

Ludovico Theo Giorgini. Score-based modeling of effective Langevin dynamics. *Physical Review E*, 114(1):L012102, 2026b. doi: 10.1103/6qpv-lqmt.

Mingming Gong, Kun Zhang, Bernhard Schölkopf, Dacheng Tao, and Philipp Geiger. Discovering temporal causal relations from subsampled data. In *Proceedings of the 32nd International Conference on Machine Learning*, volume 37 of *Proceedings of Machine Learning Research*, pp. 1898–1906, 2015. URL https://proceedings.mlr.press/v37/gongb15.html.

Wenbo Gong, Joel Jennings, Cheng Zhang, and Nick Pawlowski. Rhino: Deep causal temporal relationship learning with history-dependent noise. In *The Eleventh International Conference on Learning Representations*, 2023. URL https://openreview.net/forum?id=i_1rbq8yFWC.

C. W. J. Granger. Investigating causal relations by econometric models and cross-spectral methods. *Econometrica*, 37(3):424–438, 1969.

Vincent Guan, Joseph Janssen, Hossein Rahmani, Andrew Warren, Stephen Zhang, Elina Robeva, and Geoffrey Schiebinger. Identifying drift, diffusion, and causal structure from temporal snapshots. *arXiv preprint arXiv:2410.22729*, 2024.

Benjamin Herdeanu, Juan Nathaniel, Carla Roesch, Jatan Buch, Gregor Ramien, Johannes Haux, and Pierre Gentine. CausalDynamics: A large-scale benchmark for structural discovery of dynamical causal models. In *Advances in Neural Information Processing Systems*, volume 38, 2025. doi: 10.52202/085713-1349. URL https://proceedings.neurips.cc/paper_files/paper/2025/hash/39a3f7dc00e4723a5ca7808b04fc4d4f-Abstract-Datasets_and_Benchmarks_Track.html.

Jonathan Ho, Ajay Jain, and Pieter Abbeel. Denoising diffusion probabilistic models. In *Advances in Neural Information Processing Systems*, volume 33, pp. 6840–6851, 2020.

Aapo Hyvärinen. Estimation of non-normalized statistical models by score matching. *Journal of Machine Learning Research*, 6(24):695–709, 2005. URL http://jmlr.org/papers/v6/hyvarinen05a.html.

Aapo Hyvärinen, Kun Zhang, Shohei Shimizu, and Patrik O. Hoyer. Estimation of a structural vector autoregression model using non-gaussianity. *Journal of Machine Learning Research*, 11: 1709–1731, 2010.

Savik Kinger, Johannes Bertram, Luciano Dyballa, Eviatar Yemini, and Steven W. Zucker. Inferring active neural circuits using diffusion scores, 2026. URL https://arxiv.org/abs/2605.02852.

Steffen L. Lauritzen. *Graphical Models*. Oxford University Press, May 1996. ISBN 9780198522195. doi: 10.1093/oso/9780198522195.001.0001. URL https://doi.org/10.1093/oso/9780198522195.001.0001.

Baojuan Li, Jean Daunizeau, Klaas E. Stephan, Will Penny, Dewen Hu, and Karl Friston. Generalised filtering and stochastic DCM for fMRI. *NeuroImage*, 58(2):442–457, 2011. doi: 10.1016/j.neuroimage.2011.01.085.

Nicolai Meinshausen and Peter Buhlmann. High-dimensional graphs and variable selection with the lasso. *Annals of Statistics*, 34(3):1436–1462, 2006. doi: 10.1214/009053606000000281.

Søren Wengel Mogensen and Niels Richard Hansen. Graphical modeling of stochastic processes driven by correlated noise. *Bernoulli*, 28(4):3023–3050, 2022. doi: 10.3150/21-BEJ1446.

Roxana Pamfil, Nisara Sriwattanaworachai, Shaan Desai, Philip Pilgerstorfer, Konstantinos Georgatzis, Paul Beaumont, and Bryon Aragam. DYNOTEARS: Structure learning from time-series data. In *Proceedings of the Twenty Third International Conference on Artificial Intelligence and Statistics*, volume 108 of *Proceedings of Machine Learning Research*, pp. 1595–1605, 2020.

Jonas Peters, Dominik Janzing, and Bernhard Schölkopf. *Elements of Causal Inference: Foundations and Learning Algorithms*. Adaptive Computation and Machine Learning. The MIT Press, Cambridge, MA, USA, 2017. ISBN 9780262037310.

Paul Rolland, Volkan Cevher, Matthäus Kleindessner, Chris Russell, Dominik Janzing, Bernhard Schölkopf, and Francesco Locatello. Score matching enables causal discovery of nonlinear additive noise models. In Kamalika Chaudhuri, Stefanie Jegelka, Le Song, Csaba Szepesvari, Gang Niu, and Sivan Sabato (eds.), *Proceedings of the 39th International Conference on Machine Learning*, volume 162 of *Proceedings of Machine Learning Research*, pp. 18741–18753. PMLR, 17–23 Jul 2022.

Jakob Runge. Discovering contemporaneous and lagged causal relations in autocorrelated nonlinear time series datasets. In Jonas Peters and David Sontag (eds.), *Proceedings of the 36th Conference on Uncertainty in Artificial Intelligence (UAI)*, volume 124 of *Proceedings of Machine Learning Research*, pp. 1388–1397. PMLR, 03–06 Aug 2020.

Jakob Runge, Peer Nowack, Marlene Kretschmer, Seth Flaxman, and Dino Sejdinovic. Detecting and quantifying causal associations in large nonlinear time series datasets. *Science Advances*, 5 (11):eaau4996, 2019. doi: 10.1126/sciadv.aau4996.

Pedro Sanchez, Xiao Liu, Alison Q O'Neil, and Sotirios A. Tsaftaris. Diffusion models for causal discovery via topological ordering. In *The Eleventh International Conference on Learning Representations*, 2023. URL https://openreview.net/forum?id=Idusfje4-Wq.

Thomas Schaffter, Daniel Marbach, and Dario Floreano. GeneNetWeaver: In silico benchmark generation and performance profiling of network inference methods. *Bioinformatics*, 27(16): 2263–2270, 2011. doi: 10.1093/bioinformatics/btr373.

Shohei Shimizu, Patrik O. Hoyer, Aapo Hyvärinen, and Antti Kerminen. A linear non-gaussian acyclic model for causal discovery. *Journal of Machine Learning Research*, 7(72):2003–2030, 2006. URL http://jmlr.org/papers/v7/shimizu06a.html.

Stephen M. Smith, Karla L. Miller, Gholamreza Salimi-Khorshidi, Matthew Webster, Christian F. Beckmann, Thomas E. Nichols, Joseph D. Ramsey, and Mark W. Woolrich. Network modelling methods for fmri. *NeuroImage*, 54(2):875–891, 2011. ISSN 1053-8119. doi: 10.1016/j.neuroimage.2010.08.063.

Alexander Sokol and Niels Richard Hansen. Causal interpretation of stochastic differential equations. *Electronic Journal of Probability*, 19(100):1–24, January 2014. ISSN 1083-6489. doi: 10.1214/ejp.v19-2891. URL http://dx.doi.org/10.1214/EJP.v19-2891.

Yang Song, Jascha Sohl-Dickstein, Diederik P. Kingma, Abhishek Kumar, Stefano Ermon, and Ben Poole. Score-based generative modeling through stochastic differential equations, 2021. URL https://arxiv.org/abs/2011.13456.

Nicholas Tagliapietra, Katharina Ensinger, Christoph Zimmer, and Osman Mian. Causal structure learning for dynamical systems with theoretical score analysis. *Proceedings of the AAAI Conference on Artificial Intelligence*, 40(43):36740–36748, 2026. doi: 10.1609/aaai.v40i43.40999. URL https://ojs.aaai.org/index.php/AAAI/article/view/40999.

Alex Tank, Ian Covert, Nicholas J. Foti, Ali Shojaie, and Emily B. Fox. Neural granger causality. *IEEE transactions on pattern analysis and machine intelligence*, 44(8):4267 – 4279, 2022. doi: 10.1109/TPAMI.2021.3065601.

Yusuke Tashiro, Jiaming Song, Yang Song, and Stefano Ermon. CSDI: Conditional score-based diffusion models for probabilistic time series imputation. In Marc'Aurelio Ranzato, Alina Beygelzimer, Yann Dauphin, Percy S. Liang, and Jennifer Wortman Vaughan (eds.), *Advances in Neural Information Processing Systems*, volume 34, pp. 24804–24816. Curran Associates, Inc., 2021. URL https://proceedings.neurips.cc/paper_files/paper/2021/file/cfe8504bda37b575c70ee1a8276f3486-Paper.pdf.

Pascal Vincent. A connection between score matching and denoising autoencoders. *Neural Computation*, 23(7):1661–1674, 2011.

Benjie Wang, Joel Jennings, and Wenbo Gong. Neural structure learning with stochastic differential equations. In *The Twelfth International Conference on Learning Representations*, 2024. URL https://openreview.net/forum?id=V1GM9xDvIY.

Xun Zheng, Bryon Aragam, Pradeep Ravikumar, and Eric P. Xing. DAGs with NO TEARS: Continuous optimization for structure learning. In *Advances in Neural Information Processing Systems*, volume 31, pp. 9492–9503. Curran Associates Inc., 2018.

Appendix

A Notation, Coordinates, and Estimands

To distinguish graph supports from numerical matrices, we use the notation in Table 3 throughout. The symbol G always denotes a graph or support set, whereas A, Θ, and B_Δ denote numerical matrices to be estimated.

Table 3: Unified notation for the theory and implementation.

Symbol	Meaning
$F : \mathbb{R}^D \to \mathbb{R}^D$	Continuous-time drift; $F(x) = Ax$ in the linear case
$\Sigma = LL^\top, \Theta = \Sigma^{-1}$	Infinitesimal innovation covariance and precision
$G_{\mathrm{dyn}} = \mathrm{supp}_{\mathrm{off}}(\nabla F)$	Directed dynamics graph; positive-measure/RMS definition under π when nonlinear
$G_{\mathrm{inst}} = \mathrm{supp}_{\mathrm{off}}(\Theta)$	Undirected synchronous dependence graph
$M_\Delta = e^{A\Delta}$	Linear finite-gap transition matrix
$Q_\Delta = \int_0^\Delta e^{Au}\Sigma e^{A^\top u}\,du$	Linear finite-gap transition covariance
$Z_\Delta = (X_{t+\Delta} - X_t)/\sqrt{\Delta}$	Standardized increment
$H_\Delta = (M_\Delta - I)/\sqrt{\Delta}$, $S_\Delta = Q_\Delta/\Delta$	Linear mean coefficient and covariance of $Z_\Delta \mid X_t = x$
$s_\Delta(z \mid x) = \nabla_z \log p_\Delta(z \mid x)$	Terminal score
$J_{zz} = \partial_z s_\Delta$, $J_{zx} = \partial_x s_\Delta$	Terminal-curvature and cross blocks
$B_\Delta = -J_{zz}^{-1} J_{zx}/\sqrt{\Delta}$	Finite-gap dynamics readout; $(M_\Delta - I)/\Delta$ when linear
$\Sigma_\Delta = -J_{zz}^{-1}$, $\Theta_\Delta = \Sigma_\Delta^{-1}$	Finite-gap noise-covariance and precision readouts
$P = \mathrm{Cov}(X_t)$	Stationary OU state covariance

Superscript s marks training coordinates. If $X^s = D_x^{-1}(X - m_x)$ and $Z^s = D_z^{-1}(Z - m_z)$, then $J_F^s = D_z^{-1}(\nabla F)D_x$ and $\Theta^s = D_z \Theta D_z$. Because D_x and D_z are positive diagonal matrices, left and right diagonal scaling preserve off-diagonal zero support. Operator-norm bounds may therefore be established in standardized coordinates and mapped back through $D_z(\cdot)D_x^{-1}$.

B Proofs for Joint Identification and the Need for Two Blocks

Theorem B.1 (Joint identification of both layers from two lagged moments)**.** *Let X_t be a stable stationary OU process with $P \succ 0$, let $K_\Delta = \mathbb{E}[X_{t+\Delta} X_t^\top]$, and suppose K_Δ is invertible. If the selected branch of the matrix logarithm is unique, then $(K_\Delta, K_{2\Delta})$ jointly and uniquely determine A, Σ, and $(G_{\mathrm{dyn}}, G_{\mathrm{inst}})$.*

Proof. Since $K_\Delta = M_\Delta P$ and $K_{2\Delta} = M_\Delta^2 P$,

$$M_\Delta = K_{2\Delta} K_\Delta^{-1}, \qquad P = M_\Delta^{-1} K_\Delta, \qquad A = \Delta^{-1} \log M_\Delta.$$

The stationary Lyapunov equation gives $\Sigma = -(AP + PA^\top)$. Inverting this positive-definite matrix then determines $\mathrm{supp}_{\mathrm{off}}(\Sigma^{-1})$. Thus the two lagged moments identify not only A but also the synchronous layer: both layers are jointly identifiable from the same population moments. □

Proposition 3.1, restated. For every $\Delta > 0$, there exist two stable two-dimensional OU models with identical J_{zx} but different G_{dyn} and G_{inst}. Joint observation of (J_{zz}, J_{zx}) identifies both layers when the selected branch of the matrix logarithm is unique.

Proof. Choose any symmetric positive-definite contraction $0 \prec M \prec I$, and set

$$Q = \sqrt{\Delta}(I - M), \qquad A = \Delta^{-1} \log M.$$

In a common eigenbasis of M, write its eigenvalues as $m_i \in (0, 1)$ and define

$$\sigma_i = \sqrt{\Delta}(1 - m_i)\frac{2 \log m_i}{\Delta(m_i^2 - 1)} > 0.$$

Direct integration verifies that $\int_0^\Delta e^{2A_i u}\sigma_i\,du = \sqrt{\Delta}(1-m_i)$, so every M corresponds to a valid $\Sigma \succ 0$ and the chosen Q. Hence

$$S_\Delta = \frac{I-M}{\sqrt{\Delta}}, \qquad H_\Delta = \frac{M-I}{\sqrt{\Delta}} = -S_\Delta, \qquad J_{zx} = S_\Delta^{-1}H_\Delta = -I.$$

Take distinct eigenvalues $m_1 \neq m_2$. The diagonal matrix $M_1 = \mathrm{diag}(m_1, m_2)$ gives a diagonal A_1. Rotating it by a non-permutation orthogonal matrix U gives $M_2 = UM_1U^\top$, for which $A_2 = U(\Delta^{-1}\log M_1)U^\top$ has nonzero off-diagonal entries, although both models have $J_{zx} = -I$. The rotated Σ_2 and Θ_2 likewise have off-diagonal support.

Conversely, the two blocks jointly give

$$S_\Delta = -J_{zz}^{-1}, \quad H_\Delta = -J_{zz}^{-1}J_{zx}, \quad M_\Delta = I + \sqrt{\Delta}H_\Delta.$$

The matrix logarithm identifies A, while $Q_\Delta = \Delta S_\Delta$ is known. For stable A, vectorizing the map $\Sigma \mapsto \int_0^\Delta e^{Au}\Sigma e^{A^\top u}du$ yields eigenfactors

$$\frac{e^{(\lambda_i+\lambda_j)\Delta}-1}{\lambda_i+\lambda_j},$$

with the continuous value Δ when the denominator is zero. Stability makes these factors nonzero, so the map is injective and Σ is identified. □

Remark B.2 (Quantifier order and the range of gaps)**.** *The quantifiers in Proposition 3.1 are "for every Δ, there exists a pair of models," rather than "there exists one pair for every Δ." In the construction, Σ is determined by $Q = \sqrt{\Delta}(I-M)$ after fixing M and the selected gap. The same pair therefore need not have identical cross blocks at other gaps. Indeed, $M_2 = UM_1U^\top$ gives $J_{zx,2}(\Delta') = UJ_{zx,1}(\Delta')U^\top$, and equality requires $J_{zx,1}(\Delta')$ to commute with U. This holds at the constructed gap because $J_{zx} = -I$, but generally fails elsewhere. The construction therefore concerns the raw cross block at one observation gap; the proposition does not determine whether cross blocks at multiple gaps restore identifiability or how demixed and approximate readouts rank at finite sample sizes.*

C Discretization, Finite Gaps, and Measurement Noise

Theorem C.1 (OU covariance cusp and leakage from differencing stencils)**.** *Let A be Hurwitz, $P \succ 0$, and $AP + PA^\top + \Sigma = 0$. For a consistent differencing stencil*

$$D_hX_t = h^{-1}\sum_{\ell=1}^m c_\ell X_{t+a_\ell h}, \quad \sum_\ell c_\ell = 0, \quad \sum_\ell c_\ell a_\ell = 1,$$

define $w_- = \sum_{a_\ell<0} c_\ell a_\ell$, $C_2 = \sum_\ell |c_\ell| a_\ell^2$, and $a_* = \max_\ell |a_\ell|$. The population regression coefficient $B_h = \mathbb{E}[D_hX_tX_t^\top]P^{-1}$ satisfies

$$B_h - A = w_-\Sigma P^{-1} + R_hP^{-1},$$

where

$$\|R_h\| \le \frac{h}{2}C_2\|P\|\|A\|^2 e^{\|A\|a_*h}.$$

Proof. The lagged covariance $K(\tau) = \mathbb{E}[X_{t+\tau}X_t^\top]$ obeys

$$K(\tau) = e^{A\tau}P\ (\tau \ge 0), \qquad K(\tau) = Pe^{-A^\top\tau}\ (\tau \le 0).$$

Thus $K'(0^+) = AP$, $K'(0^-) = -PA^\top$, and $K'(0^+) - K'(0^-) = -\Sigma$. First-order Taylor expansion from the two sides cancels the constant term because $\sum c_\ell = 0$ and yields

$$w_+AP - w_-PA^\top = AP - w_-(AP + PA^\top) = AP + w_-\Sigma.$$

Summing the second-order matrix-exponential remainders gives the stated bound on R_h; right multiplication by P^{-1} completes the proof. □

A forward stencil has $w_- = 0$, so the cusp term is absent. Centered differencing has $w_- = 1/2$, giving $B_h - A = \frac{1}{2}\Sigma P^{-1} + O(h)$. This attribution applies only to front ends that can be represented by the two-sided stencil above. It does not cover PCMCI+, VARLiNGAM, general Granger procedures, or Gaussian-process smoothers without a corresponding derivation.

Theorem C.2 (Exact two-block identities for a linear transition score)**.** *The standardized increment of a linear SDE satisfies*

$$Z_\Delta \mid X_t = x \sim \mathcal{N}(H_\Delta x, S_\Delta).$$

Its terminal-score blocks are

$$J_{zz} = -S_\Delta^{-1}, \qquad J_{zx} = S_\Delta^{-1} H_\Delta,$$

and therefore

$$B_\Delta = -\frac{J_{zz}^{-1} J_{zx}}{\sqrt{\Delta}} = \frac{M_\Delta - I}{\Delta}, \qquad \Sigma_\Delta = -J_{zz}^{-1} = \frac{Q_\Delta}{\Delta}.$$

Proof. The OU variation-of-constants formula gives $X_{t+\Delta} \mid X_t = x \sim \mathcal{N}(M_\Delta x, Q_\Delta)$. The affine transformation $X_{t+\Delta} = x + \sqrt{\Delta} Z_\Delta$ yields the stated H_Δ and S_Δ. Differentiating the Gaussian log-density with respect to z and x gives the two blocks. □

Corollary C.3 (Artificial terminal-noise levels)**.** *If only the terminal variable is perturbed as* $U = \sqrt{\bar{\alpha}} Z_\Delta + \sqrt{1-\bar{\alpha}}\varepsilon$*, then*

$$-\frac{J_{uu}^{-1} J_{ux}}{\sqrt{\bar{\alpha}\Delta}} = \frac{M_\Delta - I}{\Delta}$$

is exactly invariant for every $\bar{\alpha} \in (0, 1]$*, and the analytic covariance correction is*

$$\frac{(-J_{uu})^{-1} - (1-\bar{\alpha})I}{\bar{\alpha}} = \frac{Q_\Delta}{\Delta}.$$

Training samples $\bar{\alpha} \in (0, 1)$*; the identity extends to* $\bar{\alpha} = 1$*, which denotes the unperturbed endpoint.*

Theorem C.4 (Explicit finite-gap remainders)**.** *Let* $a = \|A\|$*. Then*

$$\left\| \frac{M_\Delta - I}{\Delta} - A \right\| \le \frac{\Delta}{2} a^2 e^{a\Delta},$$

and

$$\frac{Q_\Delta}{\Delta} = \Sigma + \frac{\Delta}{2}(A\Sigma + \Sigma A^\top) + R_Q(\Delta), \quad \|R_Q(\Delta)\| \le \frac{2}{3}\Delta^2 a^2 \|\Sigma\| e^{2a\Delta}.$$

Proof. The first inequality follows from $\|e^B - I - B\| \le \frac{1}{2}\|B\|^2 e^{\|B\|}$ with $B = A\Delta$. Let $f(u) = e^{Au}\Sigma e^{A^\top u}$. Then $f'(0) = A\Sigma + \Sigma A^\top$ and $\|f''(u)\| \le 4a^2\|\Sigma\| e^{2au}$. A second-order Taylor expansion of f, integrated over $[0, \Delta]$ and divided by Δ, gives the second result. In particular, if $A_{ji} = 0$ but $(A^2)_{ji} \neq 0$, the (j, i) entry without a direct edge still contains $\frac{\Delta}{2}(A^2)_{ji} + O(\Delta^2)$. □

Numerical verification of the finite-gap expansion. We directly compute the population readout $(e^{A\Delta} - I)/\Delta$ on twenty stable linear systems. For the $1 \to 3$ readout induced by the chain $1 \to 2 \to 3$, the leading expansion coefficient is $\frac{1}{2}(A^2)_{31} = -0.24$, and the relative error at the smallest gap is below 1%. The readout varies approximately linearly with Δ, while the remainder after subtracting the leading term varies approximately quadratically. Across all systems and gaps, nonzero spurious entries occur exactly at $\mathrm{supp}(A^2) \setminus \mathrm{supp}(A)$. These results verify the constant, order, and support location in Theorem C.4. Because the learned score does not yet resolve this small second-order term reliably, this experiment evaluates only the analytic population readout.

Assumption C.5 (Nonlinear short-time score blocks)**.** *On a compact evaluation set, suppose that for some* $\nu > 0$*,*

$$J_{zz} = -\Sigma^{-1} + E_{zz}, \qquad J_{zx} = \Sigma^{-1}\sqrt{\Delta}\nabla F(x) + E_{zx},$$

with $\|E_{zz}\| \le c_{zz}\Delta$ *and* $\|E_{zx}\| \le c_{zx}\Delta^{1/2+\nu}$*.*

This is an explicit assumption on the score Jacobian. Neither short-time expansions of conditional moments nor the function-value risk of denoising score matching (DSM) alone supplies the required C^1 convergence rate.

Proposition C.6 (Nonlinear remainder under Assumption C.5)**.** *If* $\lambda_{\max}(\Sigma)c_{zz}\Delta < 1$*, then*

$$\|B_\Delta(x) - \nabla F(x)\| \le \frac{\lambda_{\max}(\Sigma)}{1 - \lambda_{\max}(\Sigma)c_{zz}\Delta}\left(c_{zz}\Delta\|\nabla F(x)\| + c_{zx}\Delta^\nu\right),$$

and

$$\|\Sigma_\Delta - \Sigma\| \le \frac{\lambda_{\max}(\Sigma)^2 c_{zz}\Delta}{1 - \lambda_{\max}(\Sigma)c_{zz}\Delta}.$$

Proof. Write $\Theta = \Sigma^{-1}$. The Neumann condition gives

$$\|(\Theta - E_{zz})^{-1}\| \le \frac{\lambda_{\max}(\Sigma)}{1 - \lambda_{\max}(\Sigma)c_{zz}\Delta}.$$

Substituting J_{zx} into $-J_{zz}^{-1}J_{zx}/\sqrt{\Delta}$ and applying the resolvent identity $(\Theta - E)^{-1} - \Theta^{-1} = (\Theta - E)^{-1}E\Theta^{-1}$ gives both inequalities. □

Proposition C.7 (EIV bias from starting-state measurement noise)**.** *Let* $Y_t = X_t + \eta_t$*, where* $\eta_t \sim \mathcal{N}(0, R)$ *is independent across time, and let* $\mathrm{Cov}(X_t) = P$*. For* $Z_{\mathrm{obs}} = (Y_{t+\Delta} - Y_t)/\sqrt{\Delta}$*,*

$$\mathrm{Cov}(Z_{\mathrm{obs}}, Y_t) = \frac{M_\Delta P - P - R}{\sqrt{\Delta}},$$

and the dynamics readout obtained by demixing the Gaussian conditional mean satisfies

$$B_{\mathrm{obs}}(\Delta) = -\frac{R(P+R)^{-1}}{\Delta} + AP(P+R)^{-1} + O(\Delta).$$

Proof. Independence of measurement noise across times gives $\mathrm{Cov}(Y_{t+\Delta} - Y_t, Y_t) = M_\Delta P - P - R$. Multiplying the conditional-mean coefficient on the right by $(P+R)^{-1}$, dividing by $\sqrt{\Delta}$, and using $M_\Delta - I = \Delta A + O(\Delta^2)$ yields the result. The corresponding Schur complement also satisfies

$$\lim_{\Delta \to 0} \Delta\ \mathrm{Cov}(Z_{\mathrm{obs}} \mid Y_t) = 2R - R(P+R)^{-1}R,$$

so the intercept from a simple slope–intercept fit cannot be interpreted directly as $2R$. □

D From a Learned Score to Two-Layer Support Recovery

Fix a physical observation gap Δ and an artificial noise level $\bar{\alpha} \in (0, 1)$, and set $\beta = 1 - \bar{\alpha}$. In standardized coordinates, let $U = \sqrt{\bar{\alpha}}Z^s + \sqrt{\beta}\varepsilon$. For a clean held-out transition pair $W = (X^s, Z^s)$, define the population blocks

$$C_* = -\operatorname{sym}\mathbb{E}_{W,\varepsilon}[\partial_u s(U \mid X^s)], \quad R_*(w) = \mathbb{E}_\varepsilon[\partial_x s(U \mid X^s) \mid W = w],$$

and

$$B_*(w) = \frac{C_*^{-1}R_*(w)}{\sqrt{\bar{\alpha}\Delta}}, \quad \Sigma_* = \bar{\alpha}^{-1}(C_*^{-1} - \beta I), \quad \Theta_* = \Sigma_*^{-1}.$$

In the one-step Euler–Maruyama Gaussian reference with additive noise, $Z^s \mid X^s = x \sim \mathcal{N}(\sqrt{\Delta}F^s(x), \Sigma^s)$, we have $C_* = (\bar{\alpha}\Sigma^s + \beta I)^{-1}$, $B_*(w) = J_F^s(x)$, and $\Sigma_* = \Sigma^s$. These identities do not depend on the realized z or ε.

Assumption D.1 (Population finite-gap residuals)**.** *On the evaluation region* $\mathcal{K}_W$*, suppose* $C_* \succ 0$ *and there exist* $b_{\mathrm{dyn}}(\Delta), b_{\mathrm{inst}}(\Delta) \ge 0$ *such that*

$$\sup_{w=(x,z)\in\mathcal{K}_W} \|B_*(w) - J_F^s(x)\| \le b_{\mathrm{dyn}}(\Delta), \qquad \|\Sigma_* - \Sigma^s\| \le b_{\mathrm{inst}}(\Delta).$$

For linear models, Theorem C.4 controls these terms explicitly. In nonlinear models, Assumption D.1 concerns population blocks after artificial terminal perturbation and aggregation under the evaluation distribution, as used by the implementation. Assumption C.5 instead describes an unperturbed, pointwise prototype; the two assumptions therefore concern different objects.

Let $\widehat{s}$ be the learned score and suppose its first score-derivative errors satisfy

$$\sup_{x,u} \|\partial_u(\widehat{s} - s)\| \le e_z, \qquad \sup_{x,u} \|\partial_x(\widehat{s} - s)\| \le e_x.$$

An L^2 error bound on score values alone is insufficient: high-frequency perturbations can make derivatives diverge while function values converge. Combine network error, aggregation over held-out trajectory segments, finite artificial-noise Monte Carlo error, and ridge regularization into

$$\|\widehat{C}_{\mathrm{dyn}} - C_*\| \le \eta_{z,\mathrm{dyn}}, \quad \|\widehat{C}_{\mathrm{inst}} - C_*\| \le \eta_{z,\mathrm{inst}}, \quad \sup_w \|\widehat{R}(w) - R_*(w)\| \le \eta_x.$$

Theorem D.2 (From Jacobian-block error to structural-matrix error)**.** *Let $m = \lambda_{\min}(C_*)$ and $L_B = \sup_{w \in \mathcal{K}_W} \|B_*(w)\|$, and suppose $\eta_{z,\mathrm{dyn}}, \eta_{z,\mathrm{inst}} < m$. The implemented readouts*

$$\widehat{B}(w) = \frac{\widehat{C}_{\mathrm{dyn}}^{-1}\widehat{R}(w)}{\sqrt{\bar{\alpha}\Delta}}, \quad \widehat{\Sigma}^s = \bar{\alpha}^{-1}(\widehat{C}_{\mathrm{inst}}^{-1} - \beta I), \quad \widehat{\Theta}^s = (\widehat{\Sigma}^s)^{-1}$$

satisfy

$$\sup_w \|\widehat{B}(w) - J_F^s(x)\| \le \delta_{\mathrm{dyn}},$$

where

$$\delta_{\mathrm{dyn}} = b_{\mathrm{dyn}}(\Delta) + \frac{\eta_x/\sqrt{\bar{\alpha}\Delta} + \eta_{z,\mathrm{dyn}} L_B}{m - \eta_{z,\mathrm{dyn}}}.$$

Moreover,

$$\|\widehat{\Sigma}^s - \Sigma^s\| \le \delta_\Sigma, \qquad \delta_\Sigma = b_{\mathrm{inst}}(\Delta) + \frac{\eta_{z,\mathrm{inst}}}{\bar{\alpha} m (m - \eta_{z,\mathrm{inst}})}.$$

If $\mu = \lambda_{\min}(\Sigma^s)$ and $\delta_\Sigma < \mu$, then

$$\|\widehat{\Theta}^s - \Theta^s\| \le \delta_{\mathrm{inst}} := \frac{\delta_\Sigma}{\mu(\mu - \delta_\Sigma)}.$$

Proof. Weyl's inequality gives $\lambda_{\min}(\widehat{C}_{\mathrm{dyn}}) \ge m - \eta_{z,\mathrm{dyn}}$. Using $R_*(w) = C_* \sqrt{\bar{\alpha}\Delta} B_*(w)$,

$$\widehat{B} - B_* = \widehat{C}_{\mathrm{dyn}}^{-1} \left[\frac{\widehat{R} - R_*}{\sqrt{\bar{\alpha}\Delta}} - (\widehat{C}_{\mathrm{dyn}} - C_*) B_* \right].$$

Taking operator norms and adding the population residual in Assumption D.1 gives the dynamics bound. For the synchronous block, the resolvent identity gives

$$\|\widehat{C}_{\mathrm{inst}}^{-1} - C_*^{-1}\| \le \frac{\eta_{z,\mathrm{inst}}}{m(m - \eta_{z,\mathrm{inst}})}.$$

Dividing by $\bar{\alpha}$ and adding b_{inst} gives the covariance bound. A second application of Weyl's inequality and the resolvent identity yields

$$\|(\widehat{\Sigma}^s)^{-1} - (\Sigma^s)^{-1}\| \le \|\widehat{\Sigma}^{-1}\| \, \|\widehat{\Sigma} - \Sigma\| \, \|\Sigma^{-1}\|,$$

which proves the precision-matrix bound. □

Proof of Corollary 4.1. For a dynamics entry, the reverse triangle inequality in L^2 gives

$$\left| \|\widehat{B}_{ji}\|_{L^2} - \|J^s_{F,ji}\|_{L^2} \right| \le \|\widehat{B}_{ji} - J^s_{F,ji}\|_{L^2} \le \delta_{\mathrm{dyn}}.$$

The operator-norm bound on the precision matrix implies an entrywise bound. Hence every null-edge score is at most the corresponding δ, while every true-edge score is at least $\gamma - \delta$; the stated margin strictly separates the two sets. □

A finite number of evaluation trajectories. Suppose evaluation states come from E independent trajectory segments. Let $V_{e,ji}$ be the empirical average of $J^s_{F,ji}(X)^2$ on segment e, and take $w_e \geq 0$ with $\sum_e w_e = 1$. The effective number of independent blocks is $n_{\text{eff}} = 1/\sum_e w_e^2$. If $|J^s_{F,ji}| \leq H$, then a weighted Hoeffding inequality and a union bound over the $D(D-1)$ directed candidates give, with probability at least $1-\zeta$,

$$\max_{i \neq j} \left| \sum_e w_e V_{e,ji} - g^2_{ji} \right| \leq H^2 \sqrt{\frac{\log(2D(D-1)/\zeta)}{2n_{\text{eff}}}}.$$

Likewise, the aggregated curvature block is controlled by a self-adjoint matrix Bernstein inequality, and the cross block averaged over artificial-noise draws is controlled by a rectangular matrix Bernstein inequality. More artificial-noise samples reduce only conditional Monte Carlo error; they do not increase the number of independent trajectory segments. The errors in Theorem D.2 therefore decompose into population finite-gap bias, architecture-dependent C^1 score error, sampling error from trajectory segments and artificial noise, the $1/\sqrt{\bar{\alpha}\Delta}$ amplification in demixing, and the conditioning of the precision-matrix inversion.

Uncorrected synchronous readout. Theorem D.2 covers the analytically corrected synchronous covariance and precision matrices. The prespecified uncorrected curvature readout ranks edges approximately by the partial-correlation ordering of $(\bar{\alpha}\Sigma^s + \beta I)^{-1}$; the experiments evaluate its finite-sample behavior.

E Experimental Protocols and Supplementary Results

E.1 Main synthetic generator and fixed parameters

For each system seed, we first sample a node ordering and then draw $m_D = \min\{2D, D(D-1)/2\}$ edges uniformly without replacement from all forward candidates. Here "ER-style" denotes a fixed-edge-count construction under a random ordering, rather than the classical independent-edge $G(n,p)$ model. In particular, at $D = 5$, $m_D = 10$ equals the maximum number of edges allowed by the ordering. Given the indegree d_j^- of node j, the drift is

$$F_j(x) = -a_j x_j + \frac{1}{\sqrt{\max(d_j^-, 1)}} \sum_{i \to j} w_{ji}\, \phi_{ji}(2x_i), \qquad \phi_{ji}(u) \in \{u, \tanh u, u^2 - 1\}. \tag{10}$$

The three mechanisms are approximately balanced across edges, the a_j values are evenly spaced over $[1.4, 2.0]$, and edge signs are equiprobable with $|w_{ji}| \sim U[1.2, 1.8]$. The Jacobian of a quadratic mechanism changes sign with the state, testing whether state aggregation preserves quadratic edges whose derivatives change sign. For the synchronous layer, we select $\lfloor D/2 \rfloor$ disjoint correlated pairs from a random node permutation. The matrix $\widetilde{\Sigma}$ has unit diagonal, covariance 0.6 within each pair, and zero elsewhere, and Equation (1) uses $L = 0.4\ \mathrm{chol}(\widetilde{\Sigma})$. Because $\widetilde{\Sigma}$ consists of disjoint 2×2 blocks, $\mathrm{supp}_{\text{off}}(\widetilde{\Sigma}^{-1})$ equals the implanted pairing.

Table 4 summarizes the data protocol for the main recovery experiment. Each trajectory segment is initialized independently from $\mathcal{N}(0, 0.35^2 I)$ and integrated for 600 burn-in steps, after which 401 states (400 fine-step intervals) are retained. Training and held-out sets use independent random streams derived from the seed. Training gaps are first placed on an eight-point logarithmic grid over $[0.02, 0.5]$ and then rounded to integer multiples of $h = 0.002$. Transitions never cross segment boundaries. Each coordinate is clipped to $[-8, 8]$ after every Euler–Maruyama update; the clipping rate is below 10^{-3}. All methods receive the same trajectories.

E.2 Irregular-sampling protocol and architectural controls

The irregular-sampling experiment keeps the latent SDE paths fixed and changes only the observation timestamps. Training gaps are $\{0.04, 0.08, 0.10, 0.12, 0.16\}$; unseen interpolation gaps within this range are $\{0.06, 0.14\}$; and mild extrapolation gaps are $\{0.18, 0.22, 0.26\}$. Direct estimators include CAST, the matched Δ-Gaussian model, and ridge regression that receives Δ explicitly.

Table 4: Fixed generator for the main nonlinear ordered ER-style systems. The dimensional-scaling experiment uses the same graph, drift, and innovation construction, with separately prespecified dimensions, seeds, and training budgets.

Item	Fixed setting
Dimensions and main seeds	$D \in \{5, 10, 15\}$; seeds 210–219
Dynamics edges	$m_D = \min\{2D, D(D-1)/2\}$
Edge mechanisms	Approximately balanced linear / tanh / quadratic
Edge weights and self-decay	$\lvert w_{ji}\rvert \sim U[1.2, 1.8]$, equiprobable signs; a_j evenly spaced over $[1.4, 2.0]$
Synchronous innovations	$\lfloor D/2 \rfloor$ disjoint pairs, $\widetilde{\Sigma}_{ij} = 0.6$, $\Sigma = 0.4^2\widetilde{\Sigma}$
Numerical integration	Euler–Maruyama, $h = 0.002$, coordinate-wise clipping to $[-8, 8]$
Trajectory segments	100 training, 50 held out; 600 burn-in steps, 400 retained intervals
Training transitions	40,000 unique triplets; 8 log-spaced observation gaps over $[0.02, 0.5]$
Main readout	$\Delta \in \{0.076, 0.2\}$; 2,400 held-out transition pairs per gap

Two imputation controls first complete a regular grid using linear interpolation or a CSDI-style model and then apply a prespecified finite-difference score-based structure-discovery pipeline. This pipeline uses the matched two-block structural scoring rule and therefore outputs both dynamics and synchronous graphs, but its score is defined on state–derivative pairs rather than observed transition pairs.

Across training, interpolation, and extrapolation gaps, CAST obtains dynamics AUPRC of at least 0.991 and synchronous top-k F1 of 1.000. The matched Gaussian lies in the same near-ceiling range for dynamics, while its synchronous F1 decreases from 1.000 to 0.920; ridge-regression dynamics AUPRC ranges from 0.835 to 0.850. Under this protocol, where approximately 80% of base-grid timestamps are missing, the linear-interpolation and CSDI-style imputation controls obtain dynamics AUPRC ranges of 0.823–0.827 and 0.313–0.315, respectively. This comparison evaluates direct modeling at the observed gaps against impute-then-discover architectures; it does not characterize the general imputation ability of CSDI.

E.3 Implementations of external baselines

Table 5 lists the external methods included in the final numerical comparisons and their implementation sources. The Δ-Gaussian model and demixing variants are matched controls introduced here and are therefore omitted.

Table 5: Implementation source, locked version, and adaptation of each external baseline. All methods are evaluated only through the dynamics-layer scores they provide; none outputs the off-diagonal support of innovation precision as a synchronous layer.

Method	Implementation	Locked version	Adaptation
PCMCI+	Tigramite	`tigramite ==5.2.10.1`	ParCorr; the main ER experiment uses `tau_max=3` and `pc_alpha=0.05`, with the sum of absolute scores over all lags as the dynamics-edge score.
VARLiNGAM	lingam	`lingam==1.13.0`	`VARLiNGAM(lags=3)`; absolute lagged adjacency matrices are aggregated and only the dynamics layer is scored.
DYNOTEARS	CausalNex	`causalnex ==0.12.1`	`from_pandas_dynamic`, lags 1–3; independent trajectory boundaries are preserved and lag 0 is excluded.
NGM	official public implementation	Public version used in the experiments	Neural-ODE, group-lasso, and continuous edge-score readout are retained; only batch sampling is changed to within-segment windows to avoid concatenating independent trajectories.
CaDyT	official public implementation	Public version used in the experiments	DblMass uses GPScore/UGDS, AB2/AB3, 1,000 epochs, and `alpha=10`.

E.4 Demixing readouts and synchronous semantics

Table 2 in the main text uses a separate $D = 15$ mechanism-study protocol. Its ten systems use seeds 1100–1109, 2,000 independent training episodes and 500 held-out episodes, and renewal-sampled gaps $\{0.04, 0.08, 0.10, 0.12, 0.16\}$. Four of the seven synchronous dyads are constrained to overlap a dynamics dyad, while the other three do not. The $D = 15$ row of Table 1 instead uses seeds 210–219, 100 training and 50 held-out trajectory segments, readout gaps $\{0.076, 0.2\}$, and randomly paired synchronous innovations. Both protocols use 30 directed edges and the same drift and innovation parameters, but their absolute F1 values are not paired comparisons. Within the mechanism study, all readout variants share each fitted model and evaluation sample. We compare the raw cross-block readout, diagonal curvature scaling, and full curvature demixing, and also measure the fraction of ordered variable pairs connected only by a synchronous edge that are misclassified as dynamics edges. Full curvature demixing substantially reduces these errors, showing that its benefit extends beyond aggregate ranking: it also prevents synchronous relations from being read as directed dynamics edges. Table 6 further compares paired and chain graph families at $\rho = 0.8$ and reports the false-positive rate (FPR). Under the stronger indirect-path effects in the chain family, diagonal curvature scaling captures most of the ranking gain over the raw cross-block readout, while the additional benefit of full curvature demixing appears mainly in quadratic-edge recall. CAST uses full curvature demixing because Proposition 3.1 shows that diagonal curvature scaling alone cannot identify both layers under general non-diagonal innovations; the scaling corresponds to a diagonal approximation of the curvature.

Table 6: Demixing variants at $\rho = 0.8$ (paired/chain graph families on the two sides of each slash; means over ten seeds). Lower FPR is better.

Variant	Dynamics AUPRC↑	Top-k F1↑	Quadratic-edge recall↑	FPR↓
Raw cross-block readout	.887/.875	.810/.785	.821/.807	.054/.061
Diagonal curvature scaling	.921/.919	.855/.830	.867/.864	.041/.049
Full curvature demixing (CAST)	.926/.913	.855/.815	.912/.910	.041/.053

Precision semantics of the synchronous graph. To distinguish precision support from covariance support, we use three disjoint three-node precision chains. Their precision matrix contains six chain edges, while the corresponding covariance matrix also has nonzero entries between the two endpoints of each chain. A descriptive audit of the 50 saved terminal-curvature matrices treats the system seed as the independent unit and nests noise-correlation levels within each seed. Precision-support AUPRC is 0.6913, compared with 0.5977 for covariance support; the paired difference is $+0.0936$, with a seed-bootstrap 95% interval of $[+0.0386, +0.1410]$. Thus the learned support ranking more closely follows innovation precision. This audit evaluates support ranking, the estimand used for synchronous-graph recovery.

E.5 Additional design comparisons

Averaging across noise levels outperforms a fixed low-noise level and approaches the best result among all individual noise levels. This aggregation reduces dependence on any one artificial noise level. On the same transitions, raw increments, z-standardized increments, and finite-difference derivatives obtain dynamics AUPRC values of $0.905, 0.904, 0.922$, respectively. Their differences on clean data are small, while z-standardized increments give the highest quadratic-edge recall (0.917, versus $0.902/0.869$). Analytic noise correction mainly improves numerical covariance calibration and stability across noise levels, with little effect on support ranking. These results correspond to the target scaling in Section 3.2 and the stabilization choices in Section 3.3.

E.6 Published single-layer baselines

Table 7 compares discrete-time methods and a nonlinear neural ODE on the same trajectories as the main table. VARLiNGAM encounters two numerical-decomposition failures at $D = 15$, so that entry aggregates the other eight runs; every other method completes all ten. These methods output only the dynamics layer, and synchronous-layer results are therefore not reported.

Table 7: Dynamics-graph recovery for published single-layer baselines and CAST on the same trajectories. "Successful" reports completed runs out of 10. Continuous edge scores use top-k selection with the true edge count; quadratic-edge recall is computed on that selection.

D	Method	Successful	G_{dyn} top-k F1	AUPRC	nSHD	Quadratic-edge recall
5	**CAST**	10/10	**0.960**	**0.992**	**0.032**	**1.000**
	NGM	10/10	0.500	0.573	0.400	0.517
	PCMCI+	10/10	0.500	0.538	0.400	0.600
	VARLiNGAM	10/10	0.610	0.651	0.312	0.708
	DYNOTEARS	10/10	0.460	0.529	0.432	0.617
10	**CAST**	10/10	**0.850**	**0.933**	**0.060**	**0.831**
	NGM	10/10	0.340	0.344	0.264	0.257
	PCMCI+	10/10	0.235	0.268	0.306	0.160
	VARLiNGAM	10/10	0.275	0.285	0.290	0.314
	DYNOTEARS	10/10	0.170	0.231	0.332	0.179
15	**CAST**	10/10	**0.733**	**0.809**	**0.071**	**0.890**
	NGM	10/10	0.193	0.204	0.215	0.150
	PCMCI+	10/10	0.207	0.208	0.212	0.210
	VARLiNGAM	8/10	0.238	0.206	0.203	0.362
	DYNOTEARS	10/10	0.117	0.161	0.236	0.100

E.7 DblMass: dynamics-graph recovery under strong measurement noise

Under strong measurement noise, CAST requires no derivative presmoothing and achieves the highest AUPRC and lowest nSHD among the published baselines for dynamical structure learning considered here, while also outperforming the matched Δ-Gaussian. CAST averages the two readout gaps $\Delta \in \{0.8, 1.0\}$ frozen in pilot experiments, avoiding the smallest gaps most affected by the $1/\Delta$ noise amplification, as predicted by Proposition C.7. CaDyT does not output a synchronous graph, so the corresponding metric is N/A.

Table 8: Dynamics-graph recovery on DblMass under strong measurement noise ($\sigma_{\text{obs}} = 0.5$; ten fixed seeds). This benchmark contains no stochastic innovation term, so all G_{inst} entries are N/A. The matched Δ-Gaussian and CAST share the observed trajectories, gap windows, and training budget, differing only in conditional-density parameterization. The bold row is CAST's fixed configuration.

Method	G_{dyn} AUPRC ↑	nSHD ↓
CAST	**0.763**	**0.200**
Δ-Gaussian	0.662	0.325
DYNOTEARS	0.737	0.260
NGM	0.723	0.275
VARLiNGAM	0.712	0.250
PCMCI+	0.563	0.363
CaDyT	0.616	0.250

F External-System Results

Table 9 summarizes dynamics-layer comparisons on NetSim and CausalDynamics-CI. Table 10 further decomposes the two-layer CausalDynamics-CI results by innovation-precision structure.

Table 9: Dynamics-layer results on two external benchmarks. NetSim first aggregates subjects within each of three simulations and then takes an unweighted mean; CausalDynamics-CI aggregates ten system triplets. NetSim has no synchronous-layer ground truth. The innovation precision in CausalDynamics-CI is constructed for this study, so cross-dataset comparisons concern only the dynamics layer.

Benchmark	Method	Statistical unit	G_{dyn} AUPRC ↑	nSHD ↓
NetSim	CAST	3 simulations (50 subject-level fits each)	0.786	0.067
	Δ-Gaussian NLL	3 simulations (50 subject-level fits each)	**0.853**	**0.040**
	NGM	3 simulations (one pooled 50-subject fit each)	0.520	0.130
	VARLiNGAM	3 simulations (50 subject-level fits each)	0.305	0.175
	PCMCI+	3 simulations (50 subject-level fits each)	0.239	0.206
	DYNOTEARS	3 simulations (50 subject-level fits each)	0.247	0.203
CausalDynamics-CI	CAST	10 triplets (60 configurations)	0.612	0.171
	NGM	10 triplets (60 configurations)	**0.713**	**0.114**
	VARLiNGAM	10 triplets (60 configurations)	0.528	0.174
	Δ-Gaussian NLL	10 triplets (60 configurations)	0.451	0.219
	PCMCI+	10 triplets (60 configurations)	0.271	0.269
	DYNOTEARS	10 triplets (60 configurations)	0.334	0.258

Table 10: CausalDynamics-CI results by innovation-precision structure. Each entry aggregates ten system triplets and two observation gaps; the empty precision graph has no true synchronous edges. The final column is the mean number of covariance-only chain endpoints entering the top six.

Precision structure	Method	Dynamics AUPRC	G_{inst} top-k F1	G_{inst} AUPRC	Chain endpoints in top six
Empty	CAST	0.589	N/A	N/A	N/A
	Δ-Gaussian NLL	0.430	N/A	N/A	N/A
Paired	CAST	0.618	0.833	0.911	0.00
	Δ-Gaussian NLL	0.456	0.983	0.982	0.00
Chain	CAST	0.630	0.908	0.948	0.40
	Δ-Gaussian NLL	0.465	0.967	0.965	0.10

CAST has higher dynamics AUPRC than the Gaussian control under the empty, paired, and chain precision structures. The synchronous results reverse this ordering: the Gaussian control is more accurate on the paired and chain structures and selects fewer covariance-only endpoints as synchronous edges in the chain case. The two parameterizations therefore have different finite-sample advantages across structural layers. Closed-form Gaussian estimation recovers synchronous precision more accurately for these near-Gaussian conditional transitions, whereas CAST's flexible score parameterization improves dynamics recovery. Because the empty precision graph contains no true

synchronous edges, its synchronous score is used only to check for spurious edges and not for absolute cross-model performance comparisons. The drift equations come from the official external source, while the synchronous Θ is constructed for this study. This experiment therefore tests transfer of the two-block readout to external drift mechanisms while retaining constructed ground truth for the synchronous layer.

G Readout Mechanism under Multiplicative Noise

Additive innovations are central to the two-block semantics. To quantify departures from this assumption, we construct a multiplicative-noise family that preserves both graph targets while breaking only state independence of the diffusion, and measure how the fixed readout changes.

A multiplicative perturbation that preserves graph structure. For each pre-generated nonlinear ER-style system, we derive $\Sigma_0 = \Theta_0^{-1}$ from its fixed sparse precision Θ_0 and replace the additive diffusion by

$$L_\kappa(x) = D_\kappa(x)\ \mathrm{chol}(\Sigma_0), \qquad \Theta_\kappa(x) = D_\kappa(x)^{-1}\Theta_0 D_\kappa(x)^{-1},$$

where $D_\kappa(x)$ is positive diagonal and

$$d_{\kappa,i}(x) = \frac{\exp\big(\kappa \tanh(x_i/s_i)\big)}{\sqrt{\mathbb{E}_0[\exp(2\kappa \tanh(X_i/s_i))]}}.$$

The scales s_i and normalization expectations are fixed using independent additive calibration trajectories, aligning the mean innovation scale across levels so that a pure magnitude change is not mistaken for a multiplicative effect. Positive diagonal congruence preserves off-diagonal zero patterns, so the support and partial correlations of $\Theta_\kappa(x)$ are exactly invariant across all states and all κ (maximum measured deviation 1.11×10^{-16}). Thus G_{dyn} and G_{inst} remain fixed, well-defined ground truth; only state independence of the diffusion is violated. We use $\kappa \in \{0, 0.15, 0.30, 0.60\}$.

Structural recovery under multiplicative noise. Table 11 reports dynamics and synchronous AUPRC for each readout. CAST's global readout degrades only slightly on both layers, and the 95% paired-bootstrap interval for the synchronous layer relative to $\kappa = 0$ includes zero. The matched Δ-Gaussian degrades more on both layers. Pointwise inversion becomes markedly unstable for $\kappa \geq 0.30$, whereas the true-model local readout recovers the ground truth exactly throughout. The degradation therefore comes from finite-sample estimation rather than a change in the target graphs.

Table 11: Two-layer AUPRC under graph-preserving multiplicative noise (means over ten systems). "Paired difference" is the mean change from $\kappa = 0$; * marks a 95% paired-bootstrap interval that excludes zero. The pointwise scheme has numerical failures at the low-noise level for $\kappa = 0.60$.

	$\kappa = 0$	0.15	0.30	0.60
Dynamics-layer AUPRC				
CAST (global pooling)	0.928	0.922	0.915	0.900
Paired difference	–	-0.006	-0.013^*	-0.028^*
CAST (pointwise, diagnostic)	0.868	0.895	0.686	0.307
Paired difference	–	$+0.026$	-0.182	-0.561^*
Δ-Gaussian	0.747	0.749	0.732	0.692
Paired difference	–	$+0.003$	-0.015	-0.055^*
True-model local readout	1.000	1.000	1.000	1.000
Synchronous-layer AUPRC				
CAST (global pooling)	0.993	0.993	0.993	0.982
Paired difference	–	$+0.000$	$+0.000$	-0.011
Δ-Gaussian	0.993	0.993	0.987	0.954
Paired difference	–	$+0.000$	-0.006	-0.039

Modal readout under a local Gaussian reference. The demixing identity explains the limited degradation of the global readout. Under a one-step Euler–Maruyama local Gaussian reference, let $\mu(x) = \sqrt{\Delta}F(x)$. At artificial noise level $\bar{\alpha}$, the perturbed standardized increment follows $z_{\bar{\alpha}} \mid x \sim \mathcal{N}(\sqrt{\bar{\alpha}}\,\mu(x), S_{\bar{\alpha}}(x))$, where $S_{\bar{\alpha}}(x) = \bar{\alpha}\,\Sigma_\kappa(x) + (1-\bar{\alpha})I$. Hence

$$-J_{zz}^{-1}J_{zx} = \sqrt{\bar{\alpha}\Delta}\,\nabla F(x) \;-\; S_{\bar{\alpha}}(x)\,[\partial_x S_{\bar{\alpha}}(x)^{-1}]\,(z_{\bar{\alpha}} - \sqrt{\bar{\alpha}}\,\mu(x)), \tag{11}$$

where the second term is linear in displacement from the conditional mode and vanishes exactly at $z_{\bar{\alpha}} = \sqrt{\bar{\alpha}}\,\mu(x)$. Holding x fixed and varying $z_{\bar{\alpha}} - \sqrt{\bar{\alpha}}\,\mu(x)$, the analytic-score error grows linearly with displacement and κ and falls to machine precision at the mode. The learned score shows the same pattern: its modal error remains nearly constant across κ, while its displacement slope increases monotonically with multiplicative strength. Near the mode, the practical readout is therefore governed mainly by network-estimation error, with state-dependent diffusion contributing an additional error that grows with displacement.

H Dimensional-Scaling Results

Table 12 reports complete means for all six methods at each dimension. For each method, the first row gives dynamics-graph AUPRC, top-k F1 using the true edge count, and nSHD; the second gives quadratic-edge recall and synchronous-layer top-k F1. NGM, PCMCI+, VARLiNGAM, and DYNOTEARS output only the dynamics layer, so their synchronous entries are N/A.

Table 12: Complete results on the pre-generated ER-style dimensional-scaling systems. For each method, the first row reports dynamics AUPRC / top-k F1 / nSHD and the second reports quadratic-edge recall / G_{inst} top-k F1. Higher is better except for nSHD. †VARLiNGAM encounters three numerical-decomposition failures at $D = 50$; the entry averages the other seven successful runs without imputing zeros.

Method	$D = 10$	$D = 30$	$D = 50$
CAST	0.910 / 0.830 / 0.068 0.895 / 0.980	0.656 / 0.608 / 0.052 0.630 / 1.000	0.465 / 0.503 / 0.040 0.397 / 0.996
Δ-Gaussian	0.786 / 0.705 / 0.118 0.850 / 0.980	0.451 / 0.460 / 0.072 0.490 / 0.920	0.256 / 0.312 / 0.055 0.304 / 0.760
NGM	0.283 / 0.265 / 0.294 0.198 / N/A	0.113 / 0.137 / 0.115 0.130 / N/A	0.065 / 0.096 / 0.072 0.093 / N/A
PCMCI+	0.279 / 0.285 / 0.286 0.240 / N/A	0.106 / 0.120 / 0.117 0.073 / N/A	0.059 / 0.069 / 0.074 0.050 / N/A
VARLiNGAM	0.189 / 0.125 / 0.350 0.125 / N/A	0.056 / 0.033 / 0.129 0.033 / N/A	0.033 / 0.027 / 0.078 † 0.027 / N/A
DYNOTEARS	0.237 / 0.200 / 0.320 0.152 / N/A	0.075 / 0.068 / 0.124 0.065 / N/A	0.048 / 0.049 / 0.076 0.063 / N/A

I Computational Complexity

Let the fixed number of training steps and batch size be S and B, let $|\theta|$ denote the number of network parameters, and let n_{eval} be the number of readout states. CAST training costs $O(SB|\theta|)$. Computing the full score Jacobian at every state and inverting the aggregated curvature matrix costs $O(n_{\text{eval}}D|\theta| + D^3)$ at readout. The implementation uses hidden width $H = \max(192, 16D)$. At the largest setting, $D = 50$, the model has 1.45M parameters and uses approximately 137 MB of peak memory. Training remains the dominant cost over the tested range; the D^3 inversion has not yet become a bottleneck.

The procedures also differ structurally. CAST uses one network and one training stage, followed by a single Jacobian readout without graph search, pruning, or missing-value completion. This simplicity follows from the differential identity underlying the readout (Section 3.1). A single run of a linear discrete baseline is one to two orders of magnitude faster than CAST.

J SCOPE AND LIMITATIONS

The two-block semantics require the observed coordinates to form a Markov diffusion driven by additive innovations, without lagged latent confounding and with complete observations at shared timestamps. Outside this setting, the curvature and cross blocks remain computable but no longer correspond to the two structures defined here. The graph G_{inst} encodes only undirected conditional dependence among innovations; it cannot distinguish common drivers, fast latent processes, and instantaneous causation. Existing public benchmarks do not provide ground truth for non-diagonal innovation precision (Dibaeinia & Sinha, 2020; Schaffter et al., 2011; Cannoodt et al., 2021; Li et al., 2011). Cross-system evidence therefore primarily concerns the dynamics layer, while systems with exact constructed ground truth validate the synchronous layer. Finite gaps also fold multihop effects into $O(\Delta)$ spurious direct edges. These small second-order terms are verifiable in the population readout but are not yet resolved reliably by the learned score. Strong measurement noise amplifies EIV bias in small-gap dynamics readouts; task-specific derivative smoothing can be preferable when measurement noise dominates.

Population finite-gap bias and C^1 score-estimation error enter the structural readout differently. Smaller readout gaps reduce spurious multihop edges but amplify measurement noise, motivating aggregation over several prespecified gaps. The C^1 term is a finite-sample effect that can decrease as score quality improves; at the population score, it does not change the two-block identification relation or its estimands. State-dependent diffusion and channel-wise missingness remain extensions beyond the present model.

Markov compatibility of the observed state. To assess whether the observed state approximately satisfies a first-order Markov assumption, we compare held-out errors from two predictors: one predicts the next standardized increment from the current state alone, and the other additionally receives the preceding observation. Adding the preceding observation consistently reduces prediction error in all three NetSim BOLD simulations, while changing it little in CausalDynamics-CI. The current NetSim observation may therefore be insufficient as a first-order Markov state, limiting how directly that benchmark tests the two-layer semantics. CAST and the matched Δ-Gaussian use the same conditioning variables, so this phenomenon does not explain their relative ordering.