Bayesian Causal Discovery in Directed Cyclic Graphs with Closed-Form Lag Bayes Factors

Seyed Reza Maadi[1,2] **Sally Cripps**[1,2]

[1]Human Technology Institute, University of Technology Sydney, Sydney, New South Wales, Australia
[2]School of Mathematical and Physical Sciences, University of Technology Sydney, Sydney, New South Wales, Australia

Abstract

In coarsely sampled longitudinal data, feedback loops can appear as cyclic contemporaneous structure, yet most causal-discovery methods assume acyclicity. Bayesian directed cyclic graph models allow cycles by using non-Gaussian errors with lagged outcomes and exogenous covariates to identify contemporaneous structure. However, consecutive lags are often highly correlated. Existing methods place a binary indicator on each candidate lag, creating exponentially many lag patterns per edge; evidence for a true edge can split across correlated configurations and fall below selection thresholds (posterior fragmentation). We replace these indicators with one categorical variable recording edge absence or active lag, so lags compete within one state space. Under the Gaussian scale-mixture of Laplace errors, each candidate coefficient integrates out exactly, yielding closed-form Bayes factors from cached scalars with no matrix operations for lag comparison. We prove edge-and-lag selection consistency and show that the at-most-one-active-lag assumption is least costly where fragmentation is most severe; the method is a finite-sample regulariser for correlated-lag regimes, not a uniformly better alternative. Cached sufficient statistics and incremental Cholesky updates make the sampler 55–67 times faster without changing the target posterior. In the largest single-lag synthetic benchmark, the categorical encoding raises the true-positive rate from 0.83 to 0.99 and lowers structural Hamming distance from 6.0 to 0.3. On the Health and Retirement Study panel (17,883 individuals, 13 waves), both formulations share a four-edge contemporaneous core, but the categorical encoding yields larger autoregressive coefficients and a sparser contemporaneous graph than the independent encoding.

1 INTRODUCTION

Feedback loops are common in complex time-varying systems: reciprocal mechanisms may operate faster than the observation frequency, producing directed cycles in the underlying causal structure [Haavelmo, 1943, Friston, 2011, Pekrun et al., 2017, Runge et al., 2019]. Under coarse sampling, within-interval feedback appears as simultaneity, so the contemporaneous causal graph may be cyclic. Most causal discovery methods assume directed acyclic graphs (DAGs) [Spirtes et al., 2000, Pearl, 2009] and, without further assumptions such as non-Gaussianity or interventions, recover only a Markov equivalence class [Spirtes, 1995, Assaad et al., 2022, Gong et al., 2024]. Recovering the true, potentially cyclic structure matters in domains such as neuroimaging, immunology, and education [Friston, 2011, Pekrun et al., 2017, Jin et al., 2025], where interventions depend on causal direction.

Non-Gaussian errors can break this equivalence. The Linear Non-Gaussian Acyclic Model (LiNGAM) [Shimizu et al., 2006] achieves unique DAG identification; Lacerda et al. [2008] and Drton et al. [2025] extend this idea to cyclic graphs under linear non-Gaussian assumptions. Building on these results, Jin et al. [2025] proposed the Bayesian Directed Cyclic Graph (DCG) framework for longitudinal data. Its key insight is to use lagged outcomes together with exogenous covariates as identifying information for the full causal graph, including cycles, under their stated assumptions, a role distinct from the classical instrumental-variable framework for unmeasured confounding. Non-Gaussian structural vector autoregressive methods also exploit non-Gaussianity for contemporaneous identification [Lanne et al., 2017], but do not perform Bayesian structure learning with discrete lag selection.

Lag selection is therefore central in DCG: the same lagged outcomes that help identify the contemporaneous cyclic structure also enter the lagged structural equations as predictors. Jin et al. [2025] and recent Bayesian lag-selection methods [Manna and Ghosh, 2025, Dempsey and Wyse,

Proceedings of the 42nd Conference on Uncertainty in Artificial Intelligence (UAI 2026), PMLR 337:4205–4231.

2025] assign independent binary spike-and-slab indicators to each lag. This works well when the maximum lag order L_y is small; Jin et al. [2025], for example, restricted their HIV application to $L_y = 1$. But many longitudinal systems require $L_y \geq 2$, and successive lags are often highly autocorrelated. In this regime, posterior mass fragments across many lag configurations per edge, diluting inclusion probabilities, analogous to the multiplicity phenomenon analysed by Barbieri et al. [2021] for correlated covariates. Existing lag-selection methods that address multicollinearity [Nicholson et al., 2020, Zhao et al., 2024] assume acyclicity and therefore do not apply here.

The computational burden also grows quickly. Jin et al. [2025] recompute weighted cross-product matrices from the full data at every Markov chain Monte Carlo (MCMC) proposal rather than caching and incrementally updating sufficient statistics, which becomes prohibitive once the network has $Q \geq 10$ outcome variables. As Table 2 (Appendix A) summarises, no existing method addresses both the statistical and computational limitations at once. We make three contributions.

1. **Categorical lag selection with closed-form Bayes factors** (Sections 3.1–3.2; Definition 1, Theorem 1). In place of the L_y independent binary lag indicators per edge used by Jin et al. [2025], whose 2^{L_y} configurations fragment posterior mass under autocorrelated lags, we introduce a single categorical variable $z_{qp} \in \{\text{no edge}, \text{lag } 1, \dots, \text{lag } L_y\}$ that jointly encodes edge presence and lag order, collapsing the per-pair model space from 2^{L_y} to $L_y + 1$ states.[1] The Gaussian scale-mixture representation of the Laplace errors then yields closed-form Bayes factors that compare lags exactly using only cached scalar statistics.
2. **Selection consistency with a complementarity property** (Sections 3.1–3.2; Theorem 2, Proposition 1). Under standard regularity conditions, the categorical encoding recovers both edge presence and the correct lag order as the sample size grows. It relies on an at-most-one-active-lag assumption (Assumption 1), but the restriction is most benign in the regime where categorical pooling is most useful: its prediction-error cost is governed by the same inter-lag correlation that drives posterior fragmentation (Remark 1). The method is therefore a finite-sample regulariser for correlated-lag regimes rather than a uniformly superior asymptotic alternative.
3. **Scalable posterior-preserving inference** (Sections 3.3–3.4; Theorem 3). Caching sufficient statistics and applying Cholesky-based incremental updates ensures that a lag change updates only the affected cross-product terms, leaving the target posterior unchanged. This brings networks with $Q \geq 10$ variables and $L_y \geq 4$ lags within reach of a single workstation.

[1]Throughout, "categorical" refers to the latent edge-and-lag state, not to the data type of the observed variables.

2 LONGITUDINAL DCG MODEL

We adopt the Bayesian longitudinal directed cyclic graph (DCG) framework of Jin et al. [2025]. Let $Y_i^{(j)} \in \mathbb{R}^Q$ denote Q outcomes for unit i at visit $j = 1, \dots, J_i$, and $X_i^{(j)} \in \mathbb{R}^S$ exogenous covariates. The structural equations are

$$\begin{aligned} Y_i^{(j)} = \mu + B_0\, Y_i^{(j)} + \textstyle\sum_{\ell=1}^{L_y} B_\ell\, Y_i^{(j-\ell)} \\ + \textstyle\sum_{\ell=0}^{L_x} C_\ell\, X_i^{(j-\ell)} + E_i^{(j)}, \end{aligned} \tag{1}$$

where $B_0 \in \mathbb{R}^{Q\times Q}$ $(\operatorname{diag}(B_0) = 0)$ encodes contemporaneous cyclic effects, $\{B_\ell\}_{\ell\geq 1}$ are lagged outcome effects, $\{C_\ell\}$ covariate effects, and $E_i^{(j)} \in \mathbb{R}^Q$ independent non-Gaussian errors. The reduced form $(I - B_0)Y_i^{(j)} = \mu + \sum_\ell B_\ell Y_i^{(j-\ell)} + \sum_\ell C_\ell X_i^{(j-\ell)} + E_i^{(j)}$ contributes a Jacobian $|\det(I - B_0)|$ to the likelihood (Appendix C). Identifiability requires (i) causal sufficiency, (ii) non-Gaussian errors, (iii) $\rho(B_0) < 1$, and (iv) sufficient exogenous and lagged identifying variation for B_0, established by Jin et al. [2025]; we introduce no new identifiability theory. Causal sufficiency is a strong assumption inherited from the DCG framework and may not hold in observational longitudinal data. Recovered edges should be read relative to the measured variable set: under latent common causes, an edge may reflect shared dependence on unobserved factors rather than a direct mechanism. We therefore interpret discovered structure as associations consistent with the DCG likelihood, not definitive causal claims.

Each structural error is Laplace, $E_{iq}^{(j)} \sim \text{Laplace}(0, 2\sigma_q)$ (scale parameter $b = 2\sigma_q$), thereby complying with the non-Gaussianity assumption of Shimizu et al. [2006]. Inference uses a Gaussian scale-mixture augmentation introducing latent precisions $Z_{iq}^{(j)} > 0$:

$$E_{iq}^{(j)} \mid Z_{iq}^{(j)}, \sigma_q^2 \sim \mathcal{N}\big(0,\ \sigma_q^2 / Z_{iq}^{(j)}\big), \tag{2}$$

$$Z_{iq}^{(j)} \sim \text{Inv-Gamma}(1, \tfrac{1}{8}), \tag{3}$$

so that conditional on Z the likelihood is a heteroskedastic weighted Gaussian regression with precision weights $w_{iq}^{(j)} = Z_{iq}^{(j)}$ (Appendix D). Recent Bayesian causal-discovery theory also studies Laplace working models under more general error distributions; for example, Chaudhuri et al. [2025] establish Bayesian DAG-selection consistency when the true errors are scale mixtures of Gaussian. We treat the Laplace specification as a tractable non-Gaussian *working* model that yields the closed-form lag Bayes factors of Section 3.2, rather than a claim that the true errors are exactly Laplace; richer scale-mixture or mixture-of-Gaussians errors are a complementary direction beyond our focus on cyclic longitudinal DCGs and finite-sample lag fragmentation.

Index sets. Because the panels are irregular, the closed-form Bayes factors and weighted statistics throughout Sections 2–3 are defined over observation index sets, which we introduce here. For each outcome q, the valid index set $\mathcal{I}_q$ collects all unit–visit pairs (i,j) for which $Y_{iq}^{(j)}$ is observed and $j > L_y$, so that all required lags exist:

$$\mathcal{I}_q = \{(i,j) : Y_{iq}^{(j)} \text{ obs.},\ j > L_y\}, \tag{4}$$

and for a specific lag ℓ and parent p the refined set restricting to observed lag-ℓ predictors is

$$\mathcal{I}_q^{(\ell,p)} = \{(i,j) \in \mathcal{I}_q : Y_{ip}^{(j-\ell)} \text{ obs.}\} \subseteq \mathcal{I}_q. \tag{5}$$

For balanced panels $\mathcal{I}_q^{(\ell,p)} = \mathcal{I}_q$ for all (ℓ,p); full details for irregular panels are in Appendix F.

The lag fragmentation problem. In the baseline formulation, each lag $\ell \in \{1,\ldots,L_y\}$ of each ordered pair (q,p) receives an independent spike-and-slab indicator $\gamma_{\ell qp} \sim \text{Bernoulli}(\rho_{\text{ind}})$, yielding a model space of size $2^{L_y Q^2}$. When lagged predictors are serially correlated, posterior evidence for a true edge disperses across lags, producing marginal inclusion probabilities below selection thresholds even when aggregate evidence is strong, an instance of the multiplicity phenomenon of Scott and Berger [2010] and Barbieri et al. [2021]. Section 3 resolves this by replacing independent lag-wise selection with a single categorical variable per ordered pair.

3 PROPOSED METHOD

Our contribution has three components: (i) a *categorical lag prior* that collapses the per-pair model space from 2^{L_y} to L_y+1 states (Section 3.1); (ii) *closed-form lag Bayes factors* enabling an exact collapsed Gibbs update with provable selection consistency (Section 3.2); (iii) *posterior-preserving accelerations* validated by numerical posterior agreement (Section 3.4). All changes are confined to lag selection; the likelihood, contemporaneous structure, and identifiability theory follow Jin et al. [2025] without modification.

3.1 CATEGORICAL LAG FORMULATION

Assumption 1 (At-Most-One Active Lag) *For every ordered pair (q,p), at most one lag $\ell \in \{1,\ldots,L_y\}$ carries a nonzero time-lagged coefficient.*

The following result shows that Assumption 1 is most restrictive when it is least needed, and approximately costless when the categorical formulation is most needed. All statements are conditional on the Laplace augmentation weights $\{w_{iq}^{(j)} = Z_{iq}^{(j)}\}$ and variance σ_q^2, matching the conditioning set of Theorem 1.

Proposition 1 (Robustness under Multi-Lag Truth) *Suppose the true DGP violates Assumption* 1 *for a single ordered pair (q,p): lags ℓ_1 and ℓ_2 are both active with coefficients β_1^* and β_2^*, $|\beta_1^*| \geq |\beta_2^*| > 0$. Let $\varphi_{\Delta\ell}$ denote the weighted predictor correlation between $Y_{ip}^{(j-\ell_1)}$ and $Y_{ip}^{(j-\ell_2)}$ on the common index set $\mathcal{I}_q^{(\ell_1,p)} \cap \mathcal{I}_q^{(\ell_2,p)}$ (formal definition in Appendix K).*

(a) Identifiability loss. *The variance inflation factor for separating the two lag coefficients is* $\text{VIF} = 1/(1-\varphi_{\Delta\ell}^2)$. *Under an AR(1)-like marginal process, $\varphi_{\Delta\ell} \approx \varphi^{\Delta\ell}$: at φ=0.9 and $\Delta\ell$=1, the VIF is* 5.3*; separating two adjacent-lag coefficients requires* $5.3\times$ *more data than estimating a single lag.*

(b) Coefficient absorption (oracle projection). *When the categorical model collapses the two-lag truth onto the dominant lag ℓ_1, the* oracle *single-lag weighted projection has probability limit*

$$\hat{\beta} \xrightarrow{P} \beta_1^* + \beta_2^* \frac{S_{12}}{S_{11}} = \beta_1^* + \beta_2^* \, \varphi_{\Delta\ell} \sqrt{S_{22}/S_{11}}, \tag{6}$$

with the omitted-lag contribution entering through the signed projection weight S_{12}/S_{11} ($\approx \varphi_{\Delta\ell}$ under balanced designs, $S_{11} \approx S_{22}$). Edge detection is preserved whenever this limit is nonzero. This is an oracle projection limit; the full-sampler posterior coefficient exhibits the same qualitative inflation but is attenuated relative to it (Appendix N).

(c) Bounded prediction cost. *The increase in residual variance from omitting lag ℓ_2 is*

$$\Delta\sigma^2 = (\beta_2^*)^2 \, \sigma_{Y,\ell_2}^2 \, (1-\varphi_{\Delta\ell}^2), \tag{7}$$

where σ_{Y,ℓ_2}^2 is the weighted second moment of the omitted predictor.

Proof. See Appendix K. □

Remark 1 (Complementarity with Fragmentation) *The single quantity $(1-\varphi_{\Delta\ell}^2)$ controls both the cost of Assumption 1 and the severity of posterior fragmentation (Section 2):*

φ	*VIF*	$1-\varphi^2$	*Fragmentation*
0.5	1.3	0.75	*mild*
0.7	2.0	0.51	*moderate*
0.9	5.3	0.19	*severe*

When φ is large, fragmentation is severe (the independent encoding disperses posterior mass across lags), but $(1-\varphi^2)$ is small: part (a) shows the per-lag decomposition is poorly identified, part (c) shows the prediction cost of collapsing to one lag is small, and part (b) shows edge detection is *unaffected. Conversely, when φ is small, Assumption 1 is more restrictive, but fragmentation does not arise, so the independent encoding of Jin et al. [2025] suffices. Thus Assumption 1 is most reliable precisely when the categorical formulation is most needed; Section 4.2 confirms this empirically.*

Definition 1 (Categorical Lag Variable) *Under Assumption 1, define $z_{qp} \in \{0, 1, \dots, L_y\}$ for each (q, p), where $z_{qp} = 0$ denotes no lagged edge and $z_{qp} = \ell \geq 1$ the unique active lag. When $z_{qp} = \ell$, write $\beta_{qp} \equiv (B_\ell)_{qp}$, with $(B_m)_{qp} = 0$ for $m \neq \ell$. The coefficient prior is*

$$\beta_{qp} \mid z_{qp}, \sigma_q^2 \;\sim\; \begin{cases} \delta_0 & z_{qp} = 0, \\ \mathcal{N}(0,\ \nu\,\sigma_q^2) & z_{qp} \geq 1, \end{cases} \tag{8}$$

with $\nu > 0$ a slab-to-spike variance ratio (default $\nu = 1$; given a conjugate prior and updated within $\mathcal{K}_\sigma$). The prior on z_{qp} decomposes into an existence-first *factorisation,*

$$\pi_0 = 1 - \rho_{\mathrm{lag}}, \qquad \pi_\ell = \rho_{\mathrm{lag}}\,\tilde{v}_\ell, \quad \ell \geq 1, \tag{9}$$

with $\tilde{v}_\ell = v_\ell / \sum_m v_m$, $v_\ell = d^{\ell-1}$ ($d = 0.8$, geometric decay favouring shorter lags), $\rho_{\mathrm{lag}} \sim \mathrm{Beta}(\frac{1}{2}, \frac{1}{2})$. A symmetric $\mathrm{Dirichlet}(\alpha)$ hyperprior ($\alpha = 0.5$) is placed on the normalised lag weights $\tilde{v}$, allowing the allocation to adapt via conjugate updates; the geometric initialisation serves as a diffuse starting point.

The categorical encoding reduces the per-pair space from 2^{L_y} to $L_y + 1$ and the total lag-selection space from $2^{L_y Q^2}$ to $(L_y + 1)^{Q^2}$.

Remark 2 (Prior Calibration) *In the baseline, independent indicators $\gamma_{\ell qp} \sim \mathrm{Bernoulli}(\rho_{\mathrm{ind}})$ give a prior edge-existence probability $P(\exists \ell : \gamma_{\ell qp} = 1) = 1 - (1 - \rho_{\mathrm{ind}})^{L_y}$. The categorical prior sets this probability directly to ρ_{lag}. For fair comparison, we calibrate $\rho_{\mathrm{ind}} = 1 - (1 - \rho_{\mathrm{lag}})^{1/L_y}$ so that both encodings share the same marginal prior on edge existence. Differences in posterior recovery are then attributable solely to the competition mechanism, not prior sparsity mismatch.*

3.2 CLOSED-FORM LAG BAYES FACTORS

The categorical encoding enables β_{qp} to be integrated out analytically, yielding a closed-form Bayes factor for exact lag comparison. For outcome q and candidate parent p, let $r_{iq}^{(j,-p)}$ denote the partial residual with parent p's contribution removed (Appendix E), and define the weighted sufficient statistics over the lag-specific index set $\mathcal{I}_q^{(\ell,p)} \subseteq \mathcal{I}_q$:

$$S_{yy}^{(\ell)} = \textstyle\sum_{(i,j) \in \mathcal{I}_q^{(\ell,p)}} w_{iq}^{(j)} \left(Y_{ip}^{(j-\ell)}\right)^2, \tag{10}$$

$$S_{yr}^{(\ell)} = \textstyle\sum_{(i,j) \in \mathcal{I}_q^{(\ell,p)}} w_{iq}^{(j)}\, Y_{ip}^{(j-\ell)}\, r_{iq}^{(j,-p)}. \tag{11}$$

Theorem 1 (Closed-Form Lag Bayes Factor) *Consider the DCG model* (1) *with the categorical lag prior (Definition 1). Conditional on the Laplace augmentation $\{Z_{iq}^{(j)}\}$ and variance σ_q^2, the conjugate slab prior* (8) *yields the following.*

(i) Log Bayes factor. *Let $\kappa_\ell \equiv \nu\, S_{yy}^{(\ell)}$, with $S_{yy}^{(\ell)}, S_{yr}^{(\ell)}$ from* (10)–(11)*. The log marginal likelihood ratio for lag ℓ versus no edge ($\beta_{qp} \equiv 0$) is*

$$\log \mathrm{BF}_\ell = \frac{1}{2}\left[\frac{\nu\,\left(S_{yr}^{(\ell)}\right)^2}{\sigma_q^2\,(1 + \kappa_\ell)} \;-\; \log(1 + \kappa_\ell)\right]. \tag{12}$$

The first term measures data fit; the second is an Occam penalty, strictly decreasing in $S_{yy}^{(\ell)}$. The entire lag comparison uses only cached scalars, requiring no matrix operations.

(ii) Coefficient posterior. *Conditional on $z_{qp} = \ell \geq 1$,*

$$\beta_{qp} \mid z_{qp} = \ell,\ Z,\ \sigma_q^2 \;\sim\; \mathcal{N}\!\left(\frac{\nu\, S_{yr}^{(\ell)}}{1 + \kappa_\ell},\ \ \frac{\nu\,\sigma_q^2}{1 + \kappa_\ell}\right). \tag{13}$$

Proof sketch; *full derivation in Appendix* G. Completing the square in β in the joint $\mathcal{N}(r^{(-p)};\ \ \beta Y_\ell, \sigma_q^2 W^{-1}) \cdot \mathcal{N}(\beta; 0, \nu\sigma_q^2)$ gives a Gaussian integral with precision $(1 + \kappa_\ell)/(\nu\sigma_q^2)$. Evaluating and dividing by the null ($\beta = 0$) gives (12); the posterior (13) is the completed-square mean and variance. □

Selection consistency. We now state sufficient conditions under which (12) correctly recovers the true graph within the conditional regression block (given Z and σ_q^2), paralleling regularity assumptions from Bayesian variable selection [Narisetty and He, 2014, Yang et al., 2016]. Throughout, Q and L_y are treated as fixed; asymptotics are in the total observation count $T = \sum_i J_i \to \infty$.

Assumption 2 (Signal Strength) *For each true edge $(q \leftarrow p) \in \mathcal{E}^*$ with coefficient β_{qp}^* at true lag ℓ^*, there exists $\lambda_T \to \infty$ such that $(\beta_{qp}^*)^2 S_{yy}^{(\ell^*)} / \sigma_q^2 \geq \lambda_T \log T$. This β-min condition ensures the signal term dominates the $O(\log T)$ Occam penalty in Lemma 1.*

Assumption 3 (Lag Separation) *For every true edge with true lag ℓ^*, the signal at any $\ell \neq \ell^*$ is strictly dominated:*

$$\left(S_{yr}^{(\ell)}\right)^2 / S_{yy}^{(\ell)} \;<\; \left(S_{yr}^{(\ell^*)}\right)^2 / S_{yy}^{(\ell^*)} -\; c\,\sigma_q^2 \log T \tag{14}$$

for some $c > 0$, where $T = \sum_i J_i$ denotes the total observation count.

Assumption 4 (Design Regularity) *There exist $0 < c_{\min} \leq c_{\max} < \infty$ such that $c_{\min} |\mathcal{I}_q^{(\ell,p)}| \leq S_{yy}^{(\ell)} \leq c_{\max} |\mathcal{I}_q^{(\ell,p)}|$ for all (q, p, ℓ).*

Assumption 5 (Sparsity) *The number of true lagged edges $s = |\mathcal{E}^*|$ satisfies $s = o(T/\log(Q^2 L_y))$.*

Lemma 1 (Signal Detection) *Under Assumptions 2–4:* (a) *for every true edge,* $\log \mathrm{BF}_{\ell^*} \geq (\lambda_T/4)\log T \to \infty$*;* (b) *for every false pair,* $\max_\ell \log \mathrm{BF}_\ell = O_P(1)$*. Full proof in Appendix* I.

Remark 3 *Under Design Regularity (Assumption 4), the Occam penalty gives a tighter bound for false pairs: $\log \mathrm{BF}_\ell = -\frac{1}{2}\log T + O_P(1)$ uniformly in ℓ (proof in Appendix I), providing faster exclusion.*

Theorem 2 (Edge Selection Consistency) *Under Assumptions 1 and 2–5, as $T \to \infty$:* (a) $P(z_{qp} > 0 \mid \mathcal{D}) \to 1$ *for every true edge;* (b) $P(z_{qp} = \ell^* \mid \mathcal{D}) \to 1$*;* (c) $P(z_{qp} > 0 \mid \mathcal{D}) \to 0$ *for every false pair. All in probability.*

Proof sketch; *full proof in Appendix* I. **(a)** Lemma 1(a) gives $\mathrm{BF}_{\ell^*} \to \infty$; since $\pi_{\ell^*}/\pi_0 > 0$ almost surely under the continuous Beta posterior of ρ_{lag}, the posterior odds diverge. **(b)** Assumption 3 gives $\log(\mathrm{BF}_\ell/\mathrm{BF}_{\ell^*}) \leq -(c/2)\log T + O(1) \to -\infty$ for $\ell \neq \ell^*$. **(c)** By Remark 3, $\max_\ell \mathrm{BF}_\ell = O_P(T^{-1/2}) \to 0$ for every false pair. □

3.3 POSTERIOR INFERENCE

We construct a blocked Gibbs sampler whose transition kernel is a composition of reversible sub-kernels.

Theorem 3 (Posterior Invariance) *Let π denote the joint posterior under* (1) *with the categorical lag prior (Definition 1) and Laplace augmentation* (2)*. Define $\mathcal{K} = \mathcal{K}_Z \circ \mathcal{K}_\rho \circ \prod_q (\mathcal{K}_{z,q} \circ \mathcal{K}_{z^\dagger,q}) \circ \mathcal{K}_\sigma \circ \mathcal{K}_{B_0}$, where $\mathcal{K}_\sigma$ collects the conjugate updates for $\{\mu_q, \sigma_q^2, C_\ell, \nu\}$. Then $\pi\mathcal{K} = \pi$.*

Proof sketch; *formal verification in Appendix* H. Each sub-kernel is either a Gibbs step from an exact full conditional ($\mathcal{K}_Z$, $\mathcal{K}_\rho$, $\mathcal{K}_\sigma$) or a collapsed-Gibbs/MH step ($\mathcal{K}_{z,q}$, $\mathcal{K}_{z^\dagger,q}$, $\mathcal{K}_{B_0}$) satisfying the conditions of Liu [1994]. Their composition preserves π by Tierney [1994, Thm 1]. □

The existence probability from the collapsed Gibbs step is

$$P(z_{qp} = \ell \mid \text{rest}) \;\propto\; \pi_\ell \cdot \mathrm{BF}_\ell \quad (\mathrm{BF}_0 \equiv 1), \tag{15}$$

implemented via log-sum-exp; a provisional β_{qp} is drawn from (13). For active edges, a lag swap $\ell_{\mathrm{cur}} \to \ell_{\mathrm{prop}}$ is proposed uniformly; the symmetric proposal gives MH ratio

$$\alpha = \min\Bigl(1,\ \frac{\pi_{\ell'}\,\mathrm{BF}_{\ell'}}{\pi_\ell\,\mathrm{BF}_\ell}\Bigr), \tag{16}$$

Algorithm 1 Categorical Lag DCG: One MCMC Iteration

```
K_Z: Update Z_iq^(j) (App. D)
K_ρ: ṽ | · (Dirichlet);  ρ_lag | ·;  set π_0, ..., π_{L_y} via (9)
for sweep s = 1, ..., S_sweep do                      ▷ S_sweep = 2
    for q = 1, ..., Q do
        Ph. 1 (K_{z,q}): For each p: log BF_ℓ (12);  draw
        z_qp (15), provisional β_qp (13);  update residuals
        Ph. 1.5 + 2 (K_{z†,q}): optional M_swap lag-swap
        MH (16) (M_swap = 0 in reported scaling experiments);
        joint K_q-variate β resample supersedes Ph. 1 draws
        (App. H)
    end for
end for
K_σ: σ_q^2, μ_q, C_ℓ, ν (conjugate)
K_{B_0}: MH (ρ(B_0) < 1) as in Jin et al. [2025]
```

where $\ell = \ell_{\mathrm{cur}}$ and $\ell' = \ell_{\mathrm{prop}}$. The K_q active coefficients are then redrawn jointly from a K_q-variate Gaussian conditional on the final z_q (Appendix H). Algorithm 1 summarises one full MCMC iteration; the implementation-level version with caching and index-set bookkeeping is given as Algorithm 2 in Appendix H.

3.4 SCALABLE IMPLEMENTATION

All accelerations below alter wall-clock time only; the target posterior is unchanged.

Proposition 2 (Per-Iteration Complexity) *Under cached sufficient statistics, one sweep costs $O\bigl(\sum_{q,p,\ell} |\mathcal{I}_q^{(\ell,p)}|\bigr) \leq O(Q^2 L_y T)$ in the dense case. Phase 2 adds $O(K_q^3)$ per outcome for Cholesky factorisations. A naïve implementation incurs an additional $O(QL_y)$ factor from full residual recomputation.*

The implementation combines three strategies, each preserving sampler exactness (Theorem 3): **(1)** Cholesky factorisation $\hat{P} = LL^\top$ with triangular solves, sampling via $\theta = \mu + \sigma_q L^{-\top} z$; **(2)** cached base residuals with incremental updates: when an edge changes lag, only the affected column's cross-products are updated; **(3)** sparse index-set evaluation over $\mathcal{I}_q^{(\ell,p)}$ for irregular panels (Appendix F). All kernels are JIT-compiled (Numba); runtimes are reported in steady state.

4 EXPERIMENTS

We evaluate the proposed method along three axes. Section 4.1 verifies that the Cholesky-based implementation preserves the posterior distribution exactly while delivering substantial speedups. Section 4.2 tests the three predictions of Theorem 2 at finite sample size, isolating the effect of categorical lag pooling against the independent spike-and-slab baseline as the model dimension Q grows. Section 4.3 applies both methods to longitudinal Health and Retirement Study panel data.

Evaluation protocol. Throughout, CAT denotes the proposed categorical lag encoding and IND the independent binary spike-and-slab baseline of Jin et al. [2025]. We assess recovery at two levels. At the aggregate edge level, CAT declares a directed lagged edge (q,p) present when the edge-existence inclusion probability $P(z_{qp}{>}0 \mid \mathcal{D})$ exceeds 0.5; IND uses the corresponding median-model detection statistic $\max_\ell P(\gamma_{\ell qp}{=}1 \mid \mathcal{D})$, declaring the edge present when at least one lag-specific PIP exceeds 0.5 [median probability model; Barbieri and Berger, 2004]. For the lag-level recovery reported in Section 4.2 (Table 1, Figure 1), the estimated set consists of directed triples (q,p,ℓ): CAT selects (q,p,ℓ) when $P(z_{qp}{=}\ell \mid \mathcal{D}) > 0.5$ and IND when $P(\gamma_{\ell qp}{=}1 \mid \mathcal{D}) > 0.5$; contemporaneous edges are selected analogously through the B_0 inclusion indicator. Let E^* and $\hat{E}$ denote the true and estimated sets at the relevant level, with $\mathrm{TP}{=}|\hat{E} \cap E^*|$, $\mathrm{FP}{=}|\hat{E} \setminus E^*|$, and $\mathrm{FN}{=}|E^* \setminus \hat{E}|$. We report four standard structure-learning metrics: the true positive rate $\mathrm{TPR} = \mathrm{TP}/(\mathrm{TP}+\mathrm{FN})$, false discovery rate $\mathrm{FDR} = \mathrm{FP}/(\mathrm{FP}+\mathrm{TP})$, their harmonic summary $\mathrm{F1} = 2\,\mathrm{TPR}(1{-}\mathrm{FDR})/(\mathrm{TPR}+1{-}\mathrm{FDR})$, and the structural Hamming distance $\mathrm{SHD} = \mathrm{FP}+\mathrm{FN}$ [Tsamardinos et al., 2006].

4.1 POSTERIOR-PRESERVING ACCELERATION

We verify that the Cholesky-based optimisations of Section 3.4 accelerate the sampler without altering its target distribution (discrepancies limited to floating-point arithmetic) on the two benchmark designs of Jin et al. [2025] (R=100 replications each):[2] Scenario I (Q=4, no lagged effects, N=1,000) and Scenario II (Q=3, mirroring the WIHS cohort [Adimora et al., 2018], $\eta \in \{0.50,\ 0.75,\ 1.00\}$). Both set L_y=2, 5,000 MCMC iterations with 2,500 burn-in and thinning of 5, matching Jin et al. [2025].

Table 3 (Appendix B) reports the results. The optimised sampler is 55–67× faster, reducing 5,000 iterations from roughly 44 minutes to under 40 seconds. Mean absolute PIP differences remain below 0.048; thresholded recovery differs by at most 0.012 on every metric. Recovery is therefore reported for the optimised implementation only. The elevated FDR in Scenario I (0.29) reflects observationally equivalent cyclic orientations: the posterior splits mass between two graphs in the same equivalence class [Jin et al., 2025, Figure 5].

[2] All runs: AMD EPYC 9354P, 32 cores, 188 GiB; Python 3.12, OpenBLAS 0.3.31; single-threaded BLAS. Timings exclude data generation and Numba JIT warm-up.

Table 1: Lag-level edge recovery on synthetic panels (N=200, J=10, L_y=4; R=31 replications; PIP>0.5). **Bold** marks the better value in each CAT/IND pair. Lag metrics evaluate all directed lagged edges, including autoregressive (true lags in $\{1,2,3\}$; lag 4 is a decoy); B_0 F1 is the contemporaneous control.

		LAG-LEVEL				CONTROLS	
Q	Method	TPR	FDR	F1	SHD	B_0 F1	s/iter
Single-lag DGP (Assumption 1 satisfied)							
10	CAT	**1.00**	0.004	**1.00**	**0.1**	1.00	0.15
	IND	0.97	**0.000**	0.98	0.4	1.00	**0.05**
15	CAT	**1.00**	0.013	0.99	0.4	1.00	0.32
	IND	0.99	**0.000**	0.99	**0.3**	1.00	**0.11**
20	CAT	**0.99**	0.004	**1.00**	**0.3**	0.98	0.55
	IND	0.83	**0.000**	0.90	6.0	0.98	**0.21**
Mixed-lag DGP (edges at lags $\{1,2\}$ or $\{1,2,3\}$ by Q)							
10	CAT	**0.99**	0.012	**0.99**	**0.3**	1.00	0.16
	IND	0.90	**0.000**	0.95	1.6	1.00	**0.05**
15	CAT	**0.99**	0.006	**0.99**	**0.3**	1.00	0.32
	IND	0.86	**0.000**	0.92	3.2	1.00	**0.11**
20	CAT	**0.99**	0.026	**0.98**	**1.4**	0.98	0.56
	IND	0.83	**0.000**	0.91	5.8	0.98	**0.21**

4.2 EMPIRICAL VALIDATION OF CATEGORICAL LAG SELECTION

Theorem 2 makes three testable predictions as $T \to \infty$: (a) $P(z_{qp}{>}0 \mid \mathcal{D}) \to 1$ for every true edge, (b) $P(z_{qp}{=}\ell^* \mid \mathcal{D}) \to 1$ for the true lag, and (c) $P(z_{qp}{>}0 \mid \mathcal{D}) \to 0$ for every false pair. We evaluate these at finite sample size $T{=}NJ{=}$2,000, varying Q to amplify fragmentation. The independent encoding serves as an ablation: both methods share the same likelihood and contemporaneous kernel $\mathcal{K}_{B_0}$ (Theorem 3), with the marginal edge-existence prior calibrated to match across the two encodings (Remark 2); they differ only in the lag-selection mechanism, isolating the effect of categorical pooling (15).

Data-generating process. We simulate balanced panels (N=200, J=10, L_y=4) for $Q \in \{10, 15, 20\}$; true causal lags lie in $\{1,2,3\}$ and lag 4 is a decoy. The contemporaneous graph B_0 uses sign-cancelling cyclic structure with spectral radius calibrated per Q ($\rho(B_0)$=0.40, 0.35, 0.30 at Q=10, 15, 20) to ensure companion stability. Cross-variable lagged edges have 4–6% density, each assigned to exactly one true lag; approximately half have weak coefficients ($|\beta| \in [0.05,\ 0.12]$) and the remainder strong ($|\beta| \in [0.20,\ 0.40]$). SINGLE scenarios at $Q \leq 15$ use strong coefficients only. The persistence parameter ϕ=0.9 yields AR coefficients $\bar{\phi}_q \in [0.48,\ 0.66]$ on the diagonal of B_1 across the R=31 replications, inducing the predictor

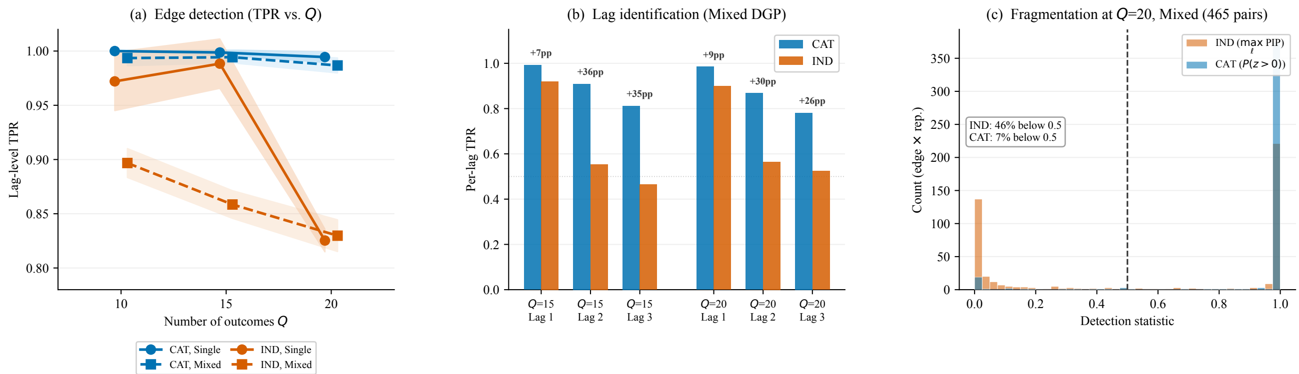

Figure 1: Scaling experiments (N=200, J=10, L_y=4, R=31; true lags in $\{1, 2, 3\}$, lag 4 decoy). **(a)** Lag-level TPR vs. Q ($\pm\,1.96$ SE bands). CAT remains above 0.98; IND degrades to ≈ 0.83 at Q=20 in both scenarios. **(b)** Per-lag TPR in the Mixed DGP at $Q \in \{15, 20\}$ (Q=10 omitted: truth lags restricted to $\{1, 2\}$). Annotations show the CAT–IND gap in percentage points; the gap concentrates at longer lags where autocorrelation disperses posterior evidence across neighbouring lagged predictors. **(c)** Detection-statistic histograms for all true lagged edges at Q=20 (Mixed; 15 edges $\times$ 31 reps = 465 pairs): $P(z_{qp}{>}0 \mid \mathcal{D})$ for CAT, $\max_\ell P(\gamma_{\ell qp}{=}1 \mid \mathcal{D})$ for IND. Under IND, 46% fall below 0.5; under CAT, 7%. The bimodal IND distribution is the empirical fingerprint of lag fragmentation.

collinearity central to the fragmentation pathology of Section 3.1: adjacent lagged predictors correlate at $r \approx \bar{\phi}^{|\ell-\ell'|}$, so the data-fit term in (12) is non-negligible even at non-true lags. Innovations are Laplace with error variance $\sigma^2_{\varepsilon,q} \sim$ Uniform$(0.08,\,0.15)$, corresponding to $\sigma_q^2 = \sigma^2_{\varepsilon,q}/8$ under the model convention $\varepsilon \sim$ Laplace$(0, 2\sigma_q)$. Balanced panels yield $|\mathcal{I}_q^{(\ell,p)}| = N(J{-}L_y) = 1{,}200$, satisfying Assumptions 4–5. Full DGP specification is in Appendix M.

We run two DGP scenarios: SINGLE, where every cross-variable edge lies at lag 1 (Assumption 1 satisfied exactly), and MIXED, where edges are distributed across multiple true lags (Q=10: lags $\{1, 2\}$; $Q \in \{15, 20\}$: lags $\{1, 2, 3\}$); lag 4 is always a decoy. Each scenario is evaluated over R=31 replications using the PIP > 0.5 median-model decision rule [Barbieri and Berger, 2004], with 50,000 MCMC iterations (25,000 burn-in, thinning every 5).

Results. Table 1 and Figure 1 summarise edge recovery along the three axes of Theorem 2. CAT maintains TPR $\geq$ 0.99 across all Q in both scenarios, consistent with $P(z_{qp}{>}0 \mid \mathcal{D}) \to 1$, while IND nearly matches at $Q \leq 15$ in SINGLE but drops to 0.83 at Q=20 (a 16-point deficit) where each equation entertains $(Q{-}1)L_y$=76 candidate lagged regressors. In MIXED, the gap appears already at Q=10 (ΔTPR=0.09) and widens to 0.16 at Q=20 (Figure 1a). The mechanism is visible in Figure 1(c): at Q=20 (MIXED), 46% of true-edge detection statistics fall below 0.5 under IND versus only 7% under CAT, confirming that ϕ-induced collinearity disperses evidence across lags so that each individual $P(\gamma_{\ell qp}{=}1 \mid \mathcal{D})$ remains sub-threshold despite strong aggregate signal, the fragmentation pathology of Section 3.1. Stratifying by lag order (Figure 1b) reveals that the gap concentrates at longer lags: CAT recovers 91% of lag-2 and 81% of lag-3 edges at Q=15 versus 55% and 47% for IND, because, under strong serial persistence, the evidence for a true delayed effect is dispersed across the correlated neighbouring lagged predictors, so each individual lag receives weaker posterior support. Moderate per-lag evidence (12) that individually falls below the median-model threshold can collectively drive $P(z_{qp}{>}0 \mid \mathcal{D})$ above it via the log-sum-exp in (15) (Appendix J).

IND achieves FDR=0 across all settings (Remark 3): the Occam penalty gives $\mathrm{BF}_\ell \approx |\mathcal{I}|^{-1/2} \approx 0.029$ at $|\mathcal{I}|$=1,200 for noise-only lag candidates, too small for their per-lag PIPs to exceed 0.5. IND's missed detections arise from fragmentation, not conservatism. CAT accepts a modest FDR ≤ 0.03 for substantially higher recall (SHD 0.3 vs. 6.0 at Q=20, SINGLE). Because both methods share $\mathcal{K}_{B_0}$ (Theorem 3), the contemporaneous block is updated by the same kernel in both samplers. In these synthetic settings B_0 recovery is essentially unchanged across encodings (B_0 F1 ≥ 0.98; Table 1), indicating that the performance gaps in Table 1 are driven by lag selection rather than by the contemporaneous sampler. The categorical sampler incurs 2.6–3.2× overhead per iteration (0.56 s at Q=20), because each ordered-pair update evaluates a Bayes factor for all L_y candidate lags and the active coefficients are resampled jointly via a Cholesky factorisation, whereas the independent encoding performs independent univariate spike-and-slab updates.

Same-pair multi-lag violation. To probe robustness when Assumption 1 fails, we add a SAME-PAIR DGP in which 30% of active cross-variable ordered pairs carry two simultaneously active lags (the secondary coefficient half the dominant, same sign, both lags drawn from $\{1, 2\}$), with Q=15, L_y=4, $\rho(B_0)$=0.10, over $\varphi \in \{0.5, 0.7, 0.9\} \times N \in \{200, 500, 2000\}$ (R=31; $\approx$4 violating pairs per replication). The realized inter-lag predictor correlation is $\varphi_{\Delta\ell} \approx 0.55$, 0.76, 0.92 at $\varphi=0.5, 0.7, 0.9$ (VIF 1.4, 2.4, 6.6). Because Assumption 1 is violated by construction, this regime lies *outside* Theorem 2 and tests Proposition 1. Cross-variable edge-level recovery (lag-agnostic, self-loops excluded) is summarised in Table 4; lag-specific recovery on violating pairs is reported by Any, Both, and Dom. Across all nine cells CAT attains TPR 0.88–1.00 and SHD 0.71–4.97 versus IND's TPR 0.55–0.65 and SHD 5.13–6.71.

CAT detects every violating pair and selects the dominant lag in 85–96% of cases, but never recovers both lags, as Assumption 1 requires. IND recovers both lags more often as N grows, but its edge-level recovery remains lower because per-lag inclusion probabilities dilute below threshold: at $\varphi=0.9$ (VIF 6.6) and $N=200$, IND fails to recover both true lags for 25% of violating pairs, whereas the pooled existence probability (15) detects every pair (SHD 0.71 versus 5.81). CAT attains lower SHD than IND in *every* cell; as N grows the gap narrows: IND's two-lag recovery improves while CAT's cross-variable FDR rises to 0.24 at $\varphi=0.9$, $N=2000$, a single-lag misspecification bias–variance signature, with the crossover approached but not reached.

Thus CAT is a finite-sample regulariser, not a uniformly superior alternative: it preserves pair-level detection in correlated-lag regimes, whereas IND is preferable when exact multi-lag decomposition is the target and the sample is large enough for per-lag evidence to clear the threshold. The elevated cross-variable FDR at $N=2000$ does not contradict Theorem 2(c), because the finite-sample SAME-PAIR design deliberately violates Assumption 1. The oracle single-lag projection (computed from the true coefficients and the exact partial residual, not sampler output) of Eq. (6) is verified in Appendix N.

4.3 REAL-DATA APPLICATION

We apply both the categorical and independent formulations to the Health and Retirement Study [HRS; Sonnega et al., 2014], a nationally representative biennial panel of Americans aged 50+. Using the DCG framework of Jin et al. [2025], we model $Q=5$ standardised outcomes: depressive symptoms (CES-D), self-rated health (SRH), body mass index (BMI), cognition, and mobility limitations, with age as an exogenous covariate ($S=1$) and maximum lag $L_y=2$ (corresponding to 2- and 4-year delays). The panel comprises $N=17{,}883$ individuals observed across 13 waves (1996–2020, $\bar{J}_i=8.9$), yielding $\sum_i (J_i-2)=123{,}422$ effective observations after the lag-2 burn-in. Both samplers run 20,000 iterations (10,000 burn-in) over $R=31$ independent replications on identical data.

Figure 2 shows the high-confidence display graph (selection rule in the caption; full posterior matrices in Appendix L). Both formulations share four contemporaneous edges: a directed association from mobility limitations to depressive symptoms (Mobility $\to$ CES-D), and contemporaneous edges into cognition from CES-D, SRH, and mobility. Beyond this shared core the two graphs diverge. The independent encoding recovers a denser contemporaneous graph (eight edges versus five), including three effects on SRH (CES-D, BMI, Mobility $\to$ SRH, $|\hat{\beta}_0| \in [0.12, 0.19]$) absent under the categorical model, while the categorical encoding uniquely identifies an SRH $\to$ CES-D contemporaneous effect. This divergence is not a deficiency of either method but a structural consequence of how each partitions temporal persistence, as we now explain.

The independent encoding recovers lag-2 autoregressive self-loops for four of five outcomes (CES-D, SRH, cognition, mobility), representing each as an AR(2) process with separate lag-1 and lag-2 coefficients (e.g., $\hat{\beta}_{qq}^{(1)}=0.55$, $\hat{\beta}_{qq}^{(2)}=0.26$ for SRH). Under Assumption 1 the categorical formulation restricts each pair to a single active lag, precluding simultaneous lag-1 and lag-2 self-loops; this yields four stable lag-1 self-loops (CES-D, SRH, BMI, mobility). The cognition self-loop attains only marginal per-lag support (lag-1 mean PIP ≈ 0.52), consistent with the near-equal strength of its lag-1 and lag-2 components under the independent encoding ($\hat{\beta}_{qq}^{(1)}=0.43$, $\hat{\beta}_{qq}^{(2)}=0.36$), which makes Assumption 1 least favourable for this outcome. Proposition 1(b) implies that collapsing two correlated lags onto one shifts mass toward the retained lag, and consistent with this the categorical lag-1 self-loop coefficients (0.57–1.00) are systematically larger than the independent lag-1 values (0.46–0.89). We read this as a *directional* observation, not a numerical test of Eq. (6): that equation is an oracle single-lag projection limit, whereas these are full-sampler posterior means, and in the near-unit-root regime the inflation can exceed what omitted-lag absorption alone predicts. By Proposition 1(c) the prediction cost of this collapse is bounded by $(\beta_2^*)^2\, \sigma_{Y,\ell_2}^2 (1-\varphi_{\Delta\ell}^2)$, small precisely when $\varphi_{\Delta\ell}$ is large (Remark 1), the high-persistence regime in which Assumption 1 is most restrictive for self-loops yet least costly.

This propagates to the contemporaneous graph: the shared $\mathcal{K}_{B_0}$ kernel (Theorem 3) conditions on residuals with lagged contributions removed, so larger categorical self-loops leave smaller B_0 residuals. SRH illustrates this: its near-unity categorical lag-1 coefficient ($\hat{\beta} \approx 1.00$) absorbs much of the serial dependence, so no contemporaneous parents satisfy the display rule, whereas distributing persistence across two lags ($0.55 + 0.26$) under the independent encoding leaves larger residual variance and three contemporaneous parents emerge (CES-D, BMI, Mobility $\to$ SRH).

Conversely, the categorical formulation surfaces cross-variable lagged structure entirely absent under the independent encoding, which recovers no cross-variable

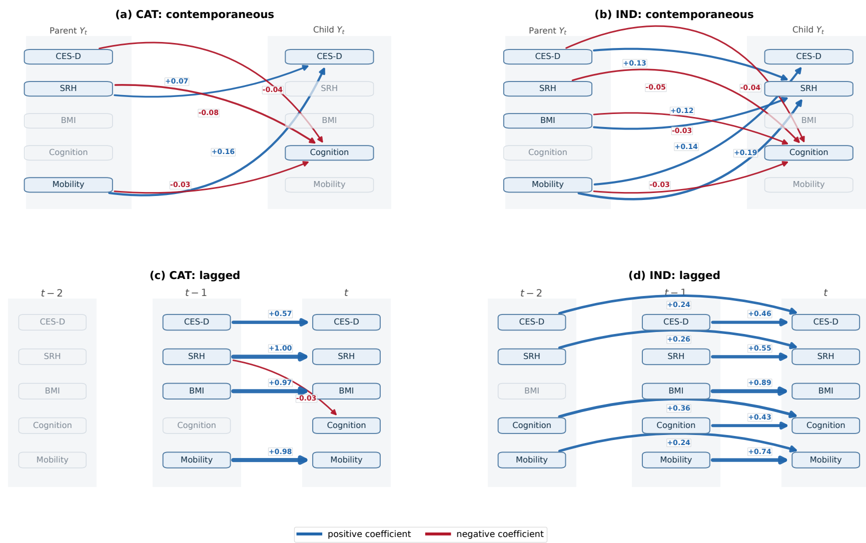

Figure 2: High-confidence display graph of estimated effects on HRS panel data (Q=5, N=17,883, L_y=2; R=31 replications). An edge is shown when its layer-specific mean posterior inclusion probability exceeds 0.5, it is present in at least 70% of replications (replication stability ≥ 0.70), and $|\hat{\beta}| \geq 0.03$; labels give the posterior mean coefficient, colour encodes sign (blue positive, red negative), and lag order is read from the source time slice. **(a, b)** Contemporaneous effects (B_0): five edges under the categorical encoding versus eight under the independent encoding. **(c)** Categorical lagged effects: four lag-1 self-loops and one cross-variable lag-1 edge (SRH $\to$ cognition); cognition has no displayed self-loop (lag-1 PIP and stability both ≈ 0.52, below the 0.70 cut). **(d)** Independent lagged effects: five lag-1 and four lag-2 self-loops, with no cross-variable lagged edges. Full posterior summaries are in Appendix L (Figure 3).

lagged edges. These categorical cross-lagged effects are posterior-supported but small: the only non-negligible one is SRH $\to$ cognition at lag 1 ($\hat{\beta} \approx -0.03$); the remainder have $|\hat{\beta}| < 0.03$ (Appendix L). This mirrors Section 4.2: weak cross-variable effects fall below the per-lag threshold under independent indicators, whereas $P(z_{qp}{>}0 \mid \mathcal{D})$ pools their evidence via (15); the small magnitudes are consistent with dominant autoregressive persistence *within* each outcome.

The recovered structure aligns with established longitudinal evidence: associations among depressive symptoms, self-rated health, BMI, cognition, and mobility are documented in the same HRS panel [Wagner and Short, 2014], and obesity–depression links are well established [Luppino et al., 2010]. Weaker cross-lagged associations, including the lag-2 Cognition $\to$ CES-D effect (Appendix L), have mixed support [van den Kommer et al., 2013, Zahodne et al., 2014]. We read all discovered directions as associations consistent with the DCG likelihood, not definitive causal claims.

5 CONCLUSION

The categorical lag formulation forces the candidate lags of each ordered pair to compete within a single edge-and-lag state, regularising Bayesian DCG lag selection in regimes where adjacent lags are highly correlated, while the inference scheme preserves the proposed posterior target (Theorem 3). It is best read as a *finite-sample regulariser for correlated-lag regimes*: categorical pooling helps most where the per-lag decomposition is weakly identified (Proposition 1, Remark 1), whereas the independent encoding of Jin et al. [2025] is preferable when simultaneous lag effects are substantively important and the data are informative enough to resolve them. In the largest single-lag synthetic benchmark, pooling raises lag-level true-positive rate from 0.83 to 0.99. The HRS application illustrates the same mechanism on real data: collapsing correlated temporal persistence into a single selected lag changes the residual variation passed to the contemporaneous graph. The limitation is intrinsic to the modelling choice: the at-most-one-active-lag assumption is least costly under high-persistence fragmentation, but cannot recover multiple active lags on the same ordered pair. Extending the closed-form update to a few jointly active lags is the natural next step.

Author Contributions

S. R. Maadi developed the method, implemented the sampler, ran the experiments, and wrote the paper. S. Cripps supervised the work and contributed to the framing and revision.

Acknowledgements

This research program relies on student participation, which was made possible through student funding provided under the Next Generation Artificial Intelligence Graduates Program (NGAIGP), conducted by CSIRO.

This research was supported by the Paul Ramsay Foundation (PRF) through the Thrive: Finishing School Well program.

During the preparation of this work, the authors used ChatGPT (OpenAI) as an AI-assisted writing tool, solely to improve the clarity, grammar, and readability of author-written text. All AI-assisted suggestions were reviewed and edited by the authors, who take full responsibility for the final content of the paper.

References

Adaora A. Adimora, Catalina Ramirez, Lorie Benning, Ruth M. Greenblatt, Mirjam-Colette Kempf, Phyllis C. Tien, Seble G. Kassaye, Kathryn Anastos, Mardge Cohen, Howard Minkoff, et al. Cohort profile: The Women's Interagency HIV Study (WIHS). *International Journal of Epidemiology*, 47(2):393–394i, 2018.

Fahimeh Arab, AmirEmad Ghassami, Hamidreza Jamalabadi, Megan A K Peters, and Erfan Nozari. Whole-brain causal discovery using fMRI. *Network Neuroscience*, 9(1):392–420, 2025. doi: 10.1162/netn_a_00438.

Charles K Assaad, Emilie Devijver, and Eric Gaussier. Survey and evaluation of causal discovery methods for time series. *Journal of Artificial Intelligence Research*, 73: 767–819, 2022.

Lazar Atanackovic, Alexander Tong, Bo Wang, Leo J Lee, Yoshua Bengio, and Jason S Hartford. DynGFN: Towards Bayesian inference of gene regulatory networks with GFlowNets. In *Advances in Neural Information Processing Systems*, volume 36, pages 74410–74428, 2023.

Maria Maddalena Barbieri and James O. Berger. Optimal predictive model selection. *The Annals of Statistics*, 32 (3):870–897, 2004. doi: 10.1214/009053604000000238.

Maria Maddalena Barbieri, James O Berger, Edward I George, and Veronika Ročková. The median probability model and correlated variables. *Bayesian Analysis*, 16(4):1085–1112, 2021.

Nilanjana Chakraborty, Kshitij Khare, and George Michailidis. Bayesian group-shrinkage based estimation for panel vector autoregressive models with mixed frequency data. *The Annals of Applied Statistics*, 19(4): 2803–2830, 2025. doi: 10.1214/25-AOAS2073.

Anamitra Chaudhuri, Anirban Bhattacharya, and Yang Ni. Consistent DAG selection for Bayesian causal discovery under general error distributions. *arXiv preprint arXiv:2508.00993*, 2025. doi: 10.48550/arXiv.2508.00993.

Daniel Dempsey and Jason Wyse. Bayesian variable selection in distributed lag models: A focus on binary quantile and count data regressions. *Communications in Statistics—Simulation and Computation*, 2025. doi: 10.1080/03610918.2025.2458578.

Mathias Drton, Marina Garrote-López, Niko Nikov, Elina Robeva, and Y. Samuel Wang. Causal discovery for linear non-Gaussian models with disjoint cycles. In *Conference on Uncertainty in Artificial Intelligence*, volume 286 of *Proceedings of Machine Learning Research*, pages 1064–1083. PMLR, 2025.

Doris Entner and Patrik O Hoyer. On causal discovery from time series data using FCI. In *Proceedings of the 5th European Workshop on Probabilistic Graphical Models*, pages 121–128, 2010.

Karl J Friston. Functional and effective connectivity: A review. *Brain Connectivity*, 1(1):13–36, 2011.

Chang Gong, Chuzhe Zhang, Di Yao, Jingping Bi, Wenbin Li, and Yongjun Xu. Causal discovery from temporal data: An overview and new perspectives. *ACM Computing Surveys*, 57(4):100:1–100:38, 2024. doi: 10.1145/3705297.

Trygve Haavelmo. The statistical implications of a system of simultaneous equations. *Econometrica*, 11(1):1–12, 1943.

Wei Jin, Yang Ni, Amanda B Spence, Leah H Rubin, and Yanxun Xu. Directed cyclic graphs for simultaneous discovery of time-lagged and instantaneous causality from longitudinal data using instrumental variables. *Journal of Machine Learning Research*, 26(22):1–62, 2025.

Gustavo Lacerda, Peter L Spirtes, Joseph Ramsey, and Patrik O Hoyer. Discovering cyclic causal models by independent components analysis. In *Conference on Uncertainty in Artificial Intelligence*, pages 366–374. AUAI Press, 2008.

Markku Lanne, Mika Meitz, and Pentti Saikkonen. Identification and estimation of non-Gaussian structural vector autoregressions. *Journal of Econometrics*, 196(2):288–304, 2017.

Jun S Liu. The collapsed Gibbs sampler in Bayesian computations with applications to a gene regulation problem. *Journal of the American Statistical Association*, 89(427): 958–966, 1994.

Floriana S. Luppino, Leonore M. de Wit, Paul F. Bouvy, Theo Stijnen, Pim Cuijpers, Brenda W. J. H. Penninx, and Frans G. Zitman. Overweight, obesity, and depression: A systematic review and meta-analysis of longitudinal studies. *Archives of General Psychiatry*, 67(3):220–229, 2010.

Alokesh Manna and Sujit K Ghosh. Bayesian models for joint selection of features and auto-regressive lags: Theory and applications in environmental and financial forecasting. *arXiv preprint arXiv:2508.10055*, 2025.

Naveen Naidu Narisetty and Xuming He. Bayesian variable selection with shrinking and diffusing priors. *The Annals of Statistics*, 42(2):789–817, 2014.

William B Nicholson, Ines Wilms, Jacob Bien, and David S Matteson. High dimensional forecasting via interpretable vector autoregression. *Journal of Machine Learning Research*, 21(166):1–52, 2020.

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

Judea Pearl. *Causality: Models, Reasoning, and Inference*. Cambridge University Press, 2nd edition, 2009.

Reinhard Pekrun, Stephanie Lichtenfeld, Herbert W Marsh, Kou Murayama, and Thomas Goetz. Achievement emotions and academic performance: Longitudinal models of reciprocal effects. *Child Development*, 88(5):1653–1670, 2017.

Martin Rohbeck, Brian Clarke, Katharina Mikulik, Alexandra Pettet, Oliver Stegle, and Kai Ueltzhöffer. Bicycle: Intervention-based causal discovery with cycles. In *Conference on Causal Learning and Reasoning*, volume 236 of *Proceedings of Machine Learning Research*, pages 209–242. PMLR, 2024.

Saptarshi Roy, Raymond K W Wong, and Yang Ni. Directed cyclic graph for causal discovery from multivariate functional data. In *Advances in Neural Information Processing Systems*, volume 36, pages 42762–42774, 2023.

Jakob Runge. Discovering contemporaneous and lagged causal relations in autocorrelated nonlinear time series datasets. In *Conference on Uncertainty in Artificial Intelligence*, volume 124 of *Proceedings of Machine Learning Research*, pages 1388–1397. PMLR, 2020.

Jakob Runge, Sebastian Bathiany, Erik Bollt, Gustau Camps-Valls, Dim Coumou, Ethan Deyle, Clark Glymour, Marlene Kretschmer, Miguel D Mahecha, Jordi Muñoz-Marí, et al. Inferring causation from time series in Earth system sciences. *Nature Communications*, 10(1):2553, 2019.

James G Scott and James O Berger. Bayes and empirical-Bayes multiplicity adjustment in the variable-selection problem. *The Annals of Statistics*, 38(5):2587–2619, 2010.

George A. F. Seber and Alan J. Lee. *Linear Regression Analysis*. Wiley, 2nd edition, 2003.

Muralikrishnna G Sethuraman, Romain Lopez, Rahul Mohan, Faramarz Fekri, Tommaso Biancalani, and Jan-Christian Hütter. NODAGS-Flow: Nonlinear cyclic causal structure learning. In *International Conference on Artificial Intelligence and Statistics*, volume 206 of *Proceedings of Machine Learning Research*, pages 6371–6387. PMLR, 2023.

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

Amanda Sonnega, Jessica D. Faul, Mary Beth Ofstedal, Kenneth M. Langa, John W.R. Phillips, and David R. Weir. Cohort profile: the Health and Retirement Study (HRS). *International Journal of Epidemiology*, 43(2): 576–585, 2014. doi: 10.1093/ije/dyu067.

Peter Spirtes. Directed cyclic graphical representations of feedback models. In *Conference on Uncertainty in Artificial Intelligence*, pages 491–498. Morgan Kaufmann, 1995.

Peter Spirtes, Clark N Glymour, and Richard Scheines. *Causation, Prediction, and Search*. MIT Press, 2nd edition, 2000.

Alex Tank, Ian Covert, Nicholas Foti, Ali Shojaie, and Emily B Fox. Neural Granger causality. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 44 (8):4267–4279, 2022.

Luke Tierney. Markov chains for exploring posterior distributions. *The Annals of Statistics*, 22(4):1701–1728, 1994.

Ioannis Tsamardinos, Laura E. Brown, and Constantin F. Aliferis. The max-min hill-climbing Bayesian network structure learning algorithm. *Machine Learning*, 65(1): 31–78, 2006.

Tessa N. van den Kommer, Hannie C. Comijs, Marja J. Aartsen, Martijn Huisman, Dorly J. H. Deeg, and Aartjan T. F. Beekman. Depression and cognition: How do they interrelate in old age? *The American Journal of Geriatric Psychiatry*, 21(4):398–410, 2013.

Diane C. Wagner and Jerome L. Short. Longitudinal predictors of self-rated health and mortality in older adults. *Preventing Chronic Disease*, 11:E93, 2014. doi: 10.5888/pcd11.130241.

Yun Yang, Martin J Wainwright, and Michael I Jordan. On the computational complexity of high-dimensional Bayesian variable selection. *The Annals of Statistics*, 44 (6):2497–2532, 2016.

Laura B. Zahodne, Yaakov Stern, and Jennifer J. Manly. Depressive symptoms precede memory decline, but not vice versa, in non-demented older adults. *Journal of the American Geriatrics Society*, 62(1):130–134, 2014.

He Zhao, Vassili Kitsios, Terence J O'Kane, and Edwin V Bonilla. Bayesian vector AutoRegression with factorised Granger-causal graphs. *arXiv preprint arXiv:2402.03614*, 2024.

Bayesian Causal Discovery in Directed Cyclic Graphs with Closed-Form Lag Bayes Factors (Supplementary Material)

Seyed Reza Maadi[1,2] **Sally Cripps**[1,2]

[1]Human Technology Institute, University of Technology Sydney, Sydney, New South Wales, Australia
[2]School of Mathematical and Physical Sciences, University of Technology Sydney, Sydney, New South Wales, Australia

A EXTENDED LITERATURE COMPARISON

Table 2 provides a comprehensive comparison of the proposed method against representative alternatives across four requirements: Bayesian posterior inference, cyclic graph structure, discrete-time lag modelling, and structured priors that pool evidence across correlated lags.

Table 2: Positioning of the proposed method against representative alternatives. SS = spike-and-slab; CI = conditional independence; VI = variational inference. "Pooled" indicates whether the prior forces candidate lags to compete rather than treating each independently.

Method	Venue	Bayesian	Cycles	Lag model	Pooled lags
DAG-based temporal methods					
PCMCI+ [Runge, 2020]	UAI 2020	×	×	per-lag CI	×
tsFCI [Entner and Hoyer, 2010]	PGM 2010	×	×	per-lag CI	×
DYNOTEARS [Pamfil et al., 2020]	AISTATS 2020	×	×	fixed	×
Neural Granger [Tank et al., 2022]	TPAMI 2022	×	×	penalised	×
Cyclic methods (cross-sectional / interventional)					
NODAGS-Flow [Sethuraman et al., 2023]	AISTATS 2023	×	✓	–	–
Bicycle [Rohbeck et al., 2024]	CLeaR 2024	partial (VI)	✓	–	–
Drton et al. [2025]	UAI 2025	×	✓	–	–
Cyclic methods (temporal)					
DynGFN [Atanackovic et al., 2023]	NeurIPS 2023	✓	✓	continuous	×
CaLLTiF [Arab et al., 2025]	Netw Neurosci 2025	×	✓	per-lag CI	×
BayesDCG [Jin et al., 2025]	JMLR 2025	✓	✓	binary SS	×
Functional DCG [Roy et al., 2023]	NeurIPS 2023	✓	✓	functional SEM	×
Bayesian lag / order selection (acyclic)					
HLag [Nicholson et al., 2020]	JMLR 2020	×	×	grouped	✓
Manna and Ghosh [2025]	arXiv 2025	✓	×	binary SS	×
Dempsey and Wyse [2025]	Comm Stat 2025	✓	×	binary SS	×
Chakraborty et al. [2025]	AOAS 2025	✓	×	group shrink	partial
PFGCG [Zhao et al., 2024]	arXiv 2024	✓	×	binary	×
Multiplicity / dilution theory					
Scott and Berger [2010]	Ann Stat 2010	Multiplicity adjustment for independent indicators			
Barbieri et al. [2021]	Bayes Anal 2021	PIP dilution under correlated predictors			
Proposed	**UAI 2026**	✓	✓	**categorical**	✓

B RUNTIME AND POSTERIOR-AGREEMENT BENCHMARK

This appendix reports the runtime and posterior-agreement benchmark referenced in Section 4.1, verifying that the Cholesky-based optimisations of Section 3.4 accelerate the sampler without altering its target distribution. Table 3 reports, for the two benchmark designs of Jin et al. [2025], the per-iteration wall-clock time of the unoptimised and Cholesky-optimised samplers together with the corresponding structure-recovery metrics; the two implementations agree to within 0.012 on every metric, so recovery is reported for the optimised implementation only.

Table 3: Runtime and recovery on the Jin et al. [2025] benchmarks (R=100 replications). Recovery is for the optimised implementation only; the unoptimised implementation differs by at most 0.012 on every metric.

		SPEED (s/iter)			RECOVERY (opt.)			
					B_0		Lag	
	η	Unopt.	Opt.	Ratio	TPR	FDR	TPR	Acc
I (Q=4)	–	0.742	0.013	55×	1.00	0.292	–	
II (Q=3)	1.0	0.525	0.008	67×	1.00	0.050	0.982	1.00
	0.75	0.525	0.008	67×	1.00	0.029	0.880	1.00
	0.50	0.525	0.008	67×	0.993	0.015	0.804	1.00

C COMPLETE LIKELIHOOD WITH JACOBIAN

From the structural equation (1) of the main text, the mapping from errors to observations is $E_i^{(j)} = (I - B_0)Y_i^{(j)} - \mu - \sum_{\ell=1}^{L_y} B_\ell Y_i^{(j-\ell)} - \sum_{\ell=0}^{L_x} C_\ell X_i^{(j-\ell)}$. The Jacobian of this transformation with respect to $Y_i^{(j)}$ is $(I - B_0)$. By the change-of-variables formula, the likelihood includes the factor $|\det(I - B_0)|$ per observation, yielding an additive log-likelihood term

$$\sum_{(i,j)\in\mathcal{V}} \log|\det(I - B_0)| = T_{\text{valid}} \log|\det(I - B_0)|, \tag{17}$$

where $\mathcal{V} = \{(i,j) : j > L_y$ and the full vector $Y_i^{(j)} \in \mathbb{R}^Q$ is observed$\}$ is the set of valid observation indices and $T_{\text{valid}} = |\mathcal{V}|$. The change-of-variables formula requires the complete Q-dimensional vector at each visit; missingness in lagged predictors $Y_{ip}^{(j-\ell)}$ is handled separately via the refined index sets $\mathcal{I}_q^{(\ell,p)}$ (Appendix F). This is well-defined when $\rho(B_0) < 1$.

D LAPLACE ERRORS, SCALE-MIXTURE AUGMENTATION, AND $Z_{iq}^{(j)}$ UPDATE

The model assumes independent Laplace errors $E_{iq}^{(j)} \sim \text{Laplace}(0, 2\sigma_q)$. The Laplace density is

$$p(e) = \frac{1}{4\sigma_q} \exp\left(-\frac{|e|}{2\sigma_q}\right). \tag{18}$$

A Gaussian scale-mixture representation is used for Gibbs sampling:

$$E_{iq}^{(j)} \mid Z_{iq}^{(j)}, \sigma_q^2 \sim \mathcal{N}\left(0, \frac{\sigma_q^2}{Z_{iq}^{(j)}}\right), \tag{19}$$

$$Z_{iq}^{(j)} \sim \text{Inverse-Gamma}(1, \tfrac{1}{8}). \tag{20}$$

Marginalising over $Z_{iq}^{(j)}$ recovers (18) exactly. Conditional on Z, each outcome equation is a heteroskedastic weighted Gaussian regression with observation weights $w_{iq}^{(j)} = Z_{iq}^{(j)}$.

Inverse-Gaussian full conditional for $Z_{iq}^{(j)}$. Although $Z_{iq}^{(j)}$ has an inverse-gamma mixing prior (20), its Gibbs full conditional is inverse-Gaussian. We use the parameterisation: for $z > 0$,

$$\mathrm{IG}(z; \mu, \lambda) = \left(\frac{\lambda}{2\pi z^3}\right)^{1/2} \exp\left(-\frac{\lambda(z-\mu)^2}{2\mu^2 z}\right). \tag{21}$$

The full conditional is

$$Z_{iq}^{(j)} \mid r_{iq}^{(j)}, \sigma_q^2 \sim \mathrm{IG}(\mu_{iq}^{(j)}, \lambda), \qquad \mu_{iq}^{(j)} = \frac{\sigma_q}{2|r_{iq}^{(j)}|}, \qquad \lambda = \frac{1}{4}. \tag{22}$$

For numerical robustness, a small lower bound $Z_{iq}^{(j)} \geq \varepsilon$ (typically $\varepsilon = 10^{-3}$) is applied.

E RESIDUAL DEFINITIONS

For each observed (i, j, q) with $(i, j) \in \mathcal{I}_q$, define the structural residual

$$\begin{aligned} r_{iq}^{(j)} &= Y_{iq}^{(j)} - \mu_q - \sum_{p \neq q} (B_0)_{qp} Y_{ip}^{(j)} - \sum_{s=1}^{S} (C_0)_{qs} X_{is}^{(j)} \\ &\quad - \sum_{\ell=1}^{L_y} \sum_{p=1}^{Q} (B_\ell)_{qp} Y_{ip}^{(j-\ell)} - \sum_{\ell=1}^{L_x} \sum_{s=1}^{S} (C_\ell)_{qs} X_{is}^{(j-\ell)}, \end{aligned} \tag{23}$$

where $\mathcal{I}_q$ is the valid index set defined in Appendix F. The contemporaneous sum excludes $p = q$ (no instantaneous self-loops); the lagged sums include $p = q$ (autoregressive effects). This residual is used in updating $Z_{iq}^{(j)}$.

For the categorical lag update of edge $(q \leftarrow p)$, the partial residual excludes the contribution of parent p at all lags:

$$r_{iq}^{(j,-p)} = r_{iq}^{(j)} + \sum_{\ell=1}^{L_y} (B_\ell)_{qp} \, Y_{ip}^{(j-\ell)}. \tag{24}$$

Under the categorical encoding, at most one term in this sum is nonzero.

F INDEX SETS FOR IRREGULAR PANELS

The valid index set $\mathcal{I}_q$ and the refined lag-specific set $\mathcal{I}_q^{(\ell,p)}$ are defined in the main text, equations (4)–(5); we collect here the additional conventions for irregular panels. Since the Jacobian term (Appendix C) requires the complete vector $Y_i^{(j)} \in \mathbb{R}^Q$ at each valid visit, we have $\mathcal{I}_q = \mathcal{V}$ for all q in practice, where $\mathcal{V} = \{(i, j) : j > L_y \text{ and } Y_i^{(j)} \in \mathbb{R}^Q \text{ fully observed}\}$. Throughout, we assume $L_x \leq L_y$; when $L_x > L_y$, replace L_y with $\max(L_y, L_x)$ in the index-set definitions. The refined set $\mathcal{I}_q^{(\ell,p)} \subseteq \mathcal{I}_q$ accounts for missingness in the lag-ℓ predictor from parent p.

All summations in (26)–(27) are taken over $(i, j) \in \mathcal{I}_q^{(\ell,p)}$. This index-set formulation avoids dense tensor traversal and enables efficient computation on sparse panels.

G CATEGORICAL LAG BAYES FACTORS

Fix outcome q and candidate parent p; we suppress the outcome subscript q on weights and residuals throughout this section (writing $w_i^{(j)}$ for $w_{iq}^{(j)}$, etc.). For lag $\ell \in \{1, \ldots, L_y\}$, let $y^{(\ell)} \in \mathbb{R}^{|\mathcal{I}_q^{(\ell,p)}|}$ have entries $y_i^{(\ell)} = Y_{ip}^{(j-\ell)}$ over $(i, j) \in \mathcal{I}_q^{(\ell,p)}$. Let $r \in \mathbb{R}^{|\mathcal{I}_q^{(\ell,p)}|}$ be the partial residual (24). With weights $w_{iq}^{(j)} = Z_{iq}^{(j)}$, the conditional regression model is

$$r_i^{(j)} = \beta \, y_i^{(\ell)} + \epsilon_i^{(j)}, \qquad \epsilon_i^{(j)} \mid Z_{iq}^{(j)}, \sigma_q^2 \sim \mathcal{N}\left(0, \frac{\sigma_q^2}{Z_{iq}^{(j)}}\right). \tag{25}$$

The slab prior is $\beta \mid \sigma_q^2 \sim \mathcal{N}(0, \nu\sigma_q^2)$.

Sufficient statistics. Define the weighted sufficient statistics

$$S_{yy}^{(\ell)} = \sum_{(i,j)\in\mathcal{I}_q^{(\ell,p)}} w_{iq}^{(j)}\,(y_i^{(\ell)})^2, \tag{26}$$

$$S_{yr}^{(\ell)} = \sum_{(i,j)\in\mathcal{I}_q^{(\ell,p)}} w_{iq}^{(j)}\,y_i^{(\ell)}\,r_i^{(j)}, \tag{27}$$

$$S_{rr} = \sum_{(i,j)\in\mathcal{I}_q^{(\ell,p)}} w_{iq}^{(j)}\,\big(r_i^{(j)}\big)^2. \tag{28}$$

Proof of Theorem 1 (Closed-Form Lag Bayes Factor). We provide the complete derivation in three steps.

Step 1: Marginal likelihood under the slab. The joint density of $\{r_i^{(j)}\}$ and β is

$$p(\{r_i^{(j)}\},\beta \mid \ell, Z, \sigma_q^2) = \Big[\prod_{(i,j)} \mathcal{N}\big(r_i^{(j)};\ \beta\,y_i^{(\ell)},\ \sigma_q^2/w_i^{(j)}\big)\Big]\,\mathcal{N}(\beta;\ 0,\ \nu\sigma_q^2). \tag{29}$$

The log of the exponent (ignoring normalisation constants) is

$$-\frac{1}{2\sigma_q^2}\left[\sum_{(i,j)} w_i^{(j)}(r_i^{(j)} - \beta\,y_i^{(\ell)})^2 + \frac{\beta^2}{\nu}\right]. \tag{30}$$

Expanding the square and collecting terms in β:

$$\begin{aligned}(30) &= -\frac{1}{2\sigma_q^2}\Big[S_{rr} - 2\beta\,S_{yr}^{(\ell)} + \beta^2\Big(S_{yy}^{(\ell)} + \tfrac{1}{\nu}\Big)\Big] \\ &= -\frac{1}{2\sigma_q^2}\left[\underbrace{\Big(S_{yy}^{(\ell)} + \tfrac{1}{\nu}\Big)\Big(\beta - \frac{S_{yr}^{(\ell)}}{S_{yy}^{(\ell)} + 1/\nu}\Big)^2}_{\text{completed square in }\beta} + S_{rr} - \frac{(S_{yr}^{(\ell)})^2}{S_{yy}^{(\ell)} + 1/\nu}\right].\end{aligned} \tag{31}$$

The posterior variance is $\sigma_q^2/(S_{yy}^{(\ell)} + 1/\nu)$ (equivalently, the posterior precision is $(S_{yy}^{(\ell)} + 1/\nu)/\sigma_q^2$), and the posterior mean is $\hat\beta = S_{yr}^{(\ell)}/(S_{yy}^{(\ell)} + 1/\nu)$. Integrating β out yields a standard Gaussian integral:

$$\begin{aligned}&p(\{r_i^{(j)}\} \mid \ell, Z, \sigma_q^2) \\ &\quad= \prod_{(i,j)} (2\pi\sigma_q^2/w_i^{(j)})^{-1/2}\ \cdot\ (2\pi\nu\sigma_q^2)^{-1/2} \\ &\qquad\times \exp\Big(-\frac{S_{rr} - (S_{yr}^{(\ell)})^2/(S_{yy}^{(\ell)} + 1/\nu)}{2\sigma_q^2}\Big) \\ &\qquad\times (2\pi\sigma_q^2)^{1/2}\,(S_{yy}^{(\ell)} + 1/\nu)^{-1/2}.\end{aligned} \tag{32}$$

Step 2: Marginal likelihood under the null. Setting $\beta = 0$, the exponent reduces to $-S_{rr}/(2\sigma_q^2)$:

$$p(\{r_i^{(j)}\} \mid \beta{=}0, Z, \sigma_q^2) = \prod_{(i,j)} (2\pi\sigma_q^2/w_i^{(j)})^{-1/2}\ \exp\big(-S_{rr}/(2\sigma_q^2)\big). \tag{33}$$

Step 3: Log Bayes factor. Taking the log ratio $\log \mathrm{BF}_\ell = \log p(\{r\} \mid \ell, Z, \sigma_q^2) - \log p(\{r\} \mid \beta{=}0, Z, \sigma_q^2)$ and cancelling the common factors $\prod(2\pi\sigma_q^2/w_i^{(j)})^{-1/2}$ and $\exp(-S_{rr}/(2\sigma_q^2))$:

$$\begin{aligned}\log \mathrm{BF}_\ell &= -\tfrac{1}{2}\log(\nu\sigma_q^2) + \frac{(S_{yr}^{(\ell)})^2}{2\sigma_q^2(S_{yy}^{(\ell)}+1/\nu)} + \tfrac{1}{2}\log\sigma_q^2 - \tfrac{1}{2}\log(S_{yy}^{(\ell)}+1/\nu) \\ &= \frac{1}{2}\left[\frac{(S_{yr}^{(\ell)})^2}{\sigma_q^2(S_{yy}^{(\ell)}+1/\nu)} - \log\big(\nu\,(S_{yy}^{(\ell)}+1/\nu)\big)\right] \\ &= \frac{1}{2}\left[\frac{\nu\,(S_{yr}^{(\ell)})^2}{\sigma_q^2\,(1+\nu S_{yy}^{(\ell)})} - \log(1+\nu S_{yy}^{(\ell)})\right]. \end{aligned} \tag{34}$$

Writing $\kappa_\ell = \nu\, S_{yy}^{(\ell)}$, this is equation (12) of the main text.

Posterior for β (Theorem 1(ii)). From (31), the posterior of β conditional on $z_{qp} = \ell$ is Gaussian with

$$\beta \mid z_{qp} = \ell, Z, \sigma_q^2, \mathcal{D} \sim \mathcal{N}\left(\frac{\nu\, S_{yr}^{(\ell)}}{1+\kappa_\ell},\ \sigma_q^2\,\frac{\nu}{1+\kappa_\ell}\right). \tag{35}$$

Occam penalty (penalty term of Theorem 1(i)). The penalty term $f(\kappa) = -\frac{1}{2}\log(1+\kappa)$ satisfies $f'(\kappa) = -1/(2(1+\kappa)) < 0$ for $\kappa > 0$, so it is strictly decreasing in $S_{yy}^{(\ell)}$. Lags with more observed data (larger $S_{yy}^{(\ell)}$) incur a larger penalty, preventing bias toward shorter lags based on sample size alone.

H COMPLETE MCMC ALGORITHM

Beta posterior for ρ_{lag}. Let M_{act} be the number of ordered pairs (q, p) with $z_{qp} > 0$. With prior $\rho_{\text{lag}} \sim \text{Beta}(a_\rho, b_\rho)$, the posterior is

$$\rho_{\text{lag}} \mid \{z_{qp}\} \sim \text{Beta}\big(a_\rho + M_{\text{act}},\ b_\rho + Q^2 - M_{\text{act}}\big)\,. \tag{36}$$

The use of Q^2 pairs (not $Q(Q-1)$) is appropriate because autoregressive terms ($q = p$) are permitted for lagged effects.

Algorithm 2 presents one post-burn-in sampler iteration with implementation-level details.

In the reported scaling experiments we set $M_{\text{swap}} = 0$ and use $S_{\text{sweep}} = 2$ post-burn-in lag-block sweeps, so the optional lag-swap sub-step is skipped and posterior invariance (Theorem 3) is unchanged.

H.1 PROOF OF THEOREM 3 (POSTERIOR INVARIANCE OF $\mathcal{K}$)

We verify that each sub-kernel leaves the joint posterior π invariant. By Tierney [1994, Theorem 1], the composition of π-invariant kernels is itself π-invariant.

Proof.

(i) $\mathcal{K}_Z$ (Alg. 2, line 2). Each $Z_{iq}^{(j)}$ is updated from its full conditional $Z_{iq}^{(j)} \mid \cdot \sim \text{IG}(\mu_{iq}^{(j)}, \lambda)$ (equation (22)). This is a Gibbs step; π-invariance is immediate.

(ii) $\mathcal{K}_\rho$ (Alg. 2, line 3). Given the current lag indicators $\{z_{qp}\}$, the lag weights are updated via $\tilde{v} \mid \cdot \sim \text{Dir}(\alpha + \text{count}_1, \ldots, \alpha + \text{count}_{L_y})$, a conjugate Gibbs step conditionally independent of (β, σ^2, B_0) given z. The full conditional for ρ_{lag} is $\rho_{\text{lag}} \mid \cdot \sim \text{Beta}(a_\rho + M_{\text{act}},\ b_\rho + Q^2 - M_{\text{act}})$ (equation (36)). Gibbs step.

(iii) $\mathcal{K}_{z,q}$ (Phase 1). For each pair (q, p), we marginalise β_{qp} by replacing the spike-and-slab conditional with the Bayes-factor-weighted marginal. The resulting conditional over $z_{qp} \in \{0, 1, \ldots, L_y\}$ is

$$P(z_{qp} = \ell \mid \text{rest}) \propto \pi_\ell \cdot \mathrm{BF}_\ell \qquad (\ell = 0, 1, \ldots, L_y,\ \ \mathrm{BF}_0 \equiv 1),$$

where "rest" denotes all parameters except (z_{qp}, β_{qp}). Sampling from this, then drawing a provisional β_{qp} from the coefficient posterior (35), constitutes a *collapsed Gibbs step* on the joint (z_{qp}, β_{qp})-space. By Liu [1994], collapsed Gibbs samplers preserve the original target distribution.

Algorithm 2 Categorical Lag DCG Sampler: one post-burn-in iteration

Define $\mathcal{I}_q$ and $\mathcal{I}_q^{(\ell,p)}$ via (4)–(5)
$\mathcal{K}_Z$: Update $\{Z_{iq}^{(j)}\}$ from the inverse-Gaussian full conditional (22) (Appendix D)
$\mathcal{K}_\rho$: Let $\text{count}_\ell = \#\{(q,p) : z_{qp} = \ell\}$; draw $\tilde{v} \mid \cdot \sim \text{Dir}(\alpha + \text{count}_1, \ldots, \alpha + \text{count}_{L_y})$; sample $\rho_{\text{lag}} \mid \cdot$ from (36); set $\pi_0 \leftarrow 1 - \rho_{\text{lag}}$ and $\pi_\ell \leftarrow \rho_{\text{lag}} \tilde{v}_\ell$, $\ell = 1, \ldots, L_y$
for sweep $s = 1, \ldots, S_{\text{sweep}}$ **do**
  for $q = 1, \ldots, Q$ **do**
    Phase 1: collapsed Gibbs ($\mathcal{K}_{z,q}$)
    for $p = 1, \ldots, Q$ **do**
      ▷ $p = q$ is allowed for autoregressive lags
      Form $r^{(-p)}$ via (24)
      for $\ell = 1, \ldots, L_y$ **do**
        Compute $S_{yy}^{(\ell)}$ and $S_{yr}^{(\ell)}$ over $\mathcal{I}_q^{(\ell,p)}$ using (26)–(27)
        Compute $\log \text{BF}_\ell$ via (34), with $\text{BF}_0 \equiv 1$
      end for
      Set $\eta_0 \leftarrow \log \pi_0$ and $\eta_\ell \leftarrow \log \pi_\ell + \log \text{BF}_\ell$, $\ell = 1, \ldots, L_y$
      Sample $z_{qp} \in \{0, 1, \ldots, L_y\}$ with $P(z_{qp} = \ell) \propto \exp(\eta_\ell)$
      if $z_{qp} = 0$ **then**
        Set $\beta_{qp} \leftarrow 0$
      else
        Sample provisional β_{qp} from (35)
      end if
      Update residual bookkeeping for the next parent p
      if $s = S_{\text{sweep}}$ **then**
        Record $P(z_{qp} > 0 \mid \text{rest}) = 1 - \exp(\eta_0 - \text{LSE}_{\ell=0}^{L_y} \eta_\ell)$ for PIP estimation
      end if
    end for
    Phase 1.5: optional lag-swap MH ($\mathcal{K}_{z^\dagger,q}$)
    if $M_{\text{swap}} > 0$ **then**
      Repeat M_{swap} times: choose an active parent p, propose a new lag uniformly from $\{1, \ldots, L_y\} \setminus \{z_{qp}\}$, recompute the relevant Bayes factors using (34), and accept using (16)
    end if
    Phase 2: joint coefficient refresh (supersedes Phase 1 provisional draws)
    Let $A_q = \{p : z_{qp} > 0\}$ and $K_q = |A_q|$
    if $K_q > 0$ **then**
      Define $\mathcal{I}_q^A = \bigcap_{p \in A_q} \mathcal{I}_q^{(z_{qp},p)}$
      Build $\tilde{X}_q$ from columns $Y_{ip}^{(j - z_{qp})}$ for $p \in A_q$ over $(i,j) \in \mathcal{I}_q^A$; set $W_q = \text{diag}(w_{iq}^{(j)})_{(i,j) \in \mathcal{I}_q^A}$
      Sample $\boldsymbol{\beta}_q$ jointly from (37)–(39) via Cholesky
      Write $\boldsymbol{\beta}_q$ into $\{(B_\ell)_{qp}\}$ according to z_{qp} and set inactive lag coefficients to zero
    end if
    Recompute residual r_q with the refreshed coefficients
  end for
end for
$\mathcal{K}_\sigma$: Update covariate effects $\{C_\ell\}_{\ell=0}^{L_x}$ and their indicators, $\{\mu_q\}$, $\{\sigma_q^2\}$, and slab variance ν from conjugate full conditionals, as in Jin et al. [2025]
$\mathcal{K}_{B_0}$: Update B_0 by Metropolis–Hastings subject to $\rho(B_0) < 1$, as in Jin et al. [2025]

After updating (z_{qp}, β_{qp}), the residuals are updated in-place: $r_i^{(j)} \leftarrow r_i^{(j)} + \beta_{\text{old}} Y_{ip}^{(j-\ell_{\text{old}})} - \beta_{\text{new}} Y_{ip}^{(j-\ell_{\text{new}})}$. This bookkeeping updates a derived quantity and does not affect stationarity.

(iv) $\mathcal{K}_{z^\dagger,q}$ **(Phases 1.5 + 2).** This combined kernel first proposes lag swaps on the β-marginalised z-space, then resamples all active coefficients from their exact full conditional given the final z_q.

Swap sub-step. For each active edge (q, r) with $z_{qr} = \ell_{\text{cur}}$, propose $\ell_{\text{prop}} \sim \text{Uniform}(\{1, \ldots, L_y\} \setminus \{\ell_{\text{cur}}\})$. The partial residual $r^{(-r)}$ is formed using the current state of all other coefficients (which may have changed since Phase 1 due to the sequential sweep over parents), and $\text{BF}_{\ell_{\text{cur}}}$, $\text{BF}_{\ell_{\text{prop}}}$ are recomputed on this residual via (34) (reusing the cached $S_{yy}^{(\ell)}$ but recomputing $S_{yr}^{(\ell)}$). With β_{qr} marginalised out, the Metropolis–Hastings acceptance ratio on the collapsed z-space is

$$\alpha = \min\left(1,\ \frac{\pi_{\ell_{\text{prop}}} \cdot \text{BF}_{\ell_{\text{prop}}}}{\pi_{\ell_{\text{cur}}} \cdot \text{BF}_{\ell_{\text{cur}}}}\right),$$

because the proposal is symmetric. This is a valid MH step on the discrete marginal z-space $\{1, \ldots, L_y\}$; detailed balance with respect to $\pi(z_{qr} \mid \cdot)$ holds. Upon acceptance, the swapped parent's provisional coefficient is set to zero and its contribution removed from the stored residual ($r_q \leftarrow r_q + \beta_{qr} Y_{ir}^{(j-\ell_{\text{cur}})}$; $\beta_{qr} \leftarrow 0$), so that subsequent swaps (when $M_{\text{swap}} > 1$) compute Bayes factors from a residual consistent with the β-marginalised z-space. All active coefficients are then redrawn jointly in Phase 2.

Coefficient resample sub-step. Let $K_q = \#\{p : z_{qp} > 0\}$ be the number of active parents for outcome q. Define the intersection index set $\mathcal{I}_q^A = \bigcap_{p \in A_q} \mathcal{I}_q^{(z_{qp}, p)}$, which ensures all active lagged predictors are observed on a common set of rows. Collect the active lag-$\ell(p)$ predictors into a matrix $\tilde{X}_q \in \mathbb{R}^{|\mathcal{I}_q^A| \times K_q}$ and the precision weights into $W_q = \text{diag}(w_{iq}^{(j)})_{(i,j) \in \mathcal{I}_q^A}$. The full conditional for the vector $\boldsymbol{\beta}_q \in \mathbb{R}^{K_q}$ is

$$\boldsymbol{\beta}_q \mid z_q, Z, \sigma_q^2 \sim \mathcal{N}\big(\hat{\boldsymbol{\mu}}_q,\ \sigma_q^2\, \hat{P}_q^{-1}\big), \tag{37}$$

where

$$\hat{P}_q = \tilde{X}_q^\top W_q \tilde{X}_q + \tfrac{1}{\nu} I_{K_q}, \tag{38}$$

$$\hat{\boldsymbol{\mu}}_q = \hat{P}_q^{-1}\, \tilde{X}_q^\top W_q\, r_q, \tag{39}$$

with r_q the residual vector over $\mathcal{I}_q^A$ with all K_q active parents removed. (For balanced panels where all lagged predictors are observed whenever $Y_i^{(j)}$ is observed, $\mathcal{I}_q^A = \mathcal{I}_q$.)

The Cholesky factorisation $\hat{P}_q = L_q L_q^\top$ is used for (a) posterior mean via forward/backward substitution, and (b) sampling via $\boldsymbol{\beta}_q = \hat{\boldsymbol{\mu}}_q + \sigma_q L_q^{-\top} \xi$, $\xi \sim \mathcal{N}(0, I_{K_q})$. This joint draw *supersedes* the provisional univariate draws from Phase 1, ensuring that the final state of $\boldsymbol{\beta}_q$ is drawn from the correct full conditional regardless of any lag changes in the swap sub-step.

Combined π-invariance. The swap sub-step is a valid MH kernel on the marginal z-space (with β integrated out); the resample sub-step then draws $\boldsymbol{\beta}_q$ from its exact conditional given the updated z_q. This "marginal-then-conditional" pattern is a standard collapsed-Gibbs construction [Liu, 1994]: any transition that first updates z while preserving $\pi(z \mid \cdot)$ and then draws $\beta \mid z$ from its full conditional preserves the joint target $\pi(z, \beta \mid \cdot)$.

(v) $\mathcal{K}_\sigma, \mathcal{K}_{B_0}$. $\mathcal{K}_\sigma$ collects the conjugate Gibbs updates for μ_q, σ_q^2, covariate coefficients $\{C_\ell\}_{\ell=0}^{L_x}$ and their selection indicators, and the slab variance ν, all from exact full conditionals. $\mathcal{K}_{B_0}$ is a Metropolis–Hastings step with spectral-radius constraint, as in Jin et al. [2025].

Each component is either a Gibbs step from an exact full conditional or a combined collapsed-Gibbs/MH step satisfying the conditions of Liu [1994]. Their composition preserves π by Tierney [1994, Theorem 1]. □

I EDGE SELECTION CONSISTENCY

This appendix provides the complete proofs of Lemma 1 (Signal Detection) and Theorem 2 (Edge Selection Consistency) stated in Section 3 of the main text.

Notation. For a true edge $(q \leftarrow p) \in \mathcal{E}^*$, let $\ell^* = \ell^*(q, p)$ denote its true lag and $\beta^* = \beta_{qp}^*$ the true coefficient. For a false pair $(q \leftarrow p) \notin \mathcal{E}^*$, $\beta_{qp}^* = 0$ at all lags. All statements are conditional on the Laplace augmentation $\{Z_{iq}^{(j)}\}$ and variance σ_q^2. We do not require pointwise concentration of the individual latent precisions; the proof relies only on the concentration of the weighted sufficient statistics $S_{yy}^{(\ell)}$ and $S_{yr}^{(\ell)}$, together with the corresponding regularity conditions for σ_q^2.

I.1 PROOF OF LEMMA 1 (SIGNAL DETECTION)

Part (a): True edge, BF divergence. At the true lag ℓ^*, the partial residual satisfies

$$r_{iq}^{(j,-p)} = \beta^* \, Y_{ip}^{(j-\ell^*)} + \tilde{\varepsilon}_i^{(j)},$$

where $\tilde{\varepsilon}_i^{(j)}$ is the residual after removing all other predictors. The cross-term sufficient statistic decomposes as

$$S_{yr}^{(\ell^*)} = \beta^* \, S_{yy}^{(\ell^*)} + \underbrace{\sum_{(i,j)} w_i^{(j)} \, Y_{ip}^{(j-\ell^*)} \, \tilde{\varepsilon}_i^{(j)}}_{\equiv R} . \tag{40}$$

By the CLT for weighted sums (the summands have mean zero and finite variance under Assumptions 2 and 4):

$$R = O_P\big(\sigma_q \sqrt{S_{yy}^{(\ell^*)}}\big). \tag{41}$$

Since Assumption 2 implies $(\beta^*)^2 S_{yy}^{(\ell^*)}/\sigma_q^2 \to \infty$, the relative remainder satisfies

$$\frac{R}{\beta^* \, S_{yy}^{(\ell^*)}} = O_P\left(\frac{\sigma_q}{|\beta^*|\sqrt{S_{yy}^{(\ell^*)}}}\right) = O_P\left(\frac{1}{\sqrt{(\beta^*)^2 S_{yy}^{(\ell^*)}/\sigma_q^2}}\right) = o_P(1). \tag{42}$$

Therefore $S_{yr}^{(\ell^*)} = \beta^* \, S_{yy}^{(\ell^*)}\big(1 + o_P(1)\big)$, and substituting into the Bayes factor formula (12) with $\kappa_{\ell^*} = \nu \, S_{yy}^{(\ell^*)} = \Theta(T)$:

$$\begin{aligned} \log \mathrm{BF}_{\ell^*} &= \frac{\nu \, (\beta^*)^2 \, (S_{yy}^{(\ell^*)})^2 \, \big(1 + o_P(1)\big)}{2\sigma_q^2(1+\kappa_{\ell^*})} - \frac{1}{2}\log(1+\kappa_{\ell^*}) \\ &= \frac{(\beta^*)^2 \, S_{yy}^{(\ell^*)}}{2\sigma_q^2} \, \big(1 + o_P(1)\big) - O(\log T), \end{aligned} \tag{43}$$

where in the second line we used $\nu (S_{yy}^{(\ell^*)})^2/(1+\nu S_{yy}^{(\ell^*)}) = S_{yy}^{(\ell^*)} \cdot \kappa_{\ell^*}/(1+\kappa_{\ell^*}) = S_{yy}^{(\ell^*)} \, (1 - O(1/T))$. Therefore:

$$\log \mathrm{BF}_{\ell^*} \geq \frac{(\beta^*)^2 \, S_{yy}^{(\ell^*)}}{4\sigma_q^2} \tag{44}$$

for all sufficiently large T (since $o_P(1) \to 0$ and the $O(\log T)$ penalty is dominated by the signal). By Assumption 2 (Signal Strength), $(\beta^*)^2 \, S_{yy}^{(\ell^*)}/\sigma_q^2 \geq \lambda_T \log T$ with $\lambda_T \to \infty$, so

$$\log \mathrm{BF}_{\ell^*} \geq \frac{\lambda_T}{4} \log T \to \infty.$$

Part (b): False pair, BF bounded. For a false pair, $\beta^* = 0$, so $S_{yr}^{(\ell)} = \sum w_i^{(j)} \, Y_{ip}^{(j-\ell)} \, \tilde{\varepsilon}_i^{(j)}$. By the CLT:

$$\frac{S_{yr}^{(\ell)}}{\sigma_q \sqrt{S_{yy}^{(\ell)}}} \xrightarrow{d} \mathcal{N}(0,1).$$

The data-fit term becomes

$$\frac{\nu \, (S_{yr}^{(\ell)})^2}{\sigma_q^2(1+\kappa_\ell)} = \frac{(S_{yr}^{(\ell)})^2}{\sigma_q^2 \, S_{yy}^{(\ell)}} \cdot \frac{\kappa_\ell}{1+\kappa_\ell} = O_P(1) \cdot O(1) = O_P(1),$$

since $(S_{yr}^{(\ell)})^2/(\sigma_q^2 \, S_{yy}^{(\ell)}) = O_P(1)$ and $\kappa_\ell/(1+\kappa_\ell) \leq 1$. The Occam penalty contributes $-\frac{1}{2}\log(1+\kappa_\ell) = -O(\log T) \leq 0$. Therefore

$$\max_\ell \log \mathrm{BF}_\ell = O_P(1).$$

Stronger bound under Design Regularity (Remark 3). Under Assumption 4, the penalty term admits a precise rate. Since $\kappa_\ell = \nu\, S_{yy}^{(\ell)}$ and $c_{\min}|\mathcal{I}_q^{(\ell,p)}| \le S_{yy}^{(\ell)} \le c_{\max}|\mathcal{I}_q^{(\ell,p)}|$ with $|\mathcal{I}_q^{(\ell,p)}| = \Theta(T)$, we have $\kappa_\ell = \Theta(T)$ uniformly in ℓ. Therefore the Occam penalty satisfies

$$-\tfrac{1}{2}\log(1+\kappa_\ell) = -\tfrac{1}{2}\log T + O(1) \quad \text{uniformly in } \ell.$$

Combining with the data-fit bound $O_P(1)$ established above:

$$\log \mathrm{BF}_\ell = \underbrace{O_P(1)}_{\text{data-fit}} + \underbrace{\left(-\tfrac{1}{2}\log T + O(1)\right)}_{\text{penalty}} = -\tfrac{1}{2}\log T + O_P(1) \quad \text{uniformly in } \ell.$$

Exponentiating, $\max_\ell \mathrm{BF}_\ell = O_P(T^{-1/2}) \to 0$, which is the rate used in the proof of Theorem 2(c).

I.2 PROOF OF THEOREM 2 (EDGE SELECTION CONSISTENCY)

Part (a): True edge recovery, $\mathrm{PIP} \to 1$**.** The posterior odds of edge existence are

$$\frac{P(z_{qp} > 0 \mid \mathcal{D})}{P(z_{qp} = 0 \mid \mathcal{D})} = \frac{\sum_{\ell=1}^{L_y} \pi_\ell\, \mathrm{BF}_\ell}{\pi_0} \ge \frac{\pi_{\ell^*}\, \mathrm{BF}_{\ell^*}}{\pi_0}. \tag{45}$$

By Lemma 1(a), $\log \mathrm{BF}_{\ell^*} \to \infty$. The prior ratio is

$$\frac{\pi_{\ell^*}}{\pi_0} = \frac{\rho_{\text{lag}}\, \tilde{v}_{\ell^*}}{1-\rho_{\text{lag}}} \; > \; 0 \quad \text{a.s.},$$

since $\rho_{\text{lag}} \in (0,1)$ a.s. under its continuous Beta posterior (whose parameters depend only on fixed Q and the current edge count). Moreover, for any $\varepsilon > 0$, $P(\rho_{\text{lag}} \ge \varepsilon \mid \{z_{qp}\})$ is bounded below by a positive constant uniformly in T. Since $\mathrm{BF}_{\ell^*} \xrightarrow{P} \infty$, the posterior odds (45) diverge in probability, giving $P(z_{qp} > 0 \mid \mathcal{D}) \to 1$.

Part (b): True lag recovery, $P(z_{qp} = \ell^* \mid \mathcal{D}) \to 1$**.** Conditional on $z_{qp} > 0$:

$$P(z_{qp} = \ell \mid z_{qp} > 0, \mathcal{D}) = \frac{\pi_\ell\, \mathrm{BF}_\ell}{\sum_{m=1}^{L_y} \pi_m\, \mathrm{BF}_m}.$$

For any $\ell \neq \ell^*$, we bound the log Bayes factor ratio. From (12):

$$\log \frac{\mathrm{BF}_\ell}{\mathrm{BF}_{\ell^*}} = \frac{1}{2}\left[\frac{\nu (S_{yr}^{(\ell)})^2}{\sigma_q^2(1+\kappa_\ell)} - \frac{\nu (S_{yr}^{(\ell^*)})^2}{\sigma_q^2(1+\kappa_{\ell^*})}\right] + \frac{1}{2}\log\frac{1+\kappa_{\ell^*}}{1+\kappa_\ell}. \tag{46}$$

We rewrite the data-fit terms using the factorisation

$$\frac{\nu (S_{yr}^{(\ell)})^2}{\sigma_q^2(1+\kappa_\ell)} = \frac{(S_{yr}^{(\ell)})^2}{\sigma_q^2\, S_{yy}^{(\ell)}} \cdot \frac{\kappa_\ell}{1+\kappa_\ell}. \tag{47}$$

By Assumption 4 (Design Regularity), $\kappa_\ell = \nu\, S_{yy}^{(\ell)} = \Theta(T)$ for all ℓ, so $\kappa_\ell/(1+\kappa_\ell) = 1 - O(1/T)$. Writing $\phi_\ell = \kappa_\ell/(1+\kappa_\ell)$ and $D(\ell) = (S_{yr}^{(\ell)})^2 / S_{yy}^{(\ell)}$, the data-fit difference becomes

$$\frac{D(\ell)\,\phi_\ell - D(\ell^*)\,\phi_{\ell^*}}{\sigma_q^2} = \frac{D(\ell) - D(\ell^*)}{\sigma_q^2} + O\!\left(\frac{\max(D(\ell), D(\ell^*))}{T\,\sigma_q^2}\right).$$

By Assumption 3: $D(\ell) < D(\ell^*) - c\,\sigma_q^2 \log T$, so the leading term is at most $-c\log T$. The correction is $O(1)$ since $D(\ell^*)/\sigma_q^2 = O(T)$ (from Assumptions 2 and 4) and the $O(1/T)$ factor absorbs it.

For the penalty ratio: since both κ_ℓ and κ_{ℓ^*} are $\Theta(T)$, $\frac{1}{2}\log[(1+\kappa_{\ell^*})/(1+\kappa_\ell)] = O(1)$.

Combining:

$$\log \frac{\mathrm{BF}_\ell}{\mathrm{BF}_{\ell^*}} \le -\frac{c}{2}\log T + O(1) \to -\infty.$$

Since L_y is fixed:

$$\sum_{\ell \neq \ell^*} \frac{\mathrm{BF}_\ell}{\mathrm{BF}_{\ell^*}} \to 0,$$

and therefore $P(z_{qp} = \ell^* \mid z_{qp} > 0, \mathcal{D}) \to 1$.

Part (c): False edge exclusion, $\mathrm{PIP} \to 0$**.** For a false pair, the conditional inclusion probability given ρ_{lag} is

$$P(z_{qp} > 0 \mid \rho_{\text{lag}}, \mathcal{D}) = \frac{\rho_{\text{lag}} \sum_\ell \tilde{v}_\ell \, \mathrm{BF}_\ell}{(1-\rho_{\text{lag}}) + \rho_{\text{lag}} \sum_\ell \tilde{v}_\ell \, \mathrm{BF}_\ell}.$$

Under Design Regularity (Assumption 4), the strengthened bound of Remark 3 gives $\log \mathrm{BF}_\ell = -\frac{1}{2}\log T + O_P(1)$ uniformly in ℓ; exponentiating, $\max_\ell \mathrm{BF}_\ell = O_P(T^{-1/2}) \to 0$. Since $\sum_\ell \tilde{v}_\ell \, \mathrm{BF}_\ell \le \max_\ell \mathrm{BF}_\ell \to 0$, the numerator vanishes while the denominator converges to $(1-\rho_{\text{lag}}) > 0$ for every $\rho_{\text{lag}} < 1$. Because ρ_{lag} has a continuous Beta posterior supported on $(0,1)$, $\rho_{\text{lag}} < 1$ holds almost surely. Therefore $P(z_{qp} > 0 \mid \rho_{\text{lag}}, \mathcal{D}) \xrightarrow{P} 0$ for $P(\rho_{\text{lag}} \mid \mathcal{D})$-almost every ρ_{lag}. Since the integrand is bounded by 1, dominated convergence yields

$$P(z_{qp} > 0 \mid \mathcal{D}) = \mathbb{E}_{\rho_{\text{lag}}}\big[P(z_{qp} > 0 \mid \rho_{\text{lag}}, \mathcal{D})\big] \;\to\; 0.$$

J EVIDENCE POOLING GAIN

Under the calibrated priors of Remark 2, the categorical encoding assigns at least as much posterior probability to edge existence as the strongest individual lag under the independent encoding whenever the best-fitting lag receives at least its average share of the prior; more generally, its advantage arises from summing evidence across the competing lag states.

Proposition 3 (Pooling Dominance) *Fix an ordered pair (q,p). Let $q_\ell = P(\gamma_{\ell qp}=1 \mid \mathcal{D})$ denote the per-lag PIP under the independent encoding with $\rho_{\text{ind}} = 1-(1-\rho_{\text{lag}})^{1/L_y}$ (Remark 2), and let $\mathrm{PIP}^{\text{cat}} = P(z_{qp}>0 \mid \mathcal{D})$ under the categorical prior. Then:*

(a) *Under uniform lag weights ($\tilde{v}_\ell = 1/L_y$ for all ℓ), $\mathrm{PIP}^{\text{cat}} \ge \max_\ell q_\ell$. More generally, the result holds whenever $\tilde{v}_{\ell^*} \ge 1/L_y$, where $\ell^* = \arg\max_\ell \mathrm{BF}_\ell$.*

(b) *The independent-encoding median-model edge-selection decision is $\widehat{E}_{qp}^{\text{ind}} = \mathbf{1}[\exists\, \ell : q_\ell > 0.5]$, a binary rule rather than a posterior probability. When posterior mass disperses across $K \ge 2$ correlated lags so that every $q_\ell < 0.5$, the independent rule misses the edge ($\widehat{E}_{qp}^{\text{ind}} = 0$), whereas $\mathrm{PIP}^{\text{cat}}$ can remain arbitrarily close to 1.*

Proof. **Part (a).** Under the independent encoding, each lag has a univariate spike-and-slab:

$$q_\ell = \frac{\rho_{\text{ind}} \, \mathrm{BF}_\ell}{1-\rho_{\text{ind}} + \rho_{\text{ind}} \, \mathrm{BF}_\ell} = \sigma\big(\log \mathrm{BF}_\ell + \log \tfrac{\rho_{\text{ind}}}{1-\rho_{\text{ind}}}\big),$$

where σ is the logistic function. The categorical PIP is

$$\mathrm{PIP}^{\text{cat}} = \frac{\sum_{\ell=1}^{L_y} \pi_\ell \, \mathrm{BF}_\ell}{\pi_0 + \sum_{\ell=1}^{L_y} \pi_\ell \, \mathrm{BF}_\ell}.$$

Under the calibrated prior, $\sum_\ell \pi_\ell = \rho_{\text{lag}}$ and $\pi_0 = 1-\rho_{\text{lag}}$. Since $f(x) = x/(\pi_0 + x)$ is strictly increasing, for the best lag $\ell^* = \arg\max_\ell \mathrm{BF}_\ell$:

$$\mathrm{PIP}^{\text{cat}} \ge \frac{\pi_{\ell^*} \, \mathrm{BF}_{\ell^*}}{\pi_0 + \pi_{\ell^*} \, \mathrm{BF}_{\ell^*}} = \frac{\rho_{\text{lag}} \tilde{v}_{\ell^*} \, \mathrm{BF}_{\ell^*}}{1-\rho_{\text{lag}} + \rho_{\text{lag}} \tilde{v}_{\ell^*} \, \mathrm{BF}_{\ell^*}}. \tag{48}$$

It suffices to show (48) $\ge q_{\ell^*}$, i.e.,

$$\frac{\rho_{\text{lag}} \tilde{v}_{\ell^*}}{1-\rho_{\text{lag}}} \;\ge\; \frac{\rho_{\text{ind}}}{1-\rho_{\text{ind}}}. \tag{49}$$

By the calibration identity, $1-\rho_{\text{ind}} = (1-\rho_{\text{lag}})^{1/L_y}$. Let $u = (1-\rho_{\text{lag}})^{1/L_y} \in (0,1)$, so $1-\rho_{\text{lag}} = u^{L_y}$. Then the left-hand side of (49) (with $\tilde{v}_{\ell^*} \ge 1/L_y$) satisfies

$$\frac{\rho_{\text{lag}} \tilde{v}_{\ell^*}}{1-\rho_{\text{lag}}} \;\ge\; \frac{1-u^{L_y}}{L_y \, u^{L_y}},$$

while the right-hand side is $(1-u)/u$. The required inequality $(1-u^{L_y})/(L_y \, u^{L_y}) \ge (1-u)/u$ simplifies to $u(1-u^{L_y}) \ge L_y \, u^{L_y}(1-u)$. Factoring $1-u^{L_y} = (1-u)\sum_{k=0}^{L_y-1} u^k$ and cancelling $(1-u) > 0$ gives $\sum_{k=1}^{L_y} u^k \ge L_y \, u^{L_y}$, which holds because $u \in (0,1)$ implies $u^k \geq u^{L_y}$ for every $k \leq L_y$. (Equality holds when $L_y = 1$, where both models coincide.) Under geometric-decay weights with $d < 1$, the condition $\tilde{v}_{\ell^*} \geq 1/L_y$ holds for early lags (which receive above-average prior mass) and may fail for later lags; the overall pooling advantage then comes from the summation in the numerator of $\mathrm{PIP}^{\mathrm{cat}}$ rather than from per-lag dominance.

Part (b). Under the median probability model [Barbieri and Berger, 2004, Barbieri et al., 2021], the independent encoding declares edge $(q \leftarrow p)$ present if and only if $\exists\, \ell : q_\ell > 0.5$. When serial correlation causes the signal to disperse so that $q_\ell < 0.5$ for all ℓ, the edge is missed. The categorical PIP suffers no such threshold failure: by Theorem 2(a), $\mathrm{PIP}^{\mathrm{cat}} \to 1$ for every true edge as $T \to \infty$ under the signal and design conditions (Assumptions 2 and 4), irrespective of the cross-lag correlation structure. Lag separation (Assumption 3) is required only for consistent recovery of the exact lag (Theorem 2(b)), not for the edge-existence pooling effect described here. $\square$

K MULTI-LAG ROBUSTNESS

We prove the three parts of Proposition 1. Suppose the true DGP for equation q has two active lags, ℓ_1 and ℓ_2, for parent p:

$$r_{iq}^{(j,-p)} = \beta_1^* Y_{ip}^{(j-\ell_1)} + \beta_2^* Y_{ip}^{(j-\ell_2)} + \varepsilon_{iq}^{(j)}, \tag{50}$$

where $|\beta_1^*| \geq |\beta_2^*| > 0$ and $\varepsilon_{iq}^{(j)}$ is zero-mean with conditional variance $\sigma_q^2 / w_{iq}^{(j)}$.

Let $\mathcal{I}^{(12)} = \mathcal{I}_q^{(\ell_1,p)} \cap \mathcal{I}_q^{(\ell_2,p)}$ denote the common observation set and define the weighted predictor cross-products over $\mathcal{I}^{(12)}$:

$$S_{11} = \textstyle\sum_{(i,j)\in\mathcal{I}^{(12)}} w_{iq}^{(j)} \left(Y_{ip}^{(j-\ell_1)}\right)^2, \tag{51}$$

$$S_{22} = \textstyle\sum_{(i,j)\in\mathcal{I}^{(12)}} w_{iq}^{(j)} \left(Y_{ip}^{(j-\ell_2)}\right)^2, \tag{52}$$

$$S_{12} = \textstyle\sum_{(i,j)\in\mathcal{I}^{(12)}} w_{iq}^{(j)}\, Y_{ip}^{(j-\ell_1)}\, Y_{ip}^{(j-\ell_2)}. \tag{53}$$

The weighted predictor correlation is $\varphi_{\Delta\ell} = S_{12}/\sqrt{S_{11}S_{22}}$.

Part (a): VIF. The standard result for two correlated predictors gives $\mathrm{VIF} = 1/(1-\varphi_{\Delta\ell}^2)$ [Seber and Lee, 2003, Ch. 3]. Under an AR(1)-like marginal process with autocorrelation φ, the weighted cross-product satisfies $\varphi_{\Delta\ell} \approx \varphi^{|\ell_1-\ell_2|}$ when the panel is balanced and weights are approximately constant across lags.

Part (b): Coefficient absorption. The categorical model fits a single lag ℓ_1. The weighted least squares estimator of β_{qp} using only lag ℓ_1 satisfies

$$\hat{\beta} = \frac{S_{1r}}{S_{11}} = \frac{\sum w_{iq}^{(j)}\, Y_{ip}^{(j-\ell_1)}\, (\beta_1^* Y_{ip}^{(j-\ell_1)} + \beta_2^* Y_{ip}^{(j-\ell_2)} + \varepsilon_{iq}^{(j)})}{S_{11}} \xrightarrow{P} \beta_1^* + \beta_2^* \frac{S_{12}}{S_{11}} = \beta_1^* + \beta_2^*\, \varphi_{\Delta\ell} \sqrt{S_{22}/S_{11}},$$

since the cross-term converges to $\beta_2^* S_{12}/S_{11} = \beta_2^*\, \varphi_{\Delta\ell}\sqrt{S_{22}/S_{11}}$; the omitted lag is thus absorbed in proportion to its signed weighted projection onto the retained lag. Under balanced designs where the two lagged predictors are on the same weighted scale, $S_{11} \approx S_{22}$ (as for standardised outcomes), this reduces to the balanced-design corollary $\beta_1^* + \beta_2^*\, \varphi_{\Delta\ell}$ corresponding to the balanced-design approximation discussed in Proposition 1. Edge detection requires $\hat{\beta} \neq 0$, which holds whenever the exact limit $\beta_1^* + \beta_2^*\, S_{12}/S_{11}$ is nonzero.

Part (c): Prediction cost. The residual variance under the single-lag model is

$$\sigma_{\text{cat}}^2 = \sigma_{\text{true}}^2 + (\beta_2^*)^2\, \sigma_{Y,\ell_2}^2\, (1 - \varphi_{\Delta\ell}^2),$$

where $\sigma_{Y,\ell_2}^2 = S_{22}/|\mathcal{I}^{(12)}|$ is the average weighted second moment of the omitted predictor. The increase $\Delta\sigma^2 = (\beta_2^*)^2\, \sigma_{Y,\ell_2}^2\, (1 - \varphi_{\Delta\ell}^2)$ follows from the Frisch–Waugh–Lovell theorem: the omitted variable bias inflates the residual variance by the variance of the omitted component not explained by the included predictor.

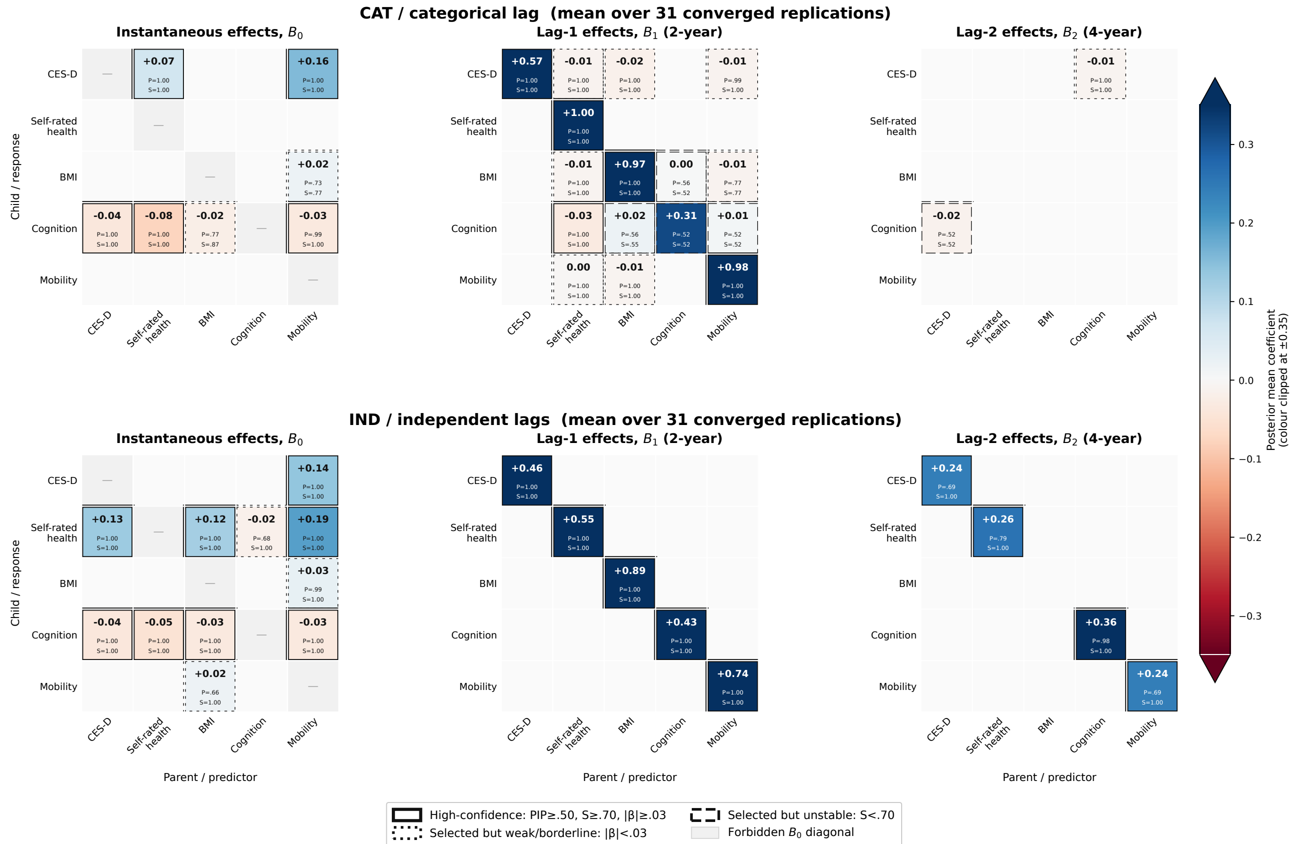

Figure 3: Full posterior coefficient matrices for the categorical (top) and independent (bottom) encodings on the HRS panel, across contemporaneous (B_0), lag-1 (B_1, 2-year), and lag-2 (B_2, 4-year) blocks; rows index the child/response and columns the parent/predictor. Each selected cell reports the posterior mean coefficient with its mean PIP (P) and replication stability (S); the border style marks selected-stable-non-negligible, selected-but-weak, and selected-but-unstable edges, and the B_0 diagonal is forbidden by construction. Colour intensity is clipped at ± 0.35 for readability; the autoregressive self-loops reach ≈ 1.0 and saturate the scale, while the cell annotations report the exact posterior means.

L FULL POSTERIOR SUMMARIES FOR THE HRS APPLICATION

Figure 2 in the main text shows a high-confidence *display* graph; here we report the complete posterior summaries that underlie it. Each matrix cell carries three quantities, averaged over the $R{=}31$ replications: the posterior mean coefficient $\hat{\beta}$, the *layer-specific* posterior inclusion probability, and the replication stability; Figure 3 annotates these for the selected cells. Unlike the edge-level detection probability used in Section 4.2 ($P(z_{qp}{>}0 \mid \mathcal{D})$, or $\max_\ell P(\gamma_{\ell qp}{=}1 \mid \mathcal{D})$ for the independent encoding), the matrix PIP is specific to each cell's lag: for contemporaneous effects it is the posterior probability that the corresponding B_0 coefficient is active; for categorical lagged effects in block B_ℓ it is $P(z_{qp}{=}\ell \mid \mathcal{D})$; and for independent lagged effects in block B_ℓ it is $P(\gamma_{\ell qp}{=}1 \mid \mathcal{D})$. The replication stability is the fraction of replications in which that same cell's within-run PIP exceeds 0.5. A cell is *selected* when its mean PIP exceeds 0.5 [Barbieri and Berger, 2004]; the main-text display additionally requires stability ≥ 0.70 and $|\hat{\beta}|{\geq}0.03$.

Mean PIP and replication stability can diverge: mean PIP is the average inclusion strength, which a few highly confident replications can push above 0.5, whereas stability counts how many replications *individually* clear the median-model threshold and so certifies that support is consistent rather than concentrated. In the HRS application the two largely agree (most selected cells have stability 1.0); the categorical cognition self-loop is the borderline case the display rule screens out: at lag-1 it is selected (cell PIP ≈ 0.52) but its stability is only ≈ 0.52, below the 0.70 cut, so it is excluded from the high-confidence display of Figure 2 while remaining visible here.

M SYNTHETIC DGP SPECIFICATION

This appendix provides full details of the data-generating process used in Section 4.2. Balanced panels have N=200 individuals, J=10 visits, and L_y=4 maximum lag.

The contemporaneous matrix B_0 is constructed with sign-cancelling cyclic blocks: for each cycle of length k, alternating positive and negative entries ensure $\rho(B_0) < 1$. Spectral radii are calibrated per Q: $\rho(B_0) = 0.40$ (Q=10), 0.35 (Q=15), 0.30 (Q=20).

Cross-variable lagged edges are placed with 4–6% density in the off-diagonal entries of B_1, B_2, B_3; lag 4 (B_4) is empty (decoy). In SINGLE scenarios, all cross-variable edges are at lag 1. In MIXED scenarios, edges are distributed: Q=10 uses lags $\{1, 2\}$; $Q \in \{15, 20\}$ uses lags $\{1, 2, 3\}$. Weak coefficients are drawn from $|\beta| \sim \mathrm{Uniform}(0.05, 0.12)$ and strong from $|\beta| \sim \mathrm{Uniform}(0.20, 0.40)$, with signs chosen uniformly.

Diagonal AR persistence: $(B_1)_{qq} \approx 0.48$–0.66, adaptively scaled from a raw value of $0.15 + 0.65\phi$ (with ϕ=0.9) so the companion-form spectral radius satisfies $\rho < 1$, plus a $\mathrm{Uniform}(-0.03, 0.03)$ jitter. Innovations are Laplace with error variance $\sigma^2_{\varepsilon,q} \sim \mathrm{Uniform}(0.08, 0.15)$, i.e. $\varepsilon^{(j)}_{iq} \sim \mathrm{Laplace}(0, \sqrt{\sigma^2_{\varepsilon,q}/2}\,)$. Under the model convention $\varepsilon \sim \mathrm{Laplace}(0, 2\sigma_q)$, this corresponds to $\sigma^2_q = \sigma^2_{\varepsilon,q}/8$. S=Q exogenous covariates follow an AR(1) process with autocorrelation 0.7 and marginal $\mathrm{Normal}(0, 1)$ distribution. Contemporaneous covariate effects are diagonal, $|(C_0)_{q,q}| \sim \mathrm{Uniform}(0.4, 0.7)$, with sparse lagged covariate effects at lags 1–2.

N SAME-PAIR MULTI-LAG VIOLATION

This appendix presents the full results for the SAME-PAIR experiment summarised in Section 4.2 of the main text, together with the oracle verification of Proposition 1(b). The experiment is fully specified by the data-generating process, metrics, and tabulated outputs below.

N.1 DGP SPECIFICATION

Q=15, L_y=4, balanced J=10. Contemporaneous B_0 uses sign-cancelling cyclic structure with $\rho(B_0)$=0.10 (smaller than Appendix M, to hold the contemporaneous structure fixed while isolating lag selection). Exactly 30% of active cross-variable ordered pairs carry two simultaneously active lags, both drawn from $\{1, 2\}$, with $|\beta^*_2| = |\beta^*_1|/2$ and matching sign; remaining edges are single-lag with the weak/strong coefficient mix of Appendix M; lags 3–4 are decoys. The persistence parameter φ is calibrated so the *realized* inter-lag predictor correlation equals $\varphi_{\Delta\ell} \approx \varphi$ (measured: 0.55, 0.76, 0.92 at φ=0.5, 0.7, 0.9); these measured values, rather than the configured φ itself, are the operative quantities for the VIF and oracle-projection checks. The diagonal autoregressive coefficient is small at low φ, with autocorrelation carried largely by the autocorrelated exogenous covariate. Laplace innovations as in Appendix M; R=31 replications.

N.2 FULL 18-CELL CROSS-VARIABLE RECOVERY TABLE

Table 4 reports all 18 cells. The headline recovery metrics are cross-variable edge-level metrics: TPR, FDR, and SHD are computed on ordered pairs (q, p) with $q \neq p$, irrespective of lag, and therefore exclude self-loops. Self-loop-inclusive FDR and SHD are given separately in Table 5. The Any, Both, and Dom. columns report lag-specific recovery on the violating pairs; Both requires both true lags to be detected, while extra selected lags do not affect that score.

Table 4: Same-pair multi-lag violation: full 18-cell cross-variable *edge-level* recovery (R=31, mean ± SE). TPR, FDR, and SHD are lag-agnostic and exclude self-loops, in contrast to the lag-level triple metrics in Table 1; each ordered pair is counted once. The Any, Both, and Dom. columns report lag-specific recovery on the violating pairs; Both is a subset criterion (both true lags must be detected, while spurious extra lags do not affect the score). This regime violates Assumption 1, so Theorem 2 does not apply.

φ	N	Meth.	TPR	FDR	SHD	Any	Both	Dom.
0.5	200	CAT	0.884 ± 0.017	0.006 ± 0.004	1.84 ± 0.25	1.000	0.000	0.952 ± 0.018
		IND	0.553 ± 0.006	0.000 ± 0.000	6.71 ± 0.08	1.000	0.742	1.000 ± 0.000
	500	CAT	0.983 ± 0.008	0.073 ± 0.012	1.52 ± 0.23	1.000	0.000	0.960 ± 0.017
		IND	0.570 ± 0.007	0.000 ± 0.000	6.45 ± 0.10	1.000	0.831	1.000 ± 0.000
	2000	CAT	1.000 ± 0.000	0.143 ± 0.015	2.68 ± 0.32	1.000	0.000	0.960 ± 0.017
		IND	0.594 ± 0.010	0.000 ± 0.000	6.10 ± 0.15	1.000	0.863	1.000 ± 0.000
0.7	200	CAT	0.910 ± 0.014	0.023 ± 0.007	1.71 ± 0.22	1.000	0.000	0.935 ± 0.020
		IND	0.596 ± 0.010	0.000 ± 0.000	6.06 ± 0.15	1.000	0.863	1.000 ± 0.000
	500	CAT	0.991 ± 0.005	0.085 ± 0.014	1.65 ± 0.28	1.000	0.000	0.927 ± 0.021
		IND	0.611 ± 0.010	0.000 ± 0.000	5.84 ± 0.15	1.000	0.879	1.000 ± 0.000
	2000	CAT	1.000 ± 0.000	0.175 ± 0.018	3.48 ± 0.47	1.000	0.000	0.919 ± 0.024
		IND	0.641 ± 0.013	0.000 ± 0.000	5.39 ± 0.20	1.000	0.944	1.000 ± 0.000
0.9	200	CAT	0.977 ± 0.008	0.029 ± 0.012	0.71 ± 0.19	1.000	0.000	0.871 ± 0.030
		IND	0.604 ± 0.011	0.000 ± 0.000	5.81 ± 0.21	0.984	0.750	0.984 ± 0.011
	500	CAT	0.998 ± 0.002	0.094 ± 0.020	1.61 ± 0.33	1.000	0.000	0.879 ± 0.035
		IND	0.635 ± 0.014	0.000 ± 0.000	5.39 ± 0.23	1.000	0.815	1.000 ± 0.000
	2000	CAT	0.998 ± 0.002	0.244 ± 0.020	4.97 ± 0.45	1.000	0.000	0.847 ± 0.032
		IND	0.652 ± 0.016	0.000 ± 0.000	5.13 ± 0.26	1.000	0.863	1.000 ± 0.000

N.3 SELF-LOOP-INCLUSIVE DISCLOSURE

At φ=0.5 the diagonal autoregressive signal is weak and autocorrelation is carried largely by the exogenous covariate channel, elevating self-loop false positives for CAT; this is outside the cross-variable phenomenon studied in the main experiment. Table 5 reports self-loop-inclusive FDR and SHD for completeness. At $\varphi \in \{0.7, 0.9\}$ the self-loop-inclusive SHD coincides with the cross-variable SHD of Table 4, because in those regimes the self-loops are recovered cleanly and contribute no additional structural errors; the two diverge only at φ=0.5, where the covariate-driven self-loop false positives inflate the self-loop-inclusive total.

Table 5: CAT self-loop-inclusive metrics for the SAME-PAIR experiment (cross-variable FDR shown for comparison).

φ	N	Self-loop-incl. FDR	Self-loop-incl. SHD	(Cross-var FDR)
0.5	200	0.033	2.23	0.006
	500	0.131	2.61	0.073
	2000	0.327	7.58	0.143
0.7	200	0.012	1.71	0.023
	500	0.046	1.65	0.085
	2000	0.099	3.48	0.175
0.9	200	0.014	0.71	0.029
	500	0.049	1.61	0.094
	2000	0.140	4.97	0.244

N.4 ORACLE VERIFICATION OF PROPOSITION 1(B)

We verify Proposition 1(b) at the oracle level by forward-simulating the data-generating process and running the single-lag oracle projection on each violating pair (regressing the exact partial residual on the dominant-lag predictor). In this balanced design the unweighted oracle regression matches the analytic balanced-design corollary $1 + \frac{1}{2}\varphi_{\Delta\ell}$ (which uses the design ratio $\beta_2^* = \beta_1^*/2$) to within simulation noise, confirming the projection algebra. The exact weighted form remains $\beta_1^* + \beta_2^* S_{12}/S_{11}$ (Appendix K); exact weighted validation would require fixing or saving the corresponding augmentation-weight sequence $\{Z_{iq}^{(j)}\}$. Proposition 1(b) is an oracle single-lag projection limit; the full-sampler posterior coefficient is expected to lie below it because of slab shrinkage, joint multivariate estimation, and selection uncertainty, so we treat the agreement between the analytic corollary and the forward-simulation regression as the verifiable check and read the sampler comparison only qualitatively. Table 6 reports the realized inter-lag correlation, the analytic balanced-design corollary, and the forward-simulation oracle regression for each φ.

Table 6: Oracle verification of Proposition 1(b): realized inter-lag correlation, analytic balanced-design corollary, and oracle regression from forward simulation (mean over violating pairs and replications, R=31). The forward-simulation projection matches the analytic corollary to simulation noise, confirming the projection algebra; the design ratio is $\beta_2^* = \beta_1^*/2$.

φ	$\varphi_{\Delta\ell}$ (realized)	Balanced corollary $1+\frac{1}{2}\varphi_{\Delta\ell}$	Oracle regression (sim.)
0.5	0.55	1.27	1.27
0.7	0.76	1.38	1.38
0.9	0.92	1.46	1.46

Relation to Theorem 2(c). The cross-variable FDR rises with N in the most collinear cells (reaching 0.24 at φ=0.9, N=2000); this does not contradict Theorem 2(c), which is an asymptotic per-pair exclusion result under Assumption 1, whereas the SAME-PAIR design is finite-sample and deliberately violates Assumption 1.