Tsinghua Science and Technology
Doi: xxxx

Research Article

Structured Sparsity of Candidate Space Prior to Mixed Embedding Reconstruction in Nonuniform Embedding

Zheng Tan, Xin Weng, and Ziyu Jia (✉)

Abstract Causal inference based on mixed embedding and conditional mutual information (CMI) is widely applied to detect directed interactions in multivariate time series. However, as the number of variables or the maximum lag order increases, the candidate embedding space expands rapidly, which may lead to instability in CMI estimation and an increased risk of selecting redundant components under limited sample conditions. This study proposes a structured candidate-space control framework prior to mixed-embedding reconstruction. Variables and their lagged components are organized into variable-level groups, and the search space of the embedding reconstruction is compressed through an inter-group sparsity strategy. Nonlinear representations are incorporated into a group-sparse screening representation to identify relevant source variables. The subsequent mixed embedding reconstruction is restricted to the original lagged components of the retained variables, while its causal criterion, stopping rule, and normalization scheme remain unchanged. Across the simulation systems, the proposed method reduced false-positive detections while preserving the true causal relationships, but its benefit depends on retaining the relevant source variables during screening. The structured sparsity suppresses redundant candidates, whereas enriched nonlinear representations improve the retention of causal sources. Empirical analyses on real financial market data further show that the strongest relationships inferred by the proposed method were more reproducible under the examined parameter perturbations. Moreover, the reconstructed causal interaction networks reveal structural shift before and after major market disruptions in application. Overall, structured candidate-space control provides a practical mechanism for improving the parsimony and reproducibility of information-theoretic nonuniform embedding, provided that the relevant causal sources are retained during screening.

Keywords causal inference; nonuniform embedding; structured sparsity; multivariate time series; high-dimensional systems

1 Introduction

The study of real-world complex systems from the perspective of complex networks has attracted increasing attention in recent years[1]. Understanding such systems crucially depends on identifying causal interactions among multivariate time series[2–4]. The measures of information theory provide a nonlinear extension of the classical Granger causality[5] framework and have led to the development of various causality measures, including transfer entropy (TE)[6] and nonuniform embedding method[7, 8]. Among these approaches, PMIME is a representative implementation of nonuniform embedding and has been widely studied these years, as it is a model-free measure of general purpose for moderately high-dimensional time series[9]. In addition, it has been extensively applied in complex systems such as physiology[10, 11] and finance[12, 13].

However, nonuniform embedding still has some key limitations that need to be improved even after development. One issue is that the curse of dimensionality makes the estimation of conditional mutual information (CMI) inaccurate as the dimension of the embedded space increases[14–16]. Another issue arises from the construction of the mixed embedding vector. For example, greedy forward selection is commonly used to sequentially select lagged variables, which can easily become trapped in suboptimal solutions and lead to unstable causal networks due to the large candidate embedding space[7, 8, 17]. These challenges are not specific to a single method but are common in nonuniform embedding frameworks, where greedy or heuristic search strategies are often used to identify causal relationships. Over the past decade, the nonuniform embedding reconstruction methods have been continuously improved. Some studies enhance its feasibility by using low-dimensional approximations of CMI and im-

• Zheng Tan is with the School of Mathematics and Statistics, University of New South Wales, Sydney, NSW 2052, Australia. E-mail:zheng.tan.math@gmail.com
• Xin Weng is with Zhongnan University of Economics and Law, Wuhan, 430073, and China.
• Ziyu Jia is with the Institute of Automation, Chinese Academy of Sciences, Beijing, 100190, and China. E-mail:ziyu.jia@ia.ac.cn
(✉) To whom correspondence should be addressed.
Manuscript received: 20xx-xx-xx; revised: 20xx-11-xx; accepted: 20xx-00-xx

Just Accepted

proved search strategies, thereby improving the sensitivity of causal detection in coupled chaotic systems[18, 19]. Other studies control false positives and improve statistical confidence by adopting more rigorous causal inference pipelines and statistical testing, leading to more accurate inference results[20]. At present, existing work mainly focuses on the modifications of embedding reconstruction process at the levels of CMI estimation, search strategy, and statistical testing. But the mentioned limitations still exist in recent reports, leading to unstable causal networks.

In contrast, constraint-based approaches such as PC-style methods address high-dimensional causal discovery by enforcing sparsity at the graph level through conditional independence testing[21–24]. This reflects a general principle in high-dimensional modeling: structural constraints are introduced to limit the search domain before detailed inference[25–27]. From this perspective, high-dimensional inference problems are often stabilized by reorganizing the candidate space rather than modifying the inference criterion itself[28, 29]. By suppressing irrelevant or weakly informative components at an early stage[30], structural control can improve robustness[31]. Therefore, it is reasonable to consider whether restricting the candidate embedding space may alleviate the challenges faced by mixed embedding reconstruction. However, structured control of the candidate embedding space remains relatively underexplored. Under high-dimensional and limited-sample conditions, the fundamental challenge arises from the fact that the candidate embedding space is composed of all variables and their lagged components, causing the size of the candidate embedding space to grow rapidly. This directly leads to the curse of dimensionality in CMI estimation and the search strategies to become trapped in suboptimal solutions[18, 19].

Therefore, the present work starts from the candidate space prior to mixed embedding reconstruction. Without changing the causality criterion, this study investigates how to perform structured control of the candidate embedding space, which help ensures non-degenerate causal inference and yields more stable causal network structures. The main contribution of this study is to propose an adaptive optimization strategy based on the characteristics of the nonuniform embedding algorithm. Specifically, by treating each target variable as the unit of analysis and constructing a sparsity-constrained search space for embedding reconstruction, the instability caused by the rapid expansion of the candidate embedding space is effectively mitigated.

The rest of this paper is organized as follows. In Section 2, the related work around nonuniform embedding methods is reviewed, where PMIME is introduced in detail as a representative baseline. Then the structured control of the nonuniform embedding candidate state space is presented in Section 3. Experimental results are reported in Section 4, where the comparison between the influence of structured candidate space control on nonuniform embedding and recent approaches is also discussed. Section 4 evaluates the proposed framework through simulated-system validation, a screening ablation study, and real-data stability tests, followed by a comparison with related time-series causal discovery approaches. In Section 5, structural changes in financial markets under external shocks are discussed as an application in finance. Discussion and future work are presented in Section 6. The study conclusion is reported in Section 7.

2 Related work

2.1 Information-theoretic nonuniform embedding approaches

Information-theoretic approaches to causal discovery are closely related to state-space reconstruction in dynamical systems. A common strategy is to represent the state of a target variable using its past observations together with lagged components of other variables. Early approaches typically adopt uniform embedding, where all variables are included with a fixed set of lags, leading to a high-dimensional representation. Within such reconstructed state spaces, information-theoretic measures extend the classical Granger causality framework by quantifying directed information flow between time series. Transfer Entropy (TE) measures whether the past of a source variable can further reduce the uncertainty of a target variable's future, given the target's own past [6, 32, 33]. In practice, TE relies on the embedding space to define its conditioning set. However, in multivariate systems, it must condition on all other variables and their lags in order to remove indirect influences. This leads to the estimation of high-dimensional joint probability distributions, which becomes unstable and unreliable as the number of variables increases.

To address these limitations, nonuniform embedding methods have been proposed to construct low-dimensional representations of system dynamics. Instead of conditioning on all variables simultaneously, these methods progressively build an embedding vector by selecting lagged components that provide the most informative contribution to the target variable's future. At each step, candidate components are evaluated based on their incremental conditional mutual information (CMI), allowing the embedding to capture nonlinear dependencies while controlling the effective dimensionality of the conditioning space[7, 8]. Therefore, unlike the "full con-

JSU Accepted

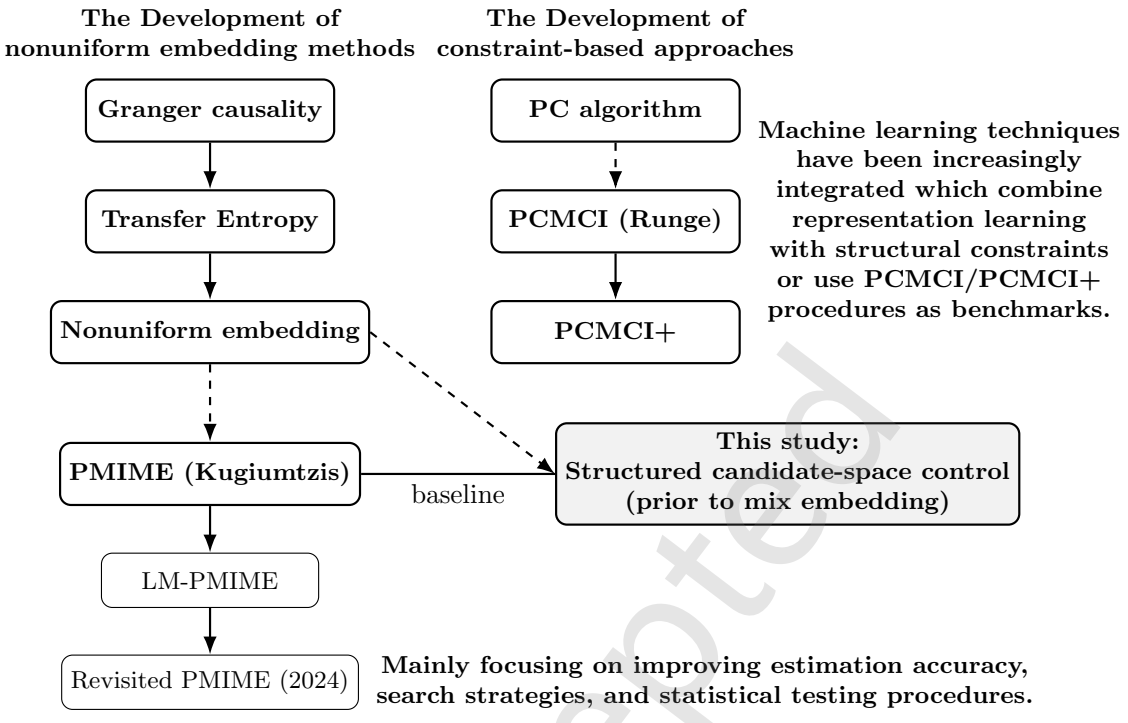


Fig. 1 Evolution of causality detection methods. The left column shows the development from Granger causality to nonuniform embedding and the PMIME family. The right column shows constraint-based approaches. The proposed method focuses on structural control of the candidate embedding space prior to reconstruction.

ditioning" strategy of TE, nonuniform embedding methods control dimensional growth through progressive embedding reconstruction and have been widely applied.

Despite these advantages, nonuniform embedding methods still face two main challenges in high-dimensional settings. First, CMI estimation becomes unstable as the embedding dimension increases. Even though the sequential strategy limits dimensional growth, the final embedding may still contain multiple lagged components, making reliable estimation difficult. Second, these methods rely on a greedy forward selection strategy. If an important component is missed early in the process, the algorithm may converge to a suboptimal embedding and stop prematurely. To alleviate instability caused by high-dimensional CMI estimation, subsequent studies introduced low-dimensional CMI approximations during the embedding selection process. In addition, improved search strategies were developed to reduce the risk of greedy algorithms becoming trapped in suboptimal embedding configurations[18]. By combining dimensionality-reduced CMI estimation with more flexible selection mechanisms, these extensions significantly enhanced causal detection performance, particularly as the length of time series increased. Subsequent studies further refined low-dimensional CMI approximations and developed more effective embedding search mechanisms, thereby improving the sensitivity and reliability of causal inference. These developments have led to substantial performance gains in coupled chaotic systems[19], and have been widely applied to complex systems such as physiological systems, financial systems, and nonlinear dynamical networks. However, these improvements have mainly focused on refining estimation and search mechanisms within the embedding reconstruction phase[34–36].

PMIME (Partial Mutual Information on Mixed Embedding) is a representative implementation of the nonuniform embedding framework. It provides a concrete realization of progressive embedding reconstruction based on CMI and has been widely adopted as a baseline method for causal discovery in multivariate time series. In recent years, rather than proposing entirely new embedding variants, some studies have revisited the PMIME causal inference pipeline. It suggest that PMIME has reached a relatively mature stage in terms of estimation and inference procedures.

In summary, the existing nonuniform embedding mechanism has reached a relatively mature stage, but the root cause of dimensional instability remains overlooked. Although these recent developments significantly enhance inference rigor and stability, the structural organization of the candidate embedding space still relies on enumerating lagged components of all variables. As system dimensionality increases,

Just Accepted

the combinatorial growth of the candidate space remains fundamentally unchanged. Consequently, these limitations remain only partially addressed. In this study, PMIME is adopted in this study as a representative baseline of nonuniform embedding-based causal inference. Its causal criterion and mixed embedding reconstruction procedure provide a concrete and well-established reference for evaluating the effect of candidate-space structural control. The proposed method does not modify the underlying information-theoretic criterion itself, but instead imposes structural organization on the candidate embedding space prior to reconstruction.

2.2 Constraint-based and hybrid approaches

2.2.1 Constraint-based methods

Constraint-based methods derived from the PC algorithm have become dominant in causal discovery of high dimension time series. The PC algorithm identifies causal structure by iteratively performing conditional independence tests and removing edges that do not satisfy statistical dependence. PCMCI extends this idea to time-series data by incorporating lagged variables and conducting conditional independence testing within a temporal framework. PCMCI+ further improves orientation rules and independence testing procedures, leading to more reliable structure estimation in practice. By introducing sparsity constraints at the graph level before detailed inference, these methods effectively reduce the search space and improve scalability in moderately high-dimensional systems. When sample sizes are sufficient, graph-level sparsification helps control false positives and achieve stable structure recovery. As a result, PCMCI and its extension methods are widely adopted as standard baseline approaches in empirical studies.

2.2.2 Learning-based extensions

In recent years, machine learning techniques have been increasingly integrated into time-series causal discovery. Methods such as STIC[37] and CD-NOTS[38] combine representation learning with structural constraints or use PCMCI/PCMCI+ procedures as benchmarks. These approaches aim to improve robustness under nonlinear dependencies, limited samples, or nonstationary conditions by combining structural regularization with data-driven feature extraction. The compatibility between graph-level sparsity and learned representations has contributed to the growing prominence of PCMCI-based frameworks. Consequently, PCMCI and its extensions are frequently used as baseline methods for evaluating newer learning-based causal discovery models[39, 40].

2.2.3 The remaining structural gap

To bridge constraint-based and embedding-based methods, a conceptually related hybrid strategy has been proposed. In this framework, graph-level pruning is first performed using the PC procedure, and information-theoretic causal measures are then computed for the retained variable pairs. Although this strategy reduces the number of candidate interactions at the graph level, the mixed embedding reconstruction within each retained pair still requires combinatorial selection over lagged components. Therefore, while sparsity is introduced at the graph level, the high-dimensional CMI estimation problem and the instability of greedy embedding selection remain at the embedding level. Overall, existing methods either impose sparsity at the graph structure level or refine estimation and search mechanisms within embedding reconstruction. Systematic structural organization of the candidate embedding space itself remains relatively underexplored.

3 Methods

3.1 Motivation: challenges in mixed embedding reconstruction

As stated in the related work, existing studies on improving nonuniform embedding methods mainly focus on modifications within the embedding reconstruction process itself, such as improvements in CMI estimation, search strategies, or statistical testing, while systematic investigations on the structural control of the candidate embedding space prior to mixed embedding reconstruction remain relatively underexplored. From a methodological perspective, the reconstruction of the mixed embedding vector involves progressively selecting lagged components from a finite but potentially large candidate space. This process can naturally be formulated as a subset selection problem in a high-dimensional candidate space with the objective of identifying[30]. During the sequential selection procedure, those components that provide additional explanatory information for the target variable. In high-dimensional settings, subset selection based on greedy forward strategies is generally sensitive to estimation errors as well as weakly relevant or redundant candidates[28]. When the candidate space lacks structural constraints, the effective search domain expands rapidly, and even if the inference criterion itself remains unchanged, the risk of trapping in suboptimal solutions is significantly increased. Similar phenomena have been widely discussed in related studies on variable selection, nonlinear delay embedding, and multivariate time series modeling.

In high-dimensional inference problems, a widely adopted strategy is to impose sparsity or structural constraints on the

Just Accepted

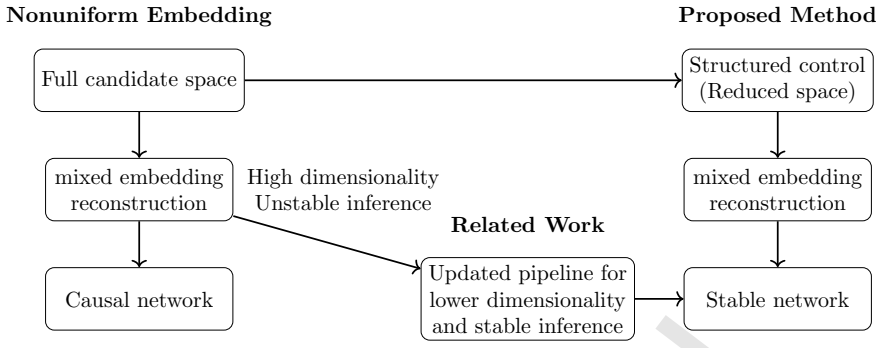


Fig. 2 Conceptual illustration of structured control in nonuniform embedding method.

candidate embedding space prior to inference, so as to reduce the effective search domain and improve the stability of the selection process[41]. This does not alter the inference criterion itself, but instead constrains the candidate embedding space, thereby allowing subsequent inference to be conducted within a more compact and information-concentrated search space[42, 43]. Therefore, this study starts with the candidate embedding space prior to mixed embedding reconstruction and introduces structured control over the candidate embedding space while keeping the causal criterion and search mechanism of mixed embedding reconstruction unchanged.

3.2 Structured control of the candidate embedding space

For each target variable, the proposed framework performs structured screening before mixed-embedding reconstruction. The purpose of this screening stage is to restrict the candidate search space while leaving the causal criterion, stopping rule, and normalization scheme of the subsequent PMIME procedure unchanged.

3.2.1 Candidate space for each target

Consider a K-dimensional multivariate time series

$$\mathbf{X}(t) = \{X_1(t), X_2(t), \ldots, X_K(t)\}. \tag{1}$$

Let $L_{\max}$ denote the maximum lag order. For a target variable X_j, the original lagged candidate space is

$$\mathcal{P}_j = \bigcup_{i=1}^{K} \{X_i(t-\ell) : \ell = 1, \ldots, L_{\max}\}. \tag{2}$$

Thus, $\mathcal{P}_j$ contains $KL_{\max}$ original lagged components. Structured screening determines which source variables should be retained before PMIME reconstructs the mixed embedding for X_j.

3.2.2 Structured nonlinear screening

When either K or $L_{\max}$ is large, the number of candidate evaluations increases rapidly. Under limited samples, this expansion may reduce the reliability of conditional mutual information estimation and increase the probability that greedy embedding reconstruction selects weak or redundant components. Structured sparsity is therefore introduced before mixed-embedding reconstruction.

For each variable X_i, define its lag vector as

$$\mathbf{x}_i(t) = \big(X_i(t-\ell)\big)_{\ell=1}^{L_{\max}} \in \mathbb{R}^{L_{\max}}, \tag{3}$$

and its squared-lag vector as

$$\mathbf{q}_i(t) = \big(X_i^2(t-\ell)\big)_{\ell=1}^{L_{\max}} \in \mathbb{R}^{L_{\max}}. \tag{4}$$

The corresponding variable-specific block is

$$\boldsymbol{\phi}_i(t) = \begin{bmatrix} \mathbf{x}_i(t) \\ \mathbf{q}_i(t) \end{bmatrix} \in \mathbb{R}^{2L_{\max}}. \tag{5}$$

For every pair $1 \leqslant i < k \leqslant K$, define the same-lag interaction block as

$$\boldsymbol{\psi}_{ik}(t) = \big(X_i(t-\ell)X_k(t-\ell)\big)_{\ell=1}^{L_{\max}} \in \mathbb{R}^{L_{\max}}. \tag{6}$$

Each interaction feature appears in exactly one pairwise block. The augmented dictionary therefore contains the original lagged terms, their squared terms, and all same-lag pairwise interactions.

For target X_j, define the screening predictor

$$\eta_j(t) = b_j + \sum_{i=1}^{K} \boldsymbol{\phi}_i^{\mathsf{T}}(t)\boldsymbol{\beta}_{ij} + \sum_{1\leqslant i<k\leqslant K} \boldsymbol{\psi}_{ik}^{\mathsf{T}}(t)\boldsymbol{\gamma}_{ik,j}. \tag{7}$$

Here, b_j is the intercept, $\boldsymbol{\beta}_{ij} \in \mathbb{R}^{2L_{\max}}$ is the coefficient vector of the variable-specific block, and $\boldsymbol{\gamma}_{ik,j} \in \mathbb{R}^{L_{\max}}$ is the coefficient vector of the interaction block.

The low-order nonlinear terms are used only to screen candidate variables. They are not intended to provide a complete model of the underlying dynamics and are not subsequently passed to PMIME.

Let

$$n = N - L_{\max} \tag{8}$$

Just Accepted

denote the effective sample size. Define the structured penalty

$$\Omega_j = \sum_{\substack{i=1\\ i\neq j}}^{K} w_i \left\|\boldsymbol{\beta}_{ij}\right\|_2 + \sum_{1\leqslant i<k\leqslant K} w_{ik} \left\|\boldsymbol{\gamma}_{ik,j}\right\|_2 . \tag{9}$$

For a candidate regularization parameter λ, the screening problem is

$$\min_{\Theta_j} \quad \frac{1}{2n} \sum_{t=L_{\max}+1}^{N} \left[X_j(t) - \eta_j(t)\right]^2 + \lambda\Omega_j, \tag{10}$$

where

$$\Theta_j = \{b_j, \boldsymbol{\beta}_{1j}, \ldots, \boldsymbol{\beta}_{Kj}, \boldsymbol{\gamma}_{12,j}, \ldots, \boldsymbol{\gamma}_{K-1,K,j}\}. \tag{11}$$

The target-history block $\boldsymbol{\beta}_{jj}$ is not penalized, so the past of the target variable remains available during subsequent embedding reconstruction. All other variable-specific and interaction blocks are penalized.

Before fitting, each predictor is standardized using the mean and standard deviation estimated from the corresponding training data. To account for differences in group size, the weights are set to

$$w_i = \sqrt{2L_{\max}}, \qquad w_{ik} = \sqrt{L_{\max}}. \tag{12}$$

The intercept is not penalized.

For a fitted model at λ, define the selected main-effect variables as

$$\mathcal{M}_j(\lambda) = \left\{ i \neq j : \left\|\widehat{\boldsymbol{\beta}}_{ij}(\lambda)\right\|_2 > 0 \right\}, \tag{13}$$

and the selected interaction pairs as

$$\mathcal{A}_j(\lambda) = \left\{ (i,k) : \begin{array}{c} 1 \leqslant i < k \leqslant K, \\ \|\widehat{\boldsymbol{\gamma}}_{ik,j}(\lambda)\|_2 > 0 \end{array} \right\}. \tag{14}$$

The retained variable set is then

$$\mathcal{S}_j(\lambda) = \{j\} \cup \mathcal{M}_j(\lambda) \cup \bigcup_{(i,k)\in\mathcal{A}_j(\lambda)} \{i,k\}. \tag{15}$$

Thus, a variable is retained if either its main-effect block is selected or it participates in a selected interaction block. In the numerical implementation, zero coefficients are identified using the convergence tolerance of the optimization solver.

3.2.3 *Selection of the regularization parameter*

The regularization parameter is selected separately for each target using F-fold blocked cross-validation. In the reported experiments, $F = 10$. The observations are divided, in their original temporal order, into contiguous validation blocks $\mathcal{I}_1, \ldots, \mathcal{I}_F$.

For the fth held-out block, define

$$e_{jf}(\lambda) = \frac{1}{|\mathcal{I}_f|} \sum_{t\in\mathcal{I}_f} \left[X_j(t) - \widehat{X}_j^{(-f)}(t;\lambda)\right]^2, \tag{16}$$

where $\widehat{X}_j^{(-f)}(t;\lambda)$ is obtained from the model fitted without block $\mathcal{I}_f$. Predictor standardization is estimated using only the corresponding training observations.

The mean cross-validation error is

$$\mathrm{CV}_j(\lambda) = \frac{1}{F}\sum_{f=1}^{F} e_{jf}(\lambda), \tag{17}$$

and its standard error is

$$\mathrm{SE}_j(\lambda) = \left\{ \frac{1}{F(F-1)} \sum_{f=1}^{F} \left[e_{jf}(\lambda) - \mathrm{CV}_j(\lambda)\right]^2 \right\}^{1/2}. \tag{18}$$

For each target, define a regularization path

$$\Lambda_j = \{\lambda_{j,1}, \ldots, \lambda_{j,M}\}, \qquad M = 100. \tag{19}$$

The values are logarithmically spaced according to

$$\lambda_{j,m} = \lambda_{j,\max}(0.01)^{(m-1)/(M-1)}, \tag{20}$$

for $m = 1, \ldots, M$. Here, $\lambda_{j,\max}$ is the smallest value for which all penalized coefficient groups are zero.

Let

$$\lambda_{j,\min} = \arg\min_{\lambda\in\Lambda_j} \mathrm{CV}_j(\lambda). \tag{21}$$

Define the set of values satisfying the one-standard-error criterion as

$$\Lambda_{j,1\mathrm{SE}} = \Big\{ \lambda \in \Lambda_j : \mathrm{CV}_j(\lambda) \leqslant \mathrm{CV}_j(\lambda_{j,\min}) + \mathrm{SE}_j(\lambda_{j,\min}) \Big\}. \tag{22}$$

The final regularization parameter is

$$\lambda_j^\star = \max \Lambda_{j,1\mathrm{SE}}. \tag{23}$$

This rule favors a more parsimonious model while retaining cross-validation performance within one standard error of the minimum.

3.2.4 *Restricted embedding search space*

After selecting $\lambda_j^\star$, define

$$\mathcal{S}_j = \mathcal{S}_j(\lambda_j^\star). \tag{24}$$

The restricted candidate space passed to PMIME is

$$\widetilde{\mathcal{P}}_j = \bigcup_{i\in\mathcal{S}_j} \{X_i(t-\ell) : \ell = 1, \ldots, L_{\max}\} \subseteq \mathcal{P}_j. \tag{25}$$

Only the original lagged components are included in $\widetilde{\mathcal{P}}_j$. The squared and interaction terms are used exclusively during structured screening and are not passed to PMIME. Mixed-embedding reconstruction is subsequently performed within $\widetilde{\mathcal{P}}_j$ using the original PMIME causal criterion, stopping rule, and normalization scheme.

Just Accepted

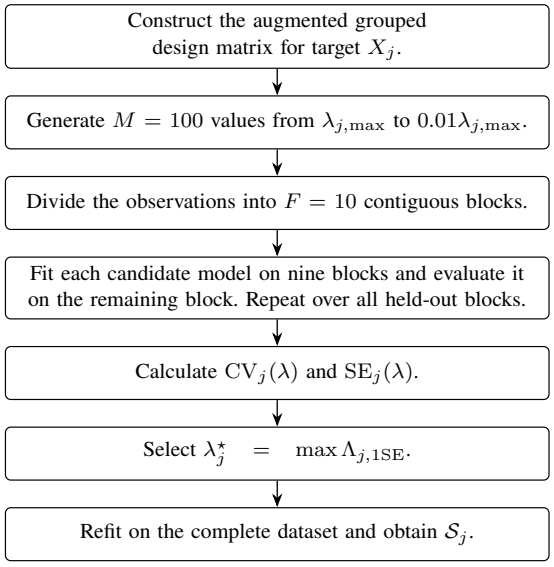


Fig. 3 Blocked cross-validation and regularization-parameter selection in the structured screening stage.

3.3 Mixed embedding within restricted search space

After constructing the restricted embedding search space $\widetilde{\mathcal{P}}_j$ for each target variable X_j in Eq. (25), mixed-embedding reconstruction is carried out within this restricted search space. The causal criterion, stopping rule, and normalization scheme remain consistent with those of the baseline PMIME method.

Specifically, the mixed embedding vector is constructed using the original nonuniform greedy selection mechanism based on CMI, while the candidate lagged components available at each embedding iteration are restricted to $\widetilde{\mathcal{P}}_j$. The squared and interaction features introduced in the screening stage are not passed to PMIME. They are used only to determine the retained variable set $\mathcal{S}_j$. The directional causal contributions of the selected lagged components are subsequently quantified using normalized CMI. By independently repeating this target-wise reconstruction procedure for all target variables, the complete information-flow matrix is obtained.

Algorithm: Structured Sparsity of Candidate Space and Mixed-Embedding Reconstruction For clarity, the proposed procedure is summarized as follows.

i. For each target variable X_j, construct the full candidate space $\mathcal{P}_j$ containing the original lagged components of all variables up to $L_{\max}$.

ii. Construct the variable-specific blocks $\phi_i(t)$ and the same-lag interaction blocks $\psi_{ik}(t)$ for structured screening.

iii. Select the regularization parameter $\lambda_j^\star = \lambda_{j,1\mathrm{SE}}$ using 10-fold blocked cross-validation, as illustrated in Fig. 3.

iv. Fit the structured screening model at $\lambda_j^\star$ and obtain the retained variable set

$$\mathcal{S}_j = \mathcal{S}_j(\lambda_j^\star).$$

v. Construct the restricted embedding search space $\widetilde{\mathcal{P}}_j$ according to Eq. (25). Only the original lagged components of the variables in $\mathcal{S}_j$ are included.

vi. Perform mixed-embedding reconstruction within $\widetilde{\mathcal{P}}_j$ using the original PMIME greedy selection procedure based on CMI.

vii. Quantify the directional causal effects using normalized CMI.

viii. Repeat the above procedure for all target variables to obtain the complete information-flow matrix.

Because the second-stage reconstruction is restricted to the selected variable set $\mathcal{S}_j$, a relevant source excluded during screening cannot be recovered subsequently. The performance of the two-stage procedure therefore depends on the screening recall.

3.4 Computational Complexity

The proposed framework consists of three computational stages: construction of the nonlinear screening dictionary, structured screening, and nonuniform mixed-embedding reconstruction. Let N denote the original time-series length, K the number of observed variables, $L_{\max}$ the maximum lag order, and

$$n = N - L_{\max} \tag{26}$$

the effective sample size after lag construction.

3.4.1 Candidate construction

For each target variable, the original PMIME candidate space contains

$$P_0 = KL_{\max} \tag{27}$$

lagged components.

The screening dictionary contains $KL_{\max}$ original lagged features, $KL_{\max}$ squared features, and $\binom{K}{2}L_{\max}$ same-lag pairwise interaction features. Its total dimension is therefore

$$\begin{aligned} P_{\mathrm{aug}} &= 2KL_{\max} + \binom{K}{2}L_{\max} \\ &= L_{\max}\left[2K + \binom{K}{2}\right]. \end{aligned} \tag{28}$$

Consequently, P_{aug} grows linearly with $L_{\max}$ and quadratically with K. Constructing the augmented design matrix

Just Accepted

requires

$$O(nP_{\mathrm{aug}}) \tag{29}$$

operations and memory. Because the same predictor dictionary can be used for all target variables, it needs to be constructed only once, although standardization parameters must be estimated separately within each cross-validation training split.

The number of structured coefficient groups is

$$G = K + \binom{K}{2}, \tag{30}$$

consisting of K variable-specific main-effect groups and $\binom{K}{2}$ interaction groups.

3.4.2 *Structured screening*

For each target variable, structured screening fits a regularization path containing M_λ candidate values and selects the final regularization parameter using F-fold blocked cross-validation. Let I_{opt} denote the number of optimization iterations required for one fit.

The precise computational complexity depends on the solver, its convergence criterion, warm starts, and active-set strategy. For a solver whose iteration requires a pass over the complete augmented design matrix, a conservative representation of the screening cost for one target is

$$O\left(FM_\lambda I_{\mathrm{opt}} n P_{\mathrm{aug}}\right). \tag{31}$$

Because screening is repeated for all K targets, the corresponding total cost is approximately

$$O\left(KFM_\lambda I_{\mathrm{opt}} n P_{\mathrm{aug}}\right). \tag{32}$$

This expression is intended to describe the dependence on the principal problem dimensions rather than provide a solver-independent exact bound. Warm starts along the regularization path and reuse of the augmented design matrix may substantially reduce the practical cost. In the reported implementation, $M_\lambda = 100$ and $F = 10$.

3.4.3 *Restricted mixed-embedding reconstruction*

For target X_j, let

$$s_j = |\mathcal{S}_j| \tag{33}$$

denote the number of variables retained by structured screening. Because only the original lagged components are passed to PMIME, the restricted search space contains

$$R_j = s_j L_{\max} \tag{34}$$

candidate components, compared with $P_0 = KL_{\max}$ components in the original unscreened search space.

It is important to distinguish P_{aug} from R_j. The former is the dimension of the nonlinear dictionary used only during screening, whereas the latter is the number of original lagged components used during PMIME reconstruction.

Suppose that the final mixed embedding for target X_j contains r_j components. Under greedy forward selection, the number of candidate evaluations is

$$\begin{aligned} Q_j &= R_j + (R_j - 1) + \cdots + (R_j - r_j + 1) \\ &= \frac{r_j(2R_j - r_j + 1)}{2}. \end{aligned} \tag{35}$$

When $r_j \ll R_j$, this quantity is of order $O(r_j R_j)$. If a large proportion of the candidates is selected, it may approach $O(R_j^2)$.

Each candidate evaluation involves mutual information or conditional mutual information estimation. Its practical cost depends on the effective sample size, the current embedding dimension, the number of nearest neighbors, and the nearest-neighbor search implementation. Let

$$C_{\mathrm{CMI}}(n, d)$$

denote the implementation-dependent cost of one information-theoretic evaluation at embedding dimension d. The reconstruction cost for target X_j may then be represented as

$$O\left(Q_j C_{\mathrm{CMI}}(n, d_j)\right), \tag{36}$$

where d_j represents the dimensions encountered during reconstruction. Across all targets, the reconstruction cost is

$$O\left(\sum_{j=1}^{K} Q_j C_{\mathrm{CMI}}(n, d_j)\right). \tag{37}$$

For comparison, the unscreened PMIME procedure begins with $P_0 = KL_{\max}$ candidates. If its final embedding for target X_j contains r_{0j} components, the corresponding number of candidate evaluations is

$$Q_{0j} = \frac{r_{0j}(2P_0 - r_{0j} + 1)}{2}. \tag{38}$$

3.4.4 *Overall computational trade-off*

Combining the three stages, the total computational burden can be represented schematically as

$$\begin{aligned} C_{\mathrm{total}} = {} & O(nP_{\mathrm{aug}}) + O\left(KFM_\lambda I_{\mathrm{opt}} n P_{\mathrm{aug}}\right) \\ & + O\left(\sum_{j=1}^{K} Q_j C_{\mathrm{CMI}}(n, d_j)\right). \end{aligned} \tag{39}$$

The first two terms are introduced by nonlinear dictionary construction and structured screening, whereas the third term represents the restricted PMIME reconstruction.

The computational effect of screening is therefore a trade-off. Its benefit is most pronounced when

$$s_j \ll K, \tag{40}$$

Just Accepted

because the reduction from $KL_{\max}$ to $s_j L_{\max}$ candidates can substantially decrease the number of repeated information-theoretic evaluations. This setting is most likely when the underlying driving structure is sparse and the original candidate space contains many irrelevant or redundant variables.

In contrast, when $s_j \approx K$, the restricted candidate space remains close to the original space. The additional cost of fitting the structured regularization path may then exceed the reduction in subsequent reconstruction. The proposed procedure should therefore not be interpreted as uniformly faster than PMIME.

The screening and reconstruction calculations are performed separately for each target variable. These target-specific calculations are independent and can, in principle, be parallelized. Nevertheless, the pairwise nonlinear dictionary causes P_{aug} to grow quadratically with K, so the screening stage may itself become computationally demanding in substantially higher-dimensional systems. More economical nonlinear dictionaries, preliminary screening rules, warm-start optimization, and parallel implementations are therefore important directions for improving scalability.

4 Experiment results

This section presents experimental tests on the structured control of candidate space and the subsequent hybrid embedding reconstruction in the constrained search system. The tests are conducted on both simulated systems and real financial systems. To evaluate the effectiveness of the proposed structured method, simulated time series from linear stochastic, nonlinear stochastic, and chaotic systems are used to compare the performance of the traditional nonuniform embedding methods. Additionally, this study includes experimental tests in real financial systems, with the baseline that detects the directional coupling in multivariable time series using the low-dimensional approximation of CMI as a comparison. This method represents an internal improvement to hybrid embedding reconstruction within nonuniform embedding methods.

4.1 Validation in Simulated Systems

4.1.1 Experimental Setup

Four simulated systems are considered: a linear vector autoregressive process, a nonlinear vector autoregressive process, coupled Hénon maps, and a coupled Lorenz chain. The first three systems are evaluated using 30 independent realizations of length $N = 512$, whereas the Lorenz system is evaluated using 50 realizations of length $N = 256$. For each system, Original PMIME, the proposed Structured method, and PCMCI-ParCorr are applied to the same data realizations.

For both PMIME-based methods, the maximum candidate lag is $L_{\max} = 5$, the prediction horizon is $T = 1$, the number of nearest neighbors is $k = 5$, and the stopping parameter is $A = 0.05$. A directed link is detected when the corresponding PMIME value is greater than zero. For PCMCI-ParCorr, the lag range is set to $\tau = 1, \ldots, 5$, with both the PC-stage and final significance levels set to 0.05. The lag-specific p-values are adjusted using the Benjamini–Hochberg procedure, and a variable-level link is declared when at least one adjusted lag-specific p-value is below 0.05.

Performance is evaluated separately for each realization after excluding self-links. Precision, sensitivity, specificity, F1-score, and structural Hamming distance (SHD) are calculated for each trial and reported as the mean $\pm$ sample standard deviation. The mean information-flow matrices are used only for visualization and not for graph estimation or performance evaluation.

4.1.2 Linear Multivariate Stochastic Process

The linear vector autoregressive (VAR) process is composed of order 4 in 5 time series (model 1 in [44]).

$$\begin{aligned}
x_{1,t} &= 0.4x_{1,t-1} - 0.5x_{1,t-2} + 0.4x_{5,t-1} + e_{1,t},\\
x_{2,t} &= 0.4x_{2,t-1} - 0.3x_{1,t-4} + 0.4x_{5,t-2} + e_{2,t},\\
x_{3,t} &= 0.5x_{3,t-1} - 0.7x_{3,t-2} - 0.3x_{5,t-3} + e_{3,t},\\
x_{4,t} &= 0.8x_{4,t-3} + 0.4x_{1,t-2} + 0.3x_{2,t-2} + e_{4,t},\\
x_{5,t} &= 0.7x_{5,t-1} - 0.5x_{5,t-2} - 0.4x_{4,t-1} + e_{5,t},
\end{aligned} \tag{41}$$

where $e_{i,t}$, $i = 1, \ldots, 5$, are independent Gaussian white noise processes.

Table 1 summarizes the results for the linear VAR system with $N = 512$. PMIME recovered nearly all true causal links, achieving a sensitivity of 0.986 ± 0.044. However, its lower specificity (0.679 ± 0.142) indicates that the unrestricted candidate space also introduced a considerable number of false-positive links. On average, PMIME produced 4.167 false positives per realization, resulting in an F1-score of 0.772 ± 0.092 and an SHD of 4.267 ± 1.929.

Structure retained all true causal links, with sensitivity increasing to 1.000 ± 0.000, while reducing the average number of false positives from 4.167 to 2.600. Consequently, specificity increased to 0.800 ± 0.108, the F1-score increased to 0.849 ± 0.071, and SHD decreased to 2.600 ± 1.404. The information-flow matrices in Fig. 4 further illustrate the reduction of spurious connections within the PMIME framework. As illustrated in Fig. 5(a), the screening stage produces a more restricted candidate space while retaining

Just Accepted

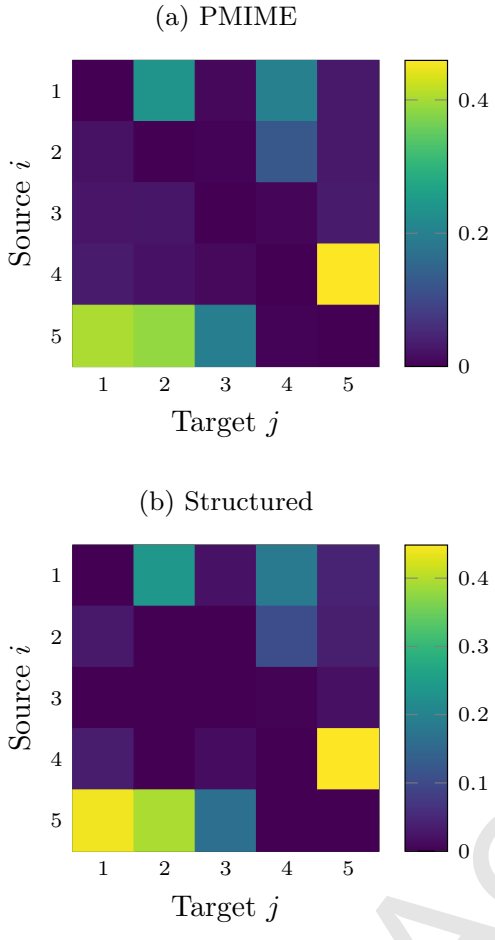


Fig. 4 Information flow matrices (5×5) on linear sys. Eq.(41). *Notes:* **(a) PMIME:** The original Partial Mutual Information from Mixed Embedding. **(b) Structured:**The proposed structured candidate-space control method in nonuniform embedding methods.

the source variables required for the evaluated true cross-variable causal links. These results suggest that structured candidate-space control improves the finite-sample specificity and structural accuracy of PMIME without sacrificing the recovery of true causal links.

PCMCI achieved the strongest overall performance in this system, with an F1-score of 0.966 ± 0.046 and an SHD of 0.533 ± 0.730. This result is expected because PCMCI employs the ParCorr conditional independence test, which is well matched to the linear Gaussian data-generating mechanism. PCMCI is therefore included as an external reference, whereas the main comparison between PMIME and Structured evaluates the effect of the proposed candidate-space control within the mixed-embedding framework.

Table 1 Performance on the linear VAR system. Results are reported as mean ± standard deviation over 30 realizations.

Method	Prec.	Sens.	Spec.	F1	SHD
PMIME	0.642 ±0.126	0.986 ±0.044	0.679 ±0.142	0.772 ±0.092	4.267 ±1.929
PCMCI	**0.937** ±0.084	**1.000** ±0.000	**0.959** ±0.056	**0.966** ±0.046	**0.533** ±0.730
Structured	0.744 ±0.109	**1.000** ±0.000	0.800 ±0.108	0.849 ±0.071	2.600 ±1.404

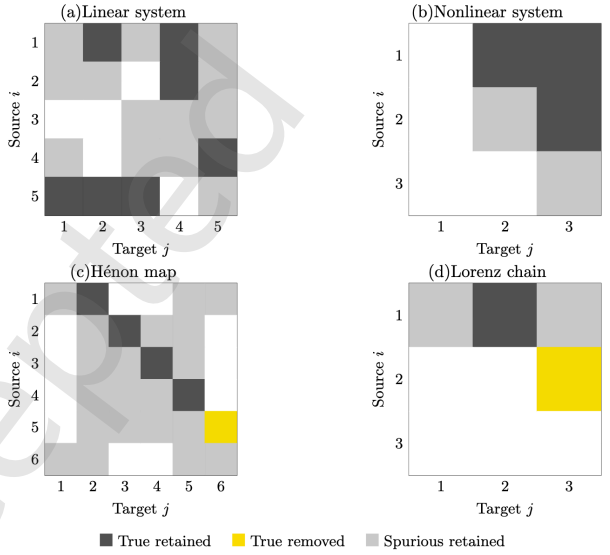


Fig. 5 Visualization of structured candidate-space control results. Each cell corresponds to a directed pair $(i \to j)$ (row = source i, column = target j).

4.1.3 *Nonlinear Multivariate Stochastic Process*

The nonlinear VAR process is of order 1 in three variables NLVAR3(1) (model 7 in [45]).

$$\begin{aligned} x_{1,t} &= f(x_{1,t-1}) + 0.4\, e_{1,t}, \\ x_{2,t} &= f(x_{2,t-1}) + 0.5\, x_{1,t-1} x_{2,t-1} + 0.4\, e_{2,t}, \\ x_{3,t} &= f(x_{3,t-1}) + 0.3\, x_{2,t-1} + 0.5\, x_{1,t-1}^2 \\ &\quad + 0.4\, e_{3,t}, \end{aligned} \tag{42}$$

where $f(x) = 3.4\, x(1-x^2)e^{-x^2}$, and $e_{i,t}$, $i = 1, 2, 3$, are independent Gaussian white noise processes.

Table 2 presents the results for the nonlinear VAR system with $N = 512$. PMIME recovered all three true causal links, achieving a sensitivity of 1.000 ± 0.000. Nevertheless, it produced an average of 0.633 false-positive links per realization, resulting in a specificity of 0.789 ± 0.255, an F1-score of 0.915 ± 0.099, and an SHD of 0.633 ± 0.765.

Structured retained all true source variables while removing the remaining false-positive detections. As shown in Fig. 5(b), the screening stage preserved the source variables associated with all three true cross-variable causal links while restricting redundant candidates. Consequently, Structured achieved pre-

Just Accepted

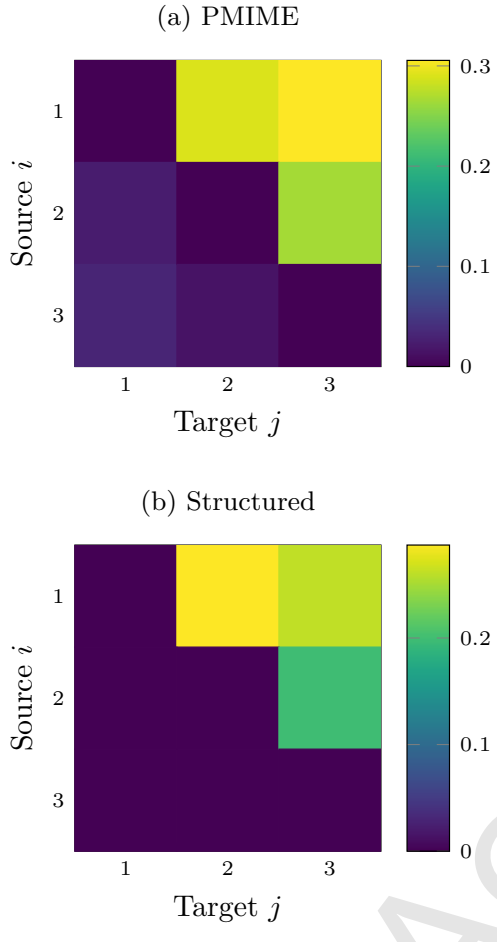


Fig. 6 Information flow matrices (3×3) on nonlinear sys. Eq.(42).

cision, sensitivity, specificity, and F1-score of 1.000 ± 0.000, together with an SHD of 0.000 ± 0.000. The information-flow matrices in Fig. 6 further show that the three true links, $X_1 \to X_2$, $X_1 \to X_3$, and $X_2 \to X_3$, were consistently recovered without additional spurious connections. These results demonstrate that structured candidate-space control can improve the specificity of PMIME while preserving both multiplicative and quadratic causal dependencies.

Table 2 Performance on the nonlinear VAR system. Results are reported as mean ± standard deviation over 30 realizations.

Method	Prec.	Sens.	Spec.	F1	SHD
PMIME	0.858 ±0.162	**1.000** ±0.000	0.789 ±0.255	0.915 ±0.099	0.633 ±0.765
PCMCI	0.967 ±0.127	0.333 ±0.000	0.978 ±0.085	0.493 ±0.025	2.067 ±0.254
Structured	**1.000** ±0.000	**1.000** ±0.000	**1.000** ±0.000	**1.000** ±0.000	**0.000** ±0.000

PCMCI achieved high precision (0.967 ± 0.127) and specificity (0.978 ± 0.085), but its sensitivity was only 0.333 ± 0.000, corresponding to the recovery of one of the three true causal links. Its lower F1-score of 0.493 ± 0.025 is primarily attributable to the two missed nonlinear relationships. This result is consistent with the use of ParCorr as the conditional independence test, since the present system contains multiplicative and quadratic dependencies that are not fully characterized by linear partial correlation. PCMCI therefore serves here as a fixed external reference rather than a nonlinear variant of the PCMCI framework.

4.1.4 *Chaotic Dynamical Systems*

The preceding stochastic-system results provide a clear observation: when the screening stage retains all relevant source variables, the proposed Structured method can reduce false-positive detections and improve the structural accuracy of PMIME without sacrificing sensitivity. Therefore, the benefit of candidate-space restriction depends critically on whether the screening stage preserves the true causal drivers.

We next examine the two chaotic dynamical systems. As shown in Fig. 5(c) and (d), the screening stage removes one true source–target relation in each system, introducing a screening false negative before mixed-embedding reconstruction. These systems allow us to examine a central question of the proposed two-stage framework: does removing a true source necessarily degrade the final causal graph, or does its effect depend on whether that link would otherwise be recoverable by PMIME? We first consider the coupled Lorenz system.

For these systems, the screening recall is defined as

$$\mathrm{SR} = \frac{|\{(i,j) \in \mathcal{E}^\star : i \in \mathcal{S}_j\}|}{|\mathcal{E}^\star|}, \tag{43}$$

where $\mathcal{E}^\star$ denotes the set of true directed links and $\mathcal{S}_j$ is the variable set retained for target X_j. Since a removed source is unavailable to the subsequent reconstruction, the attainable sensitivity of Structured is bounded above by SR.

The system of K coupled chaotic Hénon maps (model 6 in [8]) defined as

$$\begin{aligned} x_{1,t} &= 1.4 - x_{1,t-1}^2 + 0.3x_{1,t-2}, \\ x_{i,t} &= 1.4 - \left(Cx_{i-1,t-1} + (1-C)x_{i,t-1}\right)^2 \\ &\quad + 0.3x_{i,t-2}, \quad i = 2, \ldots, K. \end{aligned} \tag{44}$$

The directed connections $X_{i-1} \to X_i$, for $i = 2, \ldots, K$, constitute the true causal structure of the K coupled chaotic Hénon maps.

Table 3 presents the results for the coupled Hénon maps with $K = 6$, $N = 512$, and $C = 0.3$. PMIME recovered four of the five true directed links in every realization, giving a sensitivity of 0.800 ± 0.000. It also produced four false-positive links, resulting in a specificity of 0.840 ± 0.000, an F1-score of 0.615 ± 0.000, and an SHD of 5.000 ± 0.000.

Just Accepted

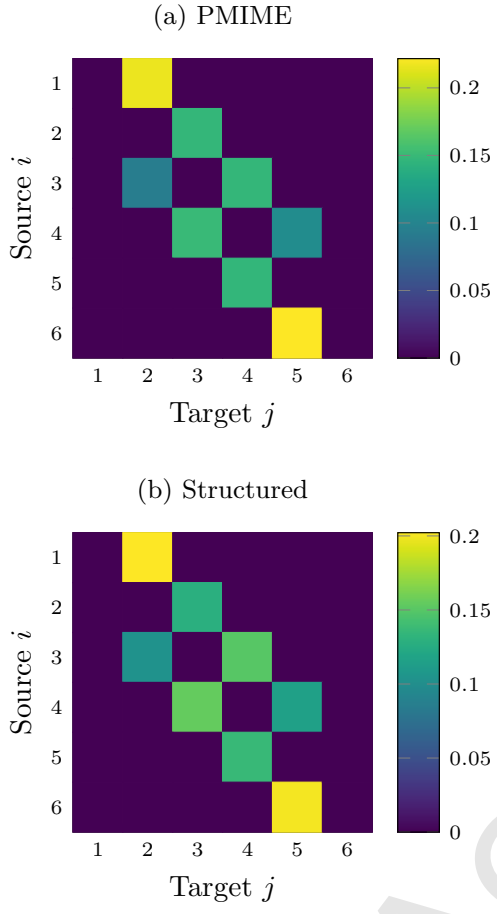


Fig. 7 Information flow matrices (6×6) on Hénon map Eq.(44).

As shown in Fig. 5(c), the screening stage retained four of the five true source–target relations but removed the link $X_5 \to X_6$. The corresponding screening recall was therefore

$$\mathrm{SR}_{\mathrm{Hénon}} = \frac{4}{5} = 0.8.$$

Despite this screening omission, the proposed Structured method produced the same final graph and performance measures as PMIME. The reason is visible in Fig. 7: the removed link $X_5 \to X_6$ was not recovered by the original PMIME either. Its removal therefore did not create an additional false negative relative to PMIME. At the same time, the retained candidate space did not eliminate the four false-positive links selected during mixed-embedding reconstruction, so no improvement in specificity or SHD was observed. This result shows that a screening false negative does not necessarily alter the final estimated graph. Its practical effect depends on whether the removed link belongs to the effective recovery set of PMIME under the given finite-sample condition. In the present Hénon system, the screening recall of 0.8 coincides with the sensitivity already attained by PMIME, and Structured consequently preserves, rather than improves, its reconstruction performance.

PCMCI obtained a lower sensitivity of 0.667 ± 0.132 and produced an average of 7.167 false-positive links, leading to an F1-score of 0.434 ± 0.085 and an SHD of 8.833 ± 2.052. Thus, under this nonlinear chaotic setting, the fixed PCMCI-ParCorr configuration was less effective than both PMIME-based methods in recovering the sparse directed chain.

Table 3 Performance on the coupled Hénon maps. Results are reported as mean ± standard deviation over 30 realizations.

Method	Prec.	Sens.	Spec.	F1	SHD
PMIME	**0.500** ±0.000	**0.800** ±0.000	**0.840** ±0.000	**0.615** ±0.000	**5.000** ±0.000
PCMCI	0.326 ±0.074	0.667 ±0.132	0.713 ±0.079	0.434 ±0.085	8.833 ±2.052
Structured	**0.500** ±0.000	**0.800** ±0.000	**0.840** ±0.000	**0.615** ±0.000	**5.000** ±0.000

The chaotic system of three coupled identical Lorenz oscillators (model 5 in [8]) is defined as

$$\begin{aligned}
\dot{x}_1 &= -10x_1 + 10y_1, \\
\dot{y}_1 &= -x_1 z_1 + 28x_1 - y_1, \\
\dot{z}_1 &= x_1 y_1 - \frac{8}{3} z_1, \\
\dot{x}_i &= -10x_i + 10y_i + C(x_{i-1} - x_i), \\
\dot{y}_i &= -x_i z_i + 28x_i - y_i, \\
\dot{z}_i &= x_i y_i - \frac{8}{3} z_i, \quad i = 2, 3.
\end{aligned} \tag{45}$$

The true directed causal connections in the three coupled Lorenz oscillators are $X_{i-1} \to X_i$, for $i = 2, 3$.

Table 4 presents the results for the coupled Lorenz system with $N = 256$, a sampling interval of 0.05, and a coupling strength of 6. PMIME produced a sparse estimated graph, with a specificity of 0.955 ± 0.097 and the lowest SHD of 1.400 ± 0.833. However, its sensitivity was only 0.390 ± 0.339, indicating that the two true directed links were not reliably recovered. Its resulting F1-score was 0.462 ± 0.365.

As shown in Fig. 5(d), the screening stage retained the link $X_1 \to X_2$ but removed $X_2 \to X_3$. Since the Lorenz chain contains only two true directed links, its screening recall was

$$\mathrm{SR}_{\mathrm{Lorenz}} = \frac{1}{2} = 0.5.$$

Structured therefore could not attain a sensitivity above 0.5, irrespective of the subsequent mixed-embedding reconstruction. Its observed sensitivity was even lower, at 0.230 ± 0.252, corresponding to an average of 0.460 recovered true links per realization.

The reduction cannot be attributed solely to the direct removal of $X_2 \to X_3$. As indicated by Fig. 8, PMIME also

Just Accepted

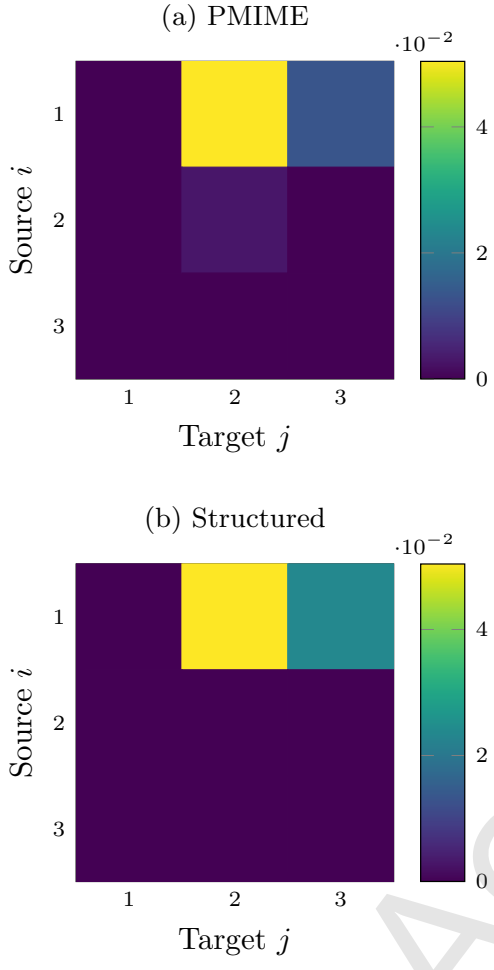


Fig. 8 Information flow matrices (3×3) on Lorenz chain Eq.(45).

assigned little information flow to this link under the present finite-sample setting. Instead, the restricted reconstruction detected the retained link $X_1 \to X_2$ less consistently than the original PMIME. Because PMIME uses a greedy conditional-information selection procedure, restricting its candidate set can alter the selected embedding and does not guarantee monotonic preservation of links that remain available. This effect is particularly relevant for the strongly coupled nonlinear Lorenz dynamics and the relatively short sample length considered here. Consequently, Structured did not provide a beneficial sensitivity–specificity trade-off in this system. Relative to PMIME, its sensitivity decreased from 0.390 to 0.230, its F1-score decreased from 0.462 to 0.290, and its SHD increased from 1.400 to 1.860. This result illustrates the principal limitation of the two-stage framework: candidate-space reduction is advantageous only when the screening stage has sufficiently high recall and the restricted search space continues to support stable mixed-embedding reconstruction.

PCMCI achieved the highest sensitivity (0.870 ± 0.264) and F1-score (0.623 ± 0.217), but this gain was accompanied by a much lower specificity of 0.520 ± 0.327 and an average of 1.920 false-positive links. Accordingly, PCMCI had a higher SHD than PMIME, despite its higher F1-score. The three methods therefore exhibit different operating characteristics in this system: PMIME favors a sparser graph with fewer structural errors, PCMCI favors recovery of the true links at the cost of additional false positives, and Structured is limited by both the screening omission and the reduced stability of the retained link.

Table 4 Performance on the coupled Lorenz system. Results are reported as mean ± standard deviation over 50 realizations.

Method	Prec.	Sens.	Spec.	F1	SHD
PMIME	**0.617** ±0.476	0.390 ±0.339	**0.955** ±0.097	0.462 ±0.365	**1.400** ±0.833
PCMCI	0.520 ±0.247	**0.870** ±0.264	0.520 ±0.327	**0.623** ±0.217	2.180 ±1.240
Structured	0.410 ±0.470	0.230 ±0.252	0.920 ±0.118	0.290 ±0.321	1.860 ±0.756

Summary. Table 5 summarizes the trial-to-trial variability of the F1-score across the four simulated systems. Structured produced an equal or lower F1-score standard deviation than PMIME in every system, reducing the cross-system average from 0.139 to 0.098. The stability improvement was most evident in the nonlinear system, where Structured recovered the same correct graph in all realizations, while it preserved the zero-variance result of PMIME in the Hénon system. Greater variability was observed for all methods in the Lorenz system because the ground-truth network contained only two links, so a change in a single detected link produced a relatively large change in F1-score. PCMCI had a slightly lower average standard deviation of 0.093, although its low variability did not always correspond to high recovery accuracy, as illustrated by its consistent false negatives in the nonlinear system. Overall, the results indicate that Structured improves the trial-to-trial stability of its direct baseline PMIME, particularly when the screening stage retains the relevant source variables.

Across the two stochastic systems, Structured achieved the strongest overall balance, with the highest average F1-score (0.925), compared with PMIME (0.844) and PCMCI (0.730). Relative to PMIME, it reduced false-positive detections and SHD while preserving all true causal links, demonstrating the benefit of structured candidate-space control when screening recall is high.

The chaotic systems showed that this benefit depends on both screening recall and the stability of the subsequent reconstruction. Structured preserved PMIME performance

Just Accepted

in the Hénon system, whereas its lower screening recall led to reduced sensitivity in the Lorenz system. PCMCI performed strongly in selected settings but exhibited more system-dependent sensitivity–specificity trade-offs. Overall, Structured is most beneficial when redundant candidates can be removed without excluding relevant source variables. To identify the specific contribution of the structured feature blocks, a focused ablation study is presented in the following section.

Table 5 Trial-to-trial variability of the F1-score across the four simulated systems. Smaller values indicate greater stability.

System	PMIME	PCMCI	Structured
Linear VAR	0.092	0.046	0.071
Nonlinear VAR	0.099	0.025	0.000
Hénon maps	0.000	0.085	0.000
Lorenz chain	0.365	0.217	0.321
Average	0.139	0.093	0.098

4.2 Ablation Study of Structured Candidate-Space Control

The proposed structured screening employs three types of features: the original lagged components, their squared terms, and same-lag pairwise interactions. This experiment investigates the contribution of these feature components.

4.2.1 Combined nonlinear stochastic system

A nine-variable nonlinear stochastic system is constructed by combining two independent dynamical blocks. The first block, comprising X_1–X_3, follows the nonlinear stochastic process used in Section 4.1 Eq.(42). The second block is a dynamically closed six-variable subsystem extracted from system S6 in [20]. Specifically, the original S6 variables $(X_2, X_3, X_4, X_6, X_7, X_9)$ are relabeled as $(X_4, \ldots, X_9)$. No causal connection is introduced between the two blocks.

The resulting system is defined as

$$x_{1,t} = f(x_{1,t-1}) + 0.4e_{1,t}, \tag{46a}$$
$$x_{2,t} = f(x_{2,t-1}) + 0.5x_{1,t-1}x_{2,t-1} + 0.4e_{2,t}, \tag{46b}$$
$$x_{3,t} = f(x_{3,t-1}) + 0.3x_{2,t-1} + 0.5x_{1,t-1}^2 + 0.4e_{3,t}, \tag{46c}$$
$$x_{4,t} = 0.3x_{4,t-1} + e_{4,t}, \tag{46d}$$
$$x_{5,t} = 0.3x_{5,t-1} + e_{5,t}, \tag{46e}$$
$$x_{6,t} = 0.3x_{6,t-1} + 0.95x_{5,t-2} + 0.95x_{4,t-3} + e_{6,t}, \tag{46f}$$
$$x_{7,t} = 0.95x_{5,t-1} + e_{7,t}, \tag{46g}$$
$$x_{8,t} = 0.3x_{8,t-1} + e_{8,t}, \tag{46h}$$
$$x_{9,t} = 0.3x_{9,t-1}x_{6,t-1} + e_{9,t}, \tag{46i}$$

where $f(x) = 3.4x(1-x^2)\exp(-x^2)$, and $e_{i,t}, i = 1, \ldots, 9$, are mutually independent standard Gaussian white-noise processes. As in the original nonlinear three-variable system, the innovations in the first block are scaled by 0.4.

The system contains the following seven cross-variable causalrelationships: $X_1 \to X_2$, $X_1 \to X_3$, $X_2 \to X_3$, $X_4 \to X_6$, $X_5 \to X_6$, $X_5 \to X_7$, $X_6 \to X_9$. Among these relationships, $X_1 \to X_3$ is generated by a squared term, whereas $X_1 \to X_2$ and $X_6 \to X_9$ are generated by same-lag interaction terms. The remaining four links are linear. Variable X_8 is included as an independent distractor. The absence of connections between the two blocks additionally introduces a large number of irrelevant source–target candidates.

4.2.2 Ablation settings

Three screening-feature configurations are compared:

(a) *Original lagged components*: only $X_i(t-\ell)$ are included;

(b) *With squared components*: both $X_i(t-\ell)$ and $X_i^2(t-\ell)$ are included;

(c) *With squared and interaction components*: the complete screening configuration additionally includes $X_i(t-\ell)X_k(t-\ell)$ for $i < k$.

For all three configurations, $N = 512$, $L_{\max} = 5$, and 10-fold blocked cross-validation are used. The regularization parameter is selected according to the one-standard-error rule. All other screening settings and the simulated realization are kept unchanged, so that the observed differences arise only from the included screening features.

For each configuration, we report the number of true relationships retained (TR), true relationships removed (TM), and spurious relationships retained (SR). Screening recall is defined as $\mathrm{TR}/(\mathrm{TR}+\mathrm{TM})$. The false-retention rate is the proportion of the 65 noncausal off-diagonal pairs that are retained, and the candidate-reduction ratio is the proportion of all 72 off-diagonal pairs removed by screening.

Table 6 Variable-level screening results for the three screening-feature configurations.

Configuration	TR	TM	SR	Recall	FRR	CRR
Original lagged	4	3	0	0.571	0.000	0.944
With squared	5	2	5	0.714	0.077	0.861
Squared and interaction	7	0	4	1.000	0.062	0.847

Notes: TR, true relationships retained; TM, true relationships removed; SR, spurious relationships retained; FRR, false-retention rate; CRR, candidate-reduction ratio.

Just Accepted

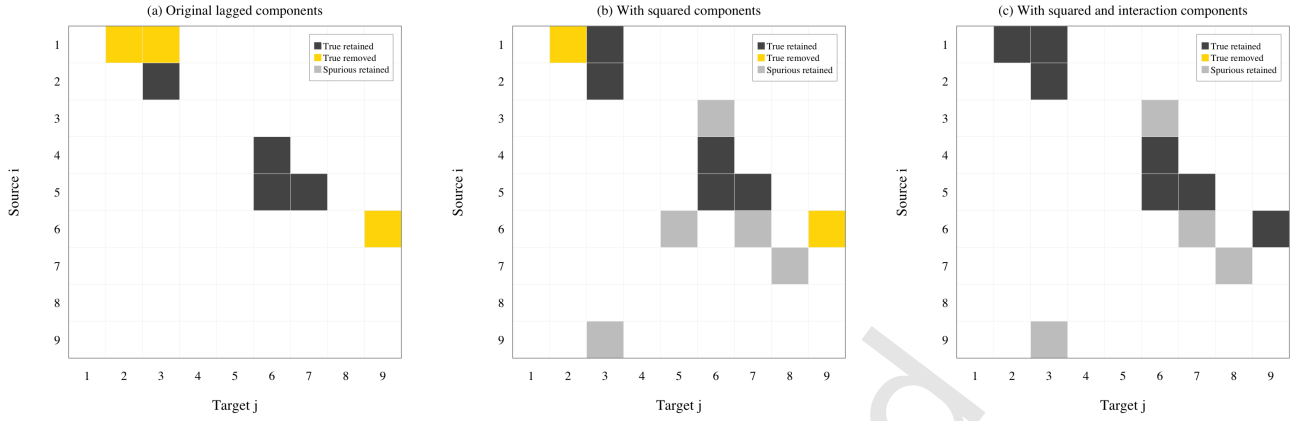


Fig. 9 Variable-level screening outcomes under three screening-feature configurations: (a) the original lagged components only; (b) the original and squared lagged components; and (c) the original lagged components, squared lagged components, and same-lag pairwise interactions. Each off-diagonal cell represents a directed source–target pair $(i \to j)$. Dark gray denotes a true relationship retained by screening, yellow denotes a true relationship removed, and light gray denotes a spurious relationship retained.

4.2.3 Results

The ablation results reveal two complementary requirements of the proposed screening framework: sufficient feature representation is needed to avoid removing relevant source variables, whereas structured sparsity is needed to prevent the expanded screening space from retaining an excessive number of irrelevant candidates.

As shown in Fig. 9(a), structured screening based only on the original lagged components produced the most parsimonious result, removing 94.4% of the possible off-diagonal candidates without retaining any spurious relationship. However, its screening recall was only 0.571. Although all four linearly driven relationships were retained, the quadratic relationship $X_1 \to X_3$ and the interaction-driven relationships $X_1 \to X_2$ and $X_6 \to X_9$ were removed. This result shows that structural regularization alone cannot compensate for an insufficient screening representation: a genuinely relevant variable may be excluded when its contribution is not adequately expressed by the available screening features.

Enriching the screening representation progressively reduced this form of model miss specification. After squared lagged components were introduced, the quadratic relationship $X_1 \to X_3$ was retained and the screening recall increased to 0.714, as shown in Fig. 9(b). The two interaction-driven relationships remained excluded until same-lag pairwise interactions were added. Under the complete configuration in Fig. 9(c), all seven true relationships were retained and the screening recall reached 1.000. Thus, the nonlinear expansion is necessary not because the method is restricted to nonlinear systems, but because the screening stage must be sufficiently expressive to avoid prematurely removing variables whose effects are not apparent through linear lagged components alone.

At the same time, enriching the representation substantially increases the feature dimension and would be undesirable without an accompanying sparsity-control mechanism. In the complete configuration, the screening design contains the original, squared, and interaction components, but the structured penalty retains only 11 of the 72 possible off-diagonal source–target pairs. The candidate-reduction ratio therefore remains 0.847, while the false-retention rate is only 0.062. This result illustrates the distinct role of structured sparsity: it controls the enlarged screening space and prevents richer feature representation from being translated into indiscriminate candidate retention.

The three configurations are fitted independently, with their regularization parameters selected separately by blocked cross-validation. Their retained sets are therefore not required to be nested, which explains the changes in the identity and number of spurious retained relationships across configurations. More importantly, the nonlinear components are used only as auxiliary features for assessing variable relevance. Once the variable set has been determined, only the original lagged components of the retained variables are passed to the subsequent reconstruction. The proposed procedure therefore does not impose a quadratic dynamical model on the analyzed system.

Overall, the two elements of the framework serve different but complementary purposes. Rich screening features reduce false exclusions caused by inadequate representation, whereas structured sparsity suppresses redundant candidates introduced by that richer representation. Because the original

Just Accepted

lagged components remain included in every configuration, linear dependencies are preserved as a nested special case rather than replaced by nonlinear terms. Together with the results from Section 4.1, the ablation indicates that the benefit of the proposed method arises from balancing representational coverage and structured candidate-space control, rather than from specialization to a particular nonlinear system.

4.3 Stability in Real Financial Systems

The simulation studies based on short-length time series demonstrate that the proposed structured approach improves causal discovery accuracy by effectively suppressing false positives. However, PMIME and its extensions are known to benefit from longer sample sizes, as information-theoretic estimators generally become more reliable with increasing data length. Therefore, to evaluate the robustness and stability of the proposed method in a more realistic setting, we consider a medium-scale dataset consisting of approximately 2000 observations. This experiment provides a moderate-scale assessment of practical feasibility rather than a systematic high-dimensional scalability study over increasing numbers of variables, lag orders, candidate features, or sample sizes.

The dataset spans roughly five years of daily observations (2019–2024) from real financial markets and includes 17 variables representing multiple interconnected asset classes. The variables cover major global equity markets across the United States, Europe, and Asia, key commodity futures such as energy and precious metals, major foreign exchange rates among leading currencies, and representative cryptocurrencies. This multi-asset composition captures a globally interconnected financial system involving traditional financial markets and digital assets. Such a system is typically characterized by weak dependencies, heterogeneous coupling strengths, and substantial market noise. Therefore, it provides a challenging testbed for information-theoretic causal inference methods.

Unlike the preceding simulation experiments, the financial application does not provide an observable ground-truth causal network. Therefore, the two parts of the empirical evaluation serve different purposes. The simulation experiments assess causal detection accuracy under known directional structures, including sensitivity, specificity, and false-positive control. In contrast, the financial application examines the cross-parameter structural reproducibility of the inferred strong-link patterns under different parameter settings. Accordingly, the retention and SSC measures reported in this section should be interpreted as indicators of structural reproducibility rather than direct evidence of causal correctness.

As baselines, we consider PMIME, the representative of traditional nonuniform embedding methods, and the low-dimensional approximation variant (L-PMIME), which approximates CMI in high-dimensional embeddings. These two baselines represent the nonuniform embedding causal detection framework and its internal refinement path, respectively. In contrast, the proposed structured method introduces a regularization step that constrains the candidate embedding space prior to reconstruction. This experimental design is intended to examine whether candidate-space control can improve the cross-parameter reproducibility of the inferred network structure in weakly coupled and noisy financial systems. It also allows the effect of candidate-space control to be compared with an internal approximation strategy based on low-dimensional CMI.

To evaluate structural reproducibility under parameter perturbations, we compare the causal-strength matrices obtained under two parameter configurations. Applying the corresponding causal-network output threshold τ_m to each matrix produces two directed edge sets, $\mathcal{E}_1$ and $\mathcal{E}_2$. Their intersection, $\mathcal{E}_s = \mathcal{E}_1 \cap \mathcal{E}_2$, defines the stable edge set. Directional retention rates quantify the proportion of edges in one threshold-derived set that are retained in the other. Detailed definitions of these metrics are provided in Appendix A. The same value of $k = 14$ is used for both PMIME and Structured to provide a size-controlled comparison of strong-link stability. The value $k = 14$ corresponds to the largest common strong-edge evaluation size supported by both PMIME-based methods across the examined reference configurations.

During the experiments, L-PMIME did not identify any directed edge above the corresponding causal-network output threshold under either examined parameter configuration. Consequently, both threshold-derived edge sets were empty, giving zero stable edges. The directional retention rates were undefined because their denominators were zero, while SSC@14 could not be computed because no top-14 positive strong-edge reference set was available. Under the adopted financial-data configuration, L-PMIME produced no directed edge above the output threshold. This result is specific to the examined data and parameter settings and does not support a general conclusion about the method.

Table 7 summarizes the reproducibility of the inferred edge sets under the two perturbation settings. Under Setting I, where A is perturbed while τ is held fixed, the two PMIME edge sets share 30 stable directed edges and achieve directional retention rates of 0.909 and 0.652, giving an average retention of 0.781. The corresponding Structured edge sets share 14 stable edges and achieve directional retention rates of 1.000

Just Accepted

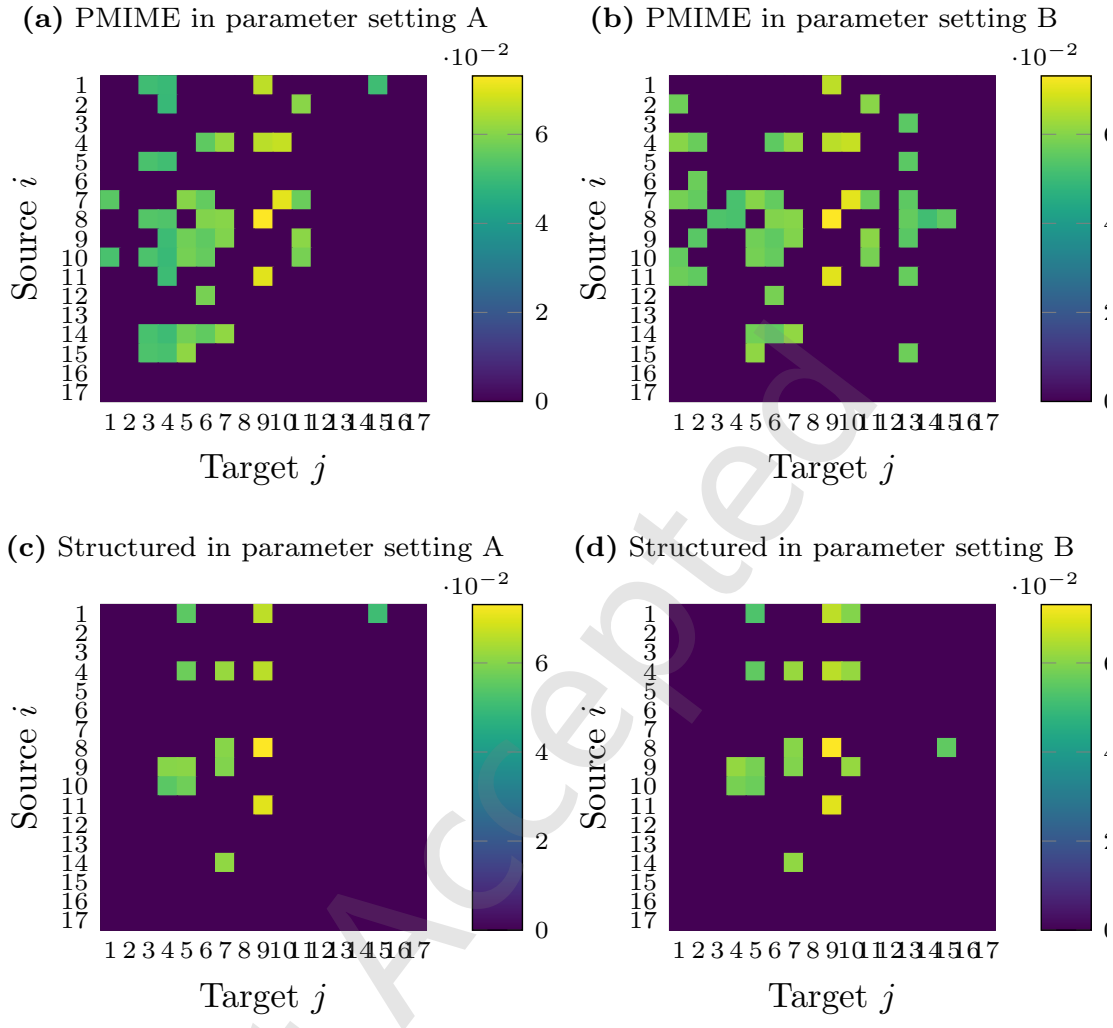


Fig. 10 Information flow matrices (17×17) in the real financial systems.

Table 7 Stability comparison under controlled parameter perturbations.

	Method	Stable edges	Retention (1→2)	Retention (2→1)	Avg. retention	SSC@14
Parameter Setting I	PMIME	30	0.909	0.652	0.781	0.880
	Structured	14	1.000	0.778	0.889	1.000
Parameter Setting II	PMIME	30	0.682	0.652	0.667	0.444
	Structured	14	0.933	0.778	0.856	1.000
Auxiliary Baseline	L–PMIME	0	N/A	N/A	N/A	N/A

Notes: I. perturbs the PMIME stopping parameter A while holding the causal-network output threshold τ fixed. II. jointly perturbs A and τ. L-PMIME produced empty threshold-derived edge sets under both examined parameter configurations.

and 0.778, with a higher average retention of 0.889.

Under Setting II, where both A and τ are perturbed, the PMIME edge sets again share 30 stable edges, but their average retention decreases to 0.667. Structured edge sets share 14 stable edges and achieve an average retention of 0.856. The absolute number of stable edges should not be compared directly across methods because PMIME produces a denser threshold-derived network. The normalized retention rates instead describe the proportion of each inferred edge set that remains reproducible under the corresponding parameter perturbation.

For a size-controlled comparison of the strongest inferred relationships, SSC is calculated using the same top-14 reference size for both methods. Under Setting (I), PMIME

Just Accepted

and Structured obtain SSC@14 values of 0.880 and 1.000, respectively. When both A and τ are perturbed in Setting (II), the PMIME value decreases to 0.444, whereas Structured maintains SSC@14 at 1.000. Thus, within the common top-14 evaluation, the strongest relationships inferred by Structured are more consistently covered by the stable edge set under the examined parameter perturbations.

The corresponding causal-strength matrices are shown in Fig. 10. PMIME produces a broader set of nonzero relationships, several of which change between parameter settings A and B. Structured yields a more concentrated output in which the dominant relationships are more consistently represented across the two configurations. The matrices visually illustrate the different reproducibility patterns quantified by the retention rates and SSC@14.

Overall, Structured produced a more reproducible strong-link structure under the examined parameter perturbations, although the retention results should be interpreted together with the different edge-set sizes and the common SSC@14 comparison.

4.4 Comparison with Related Time-Series Causal Discovery Approaches

Existing time-series causal discovery approaches improve causal identification at different stages of the analysis, including conditional-independence testing, mixed-embedding reconstruction, representation learning, nonstationarity modeling, and candidate-space control. Table 8 summarizes the primary mechanisms, typical focuses, and methodological roles of representative approaches considered in this study. Since these methods address different problems and are evaluated under different experimental settings, the comparison emphasizes their methodological characteristics rather than numerical results reported across separate studies.

PCMCI and PCMCI+ perform conditional-independence-based graph discovery, Revisited PMIME refines mixed-embedding reconstruction, STIC learns causal representations from short temporal windows, and CD-NOTS models changing causal mechanisms. Structured instead controls the candidate space before mixed-embedding reconstruction. PCMCI is evaluated directly as an external baseline, while the remaining methods clarify the methodological position of Structured.

5 Application

Financial dependence structures may change during major market shocks. We therefore examine the inferred network before and after the COVID-19 market disruption to assess how the proposed method represents window-specific structural variation in real data.

To investigate structural changes induced by this shock, the sample is divided into three phases containing 250 trading days. Phase 1 corresponds to the pre-shock period, during which global financial markets operated under relatively stable macroeconomic conditions. This phase serves as the structural baseline. Phase 2 overlaps with Phase 1 by 90% of observations but additionally includes the early stage of the COVID-19 outbreak. This period covers the rapid escalation of pandemic-related uncertainty, the February–March 2020 market sell-off, and the initial surge in volatility. The construction of this overlapping window allows us to assess both robustness under similar data and sensitivity to emerging structural perturbations. Phase 3 represents the post-shock adjustment period. This independent phase follows the peak market turbulence and includes massive monetary and fiscal policy interventions, liquidity injections by central banks, and gradual market stabilization. The selected variables form a cross-asset financial system consisting of major equity indices and sovereign bond markets across the United States and Europe, which is shown in Table 9. The system captures interactions between risk assets and safe-haven instruments, as well as cross-country and term-structure transmission mechanisms.

Before constructing the three analysis phases, the corresponding daily adjusted closing-price series were aligned according to their date indices. Closing prices were used only when adjusted closing prices were unavailable. Missing observations arising primarily from differences in market trading calendars were first filled using the last observation carried forward, and any remaining leading missing values were filled using the nearest subsequently available observation. The aligned price series were then transformed into continuously compounded log returns,

$$r_{i,t} = \log P_{i,t} - \log P_{i,t-1},$$

where $P_{i,t}$ denotes the price of variable i at time t. The log-return transformation reduces the influence of price trends and scale differences, but it does not guarantee strict stationarity. Accordingly, we do not assume that the financial system follows a single stationary causal mechanism throughout the entire observation period. The inferred networks are interpreted as window-specific structures, and differences across phases are not regarded as definitive evidence of true causal reversals.

Phase 1 exhibits significant causal interactions among Euro area government bonds of different maturities (short-,

Just Accepted

Table 8 Methodological comparison with related time-series causal discovery approaches.

Method	Primary mechanism	Typical focus	Relation to the proposed method
PCMCI/PCMCI+ [23, 24]	Conditional-independence testing with optimized conditioning sets	Lagged and contemporaneous causal-graph discovery	Provides an external constraint-based baseline, whereas Structured controls the candidate embedding space before PMIME reconstruction.
Revisited PMIME [20]	Improved statistical testing, backward revision, and examination of synergistic components	Refinement and validation of mixed-embedding reconstruction	Modifies the internal reconstruction procedure, whereas Structured retains the original PMIME causal criterion and reconstruction mechanism.
STIC [37]	Representation learning based on short-term temporal and mechanism invariance	Causal discovery from short time series using learned representations	Learns causal representations from temporal windows, whereas Structured uses explicit structured screening of the candidate embedding space.
CD-NOTS [38]	Conditional-independence discovery with a time-indexed variable	Causal discovery under nonstationarity and changing mechanisms	Explicitly models nonstationarity, whereas Structured primarily addresses redundancy in a stationary candidate embedding space.
Structured (proposed)	Group-sparse screening using original, squared, and same-lag interaction components	Candidate-space reduction before nonuniform mixed embedding	Restricts the candidate embedding space while preserving the subsequent PMIME causal criterion, stopping rule, and normalization scheme.

Table 9 Financial variables used in the stability testing.

Code	Short Name	Asset Name	Description
@ES#C	SPX	S&P 500	U.S. broad equity market benchmark
XG#C	DAX	Germany Equity	German stock market proxy
SW#C	SMI	Switzerland Equity	Swiss stock market proxy
LF#C	FTSE	UK Equity	UK large-cap equity index
EX#C	STOXX	EURO STOXX 50	Eurozone blue-chip equity index
@NQ#C	NAS	NASDAQ 100	U.S. technology-heavy equity index
@YM#C	DJ	Dow Jones	U.S. industrial equity index
@FV#C	US5Y	U.S. 5Y Treasury	Short-term U.S. government bond
@TY#C	US10Y	U.S. 10Y Treasury	Benchmark U.S. long-term yield
@US#C	US30Y	U.S. 30Y Treasury	Long-maturity U.S. government bond
EZ#C	EU2Y	Euro 1–3Y Govt Bond	Short-maturity Eurozone sovereign bond
BL#C	EU5Y	Euro 3–5Y Govt Bond	Medium-term Eurozone sovereign bond
BD#C	EU10Y	Euro 7–10Y Govt Bond	Long-term Eurozone sovereign bond
BX#C	EU30Y	Euro 15–30Y Govt Bond	Ultra-long Eurozone sovereign bond
BTP#C	IT10Y	Italy Govt Bond	Italian sovereign bond (credit-sensitive)
OAT#C	FR10Y	France Govt Bond	French sovereign bond
LG#C	UK10Y	UK Long Gilt	Long-maturity UK government bond

Notes: Daily adjusted closing-price data were obtained from Yahoo Finance through the `quantmod` package in R. The variables listed in the table were represented by their corresponding Yahoo Finance proxy series. The three reported analysis windows collectively cover the period from March 6, 2019 to March 17, 2021; the exact phase boundaries are provided in the corresponding figures.

Just Accepted

medium-, and long-term), forming transmission paths centered around the term structure. This structure reflects the dominant role of interest rate expectations, term premia, and monetary policy expectations within the bond market during relatively stable macroeconomic conditions. In contrast, the equity market occupies a relatively passive position in Phase 1, rarely acting as a primary causal source. Its fluctuations are more likely influenced by the bond market or global risk factors. This pattern is consistent with the classical financial mechanism observed prior to major shocks, where interest rates and macroeconomic expectations lead, and risk assets follow.

Phase 2 largely preserves the main inferred structure observed in Phase 1. In particular, the directional relation $EU5Y \rightarrow EU10Y$ remains unchanged across the two

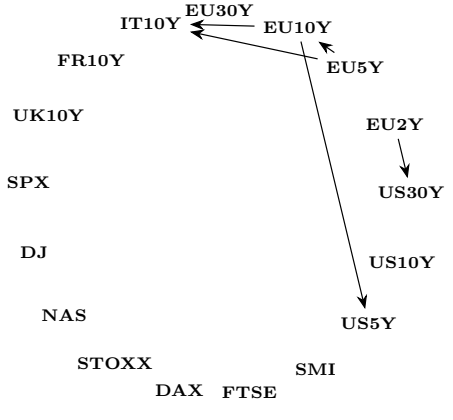


Fig. 11 **Phase 1:** 2019-03-06 - 2020-02-21

phases. This persistence is consistent with the approximately

90% overlap between the two observation windows. It indicates that replacing only a limited proportion of the pre-shock observations with observations from the early stage of the disruption does not completely alter the dominant structure inferred from the original window.

At the same time, several weaker links disappear and a small number of new directional relations emerge, such as $SMI \rightarrow DAX$. Because the observations added to Phase 2 are drawn from the early stage of the market disruption, these local changes may reflect initial adjustments in market pricing and information transmission. However, given the substantial overlap between Phases 1 and 2, these differences should be interpreted cautiously and should not be regarded as independent evidence of a structural break. Rather, Phase 2 illustrates that the dominant inferred structure is largely retained under a small change in sample composition, while some peripheral links remain sensitive to the newly introduced observations.

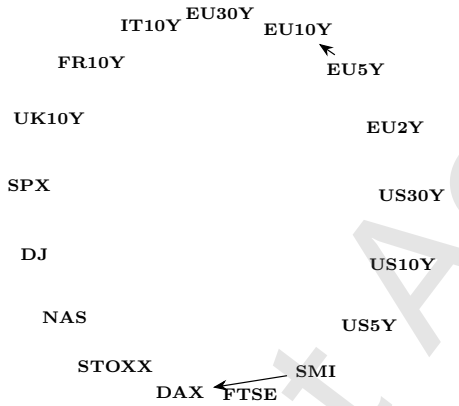


Fig. 12 **Phase 2:** 2019-04-10 – 2020-03-27

The inferred network in Phase 3 differs more substantially from those obtained in Phases 1 and 2. Several relations that were present in the earlier windows are weakened or absent, and some inferred directional patterns change. In particular, the relation $EU5Y \rightarrow EU10Y$, which is retained in both Phases 1 and 2, is replaced by $EU10Y \rightarrow EU5Y$ in Phase 3. Thus, no reversal is observed between the two highly overlapping first phases; the change in inferred direction occurs only when the later Phase 3 window is considered.

One plausible economic interpretation is that changes in monetary-policy expectations, liquidity conditions, term premia, and risk repricing during the later shock period altered the relative adjustment pattern between medium- and long-maturity Euro-area yields. Under this interpretation, the change from $EU5Y \rightarrow EU10Y$ to $EU10Y \rightarrow EU5Y$ may reflect a reorganization of information transmission along the yield curve. Nevertheless, this result should be described as a change in the inferred directional structure rather than as definitive evidence of a true causal reversal. Estimation variability, nonstationarity, and the particular choice of phase boundaries may also contribute to the observed difference.

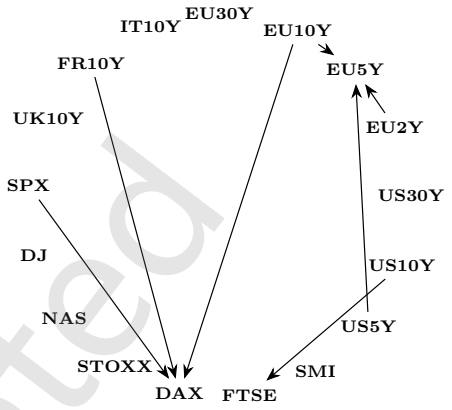


Fig. 13 **Phase 3:** 2020-03-30 – 2021-03-17

Therefore, the dominant inferred structure is largely preserved between the highly overlapping first two windows, whereas Phase 3 exhibits more substantial changes. These differences describe window-specific inferred structures and may also be influenced by estimation variability, nonstationarity, and window selection.

Overall, the phase comparison indicates that the proposed method is able to preserve the dominant structure under a small perturbation of the observation window while detecting a more substantial reconfiguration when the underlying market environment changes markedly. The observed directional change should nevertheless be interpreted as an inferred structural reversal rather than definitive proof of a true causal reversal, since estimation variability, nonstationarity, and window-selection effects may also influence the estimated network.

6 Discussion

The primary contribution of this study is the introduction of structured candidate-space control before mixed-embedding reconstruction. In conventional nonuniform embedding, the candidate space expands with the number of variables and the maximum lag, which increases the dimension of conditional mutual information estimation and the risk of selecting redundant components. The proposed framework organizes lagged components into variable-level groups and applies group sparsity before causal reconstruction. The subsequent reconstruction procedure therefore operates within a more compact candidate space, while its causal criterion, stopping rule, and normalization scheme remain unchanged.

Just Accepted

The simulation results clarify the conditions under which this candidate-space control is beneficial. In the linear and nonlinear stochastic systems, Structured reduced false-positive detections while preserving the true causal sources, leading to a favorable balance between sensitivity, specificity, and trial-to-trial stability. The chaotic systems showed that this improvement depends on screening recall. Structured preserved the PMIME result in the Hénon system, whereas the exclusion of a relevant source variable in the Lorenz system reduced the sensitivity of the subsequent reconstruction. Therefore, structured screening suppresses candidate redundancy, but its benefit depends on retaining the relevant source variables.

The ablation experiment further distinguishes the complementary functions of structural sparsity and nonlinear representation. Structured sparsity is required to remove redundant candidate variables, whereas sufficiently rich screening components are required to retain nonlinear causal sources. Squared components support the identification of quadratic effects, while same-lag interaction components support multiplicative effects. These nonlinear components are used only in the screening stage to identify relevant variables; the restricted search space subsequently contains the original lagged components of the retained variables. The feature expansion is therefore intended to improve screening coverage rather than to replace the information-theoretic nonlinear characterization performed by nonuniform embedding.

The real-data stability experiment showed that the dominant inferred relationships were reproducible under perturbations of the stopping parameter and the causal-network output threshold. The threshold-based retention rates should be interpreted together with differences in the sizes of the detected edge sets, because PMIME produced a denser inferred network than the proposed structured method. In contrast, the common SSC@14 criterion provides a size-controlled comparison of the strongest inferred relationships. Under this common criterion, Structured maintained complete weighted coverage of its reference strong-edge set under both perturbation settings, whereas the corresponding coverage of PMIME decreased when both the stopping parameter and the output threshold were perturbed.

The subsequent financial application examined temporal changes in the inferred network across the periods surrounding the COVID-19 disruption. The dominant structure was largely preserved under a small perturbation of the observation window, whereas more substantial changes appeared when the underlying market period changed markedly. Since the financial system has no known causal ground truth, the stability experiment and the application provide evidence of cross-parameter structural reproducibility and temporal structural variation, respectively, rather than validation of causal correctness.

Several limitations should be acknowledged. First, a relevant source variable removed during screening cannot be recovered by the subsequent reconstruction, making screening recall an important condition for successful candidate-space control. Second, the squared and same-lag interaction components constitute a finite-dimensional representation and may not capture higher-order or more complex nonlinear mechanisms. Finally, the selected regularization parameter affects the size of the restricted candidate space, while formal consistency and error-propagation properties of the two-stage procedure remain to be established. Future work will consider more flexible nonlinear screening representations and theoretical analysis of structured screening followed by information-theoretic reconstruction.

7 Conclusion

This study proposed a structured candidate-space control framework prior to mixed embedding reconstruction. By applying variable-level group sparsity to a screening representation containing original, squared, and same-lag interaction components, the framework restricts the candidate embedding space while preserving the subsequent reconstruction causal criterion, stopping rule, and normalization scheme. The simulation results show that this design can reduce redundant and false-positive candidates while maintaining causal recovery when the relevant source variables are retained during screening. The ablation experiment further confirms that structured sparsity and enriched nonlinear representation perform complementary functions in candidate reduction and nonlinear-source retention. The real-data stability experiment explains that the strongest relationships inferred by Structured were more consistently covered by the stable edge set than those inferred by nonuniform embedding under the examined parameter perturbations. The subsequent financial application further illustrated the preservation of the dominant inferred structure under a small change in the observation window and more substantial structural variation across markedly different market periods. Overall, the proposed framework provides a practical mechanism for controlling the candidate space of information-theoretic nonuniform embedding, with its effectiveness determined primarily by the recall of the screening stage.

Just Accepted

A Stability Metrics for Structural Robustness

To quantitatively evaluate structural reproducibility under parameter perturbations, we compare two causal-strength matrices obtained under two parameter configurations, denoted by $R^{(1)}, R^{(2)} \in \mathbb{R}^{K\times K}$. Configuration 1 is treated as the reference configuration, while Configuration 2 represents the corresponding parameter perturbation.

Threshold-derived edge sets. For configuration $m \in \{1, 2\}$, the causal-strength matrix is converted into a binary adjacency matrix using its corresponding causal-network output threshold τ_m:

$$B_{ij}^{(m)} = \mathbb{I}\left(R_{ij}^{(m)} > \tau_m\right), \qquad i \neq j, \quad m = 1, 2, \quad \text{(A1)}$$

where all diagonal entries are set to zero. Under Parameter Setting I, $\tau_1 = \tau_2$, whereas Parameter Setting II allows $\tau_1 \neq \tau_2$.

The corresponding directed edge set is

$$\mathcal{E}_m = \left\{(i, j) : B_{ij}^{(m)} = 1,\ i \neq j\right\}, \qquad m = 1, 2, \quad \text{(A2)}$$

and its size is

$$N_m = |\mathcal{E}_m| . \quad \text{(A3)}$$

Stable edges. The stable edge set is defined as the intersection of the two threshold-derived edge sets:

$$\mathcal{E}_s = \mathcal{E}_1 \cap \mathcal{E}_2. \quad \text{(A4)}$$

The number of stable edges reported in Table **??** is therefore

$$N_s = |\mathcal{E}_s| = |\mathcal{E}_1 \cap \mathcal{E}_2| . \quad \text{(A5)}$$

Directional retention rates. The directional retention rates are defined as

$$\mathrm{Ret}_{1\to2} = \frac{N_s}{N_1}, \quad \text{(A6)}$$

and

$$\mathrm{Ret}_{2\to1} = \frac{N_s}{N_2}. \quad \text{(A7)}$$

Their average is

$$\mathrm{Ret}_{\mathrm{avg}} = \frac{1}{2}\left(\mathrm{Ret}_{1\to2} + \mathrm{Ret}_{2\to1}\right). \quad \text{(A8)}$$

These quantities measure the proportion of detected edges in one parameter configuration that are reproduced in the other. A directional retention rate is undefined when its denominator is zero. Accordingly, $\mathrm{Ret}_{\mathrm{avg}}$ is also reported as undefined when both threshold-derived edge sets are empty.

Stable Strong Coverage at $k = 14$ (SSC@14). The retention rates are affected by the sizes of the complete threshold-derived edge sets. To provide a size-controlled assessment of the strongest inferred relationships, SSC is evaluated using the same value $k = 14$ for both PMIME and Structure.

Let $\mathcal{E}_{\mathrm{top},14}^{(1)}$ denote the set of the 14 strongest positive off-diagonal entries in the reference causal-strength matrix $R^{(1)}$. The averaged causal strength of each directed pair is defined as

$$W_{ij} = \frac{R_{ij}^{(1)} + R_{ij}^{(2)}}{2}. \quad \text{(A9)}$$

SSC@14 is then defined as

$$\mathrm{SSC@14} = \frac{\sum_{(i,j)\in\mathcal{E}_s\cap\mathcal{E}_{\mathrm{top},14}^{(1)}} W_{ij}}{\sum_{(i,j)\in\mathcal{E}_{\mathrm{top},14}^{(1)}} W_{ij}}. \quad \text{(A10)}$$

SSC@14 measures the strength-weighted proportion of the 14 strongest reference relationships that are contained in the stable edge set under parameter perturbation. A value of one indicates complete weighted coverage of the reference strong-edge set, whereas a lower value indicates that part of its causal strength is associated with relationships that are not retained under the perturbed configuration.

SSC@14 is undefined when fewer than 14 positive off-diagonal causal-strength values are available in the reference matrix or when the denominator is zero. This convention applies to L-PMIME in the financial-data experiment, where the threshold-derived edge sets are empty and no top-14 positive strong-edge reference set can be constructed.

Reference

[1] A. F. Siegenfeld and Y. Bar-Yam, An introduction to complex systems science and its applications, *Complexity*, vol. 2020, no. 1, p. 6105872, 2020.

[2] M. Paluš, A. Krakovská, J. Jakubík, and M. Chvosteková, Causality, dynamical systems and the arrow of time, *Chaos: An Interdisciplinary Journal of Nonlinear Science*, vol. 28, no. 7, 2018.

[3] X. S. Liang, Unraveling the cause-effect relation between time series, *Physical Review E*, vol. 90, no. 5, p. 052150, 2014.

[4] K. Hlaváčková-Schindler, M. Paluš, M. Vejmelka, and J. Bhattacharya, Causality detection based on information-theoretic approaches in time series analysis, *Physics Reports*, vol. 441, no. 1, pp. 1–46, 2007.

[5] C. W. Granger, Investigating causal relations by econometric models and cross-spectral methods, *Econometrica: journal of the Econometric Society*, pp. 424–438, 1969.

[6] T. Schreiber, Measuring information transfer, *Physical review letters*, vol. 85, no. 2, p. 461, 2000.

[7] I. Vlachos and D. Kugiumtzis, Nonuniform state-space reconstruction and coupling detection, *Physical Review*

Just Accepted

E—Statistical, Nonlinear, and Soft Matter Physics, vol. 82, no. 1, p. 016207, 2010.

[8] D. Kugiumtzis, Direct-coupling information measure from nonuniform embedding, *Physical Review E—Statistical, Nonlinear, and Soft Matter Physics*, vol. 87, no. 6, p. 062918, 2013.

[9] C. Koutlis and D. Kugiumtzis, Discrimination of coupling structures using causality networks from multivariate time series, *Chaos: An Interdisciplinary Journal of Nonlinear Science*, vol. 26, no. 9, 2016.

[10] D. Kugiumtzis and V. K. Kimiskidis, Direct causal networks for the study of transcranial magnetic stimulation effects on focal epileptiform discharges, *International journal of neural systems*, vol. 25, no. 05, p. 1550006, 2015.

[11] D. Kugiumtzis, C. Koutlis, A. Tsimpiris, and V. K. Kimiskidis, Dynamics of epileptiform discharges induced by transcranial magnetic stimulation in genetic generalized epilepsy, *International journal of neural systems*, vol. 27, no. 07, p. 1750037, 2017.

[12] A. Papana, C. Kyrtsou, D. Kugiumtzis, and C. Diks, Financial networks based on granger causality: A case study, *Physica A: Statistical Mechanics and its Applications*, vol. 482, pp. 65–73, 2017.

[13] ——, Assessment of resampling methods for causality testing: A note on the us inflation behavior, *PloS one*, vol. 12, no. 7, p. e0180852, 2017.

[14] J. Runge, J. Heitzig, V. Petoukhov, and J. Kurths, Escaping the curse of dimensionality in estimating multivariate transfer entropy, *Physical review letters*, vol. 108, no. 25, p. 258701, 2012.

[15] J. Runge, R. V. Donner, and J. Kurths, Optimal model-free prediction from multivariate time series, *Physical Review E*, vol. 91, no. 5, p. 052909, 2015.

[16] G. Brown, A. Pocock, M.-J. Zhao, and M. Luján, Conditional likelihood maximisation: a unifying framework for information theoretic feature selection, *The journal of machine learning research*, vol. 13, pp. 27–66, 2012.

[17] L. Faes, G. Nollo, and A. Porta, Information-based detection of nonlinear granger causality in multivariate processes via a nonuniform embedding technique, *Physical Review E—Statistical, Nonlinear, and Soft Matter Physics*, vol. 83, no. 5, p. 051112, 2011.

[18] Z. Jia, Y. Lin, Z. Jiao, Y. Ma, and J. Wang, Detecting causality in multivariate time series via non-uniform embedding, *Entropy*, vol. 21, no. 12, p. 1233, 2019.

[19] Z. Jia, Y. Lin, Y. Liu, Z. Jiao, and J. Wang, Refined nonuniform embedding for coupling detection in multivariate time series, *Physical Review E*, vol. 101, no. 6, p. 062113, 2020.

[20] A. Fotiadis, I. Vlachos, and D. Kugiumtzis, The causality measure of partial mutual information from mixed embedding (pmime) revisited, *Chaos: An Interdisciplinary Journal of Nonlinear Science*, vol. 34, no. 3, 2024.

[21] P. Spirtes and C. Glymour, An algorithm for fast recovery of sparse causal graphs, *Social science computer review*, vol. 9, no. 1, pp. 62–72, 1991.

[22] M. Kalisch and P. Bühlman, Estimating high-dimensional directed acyclic graphs with the pc-algorithm. *Journal of Machine Learning Research*, vol. 8, no. 3, 2007.

[23] J. Runge, P. Nowack, M. Kretschmer, S. Flaxman, and D. Sejdinovic, Detecting and quantifying causal associations in large nonlinear time series datasets, *Science advances*, vol. 5, no. 11, p. eaau4996, 2019.

[24] J. Runge, Discovering contemporaneous and lagged causal relations in autocorrelated nonlinear time series datasets, in *Conference on uncertainty in artificial intelligence*. Pmlr, 2020, pp. 1388–1397.

[25] Z. Jiang, Z. Ning, H. Miao, and L. Wang, Stdnet: A spatio-temporal decomposition neural network for multivariate time series forecasting, *Tsinghua Science and Technology*, vol. 29, no. 4, pp. 1232–1247, 2024.

[26] H. Zhao, P. Rui, J. Chen, S. Zhao, and Y. Zhang, Restage: Relation structure-aware hierarchical heterogeneous graph embedding, *Tsinghua Science and Technology*, vol. 30, no. 1, pp. 198–214, 2025.

[27] S. Li, Q. Zhao, A. R. Kivaisi, L. Wang, W. Liu, and X. Wang, Causality-aware single-source domain generalization for transferring knowledge to multiple, unknown target domain, *Tsinghua Science and Technology*, vol. 31, no. 1, pp. 518–529, 2026. [Online]. Available: https://www.sciopen.com/article/10.26599/TST.2024.9010099

[28] T. Hastie, R. Tibshirani, J. Friedman *et al.*, The elements of statistical learning, 2009.

[29] M. J. Wainwright, *High-dimensional statistics: A non-asymptotic viewpoint*. Cambridge university press, 2019, vol. 48.

[30] R. Tibshirani, Regression shrinkage and selection via the lasso, *Journal of the Royal Statistical Society Series B: Statistical Methodology*, vol. 58, no. 1, pp. 267–288, 1996.

[31] N. Meinshausen and P. Bühlmann, Stability selection, *Journal of the Royal Statistical Society Series B: Statistical Methodology*, vol. 72, no. 4, pp. 417–473, 2010.

[32] R. Vicente, M. Wibral, M. Lindner, and G. Pipa, Transfer entropy—a model-free measure of effective connectivity for the neurosciences, *Journal of computational neuroscience*, vol. 30, no. 1, pp. 45–67, 2011.

[33] J. Runge, J. Heitzig, N. Marwan, and J. Kurths, Quantifying causal coupling strength: A lag-specific measure for multivariate time series related to transfer entropy, *Physical Review E—Statistical, Nonlinear, and Soft Matter Physics*, vol. 86, no. 6, p. 061121, 2012.

[34] P. Shahsavari Baboukani, C. Graversen, E. Alickovic, and J. Østergaard, Estimating conditional transfer entropy in time series using mutual information and nonlinear prediction, *Entropy*, vol. 22, no. 10, p. 1124, 2020.

[35] A. Papana, E. Siggiridou, and D. Kugiumtzis, Detecting direct causality in multivariate time series: A comparative study,

Just Accepted

Communications in Nonlinear Science and Numerical Simulation, vol. 99, p. 105797, 2021.

[36] Z. Jia, X. Cai, Y. Hu, J. Ji, and Z. Jiao, Delay propagation network in air transport systems based on refined nonlinear granger causality, *Transportmetrica B: Transport Dynamics*, vol. 10, no. 1, pp. 586–598, 2022.

[37] R. Shen, Y. Guan, L. Liu, Y. Yang, B. Wang, C. Zhao, and J. Jiang, Causal discovery from time-series data with short-term invariance-based convolutional neural networks, *Mathematics*, vol. 13, no. 24, p. 3979, 2025.

[38] A. Sadeghi, A. Gopal, and M. Fesanghary, Causal discovery from nonstationary time series, *International Journal of Data Science and Analytics*, vol. 19, no. 1, pp. 33–59, 2025.

[39] Z. Ning, H. Miao, Z. Jiang, and L. Wang, Using multi-scale convolution fusion and memory-augmented adversarial autoencoder to detect diverse anomalies in multivariate time series, *Tsinghua Science and Technology*, vol. 30, no. 1, pp. 234–246, 2025.

[40] J. Qian, J. Miao, and X. Sun, Hierarchical causal model for analysis of complex social system, *Tsinghua Science and Technology*, vol. 30, no. 6, pp. 2606–2624, 2025.

[41] P. Bühlmann and S. Van De Geer, *Statistics for high-dimensional data: methods, theory and applications*. Springer Science & Business Media, 2011.

[42] M. Yuan and Y. Lin, Model selection and estimation in regression with grouped variables, *Journal of the Royal Statistical Society Series B: Statistical Methodology*, vol. 68, no. 1, pp. 49–67, 2006.

[43] N. Simon, J. Friedman, T. Hastie, and R. Tibshirani, A sparse-group lasso, *Journal of computational and graphical statistics*, vol. 22, no. 2, pp. 231–245, 2013.

[44] B. Schelter, M. Winterhalder, B. Hellwig, B. Guschlbauer, C. H. Lücking, and J. Timmer, Direct or indirect? graphical models for neural oscillators, *Journal of Physiology-Paris*, vol. 99, no. 1, pp. 37–46, 2006.

[45] B. Gourévitch, R. L. Bouquin-Jeannès, and G. Faucon, Linear and nonlinear causality between signals: methods, examples and neurophysiological applications, *Biological cybernetics*, vol. 95, no. 4, pp. 349–369, 2006.

Author biography



Zheng Tan is currently a Master's student in Mathematics at University of New South Wales, Australia. She has a background in mathematical modeling through participation in multiple modeling competitions. Her research interests include statistical modeling and system identification, with a focus on data-driven analysis of complex systems.



Ziyu Jia is an Associate Professor at the Institute of Automation, Chinese Academy of Sciences. His research focuses on time-series analysis methods. He has published over 60 peer-reviewed papers in venues such as IEEE Transactions on Affective Computing, IEEE Transactions on Multimedia, IEEE Transactions on Neural Systems and Rehabilitation Engineering, KDD, and ICLR. Dr. Jia currently serves as an Associate Editor or Editorial Board Member for prestigious journals including IEEE Transactions on Affective Computing and Information Fusion, and he is an Area Chair for major AI and machine learning conferences such as IJCAI and IJCNN. In addition to his academic contributions, Dr. Jia has extensive industry experience, having successfully led multiple R&D projects and secured several patents. He has received numerous honors, including the MSRA StarTrack Award, and the CIE Young Talent Award.

Just Accepted