Identifiability of VAR(1) model in a stationary setting

Bixuan LIU

Joint work with: Stéphane Robin, Viet Chi Tran and Elisa Thébault

Journées de Statistique

Clermont-Ferrand, June 1, 2026

Overview

Ecological background: model

Ecosystems are often modeled by **dynamical systems**.

For example, Robert May's model[1]

$$\frac{\mathrm{d}X_t}{\mathrm{d}t} = \Lambda^T X_t,$$

- X_t: n-dimensional vector of species abundances
- Λ: **deterministic** $n \times n$ interaction matrix
    - cooperation, competition, predation

(and the Lotka-Volterra equations[2])

Central topic in ecology: determination of Λ.

1. May (1972) 2. Akjouj et al. (2024)

Ecological background: data structure

What we would need

$X_1(t), X_2(t), \cdots, X_n(t)$: time series sampling of species $1, 2, \cdots, n$.

Too costly in practice[1].

What we have

Across many sites:

Species	1	2	$\cdots$	n	
Site 1	X_{11}	X_{12}	$\cdots$	X_{1n}	$=: X_1$
Site 2	X_{21}	X_{22}	$\cdots$	X_{2n}	$=: X_2$
$\vdots$	$\vdots$	$\vdots$	$\ddots$	$\vdots$	$\vdots$
Site N	X_{N1}	X_{N2}	$\cdots$	X_{Nn}	$=: X_N$

$X_1, X_2, \ldots, X_N$: i.i.d. samples from the stationary distribution.

Recovery of dynamic networks in ecosystems using only **static data**.

1. Legendre et al. (2010)

VAR(1) in a stationary setting

First-order Vector Autoregressive model

$$\begin{cases} X_1 = \epsilon_1, \\ X_t = \Lambda^T X_{t-1} + \epsilon_t, \quad t > 1, \\ \epsilon_t \sim \mathcal{N}(0, \omega I_n), \end{cases}$$

where Λ is a **deterministic** $n \times n$ matrix.

Associated directed graph

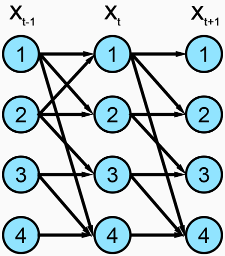
$$G = (V, \mathfrak{E}_G), \qquad V = \{1, \dots, n\}, \qquad (i,j) \in \mathfrak{E}_G \iff \Lambda_{ij} \neq 0.$$

If $\rho(\Lambda) < 1$, then X_t admits a stationary distribution

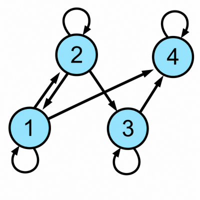
$$X_\infty \sim \mathcal{N}(0, \Sigma), \qquad \Sigma = \Lambda^T \Sigma \Lambda + \omega I_n.$$

Question: Given Σ, can we recover supp(Λ), equivalently the underlying graph G?

Graphical model

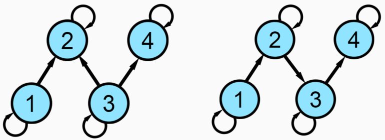


The graphical VAR(1) model

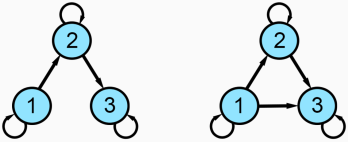


The encoded graph G

What we have shown:



Pairwise identifiable



Pairwise not identifiable

Related work: statistical approaches

Recall the stationary covariance equation:

$$\Sigma = \Lambda^T \Sigma \Lambda + \omega I_n.$$

Not directly applicable

- **Time-series analysis:** requires temporal observations.
- **Gaussian graphical models:**

$$\mathrm{supp}(\Sigma^{-1}) \neq \mathrm{supp}(\Lambda).$$

Existing condition

Young's condition[1] counts the number of free parameters.

A necessary condition for identifiability is:

$$|\,\mathrm{supp}_{\mathrm{off}}(\Lambda)| \leq \frac{n(n-3)}{2}, \;\; n \geq 5.$$

Limitation: Young's condition is only necessary; it is too weak and often satisfied in practice.

1. Young et al. (2019)

Related work: algebraic approaches

This line of work typically relies on methods from *algebraic statistics*[1].

Linear structural equation models (SEMs)

$$X = \Lambda^T X + \epsilon, \qquad \epsilon \sim \mathcal{N}(0, \Omega).$$

When $I - \Lambda$ is invertible, $X = (I - \Lambda)^{-T}\epsilon$. The covariance matrix Σ of X satisfies

$$K = \Sigma^{-1} = (I - \Lambda)\Omega^{-1}(I - \Lambda)^T.$$

Drton's work[2]: Studying algebraic properties of the Jacobian matrix of the map $(\Lambda, \Omega) \mapsto K$ yield graphical conditions for identifiability.

Other algebraic approaches are also developed for both linear SEMs[3,4] and continuous Lyapunov models[5,6].

1. Sullivant (2018) 2. Drton et al. (2025) 3. Drton (2018) 4. Bongers et al. (2021) 5. Améndola et al. (2025) 6. Recke et al. (2026)

Overview

Statistical model and identifiability definition

Stationary VAR(1) model associated with G

$$\mathbf{M}_G = \left\{\Sigma \;\middle|\; \Sigma = \Lambda^T \Sigma \Lambda + \omega I_n,\ \mathrm{supp}(\Lambda) = G,\ \omega > 0\right\}.$$

Identifiability of two candidate graphs

Global identifiability

$$\mathbf{M}_1 \cap \mathbf{M}_2 = \emptyset$$

No covariance matrix is compatible with both graphs.

Too strong in practice.

Generic identifiability

$$\dim(\mathbf{M}_1 \cap \mathbf{M}_2) < \max\{\dim(\mathbf{M}_1), \dim(\mathbf{M}_2)\}$$

The intersection is a Lebesgue measure zero subset.

Ambiguity is negligible.

Jacobian matrix and model dimension

The parametrization map:

$$\phi_G : (\Lambda, \omega) \mapsto \Sigma, \text{ s.t. } \Sigma = \Lambda^T \Sigma \Lambda + \omega I_n.$$

The Jacobian matrix of the model is

$$\mathbf{J}_G = \left(\frac{\partial \left(\phi_G\right)_j}{\partial \theta_i} \right), \quad 1 \leq i \leq E_G + 1, \quad 1 \leq j \leq \frac{n(1+n)}{2},$$

where E_G is the number of edges in G, $\left(\phi_G\right)_j$'s are the distinct entries of Σ, and θ_i's are the entries of Λ and ω.

Proposition

$$\dim\left(\mathbf{M}_G\right) = \operatorname{rank}\left(\mathbf{J}_G\right) \text{ at a generic point}^{(\star)}$$

Proof: It's sufficient to prove that ϕ_G is a rational map.

$\Rightarrow$ A tool to calculate model dimension and test identifiability.

$(\star)$ a point in a **general position**.

Linear matroid

Linear matroid of a matrix A

The pair $\{E, \mathcal{I}\}$, where $E = \{1, \cdots, n\}$, representing the n columns of A, and

$$\mathcal{I} = \left\{ S \subseteq E \mid A^S \text{ are linearly independent} \right\}.$$

Example: Let

$$A = \begin{bmatrix} 2 & 0 & 2 & 4 \\ 0 & 1 & 1 & 0 \\ 0 & 0 & 1 & 0 \\ 4 & 0 & 4 & 8 \end{bmatrix},$$

then the linear matroid of A is $\{E, \mathcal{I}\}$, where

- $E = \{1, 2, 3, 4\}$;
- $\mathcal{I} = \{\emptyset, \{1\}, \{2\}, \{3\}, \{4\}, \{1,2\}, \{1,3\}, \{2,3\}, \{2,4\}, \{3,4\}, \{1,2,3\}, \{2,3,4\}\}$.

Jacobian matroid and identifiability criteria

Jacobian matroid $\mathcal{J}(\mathbf{M}_G)$ = the linear matroid of the Jacobian matrix.

Proposition[1]

Assume that $\dim(\mathbf{M}_1) \geq \dim(\mathbf{M}_2)$. If there exists a subset S of the coordinates such that

$$S \in \mathcal{J}(\mathbf{M}_2) \setminus \mathcal{J}(\mathbf{M}_1),$$

then $\dim(\mathbf{M}_1 \cap \mathbf{M}_2) < \min\{\dim(\mathbf{M}_1), \dim(\mathbf{M}_2)\}$.

A direct result: if $\dim(\mathbf{M}_1) = \dim(\mathbf{M}_2)$, then the condition can be transformed to

$$S \in \mathcal{J}(\mathbf{M}_1) \setminus \mathcal{J}(\mathbf{M}_2) \text{ or } S \in \mathcal{J}(\mathbf{M}_2) \setminus \mathcal{J}(\mathbf{M}_1).$$

1. Sullivant (2023)

Overview

1. Introduction

2. Identifiability framework

3. Main results

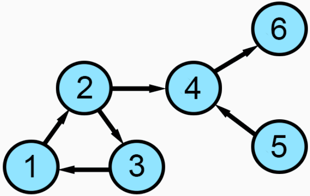
Maximal class

The set of reachable nodes from a Strongly Connected Component (SCC) with in-degree zero.

Example:

- SCCs: $\{1,2,3\},\{4\},\{5\},\{6\}$;
- Sources: $\{1,2,3\}$ and $\{5\}$ (SCCs with in-degree zero);
- Set of *maximal classes* (unique):

$$\{\{\mathbf{1},\mathbf{2},\mathbf{3},4,6\},\{4,\mathbf{5},6\}\}\,.$$



Maximal class and Jacobian nullity

A column of the Jacobian matrix $\mathbf{J}_G^{[\sigma_{ij}]}$ is zero if and only if nodes i and j belong to distinct maximal classes.

Model dimension

Theorem 1: model dimension

If G *does not contain multi-edges*, then

$$\dim(\mathbf{M}_G) = \min\{n_r, n'_c\},$$

where

$$n_r = E_G + 1;$$

$$n'_c = |\{\{a,b\} \mid a, b \in [n], a, b \text{ belong to the same maximal class}\}|.$$

Idea of proof: We prove that the submatrix $\mathbf{J}'_G$ containing all non-zero columns of the Jacobian matrix is full rank generically by showing

1. **Polynomial:** Prove that $\det(\mathbf{J}'_G)$ is a polynomial.
2. **Non-triviality:** Identify a specific parameter point where the determinant is non-zero.

Maximal class characterization

Theorem 2: Models with the same dimension

$\mathbf{M}_1$ and $\mathbf{M}_2$ with the same dimension are generically identifiable if G_1 and G_2 have different maximal classes.

Idea of proof: For models of the same dimension, generic identifiability is established by showing

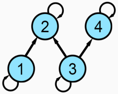
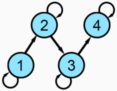
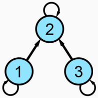
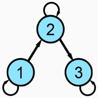
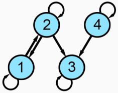
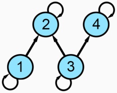
$$\exists S_1 \in \mathcal{J}(\mathbf{M}_1) \setminus \mathcal{J}(\mathbf{M}_2) \text{ or } \exists S_2 \in \mathcal{J}(\mathbf{M}_2) \setminus \mathcal{J}(\mathbf{M}_1).$$

The result follows by a consequence of the main proposition ensuring that $\exists i, j \in [n]$ s.t.

$$\mathbf{J}_1^{[\sigma_{ij}]} = \mathbf{0} \text{ and } \mathbf{J}_2^{[\sigma_{ij}]} \neq \mathbf{0}.$$

Summary

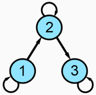
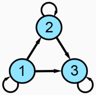
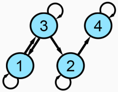
$\mathbf{M}_1$ and $\mathbf{M}_2$ are generically identifiable if

Condition		Example	
G_1 and G_2 do not contain multi-edges	$\dim(\mathbf{M}_1) = \dim(\mathbf{M}_2)$, and $\mathfrak{MC}_1 \neq \mathfrak{MC}_2$	G_1 	G_2 
	$\dim(\mathbf{M}_1) \neq \dim(\mathbf{M}_2)$	G_1 	G_2 
G_1 or G_2 contains multi-edges	Theorem 4[1]	G_1 	G_2 

1. Liu (2025)

Summary

$\mathbf{M}_1$ and $\mathbf{M}_2$ do not satisfy the sufficient identifiability criteria in this paper if

Condition		Example	
G_1 and G_2 do not contain multi-edges	$\dim(\mathbf{M}_1) = \dim(\mathbf{M}_2)$ and $\mathfrak{MC}_1 = \mathfrak{MC}_2$	G_1 	G_2 
G_1 or G_2 contains multi-edges	Theorem 4[1] is not satisfied	G_1 	G_2 is any graph

1. Liu (2025)

Thank you for your attention!

References i

R. M. May.
Will a large complex system be stable?
Nature, 238(5364):413–414, 1972.

I. Akjouj, M. Barbier, M. Clenet, W. Hachem, M. Maïda, F. Massol, J. Najim, and V. C. Tran.
Complex systems in ecology: a guided tour with large Lotka–Volterra models and random matrices.
Proceedings of the Royal Society A, 480(2285):20230284, 2024.

P. Legendre, M. De Cáceres, and D. Borcard.
Community surveys through space and time: testing the space–time interaction in the absence of replication.
Ecology, 91(1):262–272, 2010.

References ii

W. C. Young, K. Y. Yeung, and A. E. Raftery.
Identifying dynamical time series model parameters from equilibrium samples, with application to gene regulatory networks.
Statistical modelling, 19(4):444–465, 2019.

S. Sullivant.
Algebraic statistics.
American Mathematical Society, 194, 2018.

M. Drton, B. Hollering, and J. Wu.
Identifiability of homoscedastic linear structural equation models using algebraic matroids.
Advances in Applied Mathematics, 163:102794, 2025.

M. Drton.
Algebraic problems in structural equation modeling.
In *The 50th Anniversary of Gröbner Bases*, 77:35–87, Mathematical Society of Japan, 2018.

References iii

S. Bongers, P. Forré, J. Peters, and J. M. Mooij.
Foundations of structural causal models with cycles and latent variables.
The Annals of Statistics, 49(5):2885–2915, 2021.

C. Améndola, T. Boege, B. Hollering, and P. Misra.
Structural identifiability of graphical continuous Lyapunov models.
arXiv preprint arXiv:2510.04985, 2025.

C. O. Recke and N. R. Hansen.
Identifiability and estimation in continuous Lyapunov models.
arXiv preprint arXiv:2603.17142, 2026.

B. Liu.
Identifiability of VAR(1) model in a stationary setting.
arXiv preprint arXiv:2504.03466, 2025.