



Identifiable Markov Switching Models with Instantaneous Effects and Exponential Families

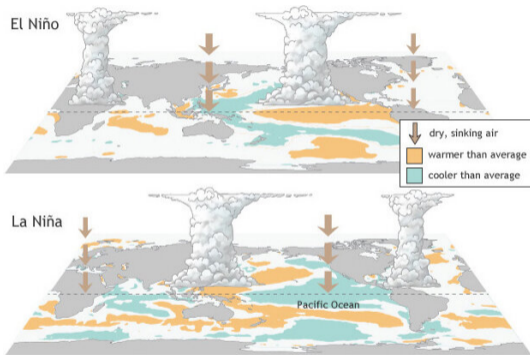
ICML 2026

Roel Hulsman^{1,2}, Carles Balsells-Rodas³, Sara Magliacane^{1,4}

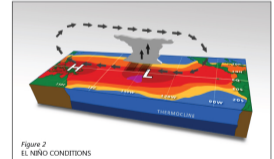
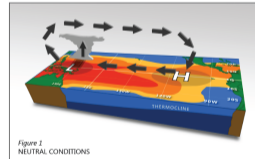
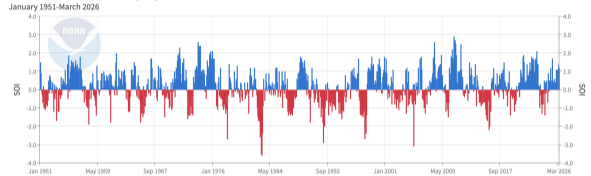
¹University of Amsterdam ²Adyen ³Imperial College London ⁴Saarland University

AMLab seminar – May 12th, 2026

Regimes in El Niño-Southern Oscillation (ENSO)^{1,2,3}



Southern Oscillation Index (SOI)

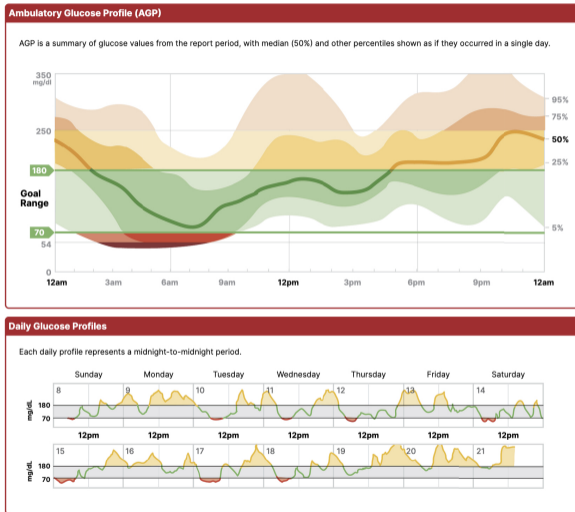


¹Michelle L'Heureux. *What is the El Niño-Southern Oscillation (ENSO) in a nutshell?* 2014. www.climate.gov/blogs/what-el-nino-nutshell.

²<https://www.ncei.noaa.gov/access/monitoring/enso/soi>.

³Emily Becker. *EN... SO?* 2014. www.climate.gov/blogs/en-so.

Instantaneous effects in glucose levels of type-1 diabetes patients⁴



⁴American Diabetes Association. *Glycemic targets: standards of medical care in diabetes*. Diabetes Care, 2025.

Our work in short

Main contribution: We establish *identifiability* for a broad class of regime-switching (nonlinear) structural causal models (SCMs) under independent exponential family noise.

Identifiability theory characterises when the data likelihood *uniquely determines* latent regimes.

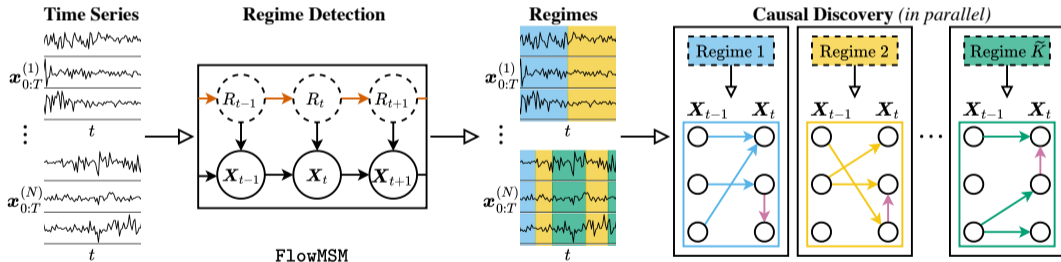


Figure 1: We detect *latent regimes* $R_{0:T}$ and *window graphs* from time series $\mathbf{X}_{0:T}$, with dependencies between regimes (red edges) and *instantaneous effects* between observed variables (pink edges).

SCMs and MSMs

What is a **structural causal model**⁵ (SCM)?

A **structural causal model** encodes the causal relations between
endogenous variables $\mathbf{X} = (X_1, \dots, X_D) \in \mathcal{X}$ (in this work continuous: $\mathcal{X} \subseteq \mathbb{R}^D$);
their causal parents $\text{Pa}(X_i) \subset \mathbf{X} \setminus X_i$; and
exogenous noises $\epsilon = (\epsilon_1, \dots, \epsilon_D) \in \mathcal{X}_\epsilon$ with $\epsilon \sim p(\epsilon)$.

We assume the data generating process can be described by a set of **structural equations**:

$$\begin{aligned} X_1 &\leftarrow f_1(\text{Pa}(X_1), \epsilon_1), \\ &\vdots \\ X_D &\leftarrow f_D(\text{Pa}(X_D), \epsilon_D). \end{aligned} \tag{1}$$

Typically, we assume causal parents are **acyclic**, such that the structural equations are *recursive*.

⁵Pearl. *Causality: Models, Reasoning, and Inference*. 2009.

The causal structure can be visualised in a **causal graph**

Assuming **causal sufficiency**, we exclude latent confounders between observed variables.
That is, all relevant variables are observed.

It allows us to visualise the causal structure in a **directed acyclic graph** (DAG).

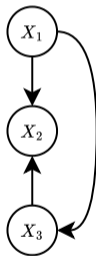


Figure 2: Example DAG with $D = 3$ variables.

(But we partly relax this later...)

A temporal SCM

Say we observe time series $\mathbf{X}_{0:T} \in \mathcal{X}^{\times(T+1)}$.

In vector-notation, for (instantaneous) parents $\mathbf{Pa}(\mathbf{X}_t) \subset \mathbf{X}_{t-1:t}$, we obtain **initial** equations,

$$\mathbf{X}_0 \leftarrow \mathbf{f}^0(\mathbf{Pa}(\mathbf{X}_0), \epsilon_0), \quad (2)$$

and **transition** equations for $t \in \{1, \dots, T\}$,

$$\mathbf{X}_t \leftarrow \mathbf{f}(\mathbf{Pa}(\mathbf{X}_t), \epsilon_t). \quad (3)$$

For simplicity, causal parents are of maximum lag $L = 1$, but all our results generalise.

We typically assume **causal stationarity**, meaning functions \mathbf{f}^0 , \mathbf{f} and causal parents \mathbf{Pa} are invariant across time. (*But not in this work...*)

A temporal causal structure

Under **causal stationarity**, we observe a repeating causal structure. (*But not in this work...*)

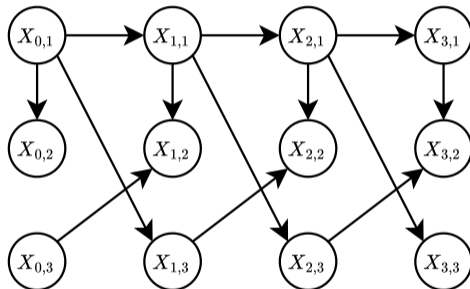


Figure 3: Example temporal DAG with $D = 3$ variables.

A regime-switching SCM

Consider time series $\mathbf{X}_{0:T}$ that is causally stationary only in **discrete segments of time**.

This is commonly modelled using discrete latent **regimes** $R_t \in \mathcal{A}_K = \{1, \dots, K\}$, where $K < \infty$. The set \mathcal{A}_K is a finite subset of \mathcal{A} , e.g. $\mathcal{A} = \mathbb{N}$, indexing all regimes in a model class.

The SCM becomes

$$\begin{aligned}\mathbf{X}_0 &\leftarrow \mathbf{f}^0(\text{Pa}(\mathbf{X}_0, R_0), R_0, \epsilon_0), \\ \mathbf{X}_t &\leftarrow \mathbf{f}(\text{Pa}(\mathbf{X}_t, R_t), R_t, \epsilon_t), \quad t \in \{1, \dots, T\}.\end{aligned}\tag{4}$$

The equations governing regimes $R_{0:T}$ can be anything, although they cannot depend on $\mathbf{X}_{0:T}$.

We often write shorthand $\mathbf{f}_a \triangleq \mathbf{f}|_{R_t=a}$ when we fix the regime to some value $a \in \mathcal{A}$.

A regime-dependent temporal causal graph

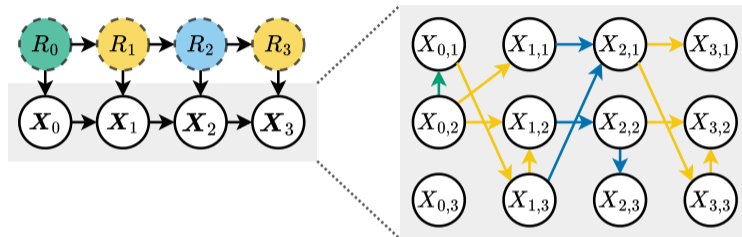


Figure 4: Example DAG given regimes $R_{0:3}$, with coloured edges belonging to **initial** and **window** graphs⁶.

R_t is a *latent confounder* for the variables X_t , inducing regime-dependent causal structures.

The causal graphs need *not* be unique across regimes.

⁶Assaad, Devijver, Gaussier. *Survey and evaluation of causal discovery methods for time series*. Journal of Artificial Intelligence Research, 2022.

An overview of our assumptions

To summarise, we assume

- *Conditional* causal stationarity: Regimes fully govern structural changes over time;
- *Conditional* causal sufficiency: No latent confounders *within each regime*;
- Acyclicity: Structural equations are recursive;
- Causal Markov & faithfulness: Conditional independency in data \iff d -separation in DAG.

For tractable estimation, we assume (i) regimes $\mathbf{R}_{0:T}$ follow a *first-order stationary Markov chain* and (ii) the mappings $\mathbf{f}_a^0, \mathbf{f}_a$ are *contractive* to guarantee stability.

Markov Switching Models (MSMs)

The dynamic Bayesian network induced by a regime-switching SCM is an MSM⁷.

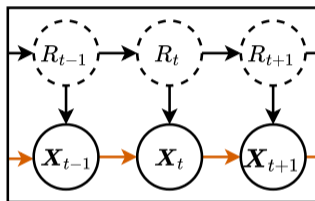


Figure 5: An MSM is an **autoregressive** Hidden Markov Model (HMM).

Classic identifiability results for finite-state HMMs^{8,9,10} do not trivially extend to MSMs.

⁷Hamilton. *A new approach to the economic analysis of nonstationary time series and the business cycle*. *Econometrica*, 1989.

⁸Kruskal. *Three-way arrays: Rank and uniqueness of trilinear decompositions, ...* *Linear Algebra and its Applications*, 1977.

⁹Allman, Matias, Rhodes. *Identifiability of parameters in latent structure models with many observed variables*. *The Annals of Statistics*, 2009.

¹⁰Gassiat, Cleyne, Robin. *Inference in finite state space non parametric hidden Markov models and applications*. *Statistics and Computing*, 2016.

The joint distribution of $\mathbf{x}_{0:T}$ can be written as a **finite mixture** over K^{T+1} regime sequences,

$$p_{\theta}(\mathbf{x}_{0:T}) = \sum_{\mathbf{r}_{0:T} \in \mathcal{A}_K^{\times(T+1)}} \underbrace{p_{\theta}(\mathbf{r}_{0:T})}_{\text{coefficients}} \underbrace{p_{\theta}(\mathbf{x}_{0:T} \mid \mathbf{r}_{0:T})}_{\text{components}}. \quad (5)$$

We factorise **components** into **initial** and **transition** distributions, drawn from families \mathcal{P}_A^0 , \mathcal{P}_A ,

$$p_{\theta}(\mathbf{x}_{0:T} \mid \mathbf{r}_{0:T}) = p_{\theta}(\mathbf{x}_0 \mid r_0) \prod_{t=1}^T p_{\theta}(\mathbf{x}_t \mid \mathbf{x}_{t-1}, r_t). \quad (6)$$

Thus, the K^{T+1} components of the mixture are drawn from the **product family** $\mathcal{P}_A^0 \otimes \mathcal{P}_A^{\otimes T}$.

¹¹Frühwirth-Schnatter. *Finite mixture and Markov switching models*. 2006.

Example SCMs (initial equations left out for brevity)

Let $\mathbf{W}_{a,0}$, $\mathbf{W}_{a,1}$ be weighted adjacency matrices, where $\mathbf{W}_{a,0}$ is acyclic and $\rho(\mathbf{W}_{a,0}), \rho(\mathbf{W}_{a,1}) < 1$.

Linear SVAR

$$\mathbf{X}_t \leftarrow \mathbf{W}_{a,0}\mathbf{X}_t + \mathbf{W}_{a,1}\mathbf{X}_{t-1} + \epsilon_t.$$

Nonlinear ANM¹²

$$\mathbf{X}_t \leftarrow \tanh(\mathbf{W}_{a,0}\mathbf{X}_t + \mathbf{W}_{a,1}\mathbf{X}_{t-1}) + \epsilon_t.$$

LSNM¹³

$$\mathbf{X}_t \leftarrow (1 + \delta - \tanh(\mathbf{W}_{a,0}\mathbf{X}_t)) \circ (1 + \delta - \tanh(\mathbf{W}_{a,0}\mathbf{X}_t + \mathbf{W}_{a,1}\mathbf{X}_{t-1})) \circ \epsilon_t,$$

where \circ denotes the Hadamard product and $\delta > 0$ is a tiny constant.

¹²Hoyer, Janzing, Mooij, Peters, Schölkopf. *Nonlinear causal discovery with additive noise models*. NeurIPS, 2008.

¹³Immer, Schultheiss, Vogt, Schölkopf, Bühlmann, Marx. *On the identifiability and estimation of causal location-scale noise models*. ICML, 2023.

A visualisation of the generated time series

MSMs with (non-)Gaussian noises and (non)linear transitions may be hard to distinguish...

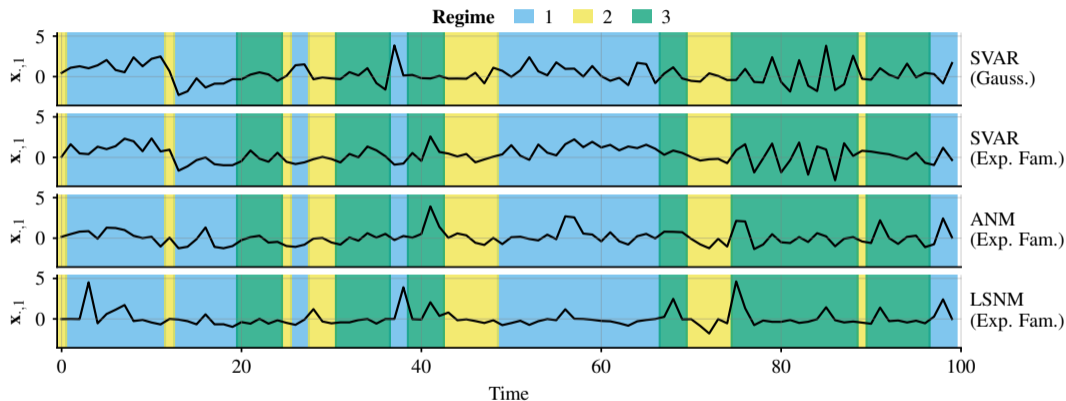


Figure 6: The only differences between the rows amounts to the structural equations and exogenous noises, using (*Gauss*) $\epsilon_{t,d} \stackrel{i.i.d.}{\sim} \mathcal{N}(0,1)$ and (*Exp. Fam.*) $\epsilon_{t,d} \stackrel{i.i.d.}{\sim} \text{Gamma}(0.25,2) - 0.5$.

Only for the linear SVAR with $\epsilon_{t,d} \stackrel{i.i.d.}{\sim} \mathcal{N}(0, 1)$, we obtain a **classic Gaussian mixture model**, e.g.,

$$\begin{aligned} p_{\theta}(\mathbf{x}_0 \mid a) &= \mathcal{N}((\mathbf{I} - \mathbf{W}_{a,0})^{-1} \boldsymbol{\mu}_a, \boldsymbol{\Sigma}_a), \\ p_{\theta}(\mathbf{x}_t \mid \mathbf{x}_{t-1}, a) &= \mathcal{N}((\mathbf{I} - \mathbf{W}_{a,0})^{-1} \mathbf{W}_{a,1} \mathbf{x}_{t-1}, \boldsymbol{\Sigma}_a), \end{aligned} \quad (7)$$

where $\boldsymbol{\Sigma}_a \triangleq (\mathbf{I} - \mathbf{W}_{a,0})^{-1} (\mathbf{I} - \mathbf{W}_{a,0})^{-T}$. Acyclicity guarantees invertibility of $\mathbf{I} - \mathbf{W}_{a,0}$.

Identifiability is guaranteed when the Gaussian parameters are distinct¹⁴.

Problem: If (i) instantaneous effects are not affine, or (ii) the noise is non-Gaussian, we no longer obtain a mixture of Gaussian distributions...

Key question: When are regimes identifiable in such non-Gaussian mixtures?

¹⁴Yakowitz, Spragins. *On the identifiability of finite mixtures*. The Annals of Mathematical Statistics, 1968.

Identifiability theory

Identifiability (up to permutation)

MSMs are **identifiable (up to permutation)** if the data likelihood $p_{\theta}(\mathbf{x}_{0:T})$ *uniquely determines* the regime prior $p_{\theta}(\mathbf{r}_{0:T})$ and mixture components $p_{\theta}(\mathbf{x}_{0:T} \mid \mathbf{r}_{0:T})$, *up to a regime relabelling*.

For brevity, in this talk we leave out “*up to permutation*”.

Thm. 3.5 – Identifiable Regime-Switching SCMs

Consider an acyclic regime-switching SCM that satisfies conditional causal stationarity, conditional causal sufficiency, and **Ass. 3.1** to **3.4**. Then the induced MSM is **identifiable**.

In the remainder, we introduce our assumptions **3.1** to **3.4** and provide a *proof sketch*.

A necessary and sufficient **global condition** for identifiability

Yakowitz and Spragins (1968)¹⁵

A necessary and sufficient condition for identifiability is **linear independence** (under finite mixtures) of the functions in the family of mixture components, *i.e.*, for any finite $\mathcal{A}_K \subset \mathcal{A}$,

$$\sum_{a \in \mathcal{A}_K} \lambda_a p_{\theta}(\mathbf{x} \mid a) = 0 \quad a.e. \quad \implies \quad \lambda_a = 0 \quad \forall a \in \mathcal{A}_K.$$

Thus, identifiability reduces to linear independence of PDFs in the **product family** $\mathcal{P}_{\mathcal{A}}^0 \otimes \mathcal{P}_{\mathcal{A}}^{\otimes T}$.

¹⁵Yakowitz, Spragins. *On the identifiability of finite mixtures*. The Annals of Mathematical Statistics, 1968.

Is linear independence in \mathcal{P}_A^0 and \mathcal{P}_A sufficient?

Short answer: It is **necessary**, but **not sufficient**.

An overlapping variable space challenges linear independence.

For $T = 1$, we have one overlapping variable \mathbf{x}_0 , i.e.,

$$\mathcal{P}_A^0 \otimes \mathcal{P}_A = \{p_{\theta}(\mathbf{x}_0 | r_0)p_{\theta}(\mathbf{x}_1 | \mathbf{x}_0, r_1)\}. \quad (8)$$

Coupling distributions through \mathbf{x}_0 might accidentally create linear dependencies in the joint space.

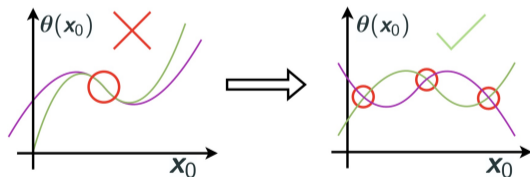


Figure 7: We allow linear dependence only on zero-measure subsets of the overlapping variable space¹⁶. A sufficiently regular space (e.g., *real-analytic*) forces discontinuities to arise solely from regime changes.

¹⁶Balsells-Rodas, Wang, Li. *On the identifiability of switching dynamical systems*. ICML, 2024.

Sufficient **local conditions** for identifiability

So, if the PDFs in the **initial** and **transition** families are **linearly independent**, and the joint variable space is **sufficiently regular**, then we obtain **identifiability up to permutation**.

Our technical contributions answer two questions:

- Can we ensure linear independence in the **initial** and **transition** families beyond Gaussians?
- Linear independence is an abstract notion, can we translate this to more fine-grained conditions on regime-switching SCMs?

For simplicity, we focus on the **transition** family in this talk, omitting the **initial** family.

Identifiable mixtures of exponential family distributions

Recall, the PDF of a **continuous** variable $\epsilon \in \mathcal{X}_\epsilon$ from a **minimal regular exponential family** is

$$p_\eta(\epsilon) = h(\epsilon) \exp(\eta \cdot \tau(\epsilon) - A(\eta)), \quad (9)$$

for P -dimensional natural parameters η , $P \geq 1$, sufficient statistic τ , base measure h and log-partition A . The support $\mathcal{X}_\epsilon \subseteq \mathbb{R}^D$ does not depend on η .

Barndorff-Nielsen (1965)¹⁷

(Finite) mixtures of a **continuous minimal regular exponential family** are **identifiable** if (i) the sufficient statistic is continuous with (ii) an image that contains an open set on its support.

Note mixtures are over the natural parameters η , keeping the sufficient statistic τ and support \mathcal{X}_ϵ fixed across regimes. Finite mixtures of multivariate Gaussian families are a special case.

¹⁷Barndorff-Nielsen. *Identifiability of mixtures of exponential families*. Journal of Mathematical Analysis and Applications, 1965.

We derive conditions on regime-switching SCMs such that we can use Barndorff-Nielsen (1965).

Ass. 3.1 – Exponential Family Noise

The exogenous noise ϵ_t is from a continuous minimal regular exponential family that satisfies:

- (a1) *Real-analytic sufficient statistic*: The sufficient statistic τ is a real-analytic function *a.e.*
- (a2) *Rich image of the sufficient statistic*: The image $\{\tau(\epsilon_t) \mid h(\epsilon_t) > 0, \epsilon_t \in \mathcal{X}_\epsilon\}$ contains a (non-empty) open set.

We strengthen *continuity* of the sufficient statistic to *real-analyticity* to avoid linear dependence in the overlapping variable space.

The condition (a2) excludes certain degenerate and curved exponential families where the image of the sufficient statistic is confined to a lower-dimensional manifold, e.g., when $P > D$.

Each regime represents a unique pushforward of the noise distribution

We aim to establish that each regime represents a **unique pushforward of the noise distribution**.

Ass. 3.2 – Functional Model Restrictions

The mappings f_a satisfy:

(b1) *Unique regimes*: For all $a, a' \in \mathcal{A}$,

$$f_a = f_{a'} \quad a.e. \quad \implies \quad a = a'; \quad (10)$$

(b2) *Pointwise diffeomorphisms*: For almost every $\mathbf{x}_{t-1} \in \mathcal{X}$, and for all $a \in \mathcal{A}$, the mapping $f_a|_{\mathbf{x}_{t-1}=\mathbf{x}_{t-1}}$ is a diffeomorphism *a.e.* in ϵ_t ;

(b3) *Jointly real-analytic transitions*: For all $a \in \mathcal{A}$, the mapping f_a is jointly real-analytic in $(\mathbf{x}_{t-1}, \epsilon_t)$ *a.e.*

Condition (b3), in combination with (a1), ensures the overlapping variable space is real-analytic.

Some ambiguous unidentifiable cases

We rule out **unidentifiable cases** that exploit symmetries, e.g., rotations of isotropic Gaussians.

Ass 3.3 – Trivial Automorphisms

At least one of the following holds:

(c1) *Trivial automorphisms of the noise*: The noise distribution has a trivial automorphism class, i.e., for any invertible mapping Φ ,

$$\Phi(\epsilon) \stackrel{d}{=} \epsilon \implies \Phi = \epsilon \text{ a.s.}; \quad (11)$$

(c2) *Monotone canonicalisation*: There exists a canonical variable order s.t. $\forall a \in \mathcal{A}$, $\partial \mathbf{f}_a / \partial \epsilon_t$ can be permuted to a lower-triangular matrix, and it has strictly positive diagonal a.e. in ϵ_t .

The stronger condition (c2) guarantees a *unique* monotone triangular transport known as the Knothe–Rosenblatt rearrangement^{18,19}.

¹⁸Knothe. *Contributions to the theory of convex bodies*. Michigan Mathematical Journal, 1957.

¹⁹Rosenblatt. *Remarks on a multivariate transformation*. The Annals of Mathematical Statistics, 1952.

The exponential family structure is preserved

The structural equations define mappings $\Phi_a | \mathbf{x}_{t-1} = \mathbf{x}_{t-1} : \mathcal{X}_\epsilon \rightarrow \mathcal{X}$ from the noise space to the observed space at each time step, for a fixed regime $a \in \mathcal{A}$.

Under **Ass. 3.1-3.3**, using the change of variables formula and the exponential family structure,

$$p_\theta(\mathbf{x}_t | \mathbf{x}_{t-1}, a) = h(\Phi_a^{-1}(\mathbf{x}_t, \mathbf{x}_{t-1})) \cdot |\det J_{\Phi_a^{-1}(\cdot, \mathbf{x}_{t-1})}(\mathbf{x}_t)| \cdot \exp\left(\boldsymbol{\eta} \cdot \boldsymbol{\tau}(\Phi_a^{-1}(\mathbf{x}_t, \mathbf{x}_{t-1})) - A(\boldsymbol{\eta})\right),$$

where $\Phi_a^{-1}(\cdot, \mathbf{x}_{t-1})$ denotes the inverse of $\Phi_a(\mathbf{x}_{t-1}, \cdot)$ with respect to ϵ_t for fixed \mathbf{x}_{t-1} .

This is still an exponential family, but each transition distribution might have a different sufficient statistic and base measure...

How can we still leverage Barndorff-Nielsen (1965)?

A reparametrisation to an exponential family with a shared sufficient statistic

Ass. 3.4 – Sufficient variability across regimes in a finite polynomial Subspace

The mappings Φ_a satisfy for some finite $O < \infty$ and all $a \in \mathcal{A}$ and almost every $\mathbf{x}_{t-1} \in \mathcal{X}$:

- (d1) *Common support*: The base measure $h \circ \Phi_a^{-1}$ can be separated into a common base measure \tilde{h} and a scaling function $b_a(\mathbf{x}_{t-1}) > 0$;
- (d2) *Finite polynomial reparametrisation*: There exists a common sufficient statistic $\tilde{\tau} : \mathcal{X} \rightarrow \mathbb{R}^{\tilde{P}}$, with $\tilde{P} \geq P$, whose components are monomials up to order O , and matrices $\mathbf{C}_a(\mathbf{x}_{t-1}) \in \mathbb{R}^{P \times \tilde{P}}$ with full row rank P , such that

$$\tau \circ \Phi_a^{-1}(\mathbf{x}_t, \mathbf{x}_{t-1}) = \mathbf{C}_a(\mathbf{x}_{t-1})\tilde{\tau}(\mathbf{x}_t) + \mathcal{R}_O(\mathbf{x}_t) \quad a.e.,$$

where the remainder is regime-invariant and can be absorbed into the base measure;

- (d3) *Injectivity of polynomial coefficients*: For all $a \neq a' \in \mathcal{A}$,

$$\mathbf{C}_a(\mathbf{x}_{t-1})^T \boldsymbol{\eta} = \mathbf{C}_{a'}(\mathbf{x}_{t-1})^T \boldsymbol{\eta} \quad \forall \boldsymbol{\eta} \in \mathbb{R}^P \quad \implies \quad \Phi_a = \Phi_{a'} \quad a.e..$$

A reparametrisation to an exponential family with a shared sufficient statistic

Ass. 3.4 guarantees an **equivalent representation** as a continuous minimal regular exponential family with distinct natural parameters and common sufficient statistic.

All regime- and history-dependent effects can be absorbed into parameters $\theta : \mathcal{X} \times \mathcal{A} \rightarrow \Theta$, i.e.,

$$p_{\theta}(\mathbf{x}_t | \mathbf{x}_{t-1}, a) = \underbrace{\tilde{h}(\mathbf{x}_t) \exp(\mathcal{R}_O(\mathbf{x}_t))}_{\bar{h}(\mathbf{x}_t)} \exp\left(\underbrace{(\mathbf{C}_a^T(\mathbf{x}_{t-1})\boldsymbol{\eta})}_{\theta_a(\mathbf{x}_{t-1})} \cdot \tilde{\tau}(\mathbf{x}_t) - \underbrace{\log b_a(\mathbf{x}_{t-1}) - \tilde{A}(\mathbf{C}_a^T(\mathbf{x}_{t-1})\boldsymbol{\eta})}_{\bar{A}_a(\theta_a(\mathbf{x}_{t-1}), \mathbf{x}_{t-1})}\right),$$

Thus, the **transition** distribution remains a minimal regular exponential family.

The same argument holds for the **initial** family.

We established finite mixtures of the **initial** and **transition** distribution families are identifiable²⁰, so the families contain **linearly independent** functions²¹.

Throughout, we made sure the overlapping variable space is **sufficiently regular**, *i.e.*, real-analytic.

Thm. 3.5 – Identifiable Regime-Switching SCMs

Consider an acyclic regime-switching SCM that satisfies conditional causal stationarity, conditional causal sufficiency, and **Ass. 3.1** to **3.4**. Then the induced MSM is **identifiable**.

²⁰Barndorff-Nielsen. *Identifiability of mixtures of exponential families*. Journal of Mathematical Analysis and Applications, 1965.

²¹Yakowitz, Spragins. *On the identifiability of finite mixtures*. The Annals of Mathematical Statistics, 1968.

Identifiable initial and window graphs

Upon identification of latent regimes, **known causal theory applies** to identify stationary causal graphs from the disentangled transition distributions.

For example, each window graph is identifiable up to a **Markov equivalence class (MEC)** of conditional independencies under *faithfulness*²².

Alternatively, the transition distributions may correspond to a single window graph when the function class of $\mathbf{f}_a^0, \mathbf{f}_a$ is restricted²³.

²²Spirtes, Glymour, Scheines. *Causation, Prediction, and Search*. 2000.

²³Peters, Mooij, Janzing, Schölkopf. *Identifiability of causal graphs using functional models*. UAI, 2011.

Method

Regime detection and causal discovery

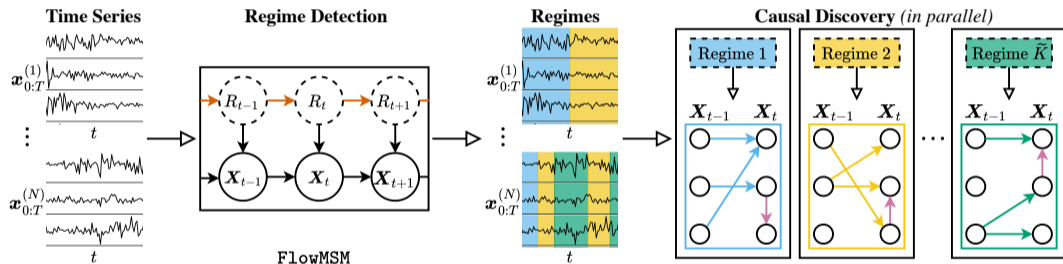


Figure 8: FlowMSM detects *latent regimes* $\mathbf{R}_{0:T}$, while a stationary causal method subsequently discovers *window graphs* $\mathbf{G}_{1:K}$ from time series $\mathbf{X}_{0:T}$.

Regime detection – FlowMSM

Input: Dataset $\{\mathbf{x}_{0:T}^{(n)}\}_{n=1}^N$ of N *i.i.d.* realisations of the time series $\mathbf{X}_{0:T}$.

Output: Estimates $\hat{p}_\theta(r_t | \mathbf{x}_{0:T}^{(n)})$ of the regime posterior likelihood.

We use **Generalised Expectation-Maximisation** (GEM), where a **neural spline flow** (NSF)²⁴ models the transition distributions, with shared parameters ψ and regime-specific embeddings $\mathcal{E}_{1:\tilde{K}}$.

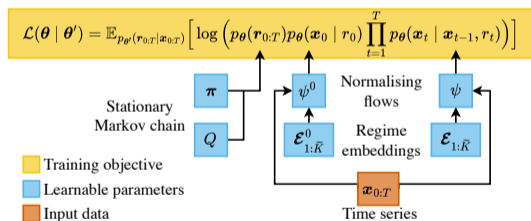


Figure 9: GEM updates $\theta^{\text{new}} \leftarrow \theta' + \alpha \nabla_{\theta} \mathcal{L}(\theta | \theta')|_{\theta=\theta'}$, guaranteed to converge to local optimum²⁵.

²⁴Durkan, Bekasov, Murray, Papamakarios. *Neural spline flows*. NeurIPS, 2019.

²⁵Dempster, Laird, Rubin. *Maximum likelihood from incomplete data via the EM algorithm*. Journal of the Royal Statistical Society, 1977.

Sample splitting scheme

We assign the sequences $\mathbf{x}_{t-1:t}^{(n)}$, $t \in \{1, \dots, T\}$, $n \in \{1, \dots, N\}$, to the MAP regime

$$\hat{r}_t^{(n)} = \arg \max_{r_t \in \mathcal{A}_{\tilde{K}}} \hat{p}_{\theta}(r_t \mid \mathbf{x}_{0:T}^{(n)}). \quad (12)$$

This creates \tilde{K} clusters with partially overlapping sliding windows.

In theory, a perfect regime assignment would lead to clusters of causally stationary windows, since the causal parents of variables $\mathbf{x}_t^{(n)}$ depend solely on the oracle regime $r_t^{(n)}$.

In practice, we show FlowMSM exhibits high accuracy in most settings.

To estimate window graphs from the clustered samples, FlowMSM can be paired with any stationary causal discovery method that allows for instantaneous effects.

- VARLiNGAM²⁶;
- DYNOTEARS²⁷;
- PCMCI⁺²⁸;
- Rhino²⁹.

²⁶Hyvärinen, Zhang, Shimizu, Hoyer. *Estimation of a structural vector autoregression model using non-gaussianity*. JMLR, 2010.

²⁷Pamfil, Sriwattanaworachai, Desai, ..., Aragam. *DYNOTEARS: Structure learning from time-series data*. AISTATS, 2020.

²⁸Runge. *Discovering contemporaneous and lagged causal relations in autocorrelated nonlinear time series datasets*. UAI, 2020.

²⁹Gong, Jennings, Zhang, Pawlowski. *Rhino: Deep causal temporal relationship learning with history-dependent noise*. ICLR, 2023.

Experiments

Synthetic SCMs (initial equations left out for brevity)

Let $\mathbf{W}_{a,0:1}$ be weighted adjacency matrices, where $\mathbf{W}_{a,0}$ is acyclic and $\rho(\mathbf{W}_{a,0}) < 1, \rho(\mathbf{W}_{a,1}) < 1$.

Linear SVAR

$$\mathbf{X}_t \leftarrow \mathbf{W}_{a,0}\mathbf{X}_t + \mathbf{W}_{a,1}\mathbf{X}_{t-1} + \epsilon_t.$$

Nonlinear ANM³⁰

$$\mathbf{X}_t \leftarrow \tanh(\mathbf{W}_{a,0}\mathbf{X}_t + \mathbf{W}_{a,1}\mathbf{X}_{t-1}) + \epsilon_t.$$

LSNM³¹

$$\mathbf{X}_t \leftarrow (1 + \delta - \tanh(\mathbf{W}_{a,0}\mathbf{X}_t)) \circ (1 + \delta - \tanh(\mathbf{W}_{a,0}\mathbf{X}_t + \mathbf{W}_{a,1}\mathbf{X}_{t-1})) \circ \epsilon_t,$$

where \circ denotes the Hadamard product and $\delta > 0$ is a tiny constant.

³⁰Hoyer, Janzing, Mooij, Peters, Schölkopf. *Nonlinear causal discovery with additive noise models*. NeurIPS, 2008.

³¹Immer, Schultheiss, Vogt, Schölkopf, Bühlmann, Marx. *On the identifiability and estimation of causal location-scale noise models*. ICML, 2023.

Synthetic experiments

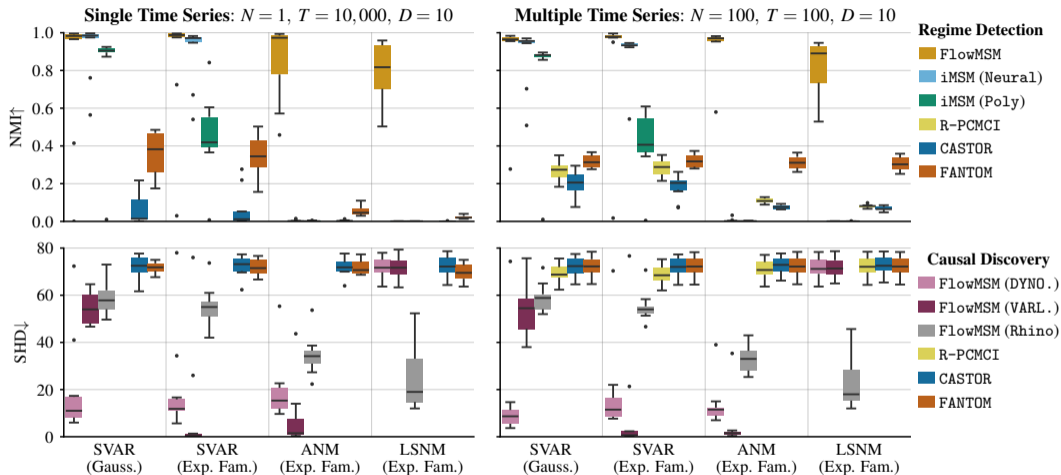


Figure 10: Performance on (*top*) regime detection and (*bottom*) causal discovery on (*left*) a single long time series and (*right*) multiple smaller time series, with $K = 3$ regimes. Each boxplot covers 10 seeds.

Fama-French five-factor asset-pricing model – Regime detection

Five risk factors that capture systemic patterns in stock returns: size, value, market risk, profitability and investment³². This is supplemented by excess returns of Apple's stock (AAPL)³³.

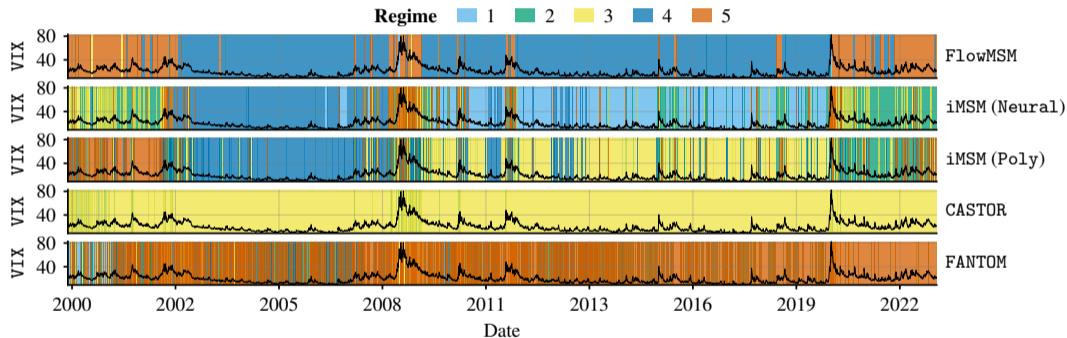


Figure 11: Estimated regimes on daily data, overlaid by the VIX volatility index (not used in training).

³²Fama, French. *A five-factor asset pricing model*. Journal of Financial Economics, 2015

³³Sadeghi, Gopal, Fesanghary. *Causal discovery from nonstationary time series*. International Journal of Data Science and Analytics, 2024.

Fama-French five-factor asset-pricing model – Causal discovery

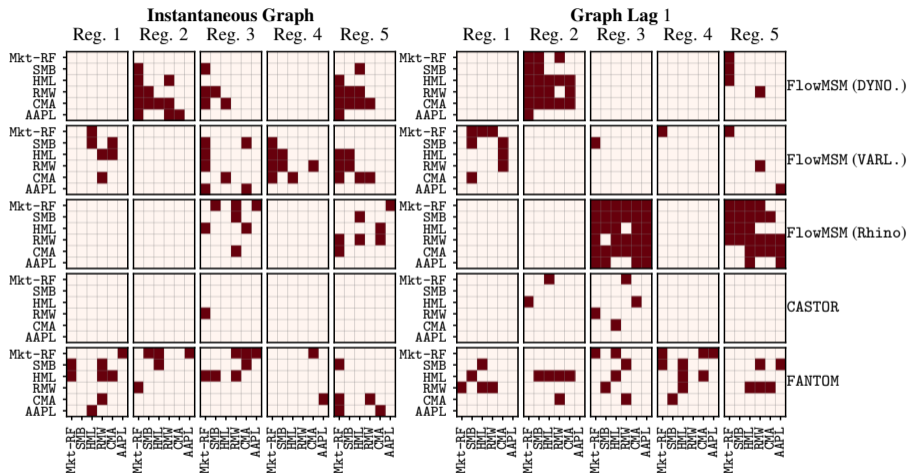


Figure 12: Adjacency matrices of the window graphs corresponding to the estimated regimes. The index (i, j, k, l) indicates the (lagged) edge $X_{t-l,i} \rightarrow X_{t,j}$ in regime k .

Conclusion

We explored regime detection and causal discovery in non-stationary time series.

We proposed novel theory on identifiability of latent regimes.

Our regime detection framework, FlowMSM, readily extends to regime-dependent causal discovery.

Future work:

- We require the regime variables $\mathbf{R}_{0:T}$ to be independent of the observations $\mathbf{X}_{0:T}$. Recent work partly relaxes this in Gaussian settings³⁴, but extending beyond is nontrivial;
- Our identifiability theory critically relies on exponential family noise and real-analytic diffeomorphisms. How to preserve linear independence in more relaxed settings?;
- Non-stationary time series might involve latent confounders other than regime variables.

³⁴Balsells-Rodas, Sumba, Narendra, Tu, Schweikert, Kjellstrom, Li. *Causal discovery from conditionally stationary time series*. ICML, 2025.