UNIVERSITY OF COPENHAGEN
DEPARTMENT OF MATHEMATICAL SCIENCES

Invariant Causal Prediction

Causal inference in the presence of latent variables

Rune Christiansen and Jonas Peters

Department of Mathematical Sciences, University of Copenhagen

Contact Information:
krunechristiansen@math.ku.dk

Keywords

causal inference invariance latent variables regression analysis hidden Markov models

Causal discovery and latent variables

Goal: Given a target Y, find the (observable) causal predictors among $X^1, \dots, X^d$.

X^1	X^2	X^3	X^4	Y
3.4	−0.3	5.8	−2.1	2.2
1.7	−0.2	7.0	−1.2	0.4
−2.4	−0.1	4.3	−0.7	3.5
3.5	−0.2	3.9	−0.9	−3.9
⋮	⋮	⋮	⋮	⋮

No latent variables

Infer set $\text{PA}(Y) = \{X^1, X^3\}$ of causal parents

Latent variables

Infer set $\text{PA}^0(Y) = \{X^3\}$ of *observable* causal parents

Problem: In general, inferring $\text{PA}^0(Y)$ from $P_{(Y,X^2,X^3,X^4)}$ is hard.

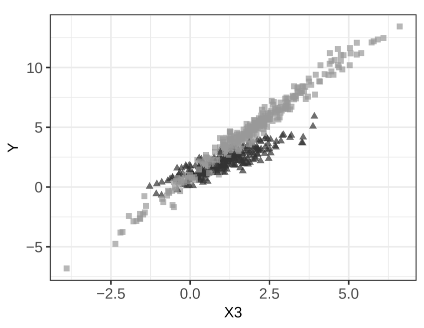
Approach: If H takes only few discrete values (here $H \in \{0,1\}$), we can exploit invariances:

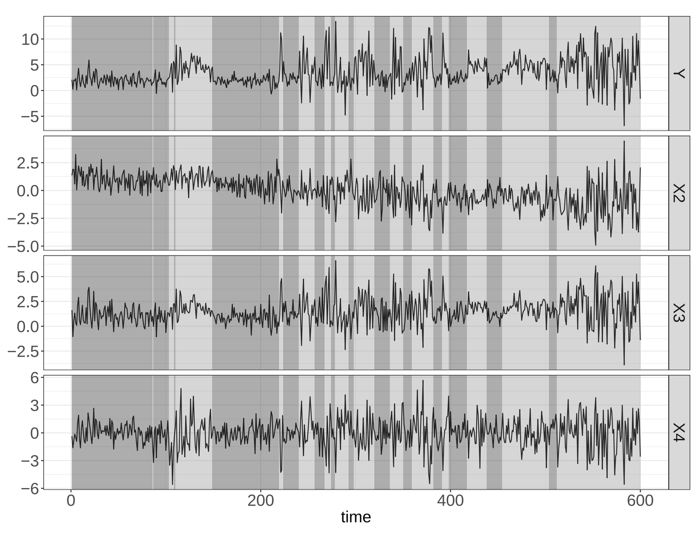
Invariant Causal Prediction

Fundamental assumption: $(X^3, H) \to Y$ is time-homogeneous. Then, for all t, x,

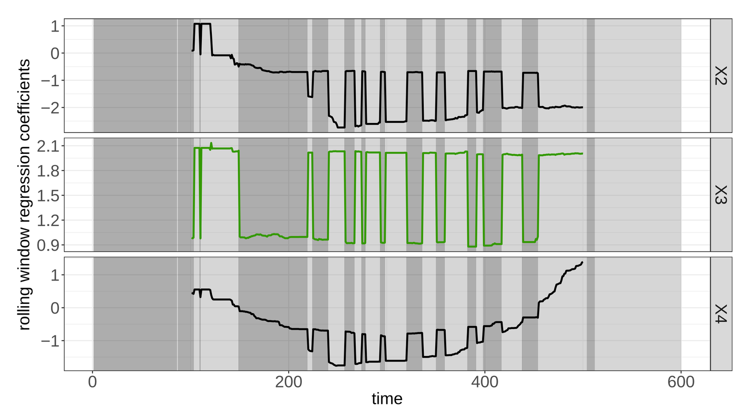
$$(1) \quad P_{Y_t \mid (X_t^3=x)} \overset{d}{=} \lambda_{xt} \underbrace{P_{Y_t \mid (X_t^3=x, H_t=0)}}_{\text{independent of } t} + (1-\lambda_{xt}) \underbrace{P_{Y_t \mid (X_t^3=x, H_t=1)}}_{\text{independent of } t}$$

■ ▲





Thanks to heterogeneity, only X^3 satisfies (1):



Key idea:

- Use sequential ordering to construct environments ▭, ▭ and ▭.
- For every subset $S \subseteq \{X^2, X^3, X^4\}$, test $H_{0,S}$: { ■, ▲ } = { ■, ▲ } = { ■, ▲ }.
- Output the estimator $\hat{S} := \bigcap_{S: H_{0,S} \text{ not rej.}} S$.

J. Peters, P. Bühlmann, and N. Meinshausen (2016). Causal inference using invariant prediction: identification and confidence intervals. *Journal of the Royal Statistical Society, Series B (with discussion)* , 78(5):947–1012.

Theorem: Assume S^* is the correct set of observable direct causes. For any test level α we obtain

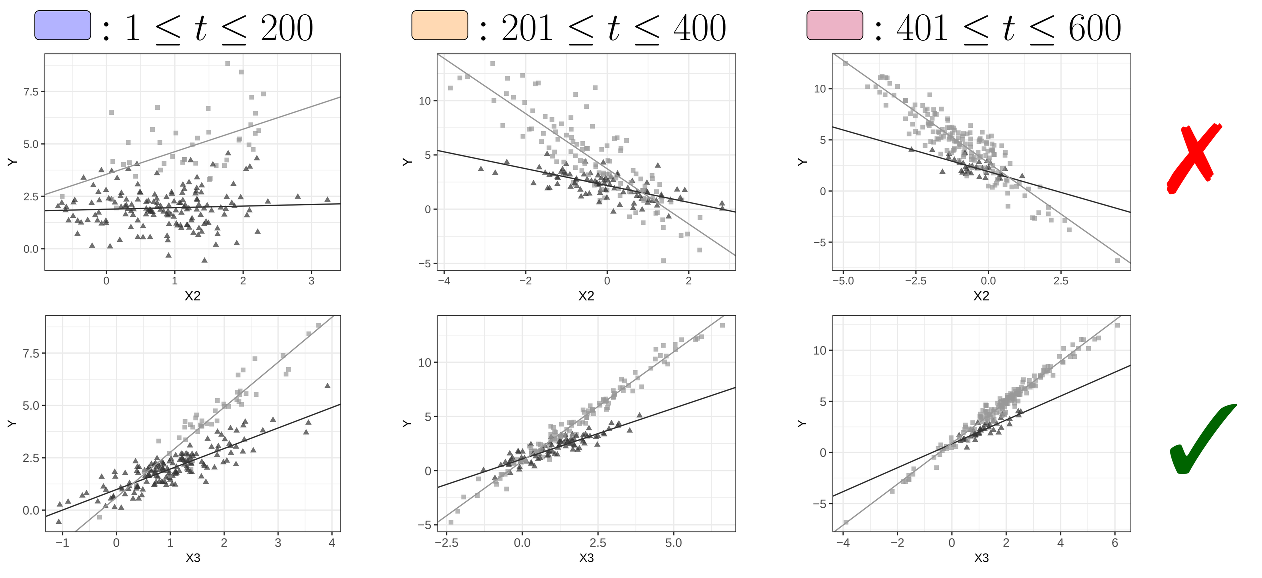
$$\lim_{n\to\infty} \mathbb{P}(\hat{S}_n \subseteq S^*) \geq 1 - \alpha.$$

Example:

▭ : $1 \leq t \leq 200$ ▭ : $201 \leq t \leq 400$ ▭ : $401 \leq t \leq 600$

$\mathbf{X^2}$:

$\mathbf{X^3}$:



S	$\emptyset$	$\{2\}$	$\{3\}$	$\{4\}$	$\{2,3\}$	$\{2,4\}$	$\{3,4\}$	$\{2,3,4\}$
$H_{0,S}$	✗	✗	✓	✗	✗	✗	✓	✗

$\Rightarrow \quad \hat{S} = \{3\}$

Asymptotically valid hypothesis tests

The false discovery control of our method relies on the asymptotic validity of the tests $(\varphi_S)_{S\subseteq\{1,\dots,d\}}$ of the hypotheses $(H_{0,S})_{S\subseteq\{1,\dots,d\}}$. We assume the true model is: for all t, x,

$$(2) \qquad P_{Y_t \mid (X_t^{S^*}=x)} = \sum_{j=1}^{\ell} \lambda_{xt}^j \mathcal{N}(x^T\beta_j, \sigma_j^2), \quad \theta = (\beta_1, \sigma_1^2, \dots, \beta_\ell, \sigma_\ell^2).$$

For every $S \subseteq \{1, \dots, d\}$, we construct φ_S by:

- For every environment $e \in \mathcal{E}$, fit model (2) to obtain an MLE $\hat{\theta}_e$ and a Fisher confidence region $C^\alpha(\hat{\theta}_e)$.
- Adjust for label permutations, i.e., construct $C^\alpha_{\text{adjusted}}(\hat{\theta}_e) := \bigcup_{\pi\in\Pi} C^\alpha(\pi(\hat{\theta}_e))$.
- Reject if $\bigcap_{e\in\mathcal{E}} C^{\alpha/|\mathcal{E}|}_{\text{adjusted}}(\hat{\theta}_e) = \emptyset$.

Under regularity conditions, these tests are asymptotically valid for any test level α:

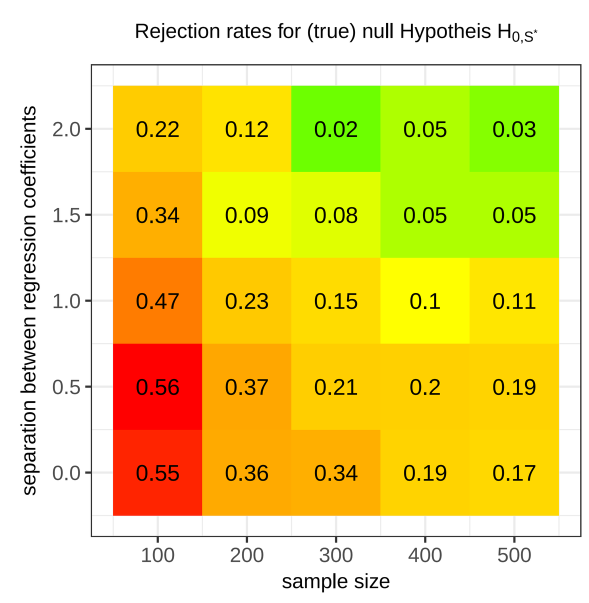
Lemma 1: The MLE is consistent up to label permutations, i.e., $\pi_n(\hat{\theta}_n) \longrightarrow \theta^0$ as $n \to \infty$

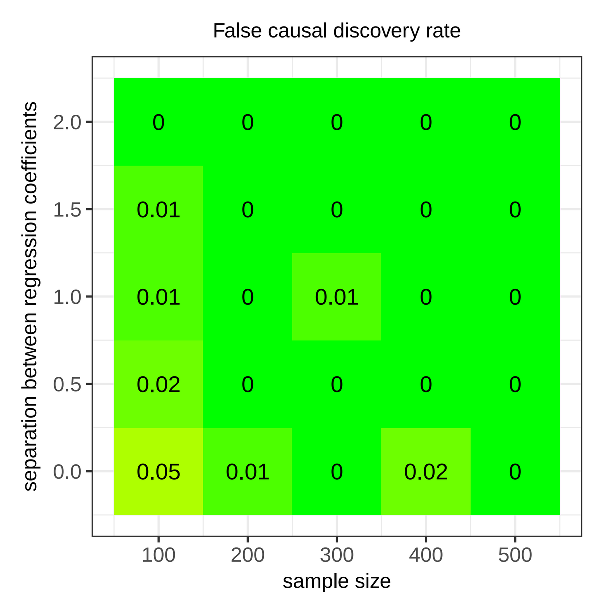
Lemma 2: The MLE is asymptotically normal up to label permuations, i.e., $\sqrt{n}(\pi_n(\hat{\theta}_n) - \theta^0) \xrightarrow{d} \mathcal{N}(0, \mathcal{I}_0)$ as $n \to \infty$

	the likelihood function		
problem:	is non-concave	is unbounded	is based on finite data
solution:	EM-algorithm / non-linear maximization	parameter constraints, e.g., $\min_j \sigma_j^2 \geq c$ or $\sigma_j^2 = \sigma_k^2$	see next section

Robustness for small sample sizes

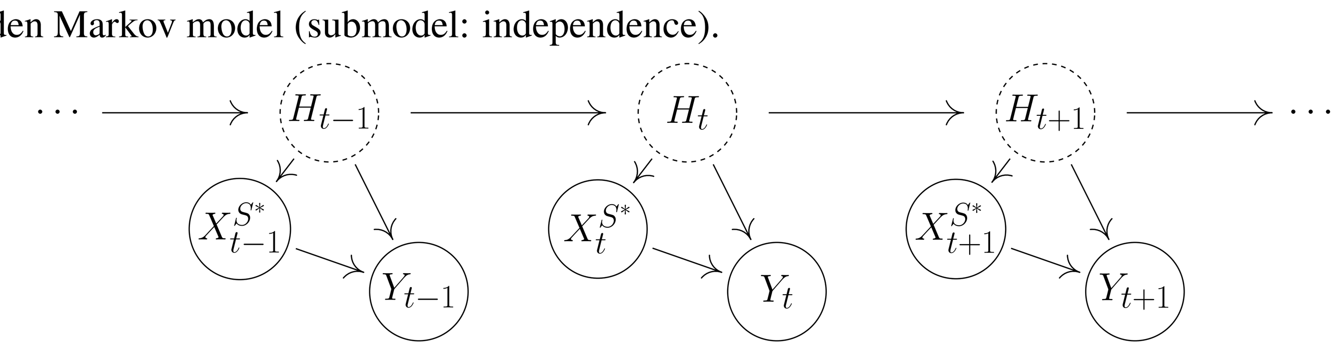
Even for small sample sizes, the overall method keeps the type I error control.





Time dependence of latent variables

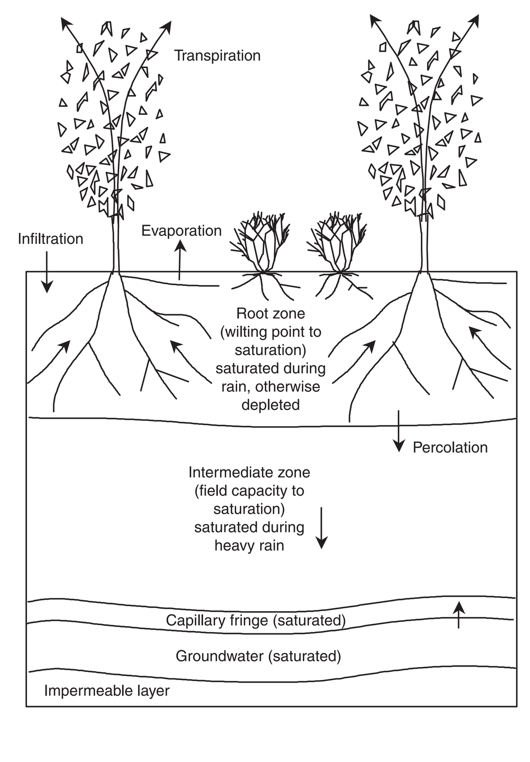
A hidden Markov model (submodel: independence).



Application in progress

A biogeochemical system:

Y: evapotranspiration, X: radiation, water pressure deficit, wind speed etc., H: soil moisture



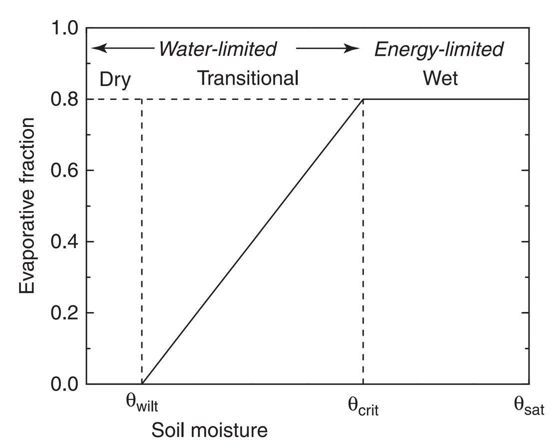


Fig. 14.6 Idealized relationship between soil water and evapotranspiration. Evapotranspiration, as a fraction of available energy, increases with wetter soil up to a maximum value at some critical water content (θ_{cr}), beyond which evapotranspiration is energy-limited. Below some minimum value (θ_{wilt}), the soil is too dry to sustain evapotranspiration. Adapted from Seneviratne et al. (2010).

G. Bonan (2015). Ecological Climatology: Concepts and Applications. *Cambridge University Press*.

Acknowledgments

This work was supported by research grant no. 18968 from VILLUM FONDEN.