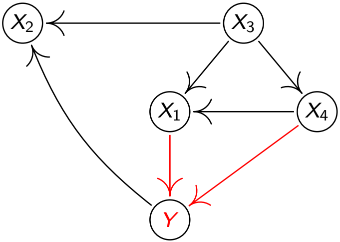
Two Ideas

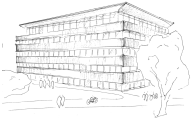


Jonas Peters, University of Copenhagen

TUM-CPH-ETH Workshop

10 October 2022

CARLSBERGFONDET
VILLUM FONDEN



Members of the Copenhagen Causality Lab:


Jeff Adams

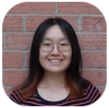

Alexander Mangulad Christgau

★ Christgau et al. (2022) 🚀


Nicola Gnecco


Niels Richard Hansen


Leonard Henckel


Shimeng Huang


Snorre Jallbjørn


Steffen L. Lauritzen


Myrto Limnios


Phillip Bredahl Mogensen

★ Mogensen & Markussen (2021) 🚀


Jonas Peters

★ Thams et al. (2021) 🚀


Niklas Pfister

★ Saengkyongam et al. (2021) 🚀


Sorawit Saengkyongam

★ Saengkyongam et al. (2021) 🚀


Nikolaj Theodor Thams

★ Thams* et al. (2022) 🚀


Sebastian Weichwald

★ Reisach et al. (2021) 🚀

Observed data: $\mathbf{X}_n$.
Observed distribution: $q \in \mathcal{Q}$.

Observed data: $\mathbf{X}_n$.
Observed distribution: $q \in \mathcal{Q}$.
Interested in: $p \in \mathcal{P}$ and $H_0^{\mathcal{P}} \subset \mathcal{P}$.

Observed data: $\mathbf{X}_n$.
Observed distribution: $q \in \mathcal{Q}$.
Interested in: $p \in \mathcal{P}$ and $H_0^{\mathcal{P}} \subset \mathcal{P}$.
Relationship: $p = \tau(q)$ (τ is known).

Observed data: $\mathbf{X}_n$.
Observed distribution: $q \in \mathcal{Q}$.
Interested in: $p \in \mathcal{P}$ and $H_0^{\mathcal{P}} \subset \mathcal{P}$.
Relationship: $p = \tau(q)$ (τ is known).

$$\exists A \subseteq \{1, \ldots, d\} : \qquad p(x^1, \ldots, x^d) \propto r(x^A) \cdot q(x^1, \ldots, x^d).$$

Observed data: $\mathbf{X}_n$.
Observed distribution: $q \in \mathcal{Q}$.
Interested in: $p \in \mathcal{P}$ and $H_0^{\mathcal{P}} \subset \mathcal{P}$.
Relationship: $p = \tau(q)$ (τ is known).

$$\exists A \subseteq \{1, \ldots, d\} : \qquad p(x^1, \ldots, x^d) \propto r(x^A) \cdot q(x^1, \ldots, x^d).$$

$$\text{E.g.:} \quad p(y, a, x^1, x^2, u) = \underbrace{\frac{p^*(a|x^2)}{q^*(a|x^1, x^2)}}_{=: r(a, x^1, x^2) \text{ (known)}} \quad q(y, a, x^1, x^2, u)$$

Observed data: $\mathbf{X}_n$.
Observed distribution: $q \in \mathcal{Q}$.
Interested in: $p \in \mathcal{P}$ and $H_0^{\mathcal{P}} \subset \mathcal{P}$.
Relationship: $p = \tau(q)$ (τ is known).

$$\exists A \subseteq \{1, \ldots, d\} : \qquad p(x^1, \ldots, x^d) \propto r(x^A) \cdot q(x^1, \ldots, x^d).$$

$$\text{E.g.: } \; p(y, a, x^1, x^2, u) = \underbrace{\frac{p^*(a|x^2)}{q^*(a|x^1, x^2)}}_{=: r(a, x^1, x^2) \text{ (known)}} \; q(y, a, x^1, x^2, u)$$

Hypothesis of interest: $\tau(q) \in H_0^{\mathcal{P}}$

Observed data: $\mathbf{X}_n$.
Observed distribution: $q \in \mathcal{Q}$.
Interested in: $p \in \mathcal{P}$ and $H_0^{\mathcal{P}} \subset \mathcal{P}$.
Relationship: $p = \tau(q)$ (τ is known).

$$\exists A \subseteq \{1, \ldots, d\} : \qquad p(x^1, \ldots, x^d) \propto r(x^A) \cdot q(x^1, \ldots, x^d).$$

$$\text{E.g.: } \; p(y, a, x^1, x^2, u) = \underbrace{\frac{p^*(a|x^2)}{q^*(a|x^1, x^2)}}_{=: r(a, x^1, x^2) \text{ (known)}} \; q(y, a, x^1, x^2, u)$$

Hypothesis of interest: $\tau(q) \in H_0^{\mathcal{P}}$
Wanted: test ψ_n with pointwise as. level:

Observed data: $\mathbf{X}_n$.
Observed distribution: $q \in \mathcal{Q}$.
Interested in: $p \in \mathcal{P}$ and $H_0^{\mathcal{P}} \subset \mathcal{P}$.
Relationship: $p = \tau(q)$ (τ is known).

$$\exists A \subseteq \{1, \dots, d\}: \qquad p(x^1, \dots, x^d) \propto r(x^A) \cdot q(x^1, \dots, x^d).$$

$$\text{E.g.:} \quad p(y, a, x^1, x^2, u) = \underbrace{\frac{p^*(a|x^2)}{q^*(a|x^1, x^2)}}_{=:r(a,x^1,x^2)\text{ (known)}} \quad q(y, a, x^1, x^2, u)$$

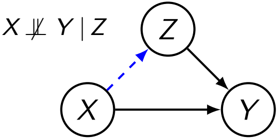
Hypothesis of interest: $\tau(q) \in H_0^{\mathcal{P}}$
Wanted: test ψ_n with pointwise as. level:
$\sup_{Q\,:\,\tau(Q) \in H_0^{\mathcal{P}}} \limsup_{n\to\infty} \mathbb{P}_Q(\psi_n(\mathbf{X}_n) = 1) \le \alpha.$

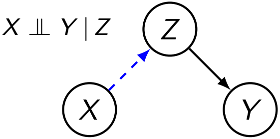
Example: Conditional Independence Testing

$$X \perp\!\!\!\perp Y \mid Z \quad + \quad \text{known conditional } q^*(z|x) \qquad \overset{distr.\ shift}{\longleftrightarrow} \qquad X \perp\!\!\!\perp Y.$$

Example: Conditional Independence Testing

$$X \perp\!\!\!\perp Y \mid Z \quad + \quad \text{known conditional } q^*(z|x) \quad \overset{distr.\ shift}{\longleftrightarrow} \quad X \perp\!\!\!\perp Y.$$

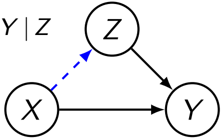




Example: Conditional Independence Testing

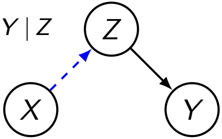
$X \perp\!\!\!\perp Y \mid Z \quad + \quad$ known conditional $q^*(z|x)$ $\quad \xleftrightarrow{\textit{distr. shift}} \quad X \perp\!\!\!\perp Y.$

$X \not\!\perp\!\!\!\perp Y \mid Z$



$X \not\!\perp\!\!\!\perp Y \quad$ in $\tau(q)$

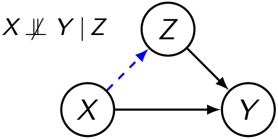
$X \perp\!\!\!\perp Y \mid Z$

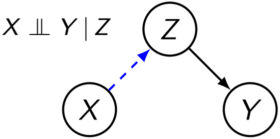


$X \perp\!\!\!\perp Y \quad$ in $\tau(q)$,

Example: Conditional Independence Testing

$$X \perp\!\!\!\perp Y \mid Z \quad + \quad \text{known conditional } q^*(z|x) \qquad \overset{distr.\ shift}{\longleftrightarrow} \qquad X \perp\!\!\!\perp Y.$$





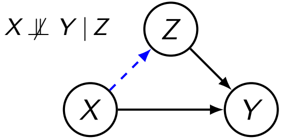
$X \not\!\perp\!\!\!\perp Y$ in $\tau(q)$ $X \perp\!\!\!\perp Y$ in $\tau(q)$,

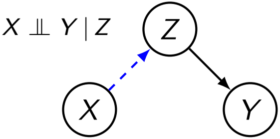
where

$$\tau(q)(x,y,z) := \frac{\phi(z)}{\underbrace{q^*(z|x)}_{=:r(z,x)\ \text{(known)}}} \cdot q(x,y,z) \qquad \text{for all } (x,y,z) \in \mathcal{Z}.$$

Example: Conditional Independence Testing

$$X \perp\!\!\!\perp Y \mid Z \quad + \quad \text{known conditional } q^*(z|x) \qquad \overset{distr.\ shift}{\longleftrightarrow} \qquad X \perp\!\!\!\perp Y.$$





$$X \not\perp\!\!\!\perp Y \quad \text{in } \tau(q) \qquad\qquad X \perp\!\!\!\perp Y \quad \text{in } \tau(q),$$

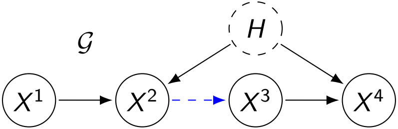
where

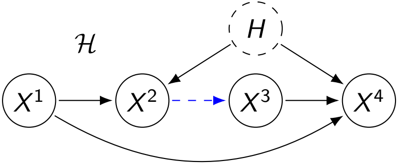
$$\tau(q)(x,y,z) := \underbrace{\frac{\phi(z)}{q^*(z|x)}}_{=:r(z,x)\text{ (known)}} \cdot q(x,y,z) \qquad \text{for all } (x,y,z) \in \mathcal{Z}.$$

Thus: test $X \perp\!\!\!\perp Y$ in $\tau(q)$.

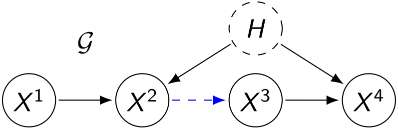
See also Candes et al. 2018, Berrett et al. 2020, Shah & Peters 2020.

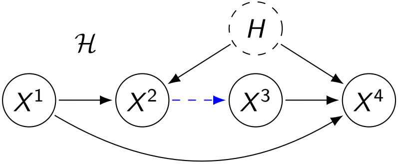
Example: Dormant independences





Example: Dormant independences



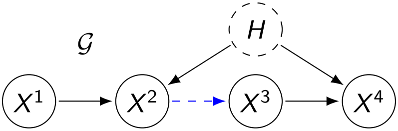


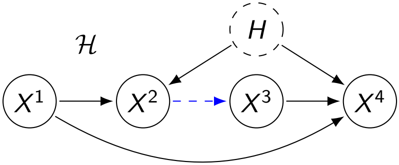
$X^1 \perp\!\!\!\perp X^4$ in $Q^{do(X^3:=N)}$

$X^1 \not\perp\!\!\!\perp X^4$ in $Q^{do(X^3:=N)}$.

Verma and Pearl 1991, Shpitser et al. 2012, Drton et al. 2009, Nowzohour et al. 2019

Example: Dormant independences





$X^1 \perp\!\!\!\perp X^4 \quad \text{in } Q^{do(X^3:=N)}$ $\qquad$ $X^1 \not\!\perp\!\!\!\perp X^4 \quad \text{in } Q^{do(X^3:=N)}.$

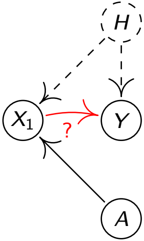
Verma and Pearl 1991, Shpitser et al. 2012, Drton et al. 2009, Nowzohour et al. 2019

Method in a nutshell:

1. Define weights $r(X_i) := p(X_i)/q(X_i)$.
2. Resample.
3. Apply existing test on target domain.

Thams, Saengkyongam, Pfister and Peters, arXiv 2021; see also SIR (Rubin, 1987; Smith and Gelfand, 1992)

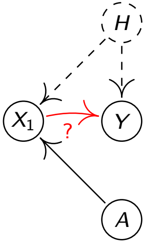
Instrumental Variables:

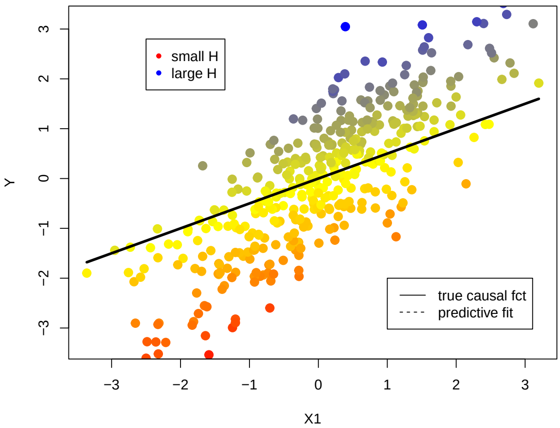


True model:

$$Y = \alpha^* X_1 + \gamma H + N_Y.$$

Instrumental Variables:

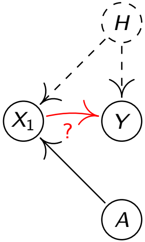




True model:

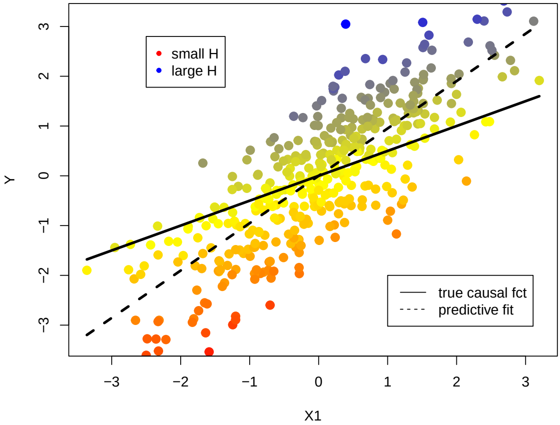
$$Y = \alpha^* X_1 + \gamma H + N_Y.$$

Instrumental Variables:

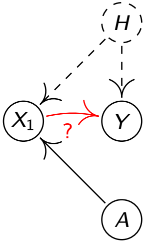


True model:

$$Y = \alpha^* X_1 + \gamma H + N_Y.$$

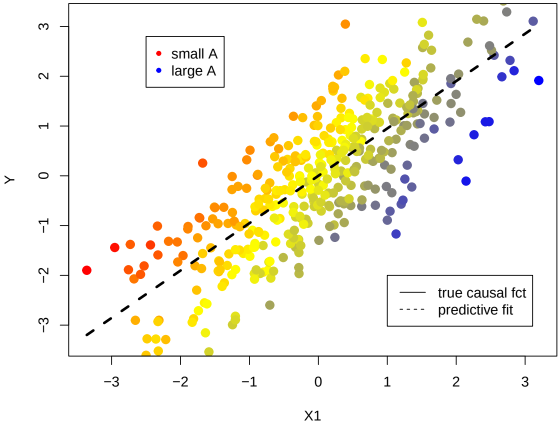


Instrumental Variables:

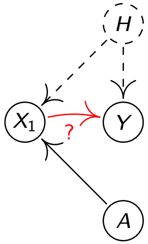


True model:

$$Y = {\color{red}\alpha^*} X_1 + \gamma H + N_Y.$$



Instrumental Variables:

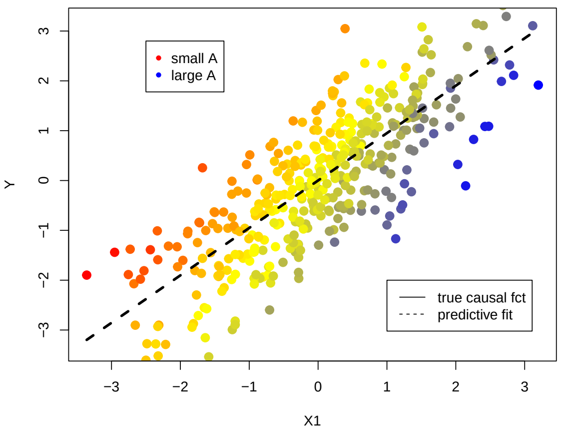


True model:

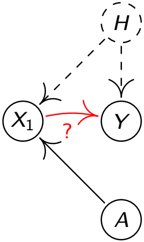
$$Y = \alpha^* X_1 + \gamma H + N_Y.$$

Find α such that

$$\mathrm{cov}[A, Y - \alpha X_1] = 0.$$



Instrumental Variables:

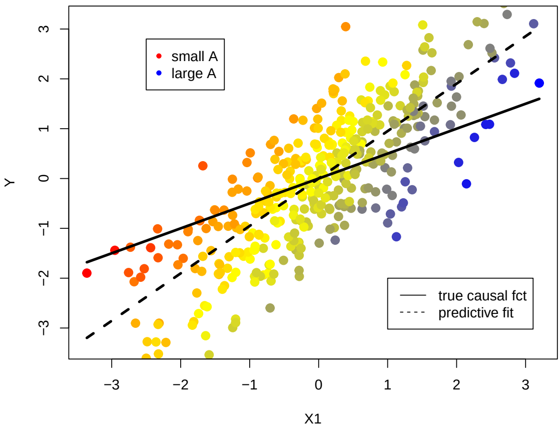


True model:

$$Y = \color{red}{\alpha^*} X_1 + \gamma H + N_Y.$$

Find α such that

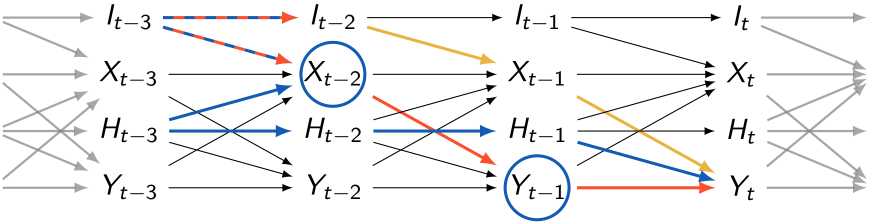
$$\text{cov}[A, Y - \alpha X_1] = 0.$$



IV: Anderson and Rubin 1949, Theil 1953, ...

Example: VAR processes.

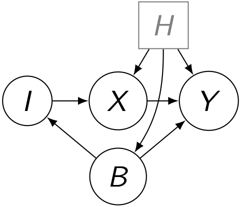
Example: VAR processes.



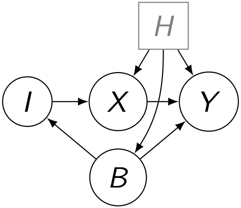
Here: $\text{cov}[I_{t-2}, Y_t - \beta^* X_{t-1}] \neq 0$ (see red path).

N. Thams, R. Nielsen, S. Weichwald, J. Peters:
Identifying Causal Effects using Instrumental Time Series: NIV and Correcting for the Past, arXiv:2203.06056

Recall: conditional IV Pearl 2009, Leo's PhD thesis 2021.

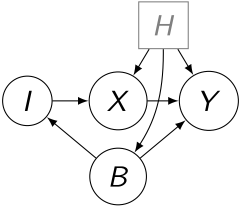


Recall: conditional IV Pearl 2009, Leo's PhD thesis 2021.



Linear SCM with DAG G and $Y = \beta^\top X + \gamma^\top W + \varepsilon$. Assume

Recall: conditional IV Pearl 2009, Leo's PhD thesis 2021.

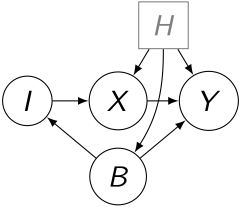


Linear SCM with DAG G and $Y = \beta^\top X + \gamma^\top W + \varepsilon$. Assume

- I and Y are d-sep by B in $G_{X \not\to Y}$.
- $B \notin \mathrm{DE}(X \cup Y)$.
- $E[cov(X, I|B)]$ has rank d_X.

Then,

Recall: conditional IV Pearl 2009, Leo's PhD thesis 2021.



Linear SCM with DAG G and $Y = \beta^\top X + \gamma^\top W + \varepsilon$. Assume

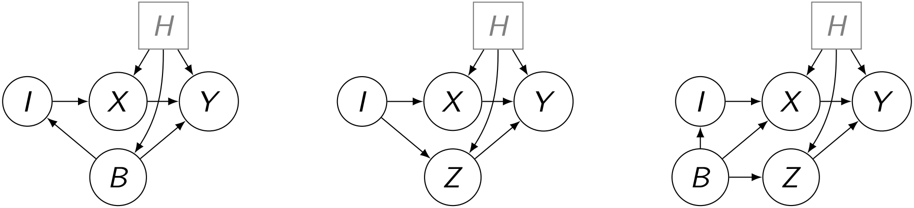
- I and Y are d-sep by B in $G_{X \not\to Y}$.
- $B \notin \mathrm{DE}(X \cup Y)$.
- $E[cov(X, I|B)]$ has rank d_X.

Then,

$$E[\mathrm{cov}(Y - \beta^\top X, I|B)] = 0.$$

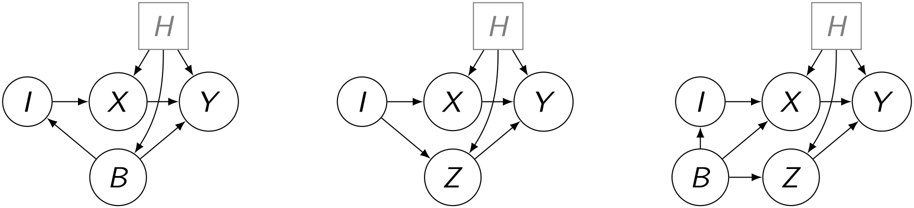
We write $\mathrm{CIV}(I|B)$.

Interested in: $X \to Y$



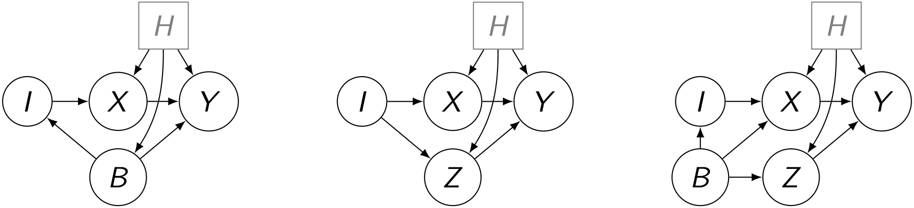
Left:

Interested in: $X \to Y$



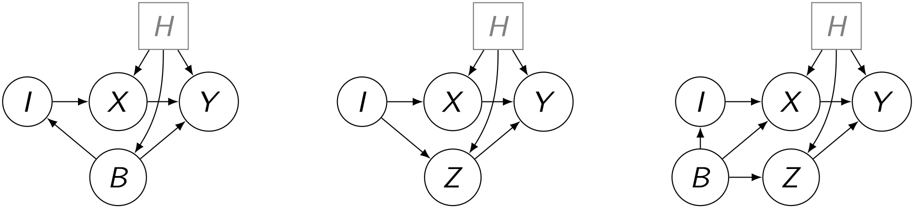
Left: $\mathrm{CIV}(I|B)$
Middle:

Interested in: $X \to Y$



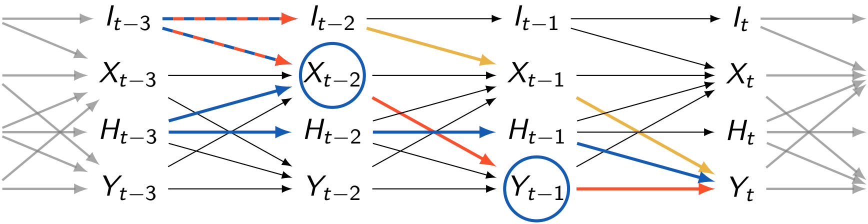
Left: $\mathrm{CIV}(I|B)$
Middle: $\mathrm{NIV}(I, Z)$
Right:

Interested in: $X \rightarrow Y$



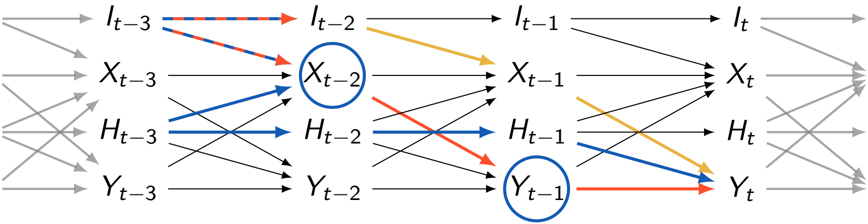
Left: $\text{CIV}(I|B)$
Middle: $\text{NIV}(I, Z)$
Right: $\text{CIV}(I|B)$ or $\text{NIV}(\{I, B\}, Z)$.

Example: VAR processes.



Here: $\mathrm{cov}[I_{t-2}, Y_t - \beta^* X_{t-1}] \neq 0$ (see red path).

Example: VAR processes.

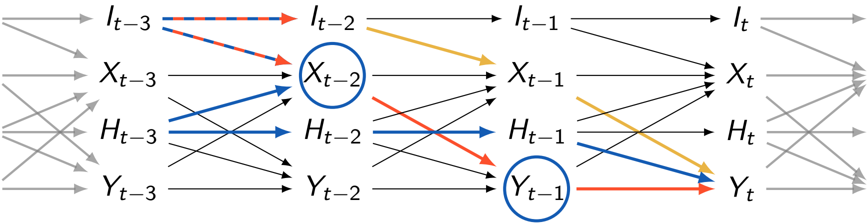


Here: $\text{cov}[I_{t-2}, Y_t - \beta^* X_{t-1}] \neq 0$ (see red path).

- conditional IV

$$\text{CIV}(I_{t-2}, \{I_{t-3}, X_{t-2}, Y_{t-1}\}) \text{ or } \text{CIV}(I_{t-2}, I_{t-3})$$

Example: VAR processes.



Here: $\mathrm{cov}[I_{t-2}, Y_t - \beta^* X_{t-1}] \neq 0$ (see red path).

- conditional IV

$$\mathrm{CIV}(I_{t-2}, \{I_{t-3}, X_{t-2}, Y_{t-1}\}) \text{ or } \mathrm{CIV}(I_{t-2}, I_{t-3})$$

- nuisance IV

$$\mathrm{NIV}\left(\begin{pmatrix} I_{t-2} \\ I_{t-3} \end{pmatrix}, Y_{t-1}\right), \text{i.e.,} \quad \mathrm{cov}\left[\begin{pmatrix} I_{t-2} \\ I_{t-3} \end{pmatrix}(Y_t - \beta^* X_{t-1} - \gamma Y_{t-1})\right] = 0$$

N. Thams, R. Nielsen, S. Weichwald, J. Peters:

Identifying Causal Effects using Instrumental Time Series: NIV and Correcting for the Past, arXiv:2203.06056

Let M be a square matrix. There exists P such that $M = PJP^{-1}$, where

$$J = \operatorname{diag}(J_{m_1}(\lambda_1), \ldots, J_{m_k}(\lambda_k)), \quad J_{m_i}(\lambda) = \begin{pmatrix} \lambda & 1 & & \\ & \lambda & \ddots & \\ & & \ddots & 1 \\ & & & \lambda \end{pmatrix} = \lambda I + N$$

Let M be a square matrix. There exists P such that $M = PJP^{-1}$, where

$$J = \mathrm{diag}(J_{m_1}(\lambda_1), \dots, J_{m_k}(\lambda_k)), \quad J_{m_i}(\lambda) = \begin{pmatrix} \lambda & 1 & & \\ & \lambda & \ddots & \\ & & \ddots & 1 \\ & & & \lambda \end{pmatrix} = \lambda I + N$$

Properties:

- Eigenvalues of M = diagonal values of J
- Geom. multipl. of λ_i (in M) = # of Jordan blocks with λ_i
- Alg. multipl. of λ_i (in M) = sum of lengths of Jordan blocks with λ_i
- M is diagonalizable if geom. and alg. multiplicity coincide.
- $M^n = PJ^nP^{-1} = P\left(\sum_{i=0}^{n}\binom{n}{i}\lambda^{n-i}N^i\right)P^{-1}$.

Data-generating process: We consider a VAR system S where

$$S_{t+1} = AS_t + \epsilon_{t+1}$$

We assume that S is stable (eigenvalues within unit circle) and that ϵ is non-singular.

$$A = \begin{array}{cc} \begin{array}{c} \mathrm{I} \\ \mathrm{H} \\ \mathrm{X} \\ \mathrm{Y} \end{array} & \left(\begin{array}{cccc} \alpha_I & 0 & 0 & 0 \\ 0 & \alpha_H & 0 & 0 \\ \left[\begin{array}{c} \gamma \\ 0 \end{array} \right. & \begin{array}{c} \nu_X \\ \nu_Y \end{array} & \left[\begin{array}{cc} \alpha_X & 0 \\ \beta & \alpha_Y \end{array} \right] & \end{array} \right) \\ & \begin{array}{cccc} M_I & & M_{XY} & \end{array} \end{array}$$

Data-generating process: We consider a VAR system S where

$$S_{t+1} = AS_t + \epsilon_{t+1}$$

We assume that S is stable (eigenvalues within unit circle) and that ϵ is non-singular.

$$A = \begin{array}{cc} \begin{array}{c} \text{I} \\ \text{H} \\ \text{X} \\ \text{Y} \end{array} & \left(\begin{array}{cccc} \alpha_I & 0 & 0 & 0 \\ 0 & \alpha_H & 0 & 0 \\ \left[\begin{array}{c} \gamma \\ 0 \end{array} \right] & \begin{array}{c} \nu_X \\ \nu_Y \end{array} & \multicolumn{2}{c}{\left[\begin{array}{cc} \alpha_X & 0 \\ \beta & \alpha_Y \end{array} \right]} \end{array} \right) \\ & \begin{array}{ccc} M_I & \qquad\qquad & M_{XY} \end{array} \end{array}$$

Required for identifiability through NIV:

$$\Sigma := E\left\{ \begin{pmatrix} X_t \\ Y_t \end{pmatrix} (I_{t-1}, \ldots, I_{t-d_X-1}) \right\} \in \mathbb{R}^{(d_X+1)\times(d_X+1)}$$

is full rank.

Consider the decomposition

$$M_{XY} = PJP^{-1}$$

with Jordan normal form

$$J = \mathrm{diag}(J_{m_1}(\lambda_1), \ldots, J_{m_k}(\lambda_k))$$

and let $a := P^{-1}M_I$. The following three statements are equivalent:

1. Σ is invertible.
2. The matrix $\begin{bmatrix} M_{XY}^0 M_I & M_{XY}^1 M_I & \cdots & M_{XY}^{d_X} M_I \end{bmatrix}$ is invertible.
3. Different Jordan blocks have different eigenvalues and for all $q \in \{1, \ldots, k\}$, the coefficient $a_{\sum_{i=1}^{q} m_i}$ is non-zero.

Summary

Testing under distributional shifts:

- Applies to many problems including cond. ind. tests, Verma constraints, model selection under covariate shift, (off-policy evaluations, two-sample testing) ...
- Black box method: resampling and test in the target domain.
- Comes with theoretical guarantees.

Summary

Testing under distributional shifts:

- Applies to many problems including cond. ind. tests, Verma constraints, model selection under covariate shift, (off-policy evaluations, two-sample testing) ...
- Black box method: resampling and test in the target domain.
- Comes with theoretical guarantees.

VAR-IV:

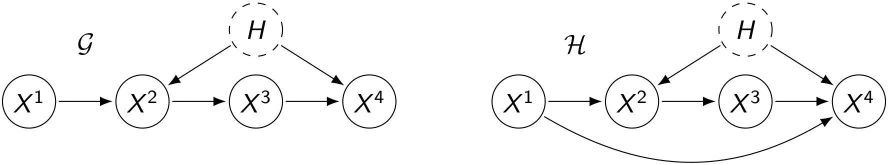
- Nuisance IV yields different identifiability results than Conditional IV.
- Both can be applied to VAR-IV settings (even multivariate).
- Identifiability results are similar to those in systems of ordinary differential equations.

S. Saengkyongam, N. Thams, JP, N. Pfister: *Invariant Policy Learning: A Causal Perspective*, arXiv 2021

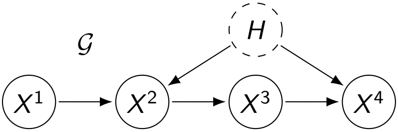
N. Thams, R. Nielsen, S. Weichwald, JP: Identif. Causal Effects using Instr. TS: Nuisance IV and Corr. for the Past, arXiv 2022

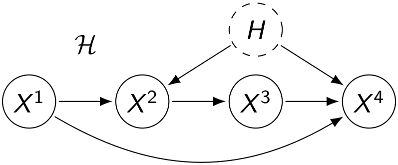
N. Thams, S. Saengkyongam, N. Pfister, JP: *Statistical Testing under Distributional Shifts*, arXiv 2021

Example: Dormant independences



Example: Dormant independences

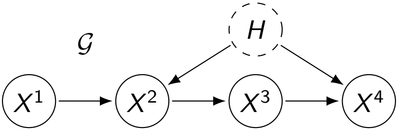


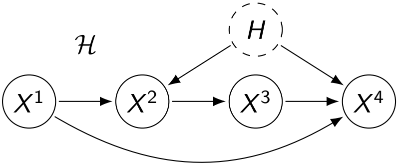


$X^1 \perp\!\!\!\perp X^4$ in $Q^{do(X^3:=N)}$

$X^1 \not\perp\!\!\!\perp X^4$ in $Q^{do(X^3:=N)}$.

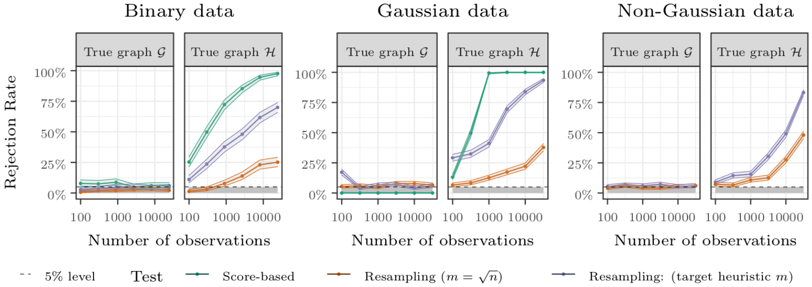
Example: Dormant independences





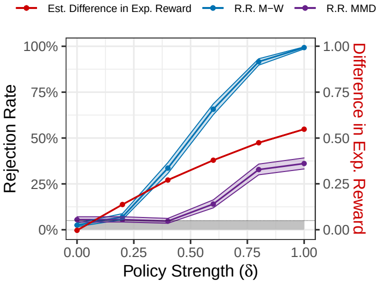
$X^1 \perp\!\!\!\perp X^4$ in $Q^{do(X^3:=N)}$

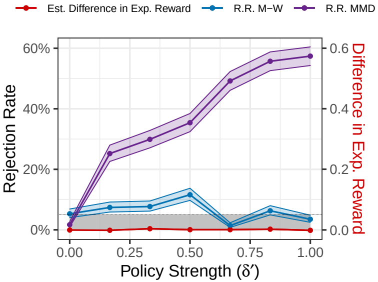
$X^1 \not\!\perp\!\!\!\perp X^4$ in $Q^{do(X^3:=N)}$.



Verma and Pearl 1991, Shpitser et al. 2012

Off-policy evaluation:





Example: Model Selection under Covariate Shift

Training distribution: $q_{\text{train}}(x, y) = q^*_{\text{train}}(x)q(y|x)$

Test distribution: $p_{\text{test}}(x, y) = p^*_{\text{test}}(x)q(y|x)$

Example: Model Selection under Covariate Shift

Training distribution: $q_{\text{train}}(x, y) = q^*_{\text{train}}(x)q(y|x)$
Test distribution: $p_{\text{test}}(x, y) = p^*_{\text{test}}(x)q(y|x)$

Goal: Model selection

Example: Model Selection under Covariate Shift

Training distribution: $q_{\text{train}}(x, y) = q^*_{\text{train}}(x)q(y|x)$

Test distribution: $p_{\text{test}}(x, y) = p^*_{\text{test}}(x)q(y|x)$

Goal: Model selection

Held out data (have): $D_{\text{heldout}} = \{(x_1, y_1), \ldots, (x_n, y_n)\} \sim q_{\text{train}}$

Held out data (wanted): $D^{sh}_{\text{heldout}} = \{(x_1, y_1), \ldots, (x_n, y_n)\} \sim p_{\text{test}}$

Example: Model Selection under Covariate Shift

Training distribution: $q_{\text{train}}(x, y) = q^*_{\text{train}}(x) q(y|x)$

Test distribution: $p_{\text{test}}(x, y) = p^*_{\text{test}}(x) q(y|x)$

Goal: Model selection

Held out data (have): $D_{\text{heldout}} = \{(x_1, y_1), \dots, (x_n, y_n)\} \sim q_{\text{train}}$

Held out data (wanted): $D^{sh}_{\text{heldout}} = \{(x_1, y_1), \dots, (x_n, y_n)\} \sim p_{\text{test}}$

'Adapted CV' (Sugiyama et al. 2007) allows us to estimate

$$\mathbf{E}_{(x,y)\sim p_{\text{test}}} \ell(x, y, \hat{f}) \quad \text{by} \quad \sum_{(x,y)\in D_{\text{heldout}}} \frac{p^*_{\text{test}}(x)}{q^*_{\text{train}}(x)} \ell(x, y, \hat{f}).$$

Example: Model Selection under Covariate Shift

Training distribution: $q_{\text{train}}(x,y) = q^*_{\text{train}}(x)q(y|x)$
Test distribution: $p_{\text{test}}(x,y) = p^*_{\text{test}}(x)q(y|x)$

Goal: Model selection

Held out data (have): $D_{\text{heldout}} = \{(x_1,y_1),\ldots,(x_n,y_n)\} \sim q_{\text{train}}$
Held out data (wanted): $D^{sh}_{\text{heldout}} = \{(x_1,y_1),\ldots,(x_n,y_n)\} \sim p_{\text{test}}$

'Adapted CV' (Sugiyama et al. 2007) allows us to estimate

$$\mathbf{E}_{(x,y)\sim p_{\text{test}}}\ell(x,y,\hat{f}) \quad \text{by} \quad \sum_{(x,y)\in D_{\text{heldout}}} \frac{p^*_{\text{test}}(x)}{q^*_{\text{train}}(x)}\ell(x,y,\hat{f}).$$

Above framework allows to test, e.g.,

$$H_0 : \mathbf{E}_{D^{sh}\sim p_{\text{test}}}\left[S(D^{sh},\hat{f}_1) - S(D^{sh},\hat{f}_2)\right] \leq 0.$$

given: $X_1, \dots, X_n \overset{\text{iid}}{\sim} q; \quad p(x) \propto r(x) \cdot q(x)$

given: $X_1, \ldots, X_n \overset{\text{iid}}{\sim} q; \quad p(x) \propto r(x) \cdot q(x)$

General and easy-to-use method based on resampling and existing test:

1. Resample m data points with weights

$$w_{(i_1,\ldots,i_m)} \propto \begin{cases} \prod_{\ell=1}^{m} r(X_{i_\ell}) & \text{if } (i_1, \ldots, i_m) \text{ is distinct and} \\ 0 & \text{otherwise.} \end{cases} \tag{1}$$

2. Apply existing test φ_m from target domain to resampled data.

given: $X_1, \ldots, X_n \overset{\text{iid}}{\sim} q$; $p(x) \propto r(x) \cdot q(x)$

General and easy-to-use method based on resampling and existing test:

1. Resample m data points with weights

$$w_{(i_1,\ldots,i_m)} \propto \begin{cases} \prod_{\ell=1}^{m} r(X_{i_\ell}) & \text{if } (i_1, \ldots, i_m) \text{ is distinct and} \\ 0 & \text{otherwise.} \end{cases} \qquad (1)$$

2. Apply existing test φ_m from target domain to resampled data.

Theorem (Thams et al., 2021)

Pointwise asymptotic level holds if:

A1 φ_m *has pointwise asymptotic level,*

A2 $m = o(\sqrt{n})$,

A3 *for all* $Q \in \tau^{-1}(H_0)$, *it holds that* $\mathbf{E}_Q[r(X_i)^2] < \infty$.

given: $X_1, \ldots, X_n \overset{\text{iid}}{\sim} q; \quad p(x) \propto r(x) \cdot q(x)$

General and easy-to-use method based on resampling and existing test:

1. Resample m data points with weights

$$w_{(i_1,\ldots,i_m)} \propto \begin{cases} \prod_{\ell=1}^{m} r(X_{i_\ell}) & \text{if } (i_1, \ldots, i_m) \text{ is distinct and} \\ 0 & \text{otherwise.} \end{cases} \tag{1}$$

2. Apply existing test φ_m from target domain to resampled data.

Theorem (Thams et al., 2021)

Pointwise asymptotic level holds if:

A1 φ_m *has pointwise asymptotic level,*

A2 $m = o(\sqrt{n})$,

A3 *for all* $Q \in \tau^{-1}(H_0)$, *it holds that* $\mathbf{E}_Q[r(X_i)^2] < \infty$.

A similar statement holds if r needs to be estimated.

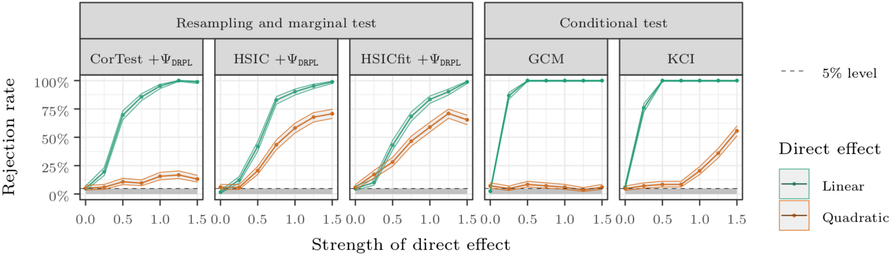
N. Thams, S. Saengkyongam, N. Pfister, JP: *Statistical Testing under Distributional Shifts*, arXiv 2021

cf. SIR, see Rubin (1987) and Smith and Gelfand (1992)

Example: Conditional Independence Testing

We sample $n = 150$ observations from

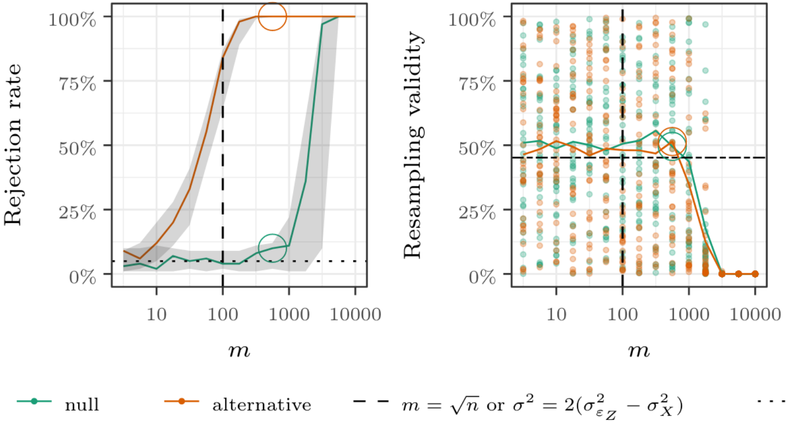
$$X := \texttt{GaussianMixture}(-2, 2) \quad Z := -X^2 + \epsilon_Z \quad Y := \sin(Z) + \theta X^\tau + \epsilon_Y,$$



Gretton et al. 2008, Zhang et al. 2011, Shah & Peters 2020

Example: Conditional Independence Testing

How can we choose *m*?



using a test by Jitkrittum et al., 2020