13 Nov 2021

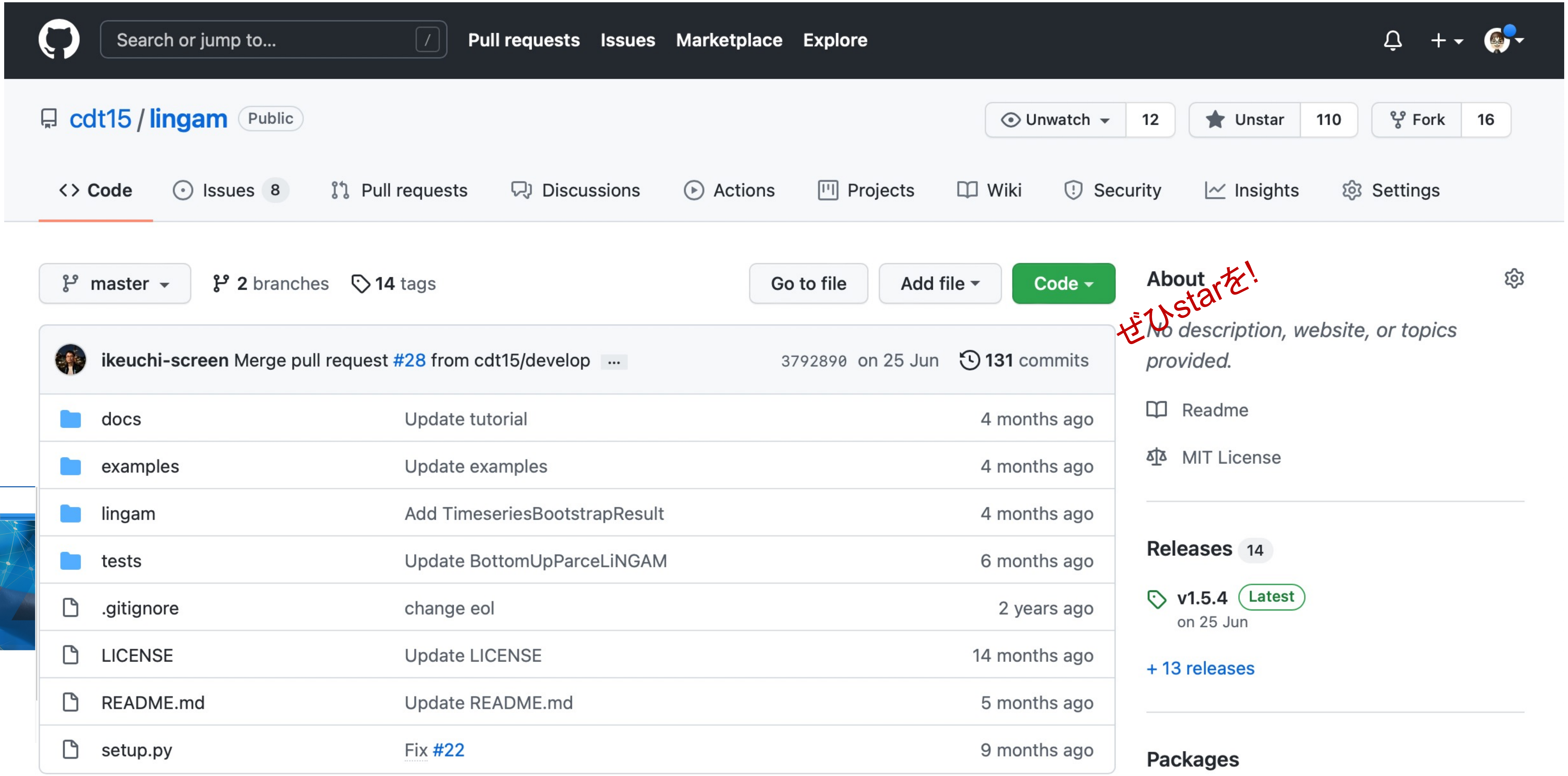
LiNGAM Python package

Shohei SHIMIZU

Shiga University & RIKEN

LiNGAM Python package

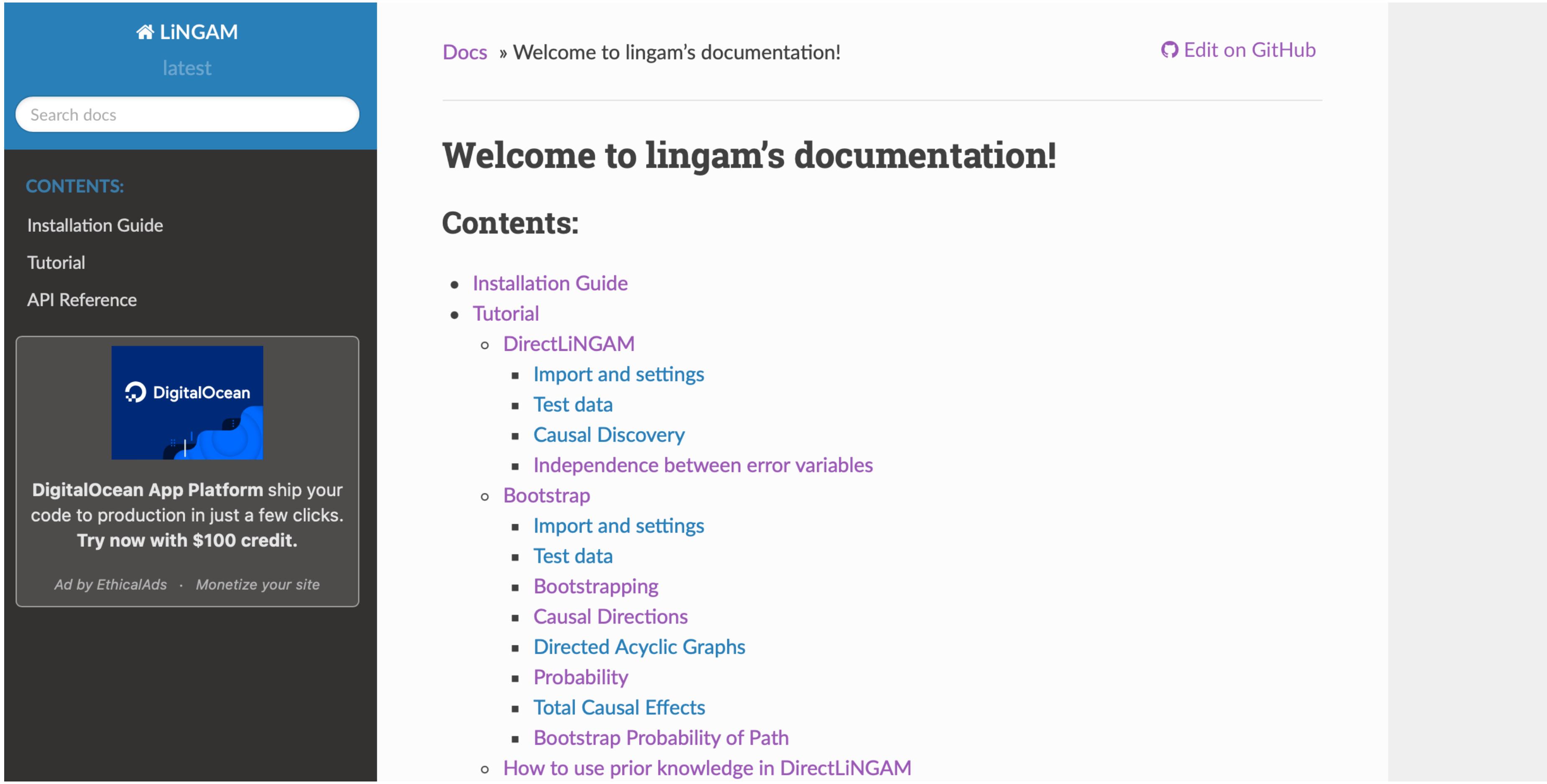
- https://github.com/cdt15/lingam



Takashi Ikeuchi
SCREEN AS

Documentation

- https://lingam.readthedocs.io/en/latest/#



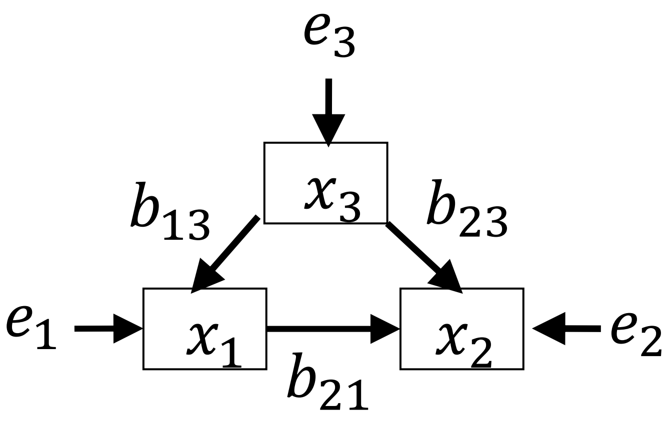
LiNGAM model is identifiable

(Shimizu, Hyvarinen, Hoyer & Kerminen, 2006)

- Linear Non-Gaussian Acyclic Model:

$$x_i = \sum_{k(j)<k(i)} b_{ij} x_j + e_i \qquad \text{or} \qquad \boldsymbol{x} = B\boldsymbol{x} + \boldsymbol{e}$$

  - $k(i)$ $(i = 1, \ldots, p)$: causal (topological) order of x_i
  - Error variables e_i are independent and non-Gaussian



Causal graph

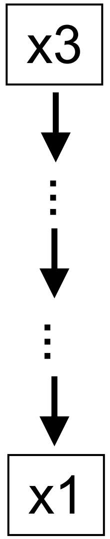
- Coefficients and causal orders identifiable
- Causal graph identifiable

Statistical reliability assessment

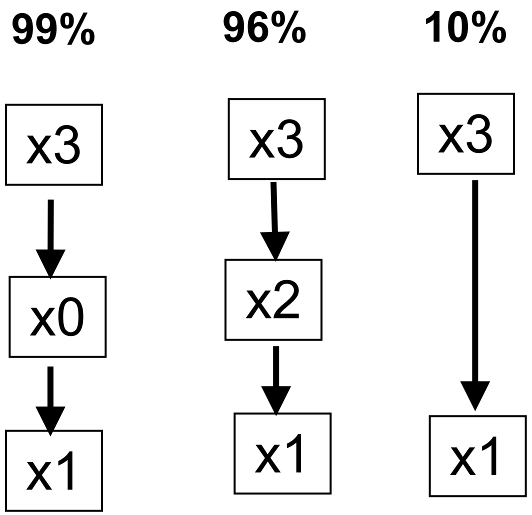
- Bootstrap probability (bp) of directed paths and edges
- Interpret causal effects having bp larger than a threshold, say 5%

	from	to	effect	probability
0	x3	x0	3.006190	1.00
1	x0	x1	3.004868	1.00
2	x2	x1	2.092102	1.00
3	x3	x1	20.931938	1.00
4	x0	x5	3.982892	1.00
5	x3	x5	12.024250	1.00
6	x2	x4	-0.887620	1.00
7	x3	x4	18.077244	1.00
8	x0	x4	7.993145	0.98
9	x3	x2	5.970163	0.96
10	x5	x1	0.011708	0.79
11	x2	x5	0.024284	0.72

Total effect:
20.9



	path	effect	probability
0	[3, 0, 1]	8.644765	0.99
1	[3, 2, 1]	11.653517	0.96
2	[3, 1]	0.102936	0.10
3	[3, 2, 0, 1]	1.007787	0.08
4	[3, 0, 5, 1]	0.889629	0.05
5	[3, 4, 0, 1]	8.419805	0.02
6	[3, 2, 4, 0, 1]	-3.802326	0.01



LiNGAM Python package: https://github.com/cdt15/lingam

Before estimating causal graphs

- Assessing assumptions by
  - Gaussianity test
  - Histograms
    - continuous?
  - Too high correlation?
    - multicollinearity?
  - Background knowledge

After estimating causal graphs

- Assessing assumptions by
  - Testing independence of error variables, for example, by HSIC (Gretton et al., 2005)
  - Prediction accuracy using Markov boundary (Biza et al., 2020)
  - Compare with the results of other datasets in which causal graphs are expected to be similar
  - Check against background knowledge

DirectLiNGAM algorithm

(Shimizu et al., 2011)

- Repeat linear regression and independence evaluation
  - https://lingam.readthedocs.io/en/latest/tutorial/lingam.html
- p>n cases (Wang & Drton, 2020)
  - https://github.com/ysamwang/highDNG

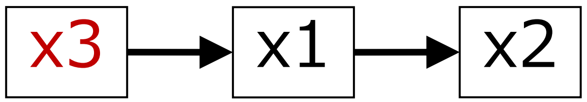
$$\begin{bmatrix} x_3 \\ x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 0 & 0 & 0 \\ 1.5 & 0 & 0 \\ 0 & -1.3 & 0 \end{bmatrix} \begin{bmatrix} x_3 \\ x_1 \\ x_2 \end{bmatrix} + \begin{bmatrix} e_3 \\ e_1 \\ e_2 \end{bmatrix} \Rightarrow \begin{bmatrix} r_1^{(3)} \\ r_2^{(3)} \end{bmatrix} = \begin{bmatrix} 0 & 0 \\ -1.3 & 0 \end{bmatrix} \begin{bmatrix} r_1^{(3)} \\ r_2^{(3)} \end{bmatrix} + \begin{bmatrix} e_1 \\ e_2 \end{bmatrix}$$

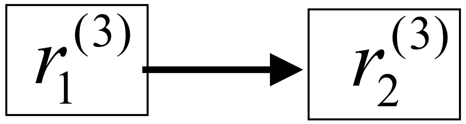
$$\boxed{x3} \rightarrow \boxed{x1} \rightarrow \boxed{x2} \qquad \boxed{r_1^{(3)}} \rightarrow \boxed{r_2^{(3)}}$$

Prior knowledge

https://lingam.readthedocs.io/en/latest/tutorial/pk_direct.html

- Prior knowledge about topological orders: k(3) < k(1) < k(2)

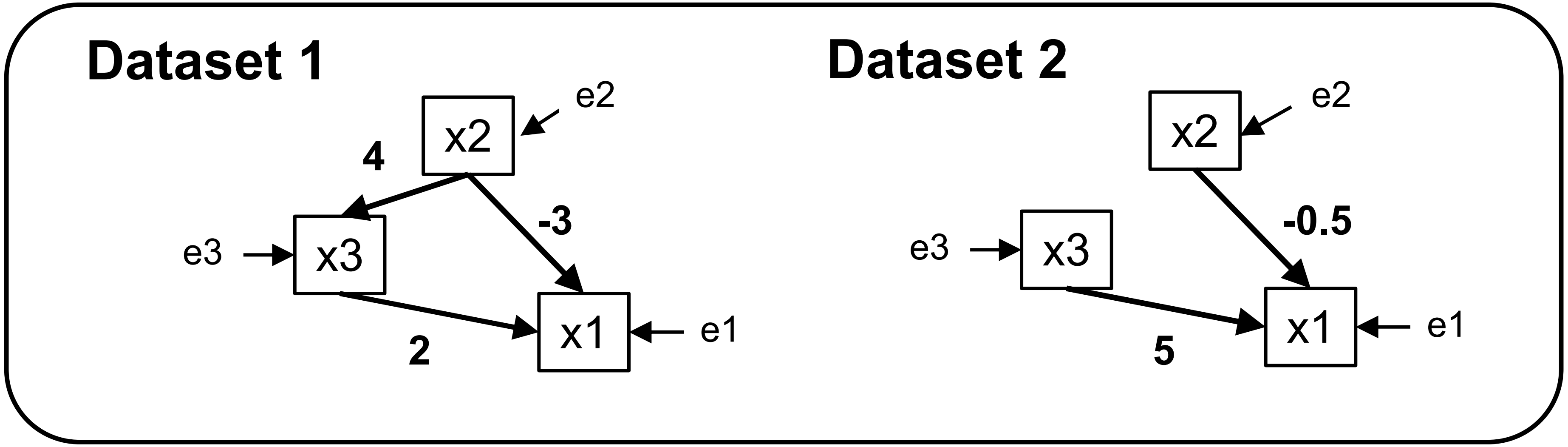




- Use prior knowledge in estimating topological causal orders and in pruning redundant edges

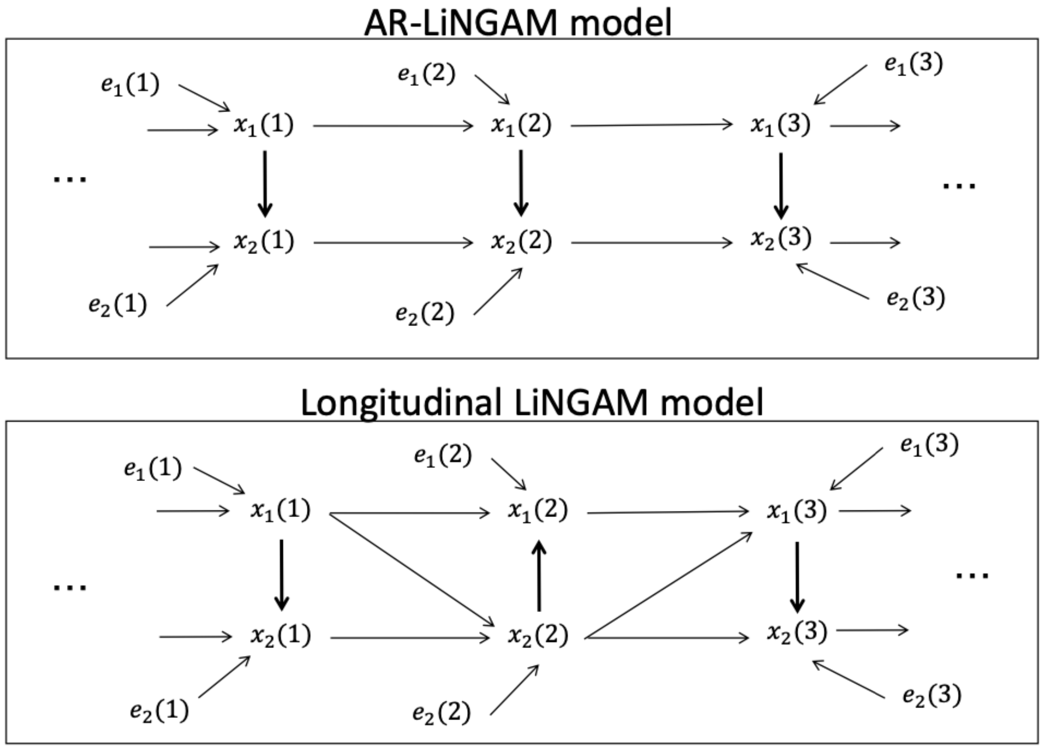
Multiple datasets

- Simultaneously analyze different datasets to use similarity (Ramsey et al. 2011; Shimizu, 2012)
  - Similarity: Causal orders same, distributions and coefficients may differ
  - https://lingam.readthedocs.io/en/latest/tutorial/multiple_dataset.html



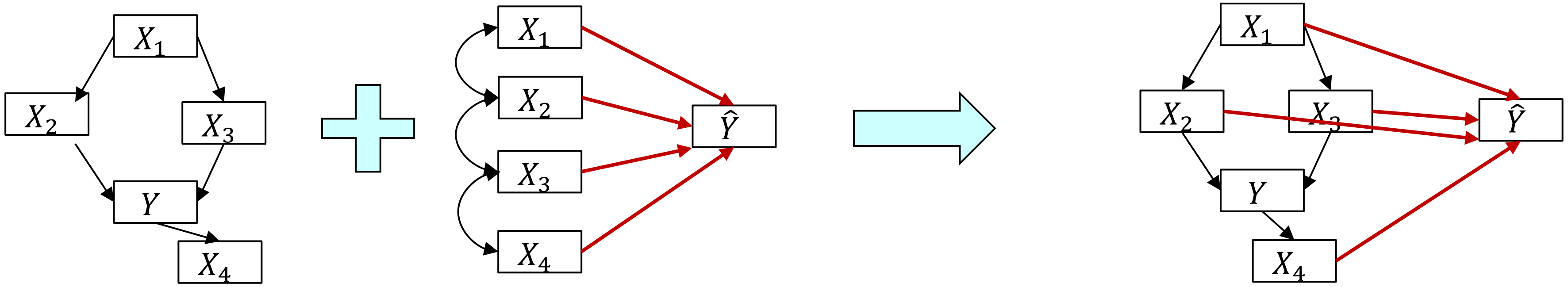
Multiple datasets: Longitudinal data

- Longitudinal data consist of multiple samples collected over a period of time (Kadowaki et al., 2013)
- https://lingam.readthedocs.io/en/latest/tutorial/longitudinal.html



Analysis of predictive mechanisms

- Combine the causal model and predictive model to model the prediction mechanism

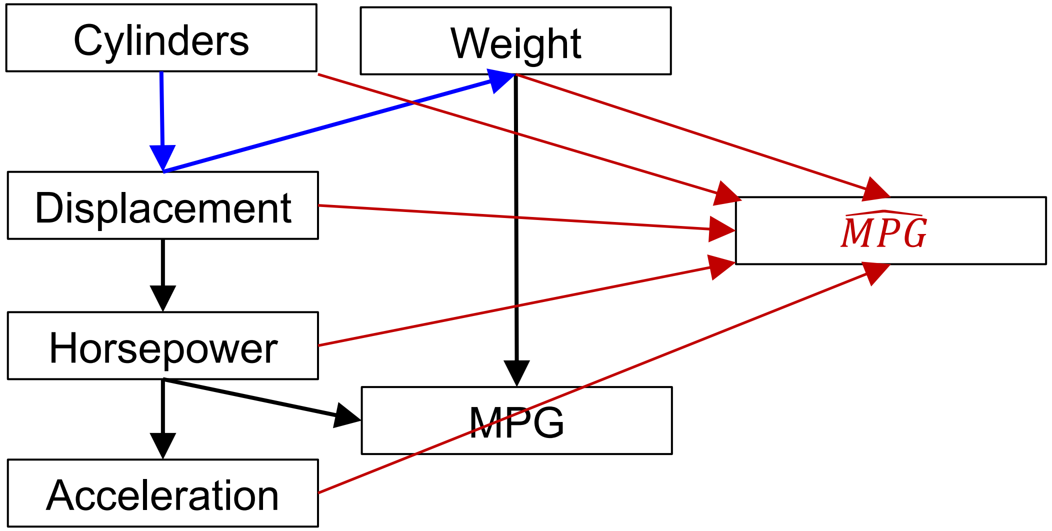


Causal model	Predictive model	Prediction mechanism model
$x_4 = f_4(y, e_4)$	$\hat{y} = f(x_1, x_2, x_3, x_4)$	$E(\hat{y} \mid do(x_i = c))$

https://lingam.readthedocs.io/en/latest/tutorial/causal_effect.html#identification-of-feature-with-greatest-causal-influence-on-prediction

Illustrative example

- Auto-MPG (miles per gallon) dataset
- Linear regression
- Which variable has the greatest intervention effect on MPG prediction?
- Which variable should be intervened on to obtain a certain MPG prediction? (Control)



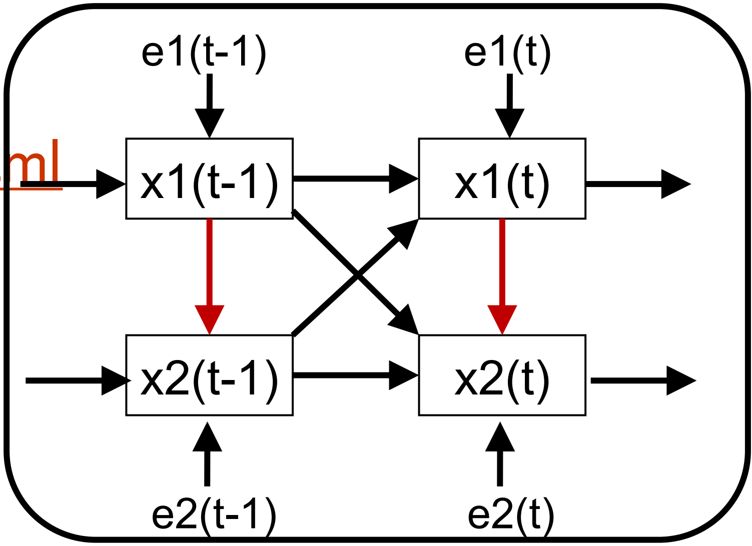
Desired MPG prediction	Suggested intervention on cylinders
15	8
21	6
30	4

Time series model

- Subsampling data:
  - SVAR: Structural Vector Autoregressive model (Swanson & Granger, 1997)

$$\mathbf{x}(t) = \sum_{\tau=0}^{k} \mathbf{B}_\tau \mathbf{x}(t-\tau) + \mathbf{e}(t)$$

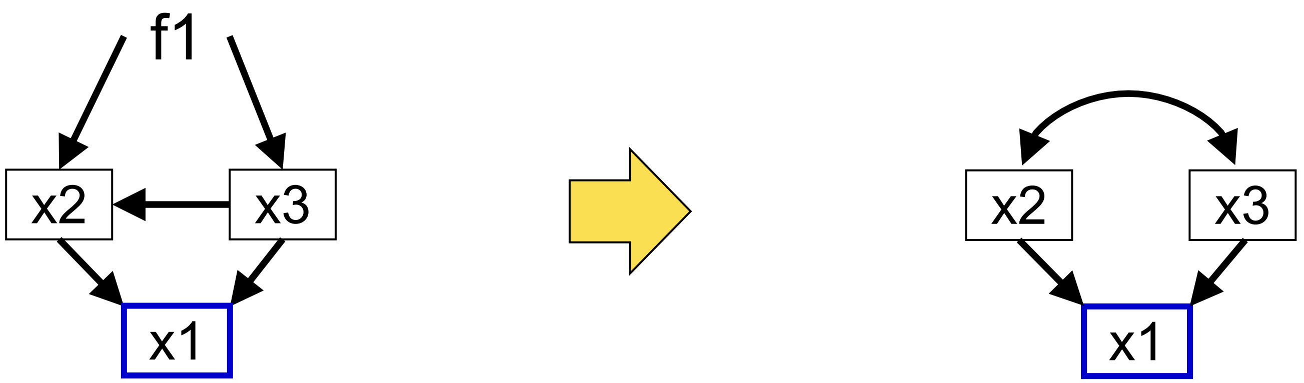
  - Identifiability using non-Gaussianity (Hyvarinen et al., 2010)
    - https://lingam.readthedocs.io/en/latest/tutorial/var.html
  - VARMA instead of VAR (Kawahara et al., 2011)
    - https://lingam.readthedocs.io/en/latest/tutorial/varma.html



- Nonstationarity
  - Assumption: Differences are stationarity (Moneta et al., 2013)

Hidden common cause (1)

- Assumption: only exogenous variables allow hidden common causes

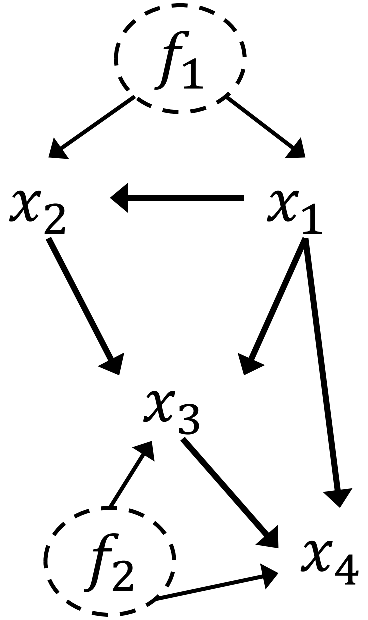


https://lingam.readthedocs.io/en/latest/tutorial/bottom_up_parce.html

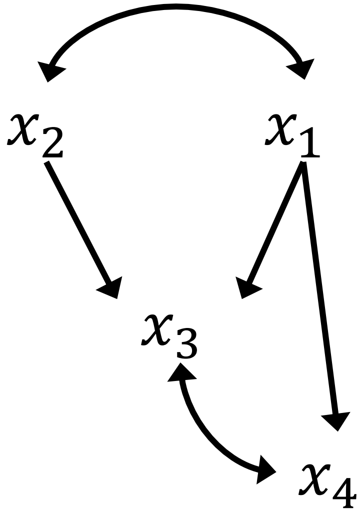
Hidden common cause (2) RCD

- For unconfounded pairs with no hidden common causes, estimate the causal directions
- For confounded pairs with hidden common causes, let them remain unknown

Underlying model



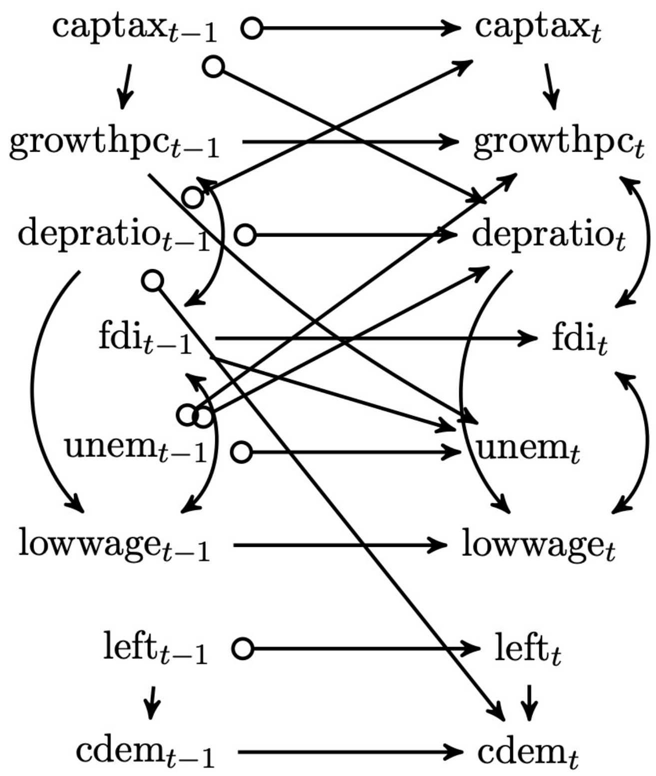
Output



https://lingam.readthedocs.io/en/latest/tutorial/rcd.html

Time series model with hidden common causes

- SVAR with hidden common causes
  - Malinsky and Spirtes (2018)
  - Gerhardus and Runge (2020)
  - Nonparametric
  - Conditional independence
  - Python: https://github.com/jakobrunge/tigramite



Nonlinear model

- Additive noise model:
- R code: http://web.math.ku.dk/~peters/code.html

$$x_i = f_i(\mathrm{par}(x_i)) + e_i$$

Home | Teaching | Research | Code | Elements of Causal Inference | NEW BOOK: The Raven's Hat

Packages:

- Python code for **CausalKinetiX** can be downloaded from github **CausalKinetiX**. Paper: N. Pfister, S. Bauer, J. Peters: Identifying Causal Stru https://arxiv.org/abs/1810.11776, 2018.
- R code for **CausalKinetiX** can be downloaded from CRAN, package name: **CausalKinetiX**. Paper: N. Pfister, S. Bauer, J. Peters: Identifying C Systems, https://arxiv.org/abs/1810.11776, 2018.
- R code for **sequential ICP** can be downloaded from CRAN, package name: **seqICP**. Paper: N. Pfister, P. Bühlmann, J. Peters: Invariant Causa 2018.
- R code for **dHSIC** can be downloaded from CRAN, package name: **dHSIC**. Paper: N. Pfister, P. Bühlmann, B. Schölkopf, J. Peters: Kernel-base 2017.
- R code for **Invariant Causal Prediction** can be downloaded from CRAN, package name: **InvariantCausalPrediction**. Paper: J. Peters, P. B inference using invariant prediction: identification and confidence intervals, JRSSB, 2016.
- R code for **SID** can be downloaded from CRAN, package name: **SID**. Paper: J. Peters, P. Bühlmann: "Structural Intervention Distance (SID) fo Computation, 2015.
- R code for **CAM** can be downloaded from CRAN, package name: **CAM**. Paper: P. Bühlmann, J. Peters, J. Ernest: CAM: Causal Additive Models, Penalized Regression, Annals of Statistics, 2014.

More code:

- **R code** for simulation experiments on **Generalised Covariance Measure (GCM)**.
- **R code** for **Half-Sibling Regression** (only simulations on iid data).
- **R code** for **Timino**. Paper: J. Peters, D. Janzing, B. Schölkopf: Causal Inference on Time Series using Structural Equation Models, Advances ir 26, 2014.
- **R code** for **ANMs**. Paper: J. Peters, J. Mooij, D. Janzing, B. Schölkopf: Causal Discovery with Continuous Additive Noise Models, JMLR, 2014.
- **Matlab code** for **Cause-Effect-Pairs** (same paper).

More code can be found **here**.

Methods based on conditional independencies

- GUI: Tetrad
  - https://github.com/cmu-phil/tetrad
- Python: causal-learn (including LiNGAM variants)
  - https://github.com/cmu-phil/causal-learn
- R: pcalg
  - https://cran.r-project.org/web/packages/pcalg/index.html

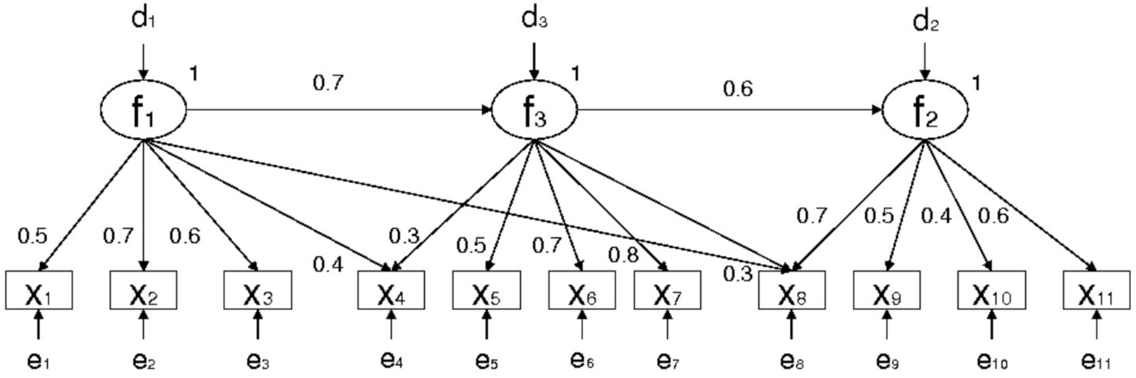
Future plan

- A nonlinear version of RCD: CAM-UV
- Latent factors
- Mixed data with continuous and discrete variables
- Overcomplete ICA based method for hidden common cause cases under development

LiNGAM for latent factors (Shimizu et al., 2009)

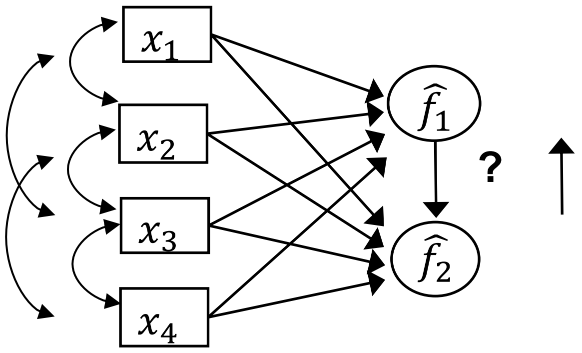
- Model:

$$\boldsymbol{f} = B\boldsymbol{f} + \boldsymbol{\epsilon}$$
$$\boldsymbol{x} = G\boldsymbol{f} + \boldsymbol{e}$$



  – Two pure measurement variables per latent factor needed to identify the measurement model (Silva et al., 2006; Xie et al., 2020)

- Estimate the latent factors and then their causal graph

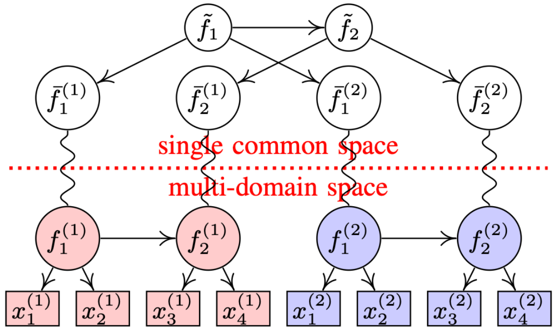


Find common and unique factors across multiple datasets (Zeng et al., 2021)

- Model

$$\boldsymbol{f}^{(m)} = B^{(m)}\,\boldsymbol{f}^{(m)} + \boldsymbol{\epsilon}^{(m)}$$
$$\boldsymbol{x}^{(m)} = G^{(m)}\,\boldsymbol{f}^{(m)} + \boldsymbol{e}^{(m)}$$

$m = 1, \ldots, M$



- Score function: likelihood + DAGness (Zheng et al., 2018)

- Feature extraction across multiple datasets + causal discovery of latent factors