Distinguishing Causes from Effects using Nonlinear Acyclic Causal Models

Kun Zhang[1] and Aapo Hyvärinen[1,2]

[1] Dept. of Computer Science & HIIT
[2] Dept. of Mathematics and Statistics
University of Helsinki

Outline

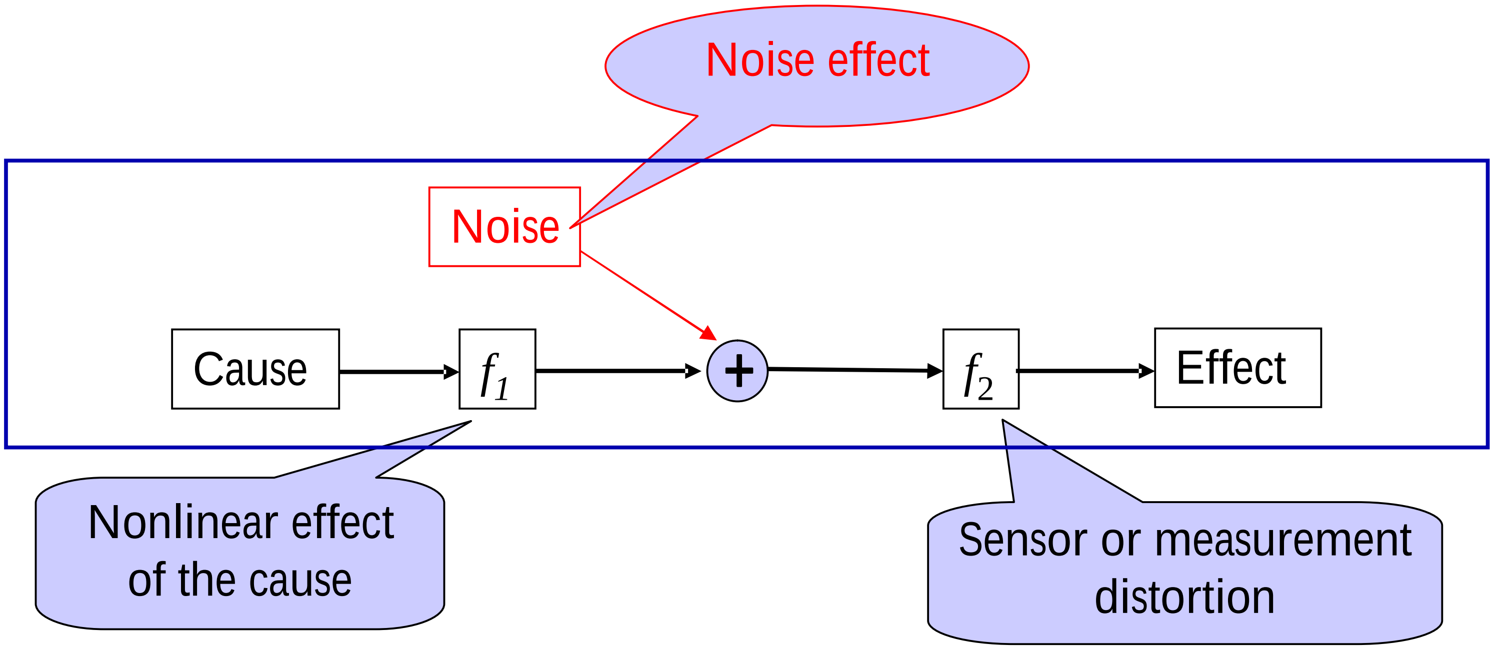
- Introduction
- Post-nonlinear causal model with inner additive noise
  - Relation to post-nonlinear independent component analysis (ICA)
  - Identification method
- Special cases
- Experiments

Methods for causal discovery

- Two popular kinds of methods
  - Constraint-based: using independence tests to find the patterns of relationships. Example: PC/IC
  - Score-based: using a score (such as BIC) to compare different causal models
- Model-based: a special case of score-based methods
  - Assumes a generative model for the data generating process
  - Can discover in what form each variable is influenced by others
  - Examples
    - Granger causality: effects follow causes in a linear form
    - LiNGAM: linear, non-Gaussian and acyclic causal model (Shimizu, et al., 2006)

Three effects usually encountered in a causal model

- Without prior knowledge, the assumed model is expected to be
  - general enough: adapted to approximate the true generating process
  - identifiable: asymmetry in causes and effects



- Represented by post-nonlinear causal model with inner additive noise

Post-nonlinear (PNL) causal model with inner additive noise

- The directed acyclic graph (DAG) is used to represent the data generating process:

pa_i: parents (causes) of x_i

$$x_i = f_{i,2}(\, f_{i,1}(pa_i) + e_i)$$

$f_{i,2}$: assumed to be continuous and invertible

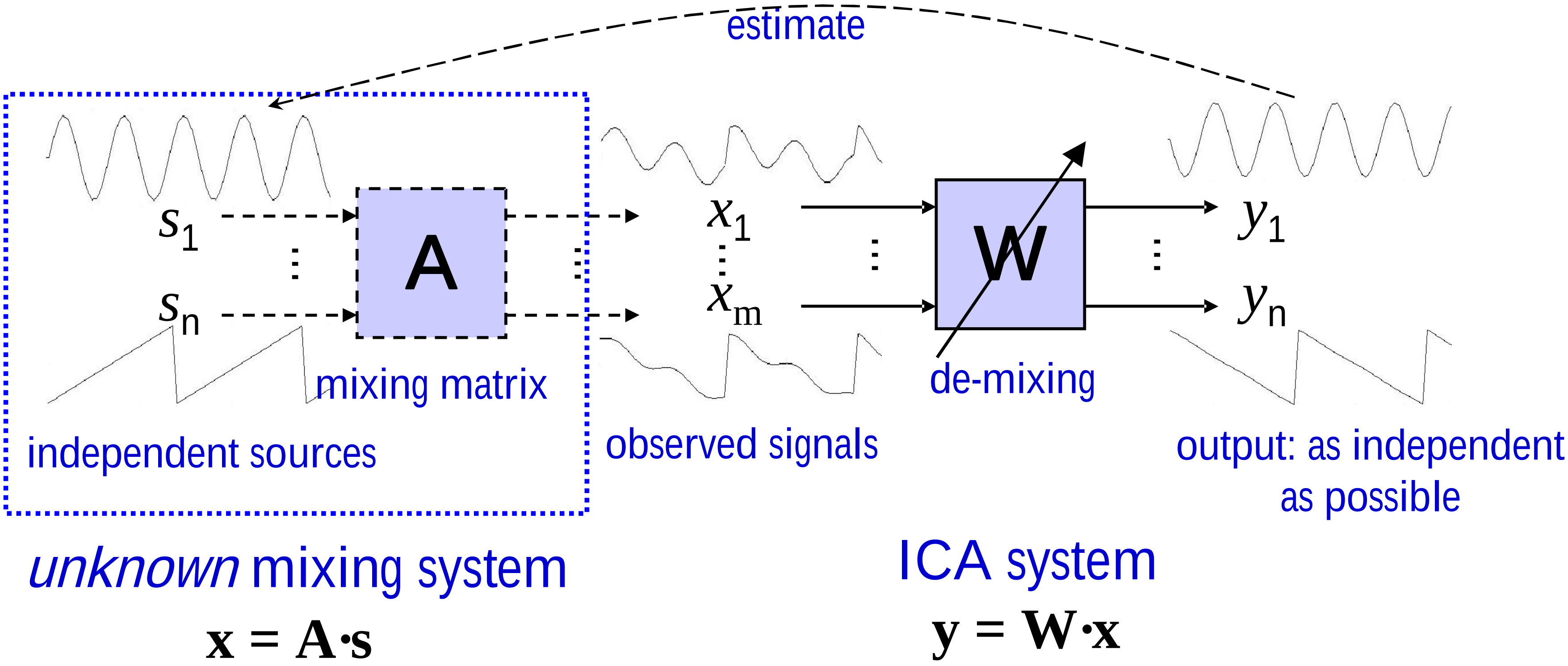
$f_{i,1}$: not necessarily invertible

e_i: noise/disturbance: independent from pa_i

- Here consider the two-variable case
  - $x_1 \rightarrow x_2$: $\quad x_2 = f_{2,2}(\, f_{2,1}(x_1) + e_2)$
- Identifiability: related to the separability of PNL mixing independent component analysis (ICA) model

Three cases of ICA: linear, general nonlinear, and PNL

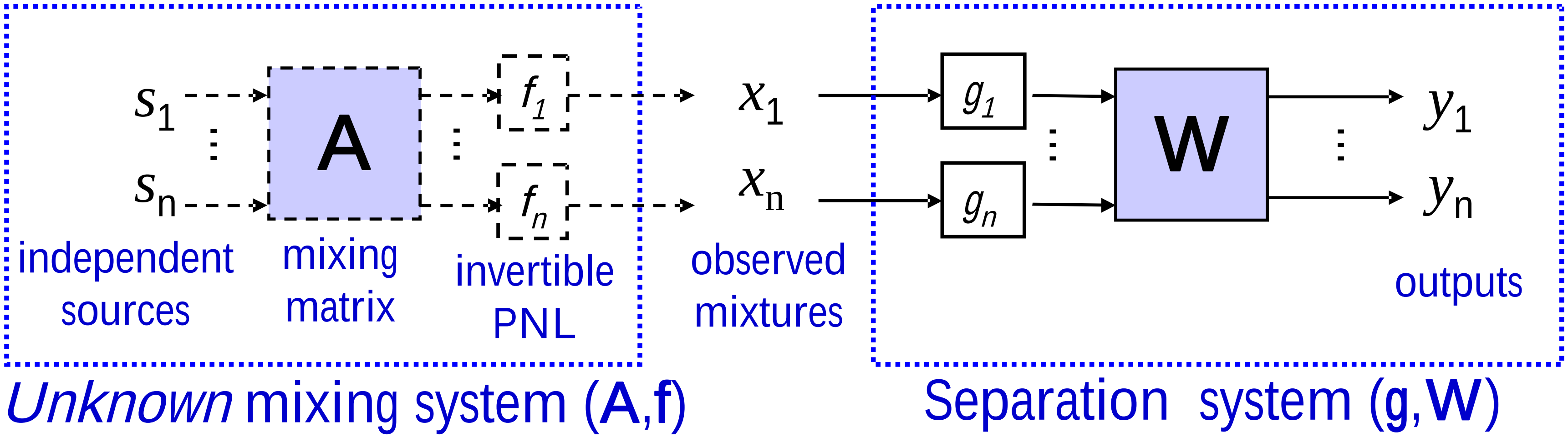
- Linear ICA: separable under weak assumptions



- Nonlinear ICA: **A** and **W** become invertible nonlinear mappings
  - not separable: y_i may be totally different from s_i

PNL mixing ICA: a nice trade-off

- Mixing system: linear transformation followed by invertible component-wise nonlinear transformation



- Separability (Taleb and Jutten, 1999): under the following conditions, y_i are independent iff $h_i = g_i \circ f_i$ is linear and y_i are a estimate of s_i
  - **A** has at least two nonzero entries per row or per column;
  - f_i are differentiable invertible function;
  - each s_i accepts a density function that vanishes at one point at least.

Identifiability of the proposed causal model

- If $f_{2,1}$ is invertible, it is a special case of PNL mixing ICA model with $\mathbf{A}$=(1, 0; 1 1): $x_2 = f_{2,2}(f_{2,1}(x_1) + e_2)$

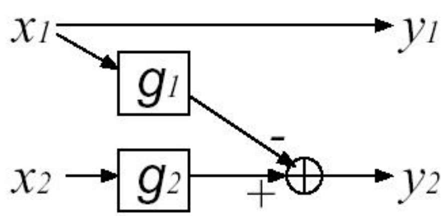
$$s_1 \triangleq f_{2,1}(x_1),\ s_2 \triangleq e_2 \Longrightarrow \begin{cases} x_1 = f_{2,1}^{-1}(s_1) \\ x_2 = f_{2,2}(s_1 + s_2) \end{cases}$$

- Identifiability: the causal relation between x_1 and x_2 can be uniquely identified if
  - x_1 and x_2 are generated according to this causal model with invertible $f_{2,1}$;
  - the densities of $f_{2,1}(x_1)$ and e_2 vanish at one point at least.
- If $f_{2,1}$ is not invertible, it is not PNL mixing ICA model. But it is empirically identifiable under very general conditions.

Identification Method

- Basic idea: which one of $x_1 \rightarrow x_2$ and $x_1 \rightarrow x_2$ can make the cause and disturbance independent ?
- Two-step procedure for each possible causal relation
  - Step 1: constrained nonlinear ICA to estimate the corresponding disturbance

Suppose $x_1 \rightarrow x_2$, i.e., $x_2 = f_{2,2}(f_{2,1}(x_1) + e_2)$. y_2 provides an estimate of e_2 , learned by minimizing the mutual information (which is equivalent to negative likelihood):



(y_2 produces an estimate of e_2)

$$I(y_1, y_2) = H(y_1) + H(y_2) + E\{\log|\mathbf{J}|\} - H(\mathbf{x})$$
$$= -E\log p_{y1}(y_1) - E\log p_{y2}(y_2) + E\{\log|\mathbf{J}|\} - H(\mathbf{x})$$

  - Step 2: uses independence tests to verify if the assumed cause and the estimated disturbance are independent

Special cases

$$x_i = f_{i,2}(\, f_{i,1}(pa_i) + e_i)$$

- If $f_{i,1}$ and $f_{i,2}$ are both linear
  - at most one of e_i is Gaussian: LiNGAM (linear, non-Gaussian, acyclic causal model, Shimizu et al., 2006)
  - all of e_i are Gaussian: linear Gaussian model
- If $f_{i,2}$ are linear: nonlinear causal discovery with additive noise models (Hoyer et al., 2009)

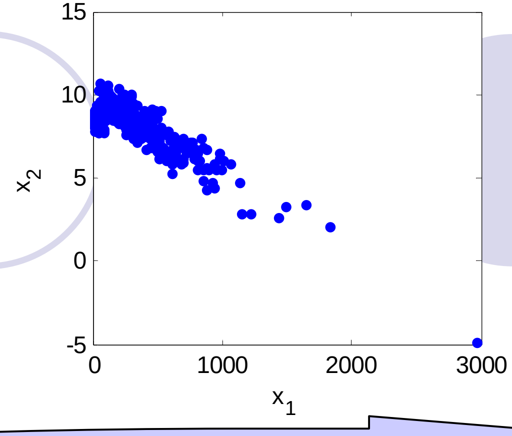
Experiments

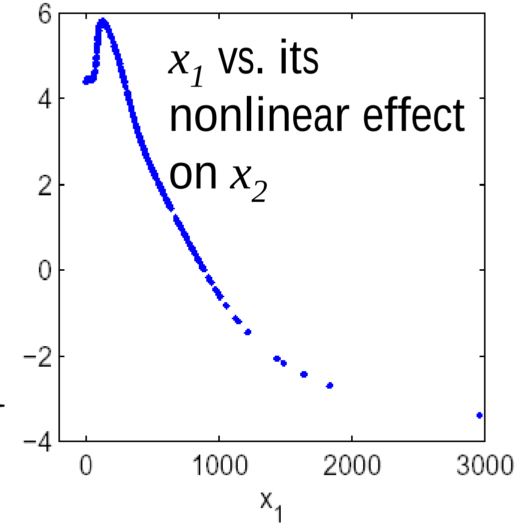
- For the CausalEffectPairs task in the Pot-luck challenge
  - Eight data sets
  - Each contains the realizations of two variables
  - Goal: to identify which variable is the cause and which one the effect
- Settings
  - g_1 and g_2 in constrained nonlinear ICA: modeled by multilayer perceptrons (MLP's) with one hidden layer
  - Different #hidden units (4~10) were tried; results remained the same
  - Kernel-based independence tests (Gretton et al., 2008) were adopted

Results

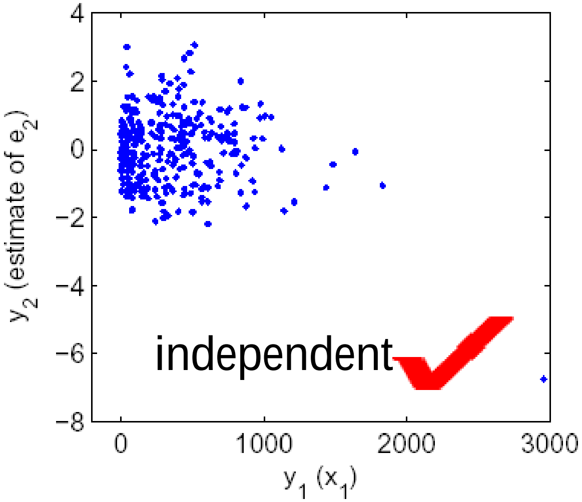
Data set	Result (direction of causality)	Remark
1	$x_1 \rightarrow x_2$	*Significant*
2	$x_1 \rightarrow x_2$	*Significant*
3	$x_1 \rightarrow x_2$	*Significant*
4	$x_2 \rightarrow x_1$	*not significant*
5	$x_2 \rightarrow x_1$	*Significant*
6	$x_1 \rightarrow x_2$	*Significant*
7	$x_2 \rightarrow x_1$	*Significant*
8	$x_1 \rightarrow x_2$	*Significant*

Data Set 1

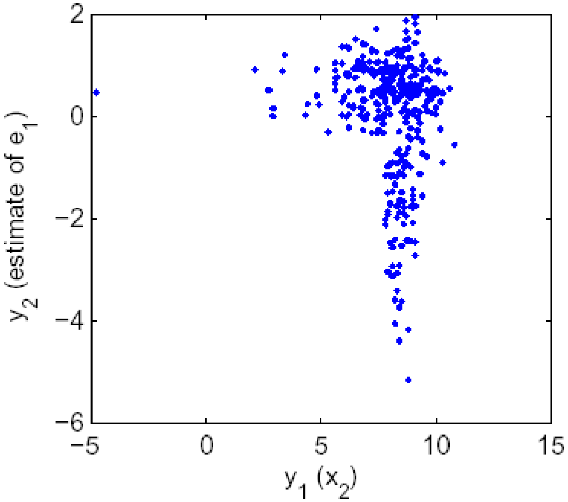


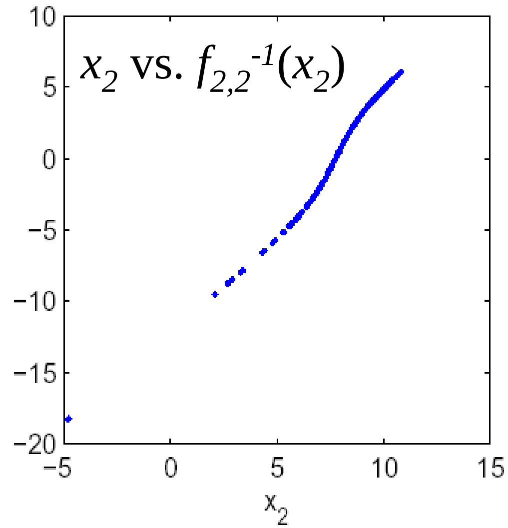


(a) y_1 vs y_2 under hypothesis $x_1 \rightarrow x_2$



(b) y_1 vs y_2 under hypothesis $x_2 \rightarrow x_1$

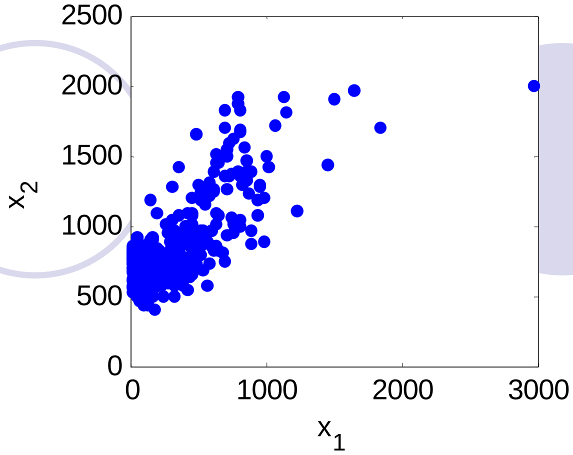




Independence test results on y_1 and y_2 with different assumed causal relations

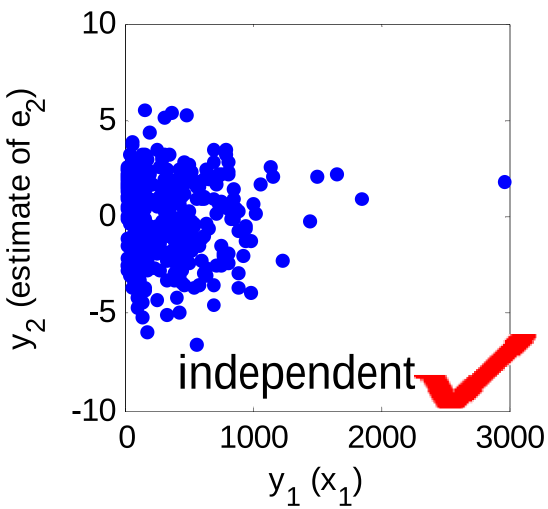
Data Set	$x_1 \rightarrow x_2$ assumed		$x_2 \rightarrow x_1$ assumed	
	Threshold ($\alpha = 0.01$)	Statistic	Threshold ($\alpha = 0.01$)	Statistic
#1	2.3×10^{-3}	1.7×10^{-3}	2.2×10^{-3}	6.5×10^{-3}

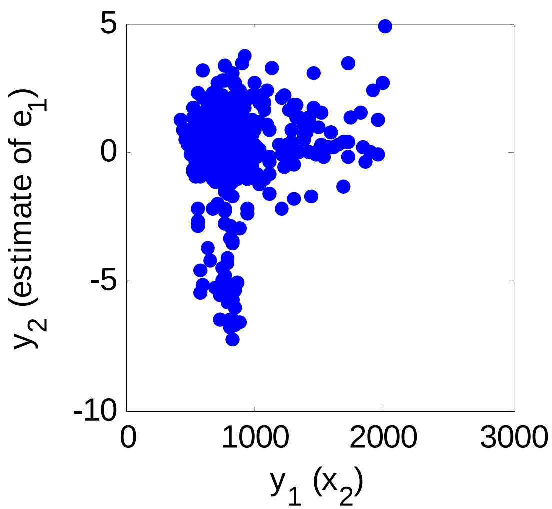
Data Set 2

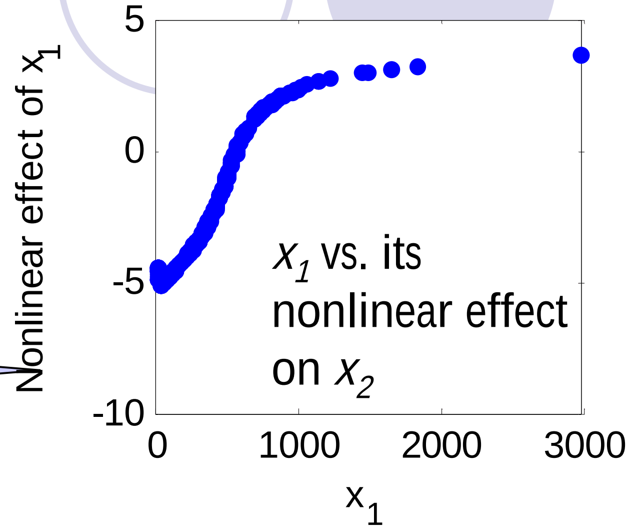


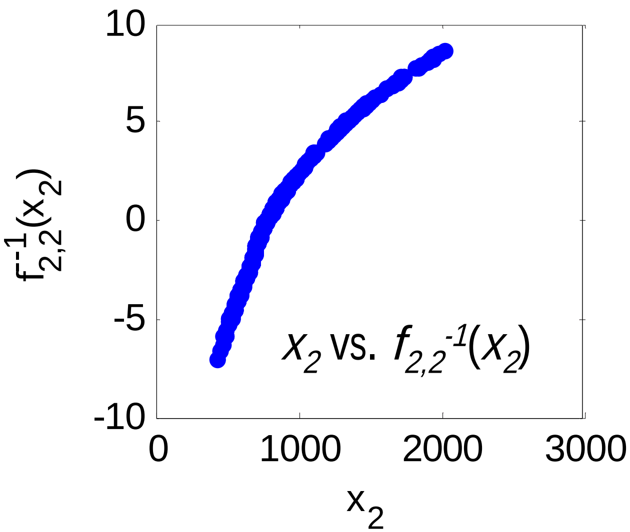
(a) y_1 vs y_2 under hypothesis $x_1 \rightarrow x_2$

(b) y_1 vs y_2 under hypothesis $x_2 \rightarrow x_1$

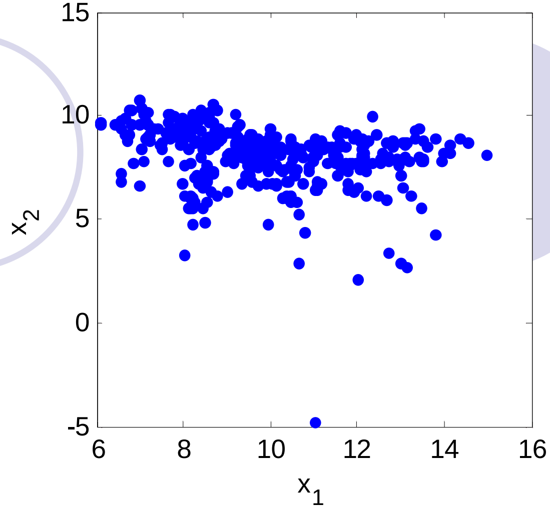


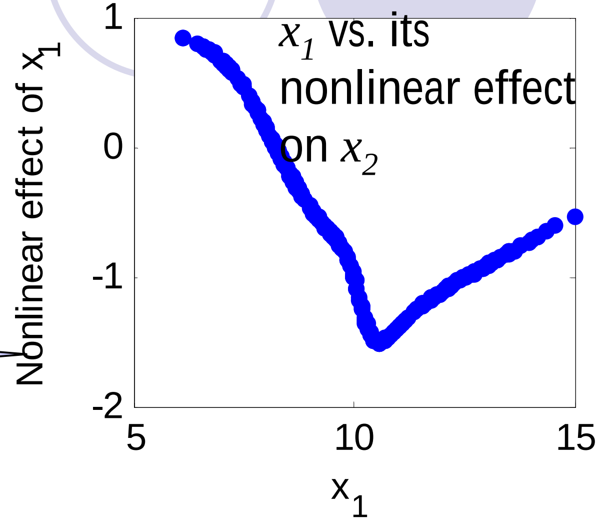






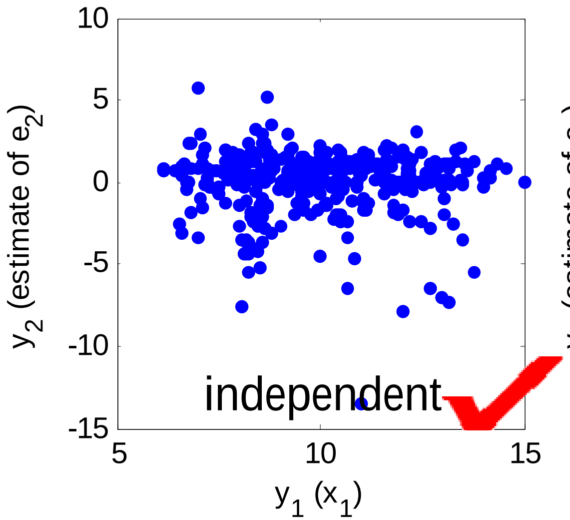
Data Set 3

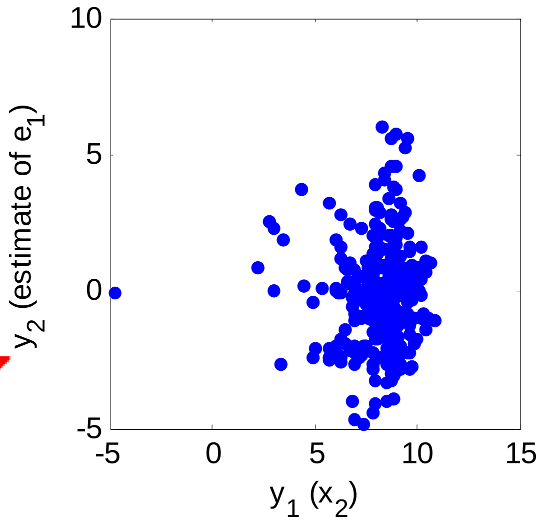


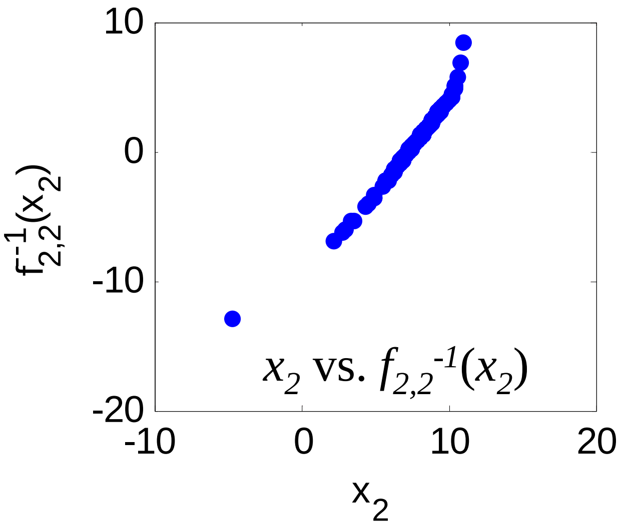


(a) y_1 vs y_2 under hypothesis $x_1 \rightarrow x_2$

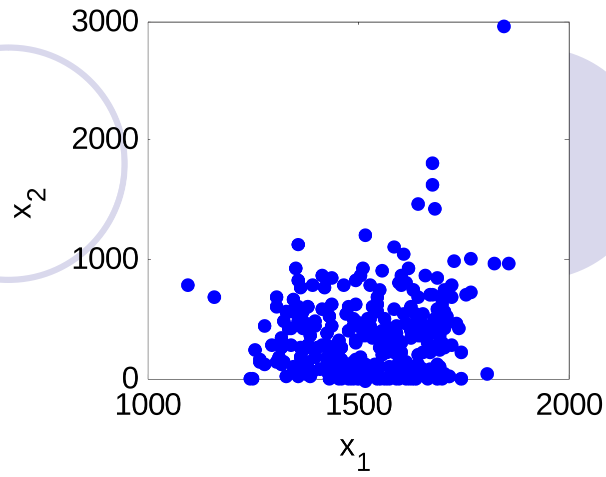
(b) y_1 vs y_2 under hypothesis $x_2 \rightarrow x_1$

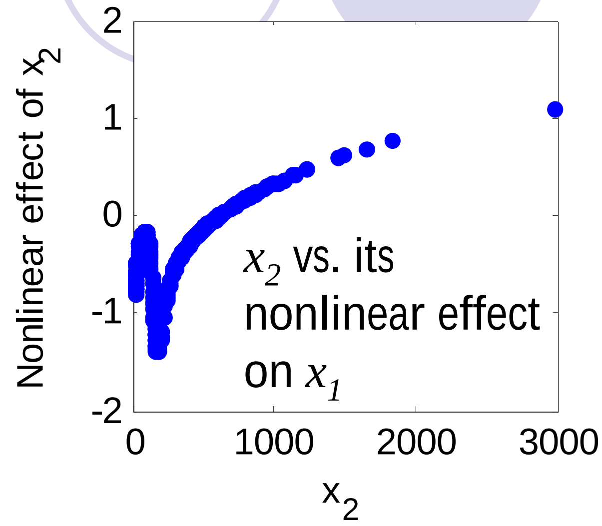




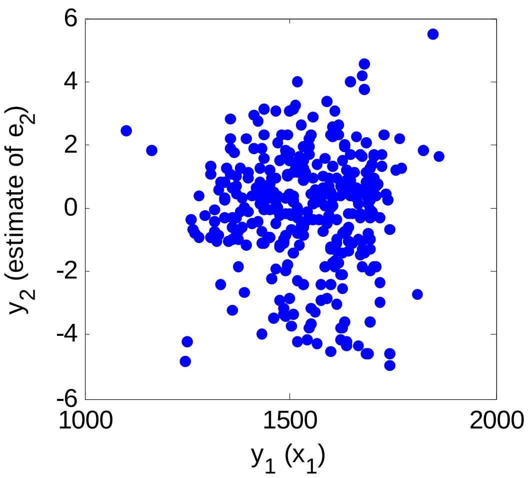


Data Set 4

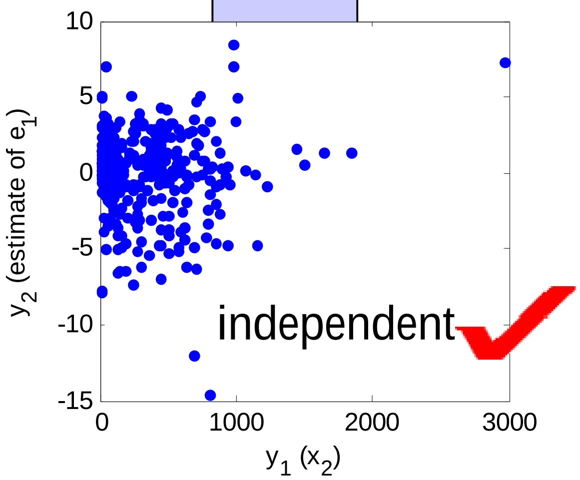


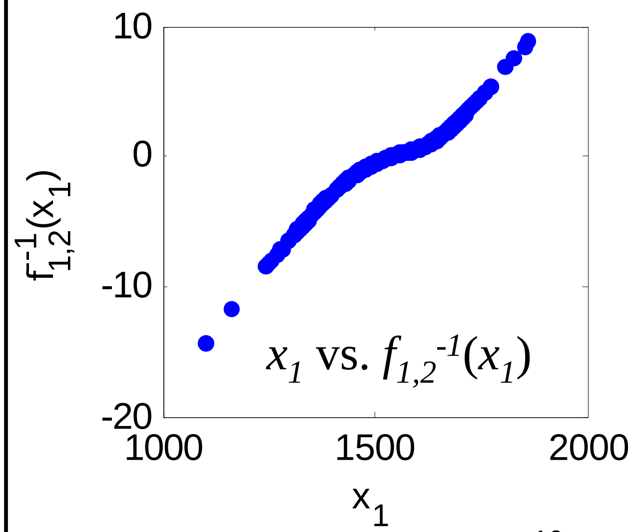


(a) y_1 vs y_2 under hypothesis $x_1 \rightarrow x_2$

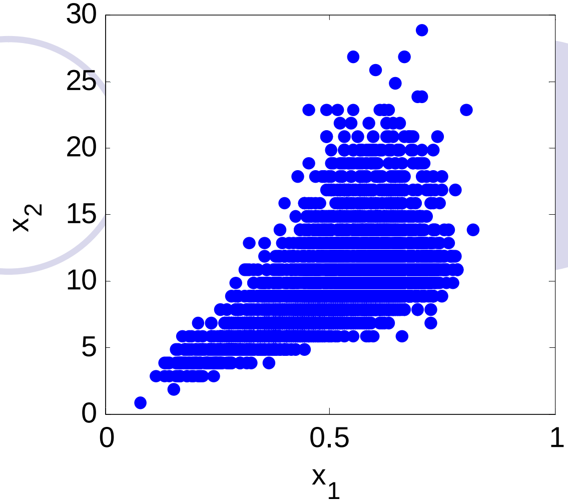


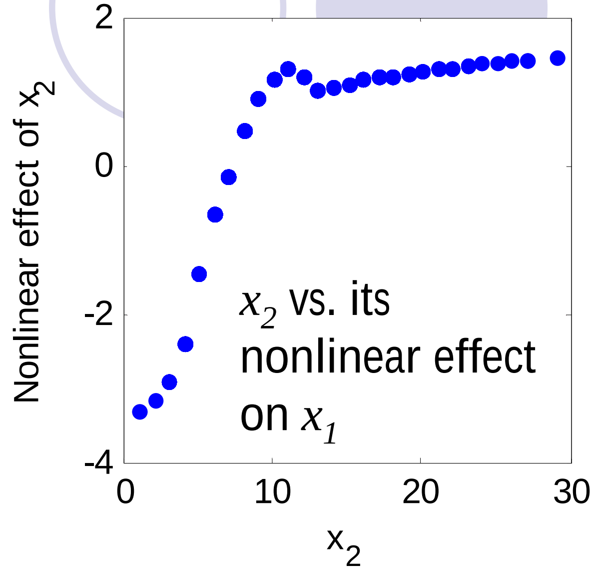
(b) y_1 vs y_2 under hypothesis $x_2 \rightarrow x_1$



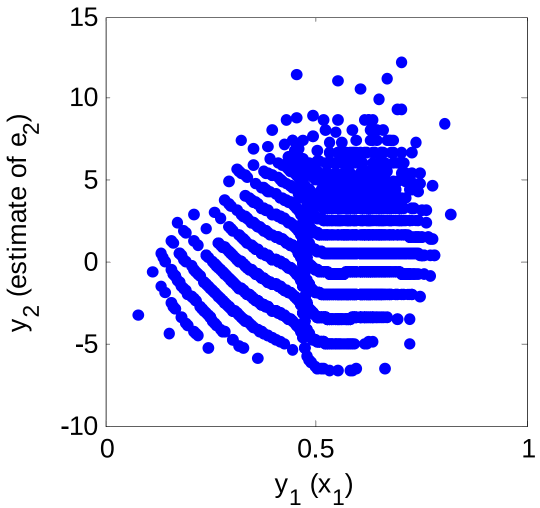


Data Set 5

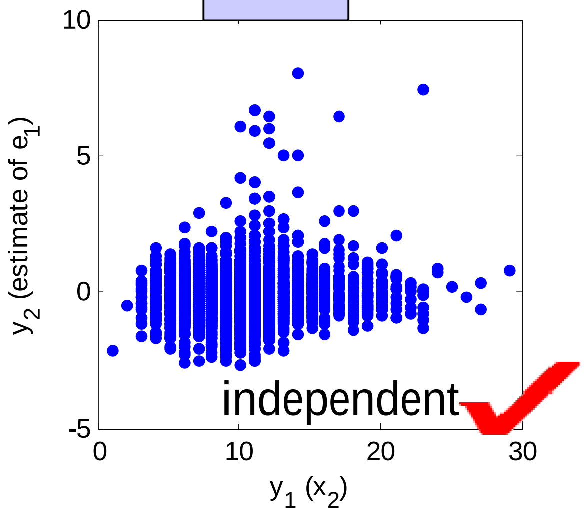


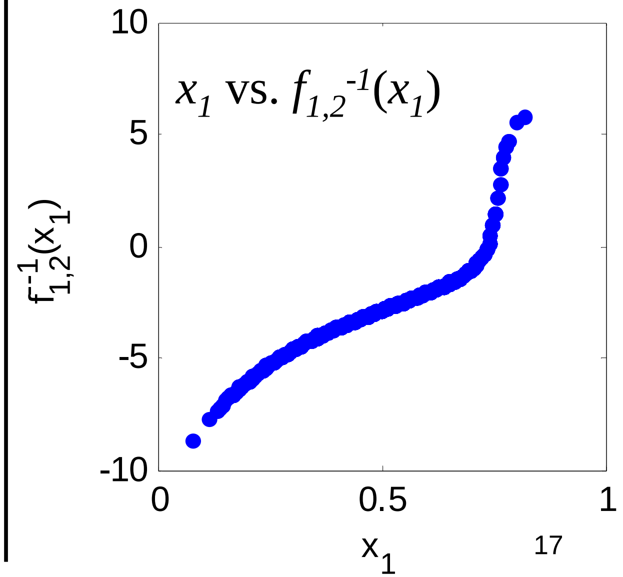


(a) y_1 vs y_2 under hypothesis $x_1 \rightarrow x_2$

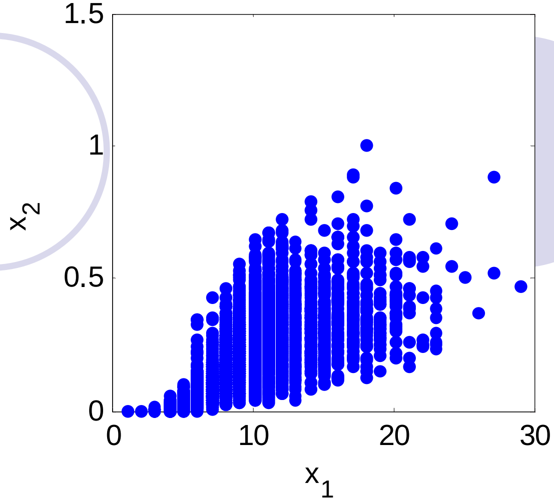


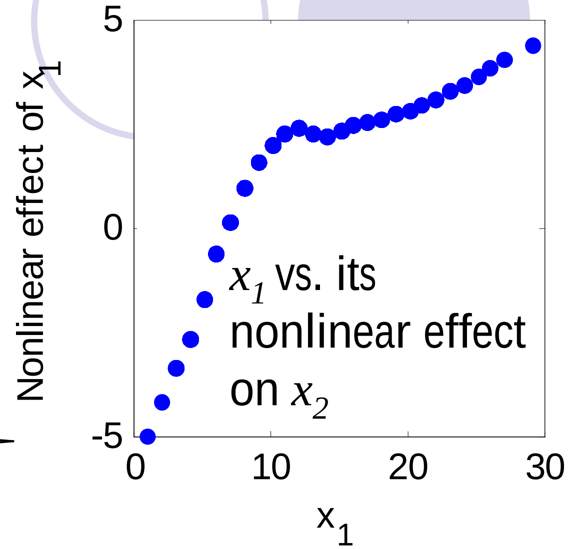
(b) y_1 vs y_2 under hypothesis $x_2 \rightarrow x_1$





Data Set 6

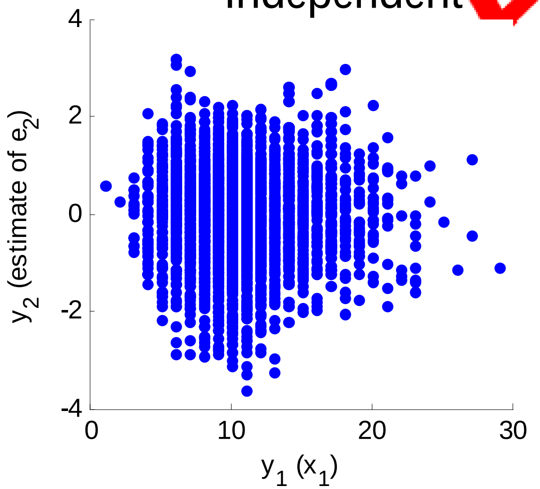


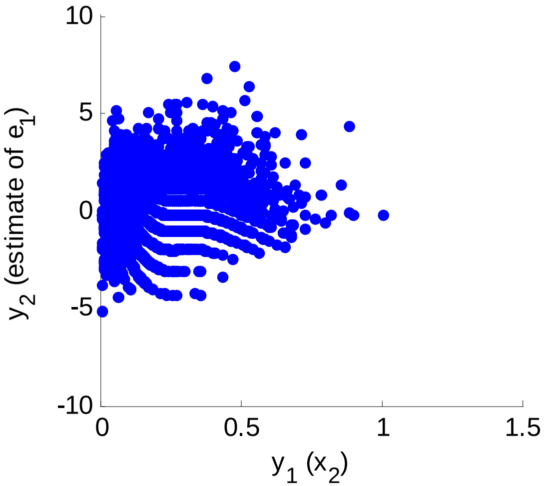


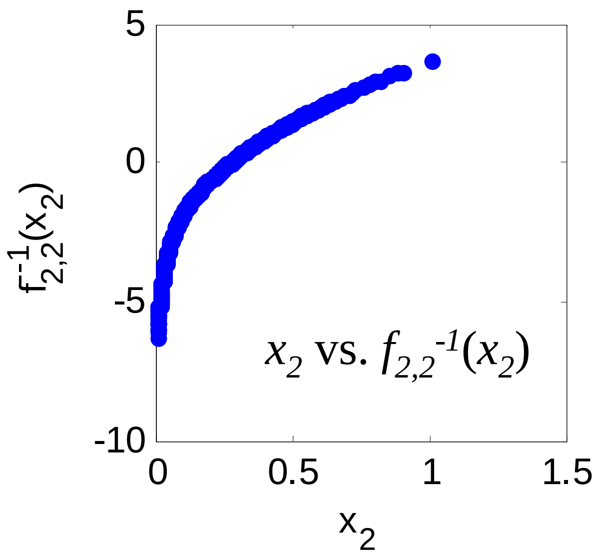
(a) y_1 vs y_2 under hypothesis $x_1 \rightarrow x_2$

(b) y_1 vs y_2 under hypothesis $x_2 \rightarrow x_1$

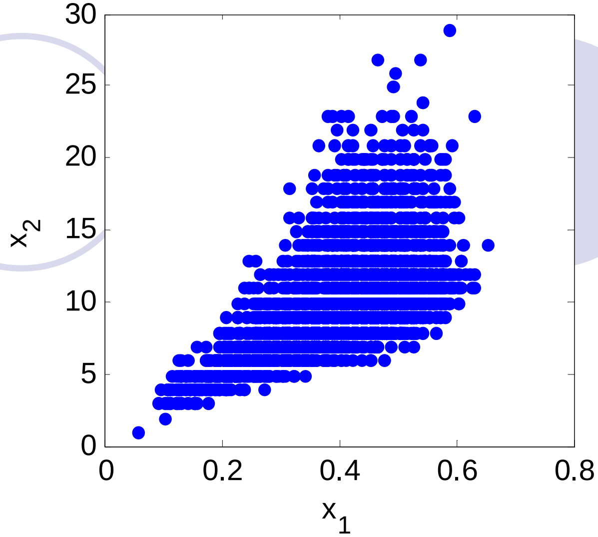
independent

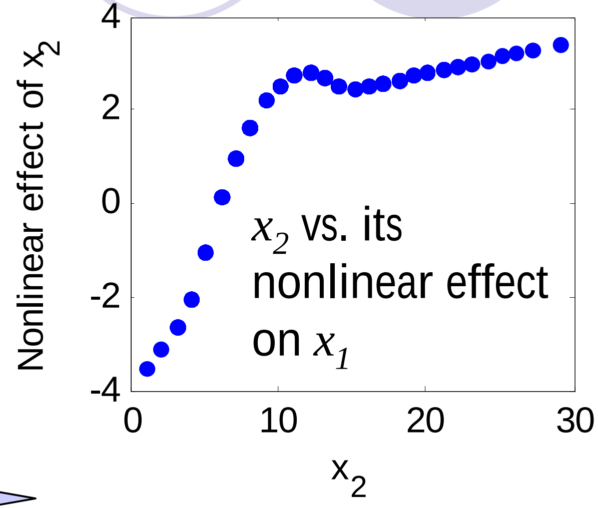




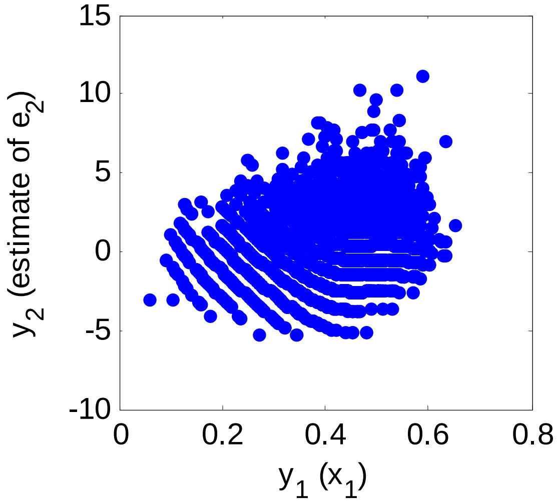


Data Set 7

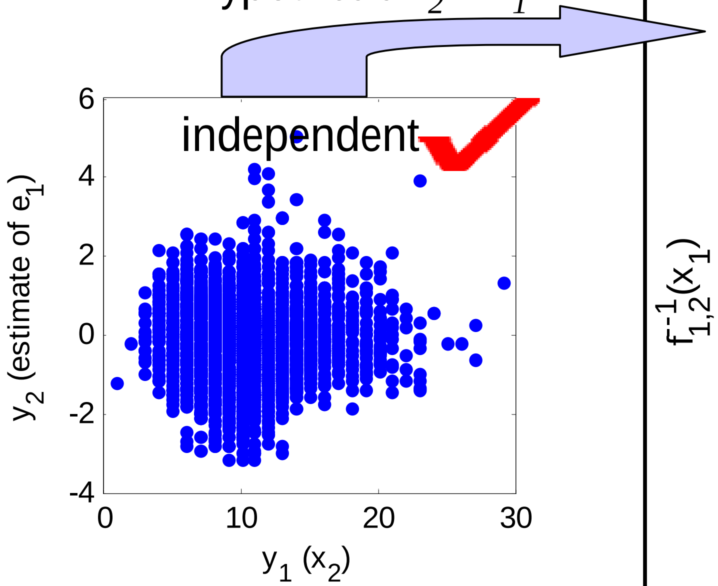


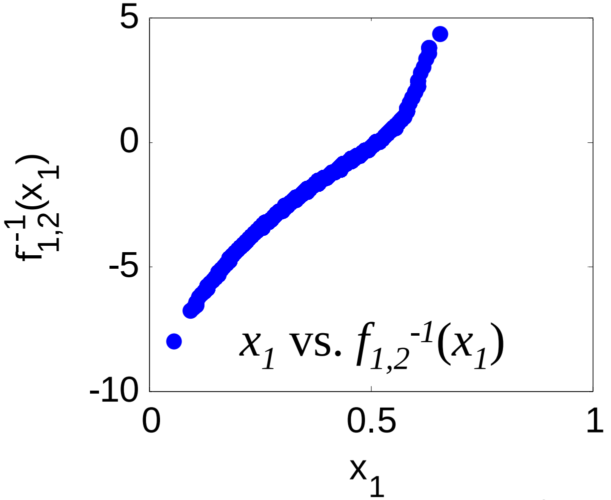


(a) y_1 vs y_2 under hypothesis $x_1 \to x_2$

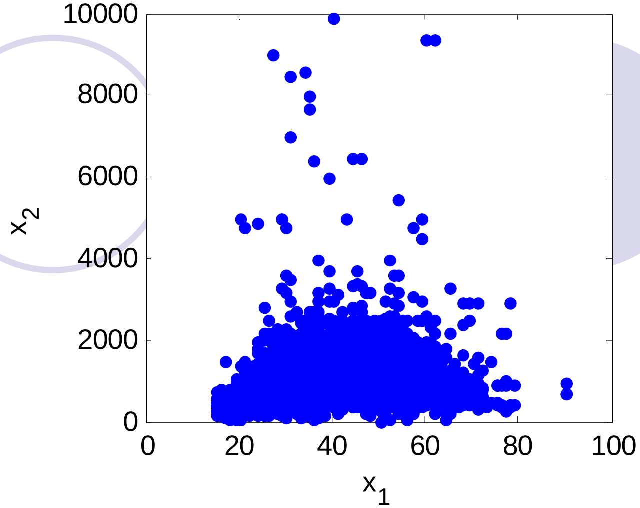


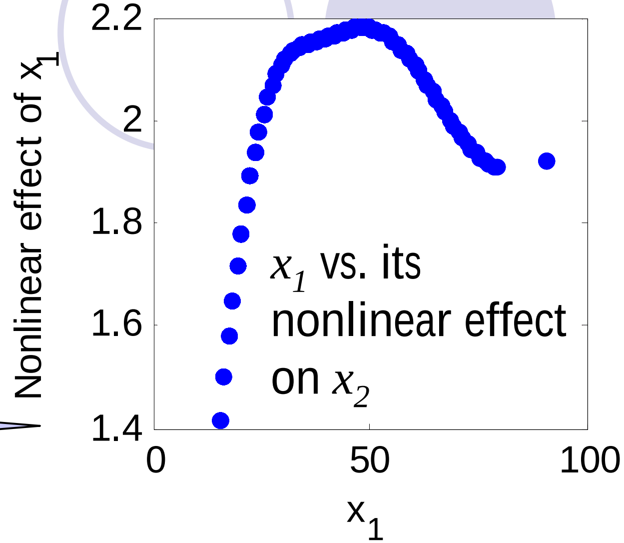
(b) y_1 vs y_2 under hypothesis $x_2 \to x_1$



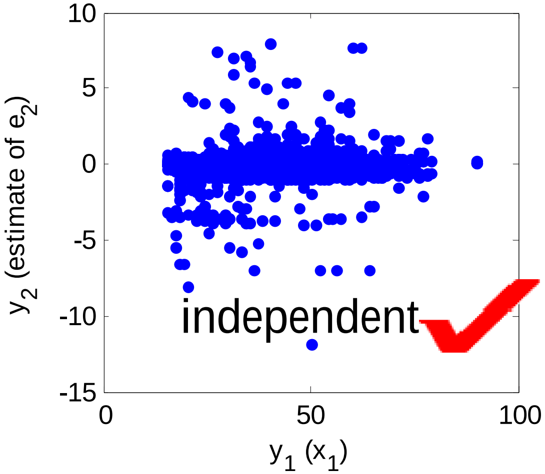


Data Set 8

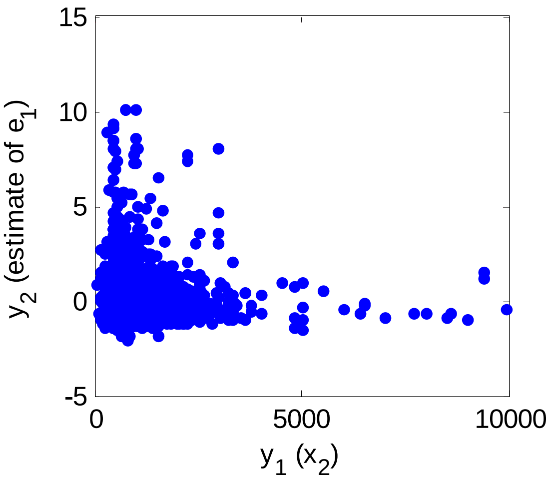


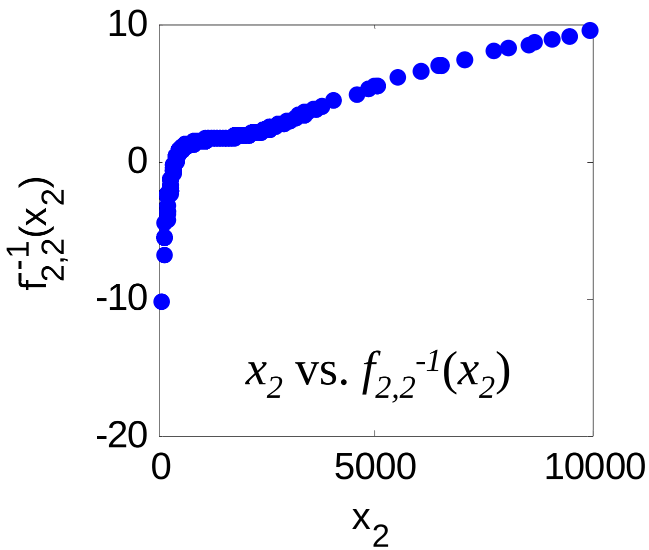


(a) y_1 vs y_2 under hypothesis $x_1 \rightarrow x_2$



(b) y_1 vs y_2 under hypothesis $x_2 \rightarrow x_1$





Conclusion

- Post-nonlinear acyclic causal model with inner additive noise
  - Very general: nonlinear effect of cause, noise effect & sensor nonlinear distortion
  - Still identifiable
- Experimental results on the CauseEffectPairs problem show its applicability for some practical problems
- Future work
  - Identifiability of this model in the general case of more than two variables
  - Efficient identification methods

References

- A. Taleb and C. Jutten. Source separation in post-nonlinear mixtures. *IEEE Trans. on Signal Processing*, 47(10): 2802—2820, 1999
- S. Shimizu, P.O. Hoyer, A. Hyvärinen, and A.J. Kerminen. A linear non-Gaussian acyclic model for causal discovery, *Journal of Machine Learning Research*, 7:2003--2030, 2006
- P.O. Hoyer, D. Janzing, J. Mooij, J. Peters, and B. Schölkopf. Nonlinear causal discovery with additive noise models. In *NIPS 21*. To appear, 2009
- A. Gretton, K. Fukumizu, C.H. Teo, L. Song, B. Schölkopf, and A.J. Smola. A kernel statistical test of independence. In *NIPS 20*, pages 585—592, 2008