

Created in July 2009

A direct method for LiNGAM

Shohei Shimizu, Osaka Univ.

Aapo Hyvarinen, Univ. Helsinki

Yoshinobu Kawahara, Tokyo Inst. Tech.

Takashi Washio, Osaka Univ.

Abstract

- Structural equation models (SEMs) are widely used in many empirical sciences (Bollen, 1989)
- A **non-Gaussian** framework has been shown to be useful for discovering SEMs (Shimizu, et al. 2006)
- Propose a **new** non-Gaussian estimation method
 - No algorithmic parameters
 - Guaranteed convergence in a fixed number of steps if the data strictly follows the model

Background

Linear **Non-Gaussian** Acyclic Model (**LiNGAM** model) (Shimizu et al. 2006)

- A SEM model, **identifiable** using non-Gaussianity
- Continuous observed random variables x_i
- Directed **acyclic** graph (DAG)
- Linearity
- Disturbances e_i are independent **and non-Gaussian**

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

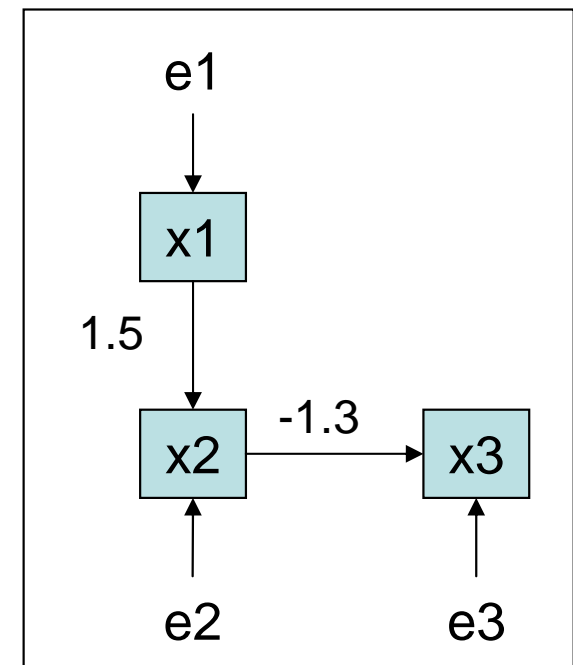
- $k(i)$ denotes an order of x_i
- \mathbf{B} can be permuted to be lower triangular by simultaneous equal row and column permutations

Example

- A three-variable model

$$\begin{aligned} x_1 &= e_1 \\ x_2 &= 1.5x_1 + e_2 \\ x_3 &= \quad \quad -1.3x_2 + e_3 \end{aligned} \quad \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \underbrace{\begin{bmatrix} 0 & 0 & 0 \\ 1.5 & 0 & 0 \\ 0 & -1.3 & 0 \end{bmatrix}}_{\mathbf{B}} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} + \begin{bmatrix} e_1 \\ e_2 \\ e_3 \end{bmatrix}$$

- Orders of variables:
 - $k(1) = 1, k(2) = 2, k(3) = 3$
 - x_2 can be influenced by x_1 , but never by x_3
- External influences:
 - x_1 is **equal to** e_1 and is **exogenous**
 - e_2 and e_3 are **errors**



Our goal

- We know
 - Data X is generated by $\mathbf{x} = \mathbf{B}\mathbf{x} + \mathbf{e}$
- We do **NOT** know
 - Connection strengths: b_{ij}
 - Orders: $k(i)$
 - Disturbances: e_i
- What we observe is data X only
- **Goal**
 - Estimate B and k using data X only!

Previous work

Independent Component Analysis

(Comon 1994; Hyvarinen et al., 2001)

$$\mathbf{X} = \mathbf{A}\mathbf{S}$$

- \mathbf{A} is an unknown square matrix
- S_i are independent **and non-Gaussian**
- Identifiable including the rotation (Comon, 1994)
- Many estimation methods
 - e.g., FastICA (Hyvarinen,99), Amari (99) and Bach & Jordan (02)

Key idea

- Observed variables x_i are linear combinations of **non-Gaussian independent** disturbances e_i

$$\mathbf{x} = \mathbf{B}\mathbf{x} + \mathbf{e}$$

$$\Rightarrow \mathbf{x} = (\mathbf{I} - \mathbf{B})^{-1} \mathbf{e}$$

$$= \mathbf{A}\mathbf{e} \quad \text{-- ICA!}$$

- ICA gives $\mathbf{W} = \mathbf{PDA}^{-1} = \mathbf{PD}(\mathbf{I} - \mathbf{B})$
 - P: Permutation matrix, D: scaling matrix
- Permutation indeterminacy in ICA can be solved
 - Can be shown that **the correct permutation is the only one which has no zeros in the diagonal** (Shimizu et al., UAI2005)

LiNGAM algorithm (Shimizu et al., 2006)

1. Do ICA (here, FastICA) and get $W = PD(I-B)$
2. Find a permutation \mathbf{P}_1 that gives no zeros on the diagonal. Then we obtain $D(I-B)$.

$$\hat{\mathbf{P}}_1 = \min_{\mathbf{P}_1} \sum_i \frac{1}{|(\mathbf{P}_1 \mathbf{W})_{ii}|}$$

3. Divide each row by its corresponding diagonal element. Then we get $I-B$, *i.e.*, B
4. Find a simultaneous row and column permutation Q so that the permuted B is as close as possible to be strictly lower triangular. Then we get $k(i)$.

$$\hat{\mathbf{Q}} = \min_{\mathbf{Q}} \sum_{i \leq j} (\mathbf{QBQ}^T)_{ij}$$

Potential problems of ICA-based LiNGAM algorithm

1. ICA is an iterative search method
 - May stuck in a local optimum if the initial guess or step size is badly chosen

2. The permutation algorithms are not scale-invariant
 - May provide different variable orderings for different scales of variables

A new method

DirectLiNGAM: A direct method ¹³

- **Basic idea:**

1. Find an exogenous variable, here x_1

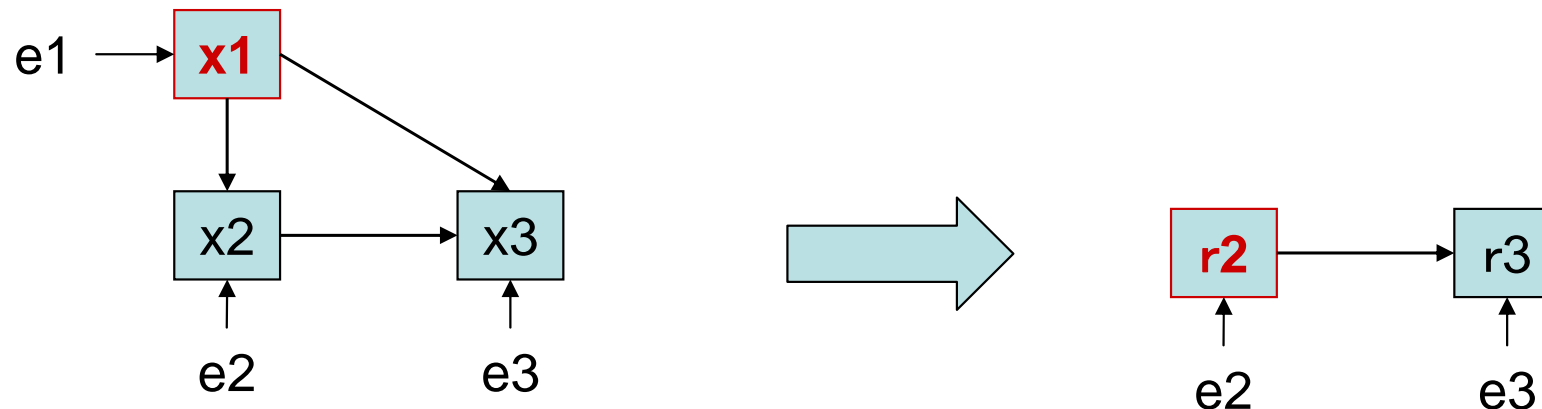
- Then we get $x_1 \rightarrow \{x_2, x_3\}$

- Remove the component x_1 from data by regression

2. Find an “exogenous” residual, here r_2

- Then we get $r_2 \rightarrow r_3$, *i.e.*, $x_2 \rightarrow x_3$

3. Eventually, we find $x_1 \rightarrow x_2 \rightarrow x_3$



Identification of an exogenous variable (two variable cases)

i) $x_1 (= e_1)$ is exogenous

$$x_2 := b_{21} x_1 + e_2$$

x_1 and e_2 are independent

x_2 and e_2 are NOT independent

$$r_2 = x_2 - \frac{\text{cov}(x_2, x_1)}{\text{var}(x_1)} x_1 = e_2$$

x_1 and r_2 are independent

ii) x_1 is NOT exogenous

$$x_1 := b_{12} x_2 + e_1$$

x_2 and e_1 are independent

x_1 and e_1 are NOT independent

$$\begin{aligned} r_2 &= x_2 - \frac{\text{cov}(x_2, x_1)}{\text{var}(x_1)} x_1 \\ &= \left\{ 1 - \frac{b_{12} \text{cov}(x_2, x_1)}{\text{var}(x_1)} \right\} x_2 - \frac{\text{cov}(x_2, x_1)}{\text{var}(x_1)} e_1 \end{aligned}$$

x_1 and r_2 are NOT independent

In general...

- **Lemma 1**: x_j and its residuals $r_i = x_i - \frac{\text{cov}(x_i, x_j)}{\text{var}(x_j)} x_j$ are independent for **all** $i \neq j \iff x_j$ is exogenous
- A technical assumption (correlation-faithfulness):
 - No combined effects of multiple pathways cancel out to be zeros and accidentally make correlations and partial correlations of variables equal to zeros

An independence measure

- Evaluate independence between a variable and a residual by a nonlinear correlation:

$$\left| \text{corr} \left\{ x_j, g \left(r_i^{(j)} \right) \right\} \right| \quad (g = \tanh)$$

- Taking the sum over all the residuals, we get:

$$T = \sum_{i \neq j} \left| \text{corr} \left\{ x_j, g \left(r_i^{(j)} \right) \right\} \right| + \left| \text{corr} \left\{ g \left(x_j \right), r_i^{(j)} \right\} \right|$$

- A fair comparison with the **FastICA**-based LiNGAM
- Can use more sophisticated measures as well (Bach & Jordan, 02; Gretton et al., 05)
- Can identify an exogenous variable by finding a variable that minimizes T

Remove the component

- **Lemma 2**: A LiNGAM model holds for the residuals r_i
- **Corollary 1**: The ordering of **the residuals** r_i is **equivalent to** that of corresponding **original observed variables** x_i

$$\mathbf{x} = (\mathbf{I} - \mathbf{B})^{-1} \mathbf{e}$$

$$= \mathbf{A} \mathbf{e}$$

$$\begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ a_{21} & 1 & 0 \\ a_{31} & a_{32} & 1 \end{bmatrix} \begin{bmatrix} e_1 \\ e_2 \\ e_3 \end{bmatrix}$$

$$x_2 \rightarrow x_3$$



$$\begin{bmatrix} r_2 \\ r_3 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ a_{32} & 1 \end{bmatrix} \begin{bmatrix} e_1 \\ e_2 \end{bmatrix}$$

$$r_2 \rightarrow r_3$$

Simulations

- Number of variables $p = 10, 20, 50, 100$
- Sample size $n = 30, 50, 80, 200, 500, 1000, 2000$
- Generated 501 datasets under each combination of p and n , randomizing:
 - Network structures
 - Connection strengths
 - Various non-Gaussian distributions including super-Gaussian and sub-Gaussian
- Note that nothing done to make the parameter values satisfy the correlation-faithfulness assumption

Evaluation measure

- For **correct** variable orders $k(i)$, the permuted **true** \mathbf{B} is (strictly) **lower-triangular**

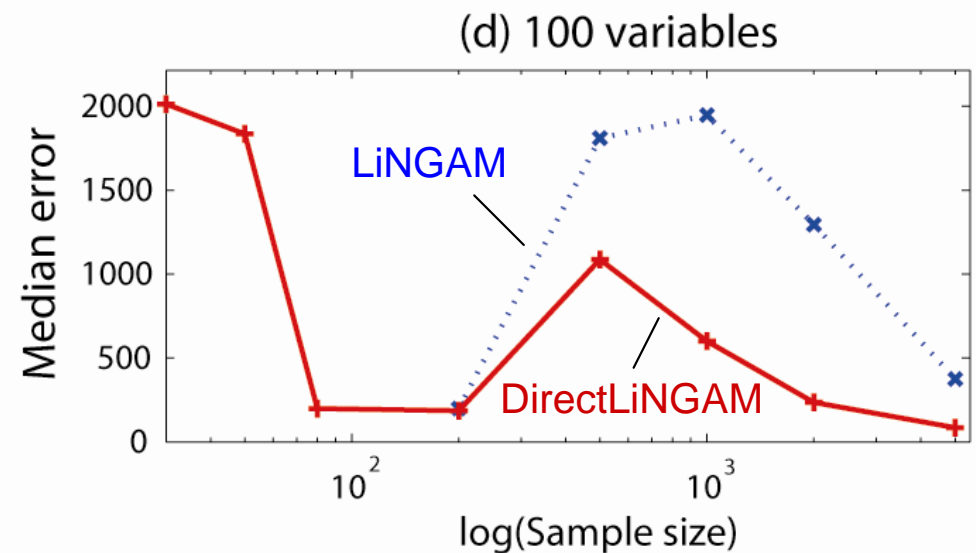
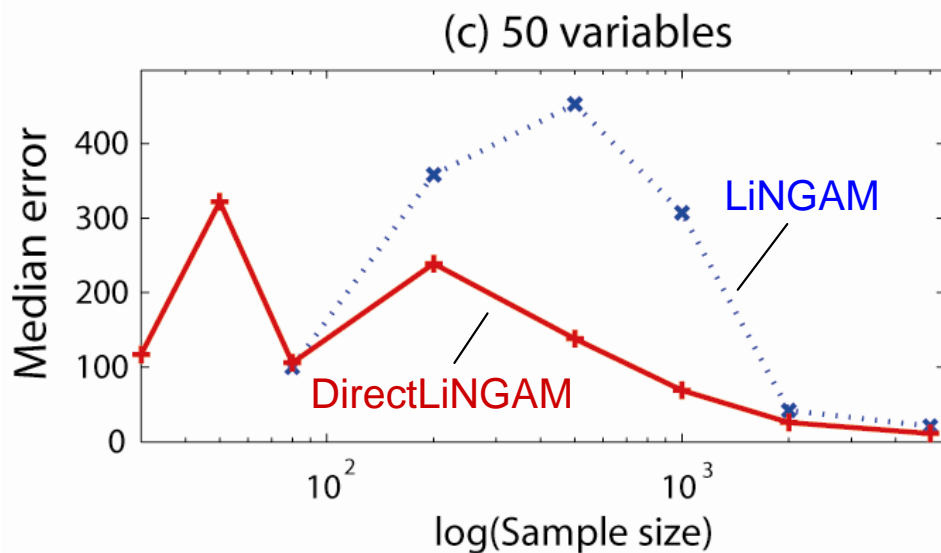
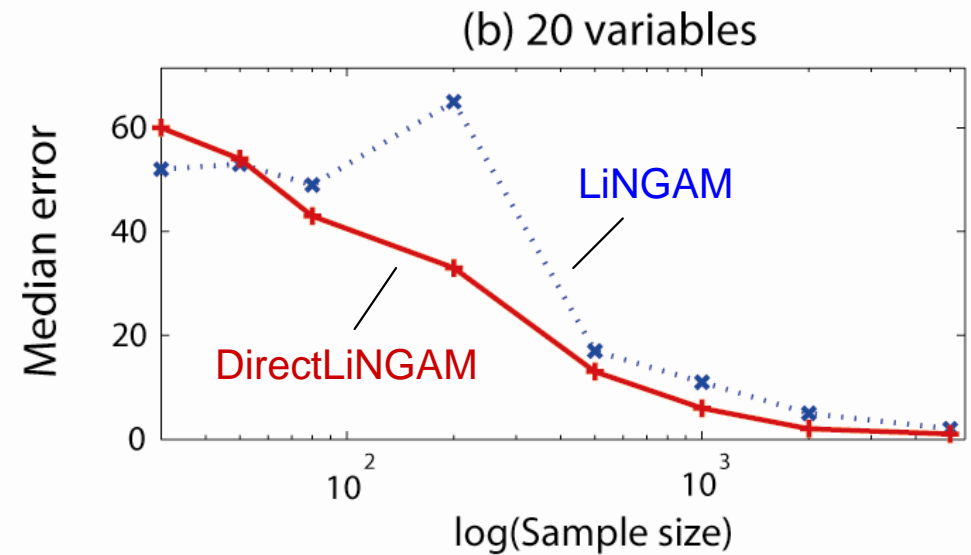
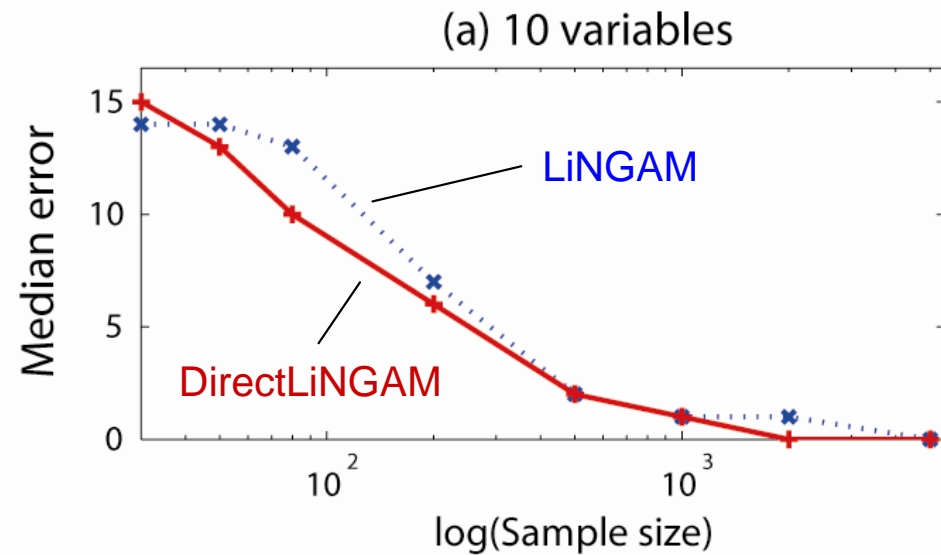
$$\mathbf{B}_{permuted}^{True} = \begin{bmatrix} 0 & 0 & 0 \\ * & 0 & 0 \\ * & * & 0 \end{bmatrix}$$

- More **errors** in estimated orders, it has **more non-zero** elements in its upper triangular part

$$\mathbf{B}_{permuted}^{True} = \begin{bmatrix} 0 & \underline{0.5} & 0 \\ * & 0 & \underline{0.8} \\ * & * & 0 \end{bmatrix}$$

- Count **how many non-zeros** in its upper triangular part

Number of non-zero elements (errors) v.s. sample size



$n = 30, 50, 80, 200, 500, 1000, 2000$

Conclusion

- Discovery of SEMs from data is an important topic of current research
- Use of non-Gaussianity is useful for model identification
- DirectLiNGAM is a direct method
 - No algorithmic parameters
 - Guaranteed convergence in a fixed number of steps (the number of variables) if the data strictly follows the model

Reference

- S. Shimizu, A. Hyvärinen, Y. Kawahara, and T. Washio.
A direct method for estimating a causal ordering in a linear non-Gaussian acyclic model.
In Proceedings of the 25th Conference on Uncertainty in Artificial Intelligence (UAI2009), 2009.