Osaka University Short-term Student Program

Independent Study-B Report

Spring2016

The Pairwise LiNGAM in Groups of Variables

By Using Boston Housing Data

Student Name: Chia-Hui Shen

Student ID Number: 09Z15018

Supervisor: Prof. Shimodaira & Prof. Shimizu
Shimodaira Laboratory

Contents

1 Introduction of Basic Linear Non-Gaussian Acyclic Model

1.1 Linear Non-Gaussian Acyclic Model

The Linear Non-Gaussian Acyclic Model is a way to estimate the connection strength and causal order based on only observed data. In contrast to the other methods, the LiNGAM approaches (Shimizu et al, UAI2005, 2006 JMLR) [1] can replace the Gaussian assumption by Non-Gaussian assumption. Moreover, we can identify the connection strength and structure.

A model with assumptions, the error variables are independent, non-Gaussian probability density function,and the causal relation is acyclic, is called the model Linear Non-Gaussian Acyclic Model (LiNGAM).

1.2 Model

$$x_i = \sum_{k(j)<k(i)} b_{ij} x_j + e_i \quad (1)$$

x_i : $observed\ continuous\ random\ variables$

b_{ij} : $the\ strengths\ of\ the\ causal\ connections\ from\ x_j\ to\ x_i$ with a causal ordering denoted by $k(i)$

e_i : $independent\ and\ non-Gaussian\ probability\ density\ functions, i = 1,2, \dots\dots, p$

The LiNGAM model with matrix form is

$$\boldsymbol{x} = \mathbf{B}\boldsymbol{x} + \boldsymbol{e} \quad (2)$$

,where $\boldsymbol{x}$ and $\boldsymbol{e}$ collect the observed variable x_i and e_i, and $\mathbf{B}$ collects the causal connection strength b_{ij}, respectively.

The LiNGAM is identifiability which means that the connection strength matrix B can be uniquely identified based only on the observed data matrix $\boldsymbol{x}$. To show the property of identifiability, the following graphs and simulations showed the difference between two models, which are the same strength of the causal connection but different arrows of causal connection. First, generating 2 observed random variables with sample size 5000 from Model (1). Assuming the strength of the causal connection is 0.8, the non-Gaussian possibility density function of error is assumed uniform distribution, and arrows of causal connection are opposite. Second, plotting the simulated data where x_1 lies on x-axiom and x_2 lies on y-axiom. The

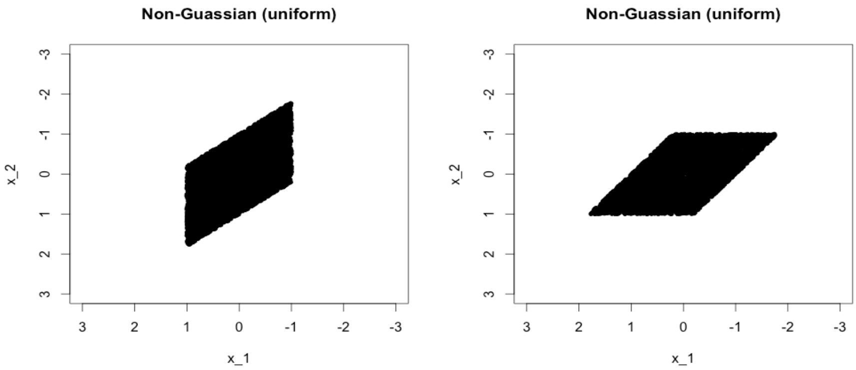
[1] Pairwise Likelihood Ratios for Estimation of Non-Gaussian Structural Equation Models, Aapo Hyvärinen (2013) Journal of Machine Learning Research 14 (2013) 111-152 Submitted 10/11; Revised 8/12; Published 1/13

Figure (1). showed the different shape of graphs and 5 simulated observations. Even though they have equal strength of the causal connection, both of them still have different arrows of causal connection and shapes. That is because they have the property of identifiability.

$x_1 \xrightarrow{0.8} x_2$ $\qquad$ $x_1 \xleftarrow{0.8} x_2$

$x_1 = e_1$ $\qquad$ $x_2 = e_2$

$x_2 = 0.8x_1 + e_2$ $\qquad$ $x_1 = 0.8x_2 + e_1$



x_1	x_2
-0.70693789	-1.17544128
0.41184763	0.89560796
-0.65912372	-0.07186119
0.09126919	-0.77004160
0.49016526	-0.08876460

x_1	x_2
-0.11643240	-0.64147063
0.42686002	-0.24151070
-0.02792522	-0.49100658
1.47609301	0.73137390
0.03694876	-0.08013971

Figure (1).

2 Basic LiNGAM in Boston Housing Data

2.1 Introduction of Boston Housing Data

Boston Housing data contains information collected by the U.S Census Service concerning housing in the area of Boston Mass. In this dataset, they measured multiple factors, which may affect the housing prices in different neighborhoods in the Boston area. The dataset has 506 observations with 14 variables. The 14 variables are CRIM, ZN, INDUS, CHAS, NOX, RM, AGE, DIS, RAD, TAX, PTRATIO, B, LSTAT, MEDV, respectively.

Here are the 14 variables in Detail:

CRIM: per capita crime rate by town

ZN: proportion of residential land zoned for lots over 25,000 sq. ft

INDUS: proportion of non-retail business acres per town

CHAS: Charles River dummy variable (= 1 if tract bounds river; 0 otherwise)

NOX: nitric oxides concentration (parts per 10 million)

RM: average number of rooms per dwelling

AGE: proportion of owner-occupied units built prior to 1940

DIS: weighted distances to five Boston employment centers

RAD: index of accessibility to radial highways

TAX: full-value property-tax rate per $10,000

PTRATIO: pupil-teacher ratio by town

B: 1000(Bk - 0.63)^2 where Bk is the proportion of blacks by town

LSTAT: % lower status of the population

MEDV: Median value of owner-occupied homes in $1000's

```
#Original data
#CRIM     ZN     INDUS CHAS NOX    RM     AGE   DIS    RAD  TAX   PTRATIO  B     LSTAT MEDV
#0.00632  18.00  2.310  0  0.5380 6.5750 65.20  4.0900  1  296.0  15.30 396.90  4.98  24.00
#0.02731   0.00  7.070  0  0.4690 6.4210 78.90  4.9671  2  242.0  17.80 396.90  9.14  21.60
#0.02729   0.00  7.070  0  0.4690 7.1850 61.10  4.9671  2  242.0  17.80 392.83  4.03  34.70
#0.03237   0.00  2.180  0  0.4580 6.9980 45.80  6.0622  3  222.0  18.70 394.63  2.94  33.40
```

Table (1). 4 samples of Boston Housing Data

2.2 Estimate the causal effect

To know the relation between those 14 variables, I used the Pairwise LiNGAM[2] (Aapo Hyvärinen, 2013) to estimate the causal connection strength.

Pairwise LiNGAM is an approach, which can measure the causal direction and causal effect between two non-Gaussian random variables. This approach uses the likelihood ratio between two non-Gaussian random variables to estimate the causal direction and causal effect. Among this method, the log likelihood is given by (Hyvärinen et al., 2010) as $\log L(x \to y) = [\sum_t G_x(X_t) + G_d(\frac{y_t - \rho x_t}{\sqrt{1-\rho^2}})] - T\log(1-\rho^2)$ and the likelihood ratio is $R = \frac{1}{T}\log L(x \to y) - \frac{1}{T}\log L(x \leftarrow y)$. Hence, if R is positive, the causal direction will be x to y. On the other hand, if R is negative, the causal direction will be y to x.

2.3 Problem in Boston Housing Data

In the Boston Housing Dataset, there is a categorical variable—CHAS. However, if we want to apply the LiNGAM on the Boston housing data to estimate the causal effect, the variables must be continuous random variables under LiNGAM assumptions. Therefore, we need to figure out how to let the Boston Housing Data be workable on LiNGAM.

According to paper--Bayesian Networks for Variable Groups (Pekka Parviainen, 2015)[3]--, they used another method to estimate the casual

---

[2] Pairwise Likelihood Ratios for Estimation of Non-Gaussian Structural Equation Models, Aapo Hyvärinen (2013). Journal of Machine Learning Research 14 (2013) 111-152 Submitted 10/11; Revised 8/12; Published 1/13

[3] Bayesian Networks for Variable GroupsP Parviainen, S Kaski (2015)arXiv preprint arXiv:1508.07753

direction in Boston Housing Data. In this paper, they found that the variable-CHAS—did not have any causal effect to other variables. Although they applied the different method to estimate the cauls direction, I still can estimate the causal effect based on such result. Therefore, I eliminate the CHAS variable from dataset. The dataset become 506 observations with 13 variables. In the following application, I will continue to use dataset which have only 13 variables to apply on the research.

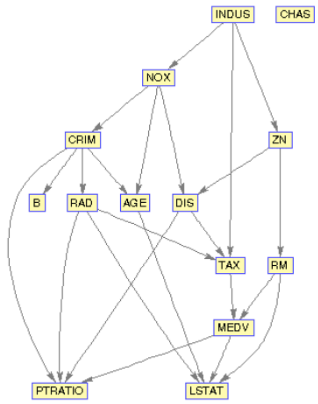


Figure (2).

2.4 Measuring the causal direction

Likelihood Ratio Matrix

	DIS	RAD	ZN	INDUS	RM	AGE	B	LSTAT	CRIM	NOX	PTRATIO	MEDV	TAX
DIS	0	-0.233	-0.46	0.0113	0.0084	-0.161	-0.294	-0.016	-0.09	0.0639	0.0605	0.0924	-0.132
RAD	0.2333	0	-0.092	0.6165	0.055	0.2195	-0.632	0.3529	0.9484	0.5884	0.2496	0.3369	0.2212
ZN	0.46	0.0924	0	0.4268	0.2175	0.4996	-0.056	0.3299	0.0471	0.4585	0.2179	0.265	0.0988
INDUS	-0.011	-0.616	-0.427	0	-0.009	-0.02	-0.422	0.0462	0.2321	0.001	0.1091	0.191	-0.444
RM	-0.008	-0.055	-0.218	0.009	0	-0.029	-0.018	-0.029	-0.053	0.0186	0.0041	-0.06	-0.013
AGE	0.1612	-0.219	-0.5	0.0202	0.0286	0	-0.235	0.1121	-0.033	0.0966	0.1099	0.2509	-0.073
B	0.2936	0.6317	0.0562	0.4223	0.018	0.2352	0	0.463	0.3904	0.4581	0.158	0.4952	0.6688
LSTAT	0.0163	-0.353	-0.33	-0.046	0.0293	-0.112	-0.463	0	0.1724	0.0248	-0.026	0.0852	-0.215
CRIM	0.09	-0.948	-0.047	-0.232	0.0533	0.0327	-0.39	-0.172	0	-0.24	0.074	0.3797	-0.705
NOX	-0.064	-0.588	-0.458	-1E-03	-0.019	-0.097	-0.458	-0.025	0.2402	0	0.0332	0.1401	-0.36
PTRATIO	-0.06	-0.25	-0.218	-0.109	-0.004	-0.11	-0.158	0.0264	-0.074	-0.033	0	-0.01	-0.081
MEDV	-0.092	-0.337	-0.265	-0.191	0.0604	-0.251	-0.495	-0.085	-0.38	-0.14	0.0099	0	-0.279
TAX	0.1319	-0.221	-0.099	0.4437	0.0129	0.0725	-0.669	0.2153	0.7047	0.3595	0.0814	0.2788	0

Table(2).

In the Table (2), we can see that the value in the 2^{nd} row and 1^{st} column is 0.233. Because 0.233 is positive, we can say that RAD has the causal effect to DIS. Similarly, the value in the 4^{th} row and 1^{st} column is -0.011which is negative. Therefore, we can say that DIS has the causal effect to INDUS.

3 LiNGAM in Groups of Variables

Although we have already used Pairwise LiNGAM to measure the causal direction between 13 random variables of Boston Housing Data, there is still left something to be desired.

If we just consider the causal effect between each single variable, some information will loss. Take the fMRI data as an example, in previous studies, aggregated data in fMRI – measurements of brain activities – have already been used to find the causal relations of single variables using the pairwise-LiNGAM. However, some information that we may not notice will loss because variables that form brain regions sometimes are heuristic averaged to single variables and this could lead to a serious bias when uncovering connections. Therefore, I want to extend the LiNGAM to the groups of variables and to receive the information that we uncover.

3.1 Groups of Variables

We denote group as the set of variables which all belong to the same random vector and denote the variable as a single random variable which belongs to one of the groups. How groups be organized is based on the background knowledge. Therefore, I will not discuss which variables belong to what groups in this report.

Take the Boston Housing Data as an example, according to Pekka Parviainen (2015)[4]; they divided the variables into 9 groups. Some of groups have only one variable but this will not affect the application of groups.

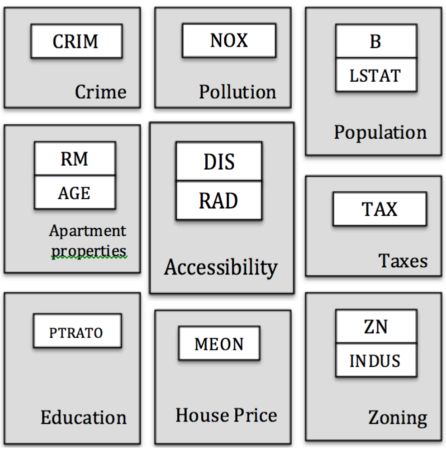


Figure (3). Group information of Boston Housing Data

[4] Bayesian Networks for Variable GroupsP Parviainen, S Kaski (2015)arXiv preprint arXiv:1508.07753

3.2 Model

Let χ_g denote group g, and let the random vector $\chi_g = (\chi_1^{(g)}, \dots, \chi_{n_g}^{(g)})^T$ which has n_g variables in group g. Assume the group can be arranged in the causal order $K = (k_1, \dots, k_G)$. All variables in χ_g are observed, and the grouping of these variables is known.

$$x_{k_i} = \sum_{j<i} B_{k_i k_j} x_{k_j} + e_{k_i},\ i = 1, \dots, G \qquad (3)$$

$B_{k_i k_j}$: aribtrary matrics of dimension $n_{k_i} \times n_{k_j}$, containing the dircet effects from groups x_{k_i} to x_{k_j}.

e_{k_i} : assume to be zero mean, and mutually independent over groups, but allowed to be dependent within each group.

4 Simulation in Groups of Variables

To know more about how groups of variables work on different LiNGAM methods, here are two simulations to display the comparison of three LiNGAM methods with different number of variables in each group. The simulation is based on the Entner's method (2012)[5]. According this paper, they estimate the causal order by regressing out the exogenous group of all groups. The exogenous group mean that if X_i is exogenous group, then for all X_j, which i is not equal to j, $B_{j,i}$ are zero.

About the simulations, the first simulation Figure (4) is 5 grouped variables with 6 variables in each group and Figure (5) is 5 grouped variables with 12 variables in each group. We simulated three different sample sizes: 200, 500, 1000 and repeat 100 times. Finally, observe the error rate of result in different LiNGAM methods. The three LiNGAM methods are Group Directed LiNGAM, Pairwise LiNGAM, and Group ICA LiNGAM, respectively.

The simulation process is that at first, I generated a random dataset based on the known groups information under non- Gaussian distribution. In the generation data, I have already given a correct causal ordering for each dataset. After that, I only applied the information of random dataset to estimate the causal order by using LiNGAM methods. Finally, comparing the result with correct causal

[5] Estimating a causal order among groups of variables in linear models, Doris Entner(2012) ICANN'12 Proceedings of the 22nd international conference on Artificial Neural Networks and Machine Learning- Volume PartII Pages84-91

ordering. If the result is the not equal to the real causal ordering, it will be an error. After repeating the 100 times, calculating the proportion of errors from every LiNGAM method in each simulation. At last, plotting the group shows in Figure (4) and Figure (5). From the Figures, we can know that the group ICA LiNGAM always has highest error rate. When the sample size increased, the error rate decreased in the Group Directd LiNGAM. About the Pairwise LiNGAM method, Pairwise LiNGMA has the lowest error rate in each simulation

Error Rate / Sample Size	Group Directed LiNGAM	Group ICA LiNGAM	Pairwise LiNGAM
200	0.397	0.151	0.049
500	0.388	0.022	0.008
1000	0.349	0.009	0.002

Table (4).

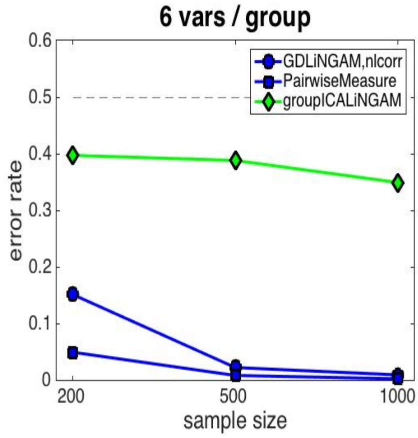


Figure (4).

Error Rate / Sample Size	Group Directed LiNGAM	Group ICA LiNGAM	Pairwise LiNGAM
200	0.401	0.183	0.020
500	0.371	0.03	0.004
1000	0.384	0.004	0

Table (5).

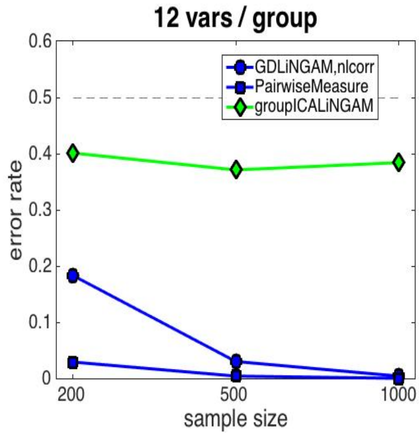


Figure (5).

5 Boston Housing Data in Groups of Variables

According to the above simulation, the Pairwise LiNGAM provides more precise estimation to the causal order. In the following section, I will apply the Pairwise LiNGAM on the real data—Boston Housing Data—which I used on the previous section to estimate the causal effect under groups of variables.

In the Section 3.1, I have already had the group information of Boston Housing Data (Consider the Figure (3)). I will use such group information to estimate the causal order within groups of variables.

In this section, I would explain the Boston Housing Data's group information in detail and apply such group information to the LiNGAM.

5.1 Groups of Variable in Boston Housing Data

Boston Housing Data has 13 single variables and 506 samples. According to the Pekka Parviainen (2015)[6], the group information of Boston Housing Data is separated into 9 groups. Every group can have different number of variables or can be single variables. The organizations of groups are: Group Accessibility consisted of variables CHAS, DIS, and RAD; Group Zoning consisted of variables ZN and INDUS; Group Apartment properties consisted of variables RM and AGE; Group Population consisted of variables B and LSTAT. Five of our groups consisted of one variable: Crime of CRIM, Pollution of NOX, Education of PTRATIO, House prices of MEDV, and Taxes of TAX. (Consider the Figure (3))

5.2 Model of Groups of Variables

Regarding to the groups of variables, I need to seem the groups of variables as the vector and the equation for each groups is in Equation (3). If we arrange the groups in a causal order $\boldsymbol{K} = (\boldsymbol{k_1}, \dots, \boldsymbol{k_G})$ and define $\boldsymbol{x} = (\boldsymbol{x_{k_1}}, \boldsymbol{x_{k_2}}, \dots, \boldsymbol{x_{k_G}})$ **and** $\boldsymbol{e} = (\boldsymbol{e_{k_1}}, \boldsymbol{e_{k_2}}, \dots, \boldsymbol{e_{k_G}})$, we can change the Equation(3) into the matrix form: $\boldsymbol{x} = \boldsymbol{Bx} + \boldsymbol{e}$ where $\boldsymbol{B}$ is a lower block triangular matrix. According to this model, we need to estimate the causal order. Assume that the number of single variable in each group is $\boldsymbol{n} = (\boldsymbol{n_{k_1}}, \boldsymbol{n_{k_2}}, \dots, \boldsymbol{n_{k_G}})$. ***I***n the matrix $\boldsymbol{x}$, the Boston Housing Data would be a 13x506 matrix and the 1st $\boldsymbol{n_{k_1}}$ row correspond to group $\boldsymbol{x_{k_1}}$,..., the last $\boldsymbol{n_{k_G}}$ row corresponds to group $\boldsymbol{x_{k_G}}$.

5.3 About the ParceLiNGAM

At the first, I applied the Pairwise LiNGAM based on Entner's method[7] and Equation (3). However, in this method, they did not provide a completed method to estimate the causal measure and find the connection strength B. To have better estimation, I combined another LiNGAM

[6] Bayesian Networks for Variable Groups Parviainen, S Kaski (2015) arXiv preprint arXiv:1508.07753

[7] Estimating a causal order among groups of variables in linear models, Doris Entner (2012) ICANN'12 Proceedings of the 22nd international conference on Artificial Neural Networks and Machine Learning- Volume PartII Pages84-91

model—ParceLiNGAM[8]—to modify the original method and hope the estimation can be workable. The model (4) is called ParceLiNGAM which has an additional condition f which is call latent confounder.

$$x = \mathbf{B}x + \Lambda f + \mathbf{e} \tag{5}$$

B : connection strength matrix
e : assume to be zero mean, and mutually independent
Λ : the connection strength λ from latent confounder f to x
f : non-Gaussian latent confounding , which is an unobserved variable with zero mean and non-zero variance.

The original purpose of this model is to prevent the wrong estimation if the LiNGAM assumptions is violated. Latent confounder is a variable which is not observed but exerts a causal influence on some of the observed variables. Therefore, to prevent the wrong estimation of connection strength and violation of assumption, I extended this model and modify the Entner's method to adjust the way I used on estimating the causal inference.

The reason why I can apply this model is the algorithm is this model is suitable to extend to groups of variable and also can estimate the connections strength. According to the algorithm3 step 6, " Estimate connection strengths b_{ij} if all the non-descendants of x_i are estimated, i.e. the i-th row of C has no zero. This can be done by doing multiple regression of x_i on all of its non-descendant x_j with $k\,(j) < k(i)$", although this algorithm did not assume groups of variable, this algorithm can be applied to Entner's method as well.

[8] ParceLiNGAM: A Causal Ordering Method Robust Against Latent Confounders, Tatsuya Tashiro ,(2013) Neural Compute. 2014 Jan;26(1):57-83. doi: 10.1162/NECO_a_00533. Epub 2013 Oct 8.

INPUT: Data matrix $\mathbf{X}$ and a threshold α

1. Given a d-dimensional random vector $\boldsymbol{x}$ and a $d \times n$ data matrix of the random vector as $\mathbf{X}$, define U as the set of variable indices of $\boldsymbol{x}$, *i.e.*, $\{1, \cdots, d\}$. initialize a $d \times d$ causal ordering matrix $\mathbf{C}$ by the zero matrix.
2. Apply Algorithm 1 on $\mathbf{X}$ using the threshold α to estimate K_{top} and K_{bttm} and update $\mathbf{C}$.
3. Let $U_{res} := U \setminus (K_{top} \bigcup K_{bttm})$. Denote by $\mathbf{C}_{res}$ the corresponding causal ordering matrix. Denote by $|U_{res}|$ the number of elements in U_{res}. Go to Step 6 if $|U_{res}| \le 2$.
4. Collect variables x_j with $j \in U_{res}$ in a vector $\boldsymbol{x}_{res}$. Collect variables x_j with $j \in K_{top}$ in a vector $\boldsymbol{x}_{top}$. Perform least squares regressions of $\boldsymbol{x}_{top}$ on the i-th element of $\boldsymbol{x}_{res}$ for all $i \in U_{res}$ and collect the residuals in the residual matrix $\mathbf{R}_{res}$ whose i-th row is given by the residuals regressed on x_i.
5. Apply Algorithm 2 on $\mathbf{R}_{res}$ using the threshold α to estimate $\mathbf{C}_{res}$. Replace every c_{ij} $(i \ne j)$ of $\mathbf{C}$ by the corresponding element of $\mathbf{C}_{res}$ if c_{ij} is zero and the corresponding element of $\mathbf{C}_{res}$ is 1 or -1.
6. **Estimate connection strengths b_{ij} if all the non-descendants of x_i are estimated, *i.e.*, the i-th row of $\mathbf{C}$ has no zero. This can be done by doing multiple regression of x_i on all of its non-descendants x_j with $k(j) < k(i)$.**

OUTPUT: A causal ordering matrix $\mathbf{C}$ and a set of estimated connection strength b_{ij}.

Figure (6).

To extend the model (4) to the groups of variables, I modified the model which is suitable to apply on the groups of variables. The model (5) is the extended model in groups of variables.

$$x_{k_i} = \sum_{j<i} B_{k_i k_j} x_{k_j} + \sum_{j<i} \Lambda_{k_i k_j} f_{k_j} + e_{k_i},\ i = 1,\ldots,G \qquad (5)$$

$B_{k_i k_j}$: matrix of dimension $n_{k_i} \times n_{k_j}$, containing the dircet effects from groups x_{k_j} to x_{k_i} and should be permuted as lower triangular and be full column rank.

e_{k_i} : assume to be zero mean, and mutually independent over groups, but allowed to be dependent within each group.

$\Lambda_{k_i k_j}$: matrix of dimension $n_{k_i} \times n_{k_j}$ with full column rank, containing the connection strength λ_{ij} from latent confounder f $_{k_j}$ to x_{k_i}.

f_{k_j} : existence of latent confounding of k_i, which is an unobserved variable and has a parent of more than one observed variable with zero mean and non-zero variance.

Rewriting the model as a matrix

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

5.4 Causal Order in Boston Housing Data

After adjusting the model, I extended the way to estimate the causal order. According to the Lemma1 from ParceLiNGAM[9], I need to extended the

[9] ParceLiNGAM: A Causal Ordering Method Robust Against Latent Confounders, Tatsuya Tashiro ,(2013) Neural Compute. 2014 Jan;26(1):57-83. doi: 10.1162/NECO_a_00533. Epub 2013 Oct 8.

algorithm and combine with the original method to have better estimation. In the Entner's method, I can estimate the causal order of K(Figure.(7)). However, after combining ParceLiNGAM, I need to transfer the vector K into C matrix (Figure (8))which used in ParceLiNGAM. Here is the definition of C:

$$\mathbf{C} = [c_{ij}],\ c_{ij} := \begin{cases} -1 & \text{if } k(i) < k(j) \\ 1 & \text{if } k(i) > k(j) \\ 0 & \text{if it is unknown whether either} \\ & \text{of the two cases above (-1 or 1) is true.} \end{cases}$$

Here are the results:

$$K = [8 \quad 6 \quad 9 \quad 5 \quad 7 \quad 2 \quad 3 \quad 4 \quad 1]$$

Figure (7).

$$C = \begin{vmatrix} 1 & 1 & -1 & 1 & 1 & 1 & 1 & 1 & 1 \\ -1 & 1 & -1 & 1 & -1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ -1 & -1 & -1 & 1 & -1 & 1 & 1 & 1 & 1 \\ -1 & 1 & -1 & 1 & 1 & 1 & 1 & 1 & 1 \\ -1 & -1 & -1 & -1 & -1 & 1 & -1 & -1 & 1 \\ -1 & -1 & -1 & -1 & -1 & 1 & 1 & -1 & 1 \\ -1 & -1 & -1 & -1 & -1 & 1 & 1 & 1 & 1 \\ -1 & -1 & -1 & -1 & -1 & -1 & -1 & -1 & 1 \end{vmatrix}$$

Figure (8).

The name of groups in order with the causal order K is: Housing Prices, Pollution, Taxes, Crime, Education, Zoning, Apartment Properties, and Accessibility, respectively.

5.5 Connection Strength in Boston Housing Data

After estimating the causal order, I used this information to estimate the connection strength which is based on the model (5). The connection strength matrix of B is a 9x9 matrix and the dataset is a 13x506 matrix where the row is the number of single variables and the column is the number of observations. To estimate the connection strength matrix, we need to seem a group of variable as a vector. As a result, the matrix X in model (5) will be a 9x506 matrix.

	Accessibility	Zoning	Apartment Properties	Population	Crime	Polluation	Education	Housing Price	Taxes
Accessibility	0	0	0	0	0	0	0	0	0
Zoning	-1.4314858	0	-1.6422317	-0.0014046	0	0	0.08244481	0	0.01695675
Apartment Properties	0.06848498	0	0	0	0	0	0	0	0
Population	3.3775789	0	-0.6805242	0	0	0	0	0	-0.2175987
Crime	-0.7363962	-0.2570536	0.35655584	-0.0117934	0	0	-0.0610974	-0.179994	0.02573807
Polluation	-0.0255785	0.00411339	0.00387835	-3.42E-05	-0.0003569	0	-0.0120534	-0.0025424	0.00018132
Education	0.03472113	0	-0.7496485	0.00071073	0	0	0	0	0.00540911
Housing Price	-0.4753479	-0.1853837	7.03504195	0.01668541	0	0	-0.9110493	0	-0.0053513
Taxes	-39.647475	0	-45.665973	0	0	0	0	0	0

Table (6).

According to the Table. (6), there is the connection strength between 9 groups of variables. For example, the Accessibility has the -1.43 causal effects on Zooming.

6 Conclusion

In this semester, I read Entner's paper and followed its steps to do the simulation of the groups of variables. After receiving the results that Pairwise LiNGAM can have better estimation of causal order, I started to apply such method on the real data.

The real data I applied is called Boston Housing Data. I tried to extend the LiNGAM to groups of variable and estimate the causal effect between different groups of variables. Boston Housing Data is a classical and simple example. In this dataset, I used Entner's method and ParceLiNGAM to estimate the causal order and causal effect with groups of variables. I extended the model to groups of variables, modified the algorithm to fit into the groups of variable, and learned how to use MATLAB to reach the goal I wanted. During the process, I met the problems in the characteristics of variables in Boston Housing Data and the model which is not suitable to estimate the causal effect. To estimate the connection strength, I modified the method and combined the ParceLiNGAM. Finally, I estimated the causal effect in groups of variables. However, if I want to apply this method on the time-series data, such as fMRI dataset, there still leave something to be desired. According to lots of different approaches, they tried to have better estimation of the causal effect in fMRI data. They usually used the aggregated variables to estimate the causal influence but some information may loss. If we can apply the fMRI dataset on the groups of variable and estimate the causal influence, we may have more knowing about fMRI data.

7 Reference

7.1 ParceLiNGAM: A Causal Ordering Method Robust Against Latent Confounders, Tatsuya Tashiro, (2013) Neural Compute. 2014 Jan;26(1):57-83. doi: 10.1162/NECO_a_00533. Epub 2013 Oct 8.

7.2 Estimating a causal order among groups of variables in linear models, Doris Entner (2012). ICANN'12 Proceedings of the 22nd international conference on Artificial Neural Networks and Machine Learning- Volume PartII Pages84-91

7.3 Bayesian Networks for Variable Groups Parviainen, S Kaski (2015). arXiv preprint arXiv:1508.07753

7.4 Pairwise Likelihood Ratios for Estimation of Non-Gaussian Structural Equation Models, Aapo Hyvärinen (2013). Journal of Machine Learning Research 14 (2013) 111-152 Submitted 10/11; Revised 8/12; Published 1/13