

Individual Causal Structure Learning from Population Data

Wei Chen¹, Xiaokai Huang¹, Zijian Li², Ruichu Cai^{1,3}*, Zhiyi Huang¹ and Zhifeng Hao^{1,4}

¹School of Computer Science, Guangdong University of Technology, Guangzhou, China

²Mohamed bin Zayed University of Artificial Intelligence, Abu Dhabi, United Arab Emirates

³Peng Cheng Laboratory, Shenzhen, China

⁴College of Mathematics and Computer Science, Shantou University, Shantou, China

{chenweidelight, ng.hioukai, leizigin, cairuichu, huangzhiyichn}@gmail.com, zfhao@gdut.edu.cn

Abstract

Learning the causal structure of each individual plays a crucial role in neuroscience, biology, and so on. Existing methods consider data from each individual separately, which may yield inaccurate causal structure estimations in limited samples. To leverage more samples, we consider incorporating data from all individuals as population data. We observe that the variables of all individuals are influenced by the common environment variables they share. These shared environment variables can be modeled as latent variables and serve as a bridge connecting data from different individuals. In particular, we propose an Individual Linear Acyclic Model (ILAM) for each individual from population data, which models the individual's variables as being linearly influenced by their parents, in addition to environment variables and noise terms. Theoretical analysis shows that model is identifiable when all environment variables are non-Gaussian, or even if some are Gaussian with an adequate diversity in the variance of noises for each individual. We then develop an individual causal structures learning method based on the Share Independence Component Analysis technique. Experimental results on synthetic and real-world data demonstrate the correctness of the method even when the sample size of each individual's data is small.

1 Introduction

Learning individual causal structures from observational data is a crucial task [Spirtes *et al.*, 2000; Shimizu *et al.*, 2006; Huang *et al.*, 2020b; Wang and Drton, 2023]. In practice, it is often difficult to collect a substantial amount of data from the same individual due to inefficiency, high costs and occasionally ethical concerns. Instead, data collected in small quantities from various individuals can form a dataset with a substantial amount of data. This data that encompasses the datasets collected from different individuals is called *Population data*. For example, in fMRI analysis, scientists typically instruct different individuals or subjects to perform

the same task while collecting data [Glasser *et al.*, 2016; Miller *et al.*, 2016]. Due to constraints such as time limitations during task execution, the amount of data collected from a single individual is generally limited. They are interested in considering all data of different individuals which forms population data, to learn the individual causal structure. The *individual causal structure* refers to model or the graphical representation describing the causal relationships among variables within an individual. In fMRI analysis, an individual causal structure is the causal relations between Regions of interest (ROIs) for an individual. Different individuals entail different causal structures [Smith *et al.*, 2011; Zhang *et al.*, 2023; Cai *et al.*, 2024], which introduces variability and complexity in causal relationships across individuals when using population data. Thus, learning causal structures for each individual from population data is challenging.

To leverage the population data, a subsequent problem is how to use the information implied in other individuals' data when learning one individual causal structure. A straightforward idea is to apply existing methods to the data of each individual. However, the effectiveness of existing methods depends on having a sufficiently large sample size. In practice, we cannot always collect enough individual data for causal discovery, leading to challenges related to inadequate sample size. An alternative approach involves directly applying methods for individual causal graphs to the aggregated individual data, without considering the different individual causal structures. Nevertheless, this strategy yields a singular causal graph for the entire population, potentially leading to inaccuracies. Recently, some methods first cluster multiple individual samples into several groups and then learn causal structures from the clustered data [Hu *et al.*, 2018; Huang *et al.*, 2019]. Some recover the shared causal mechanism [Ghassami *et al.*, 2018; Perry *et al.*, 2022], leveraging the sparsity of mechanism changes. This may ignore the specific causal relationships of each individual. The common idea of these methods is to argue the sample, and they focus on the shared causal mechanisms without utilizing the shared environment information among individuals.

Considering the generating process of the individual data, we find that the data of each individual is collected from the same environment as other individual data. This common environment may influence the variables of each individual, which can be regarded as the connection between different

*Corresponding Author

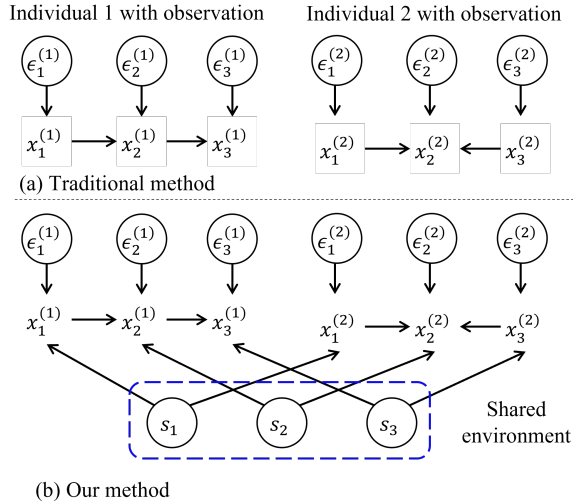


Figure 1: An example of traditional methods and our methods to learn individual causal structures from population data. In these causal graphs, $x_j^{(i)}$ denotes the j -th observed variable for individual i , $\epsilon_j^{(i)}$ denotes the j -th noise term for individual i , and s_j denote the j -th shared environment variable for two individuals. (a) Traditional methods always learn the individual causal structures separately from different individual’s data; (b) Our method considers the shared environment variables to leverage the population data to learn the causal structure for individual 1 and individual 2.

individual datasets. Take Figure 1 as an example. In fMRI analysis, two individuals (denoted as Individual 1 and Individual 2) live in the same country, and their ROIs (denoted as $x_j^{(i)}$, for $i = 1, 2$ and $j = 1, 2, 3$) are affected by the shared environment variables (denoted as s_1, s_2, s_3). These variables are latent but can be used as the connection of two individuals’ variables. That is, the shared environment information is regarded as the bridge to all individual data in the population data. Given the limited sample size for each individual, drawing inspiration from the shared information, the synthesis of these samples could contribute to the learning of individual causal relationships. Therefore, it is crucial to explore how to effectively utilize data from all individuals to learn the causal structures inherent to different individuals.

Motivated by the above example, we consider taking the shared environment information as the exogenous latent confounder for all individual data. In light of this, we propose a data-generating model for each individual, named the *Individual Linear Acyclic Model (ILAM)*. This model is designed to represent all individual data consistently. Specifically, in ILAM, each individual’s variables are primarily influenced by the common environment variables shared by all individuals and the individual-specific causal relationships from their parents, with the corresponding noises. Similar to the typical Linear Non-Gaussian Acyclic Model (LiNGAM), the assumption of non-Gaussian noises is helpful for model identification, which is also proven for our proposed model. Besides, leveraging all the individuals’ data, it is proven that the diversity of noise terms within each individual can be used to guarantee the identifiability of ILAM, which allows some

noise terms to be Gaussian. Based on the proposed model and the identification of theoretical results, we develop the Individual Causal Structure Learning (ICSL) method for estimating the individual causal structure.

2 Related Work

In this section, we investigate the related work of causal discovery from multiple datasets of different individuals, which may constitute population data. The existing work pursues three primary objectives: recovering the specific causal structure for each individual, learning the shared causal graph among (a group of) individuals, and learning the specific and shared causal relationships within groups of individuals.

For the first type, several methods [Spirites *et al.*, 2000; Shimizu *et al.*, 2006; Shimizu *et al.*, 2011; Chen *et al.*, 2021; Chen *et al.*, 2024] are proposed and applied to each dataset of each individual separately. However, their performance heavily depends on the sample sizes. The larger the sample sizes, the better performance they obtain. For the second type, certain methods [Zhang *et al.*, 2017; Huang *et al.*, 2020b] argue that some causal mechanisms may vary across different datasets or environments. Thus, they assume that there exists a single unobserved variable affecting some observed variables in causal graphs, leading to changes in causal mechanisms. CD-NOD method [Huang *et al.*, 2020b] introduces a time/domain index to model the non-stationary or heterogeneous. [Saeed *et al.*, 2020] considers that multiple datasets are generated from a mixture of K Directed Acyclic Graph (DAG) models. The MSS method [Perry *et al.*, 2022] utilizes the sparse mechanism shift (SMS) hypothesis, and introduces the Mechanism Shift Score to recover the causal graph. [Ghassami *et al.*, 2018] exploits the principle of independent changes to learn the causal structure from observational data given in multiple domains. CD-MiNi [Huang *et al.*, 2020a] considers the situation that each dataset contains a subset of all variables.

In recent years, some researchers have proposed methods for recovering the specific and shared causal relationships when using the same group of data. The Group Iterative Multiple Model Estimation approach [Gates *et al.*, 2010; Gates and Molenaar, 2012] attempts to heuristically uncover time-lagged causal relations at both group and individual levels, without providing theoretical guarantees. Some approaches characterize the distributions of causal mechanism parameters through a mixture model. The SSCM method leverages the shared and specific information to learn the causal structures and cluster the causal mechanism [Huang *et al.*, 2019]. To model the non-linear relationships, the ANM Mixture Model (ANM-MM) [Hu *et al.*, 2018] assumes that the causal mechanism parameter is drawn from a discrete distribution on a finite set. [Pashami *et al.*, 2018] integrate clustering and learning causal structures by clustering subjects into multiple groups based on the estimated causal structures. All of these methods attempt to argue the sample with the dataset of individuals within the same group, in terms of causal mechanisms. In our work, we do not emphasize the aggregation of samples with partially identical labels to cluster individuals. Instead, we focus on leveraging the shared in-

formation in population data, to learn individual causal structures by considering the data generation process.

3 Individual Linear Acyclic Model

Suppose there are m individuals in the same environment. Each individual has n observed variables. The individuals share some common information from the environment, but their causal relationships may be different. Let $\mathbf{x}^{(1)}, \mathbf{x}^{(2)}, \dots, \mathbf{x}^{(m)}$ denote the observation of m different individuals. We aim to learn the individual causal model over n variables for each individual. Similar to the Linear Non-Gaussian Acyclic Model, we assume that for each individual, the causal relations between two observed variables are linear, and the causal graph among variables is acyclic. Inspired by the variables of each individual affected by the common environment variables, we introduce latent variables to represent the shared environment variables. Then, we propose an *Individual Linear Acyclic Model* (ILAM) to model the data generation process of each individual that shares the same environment with others. In detail, the observed variable $x_j^{(i)}$ of the i -th individual, satisfy the following generation process:

$$x_j^{(i)} = \sum_{k \in \mathbf{PA}_j^{(i)}} b_{j,k}^{(i)} x_k^{(i)} + s_j + e_j^{(i)}, \quad (1)$$

for $j = 1, 2, \dots, n$, where $\mathbf{PA}_j^{(i)}$ is an index set containing the parent of $x_j^{(i)}$, $b_{j,k}^{(i)}$ is the causal strength from $x_k^{(i)}$ to $x_j^{(i)}$, \mathbf{s} and $\mathbf{e}^{(i)}$ denote the shared environment variables and the noise terms of individual i , respectively. All the environment variables and the noise terms are independent of each other.

In the form of matrix, Eq. (1) can be written as:

$$\mathbf{x}^{(i)} = \mathbf{B}^{(i)} \mathbf{x}^{(i)} + \mathbf{s} + \mathbf{e}^{(i)}, \quad (2)$$

where $\mathbf{B}^{(i)}$ is a causal strength matrix among observed variables. Since $\mathbf{B}^{(i)}$ represents a Directed Acyclic Graph (DAG), $\mathbf{B}^{(i)}$ could be permuted to a strictly lower triangular matrix. Without loss of generality, in ILAM, we assume that the noise terms for each individual have zero mean, i.e., $\forall i \in \{1, 2, \dots, m\}$, $\mathbf{e}^{(i)} \sim \mathcal{N}(\mathbf{0}, \Sigma^{(i)})$, where the $\Sigma^{(i)}$ are diagonal and positive matrices. Besides, the shared environmental noises are assumed to be unit variance, i.e., $\mathbb{E}[\mathbf{s}\mathbf{s}^\top] = \mathbf{I}$. Note that these assumptions are common in other methods, and can be achieved by normalizing the data.

Our goal is to infer the individual causal structures $\mathbf{B}^{(1)}, \mathbf{B}^{(2)}, \dots, \mathbf{B}^{(m)}$ for all individuals, given population data that consists of all observed individual datasets, which are generated by ILAM.

4 Identifiability

Based on the model (2), to identify the causal strength matrices, we find that the unknown influence is from the latent environment variables and the noise terms. Inspired by the LiNGAM, we would like to transfer the model to represent the mapping from the latent variables part to the observed variables part. That is, for $i \in \{1, 2, \dots, m\}$, we can transfer model (2) to the following equation:

$$\begin{aligned} \mathbf{x}^{(i)} &= (\mathbf{I} - \mathbf{B}^{(i)})^{-1} (\mathbf{s} + \mathbf{e}^{(i)}) \\ &= \mathbf{A}^{(i)} (\mathbf{s} + \mathbf{e}^{(i)}), \end{aligned} \quad (3)$$

where $\mathbf{A}^{(i)} = (\mathbf{I} - \mathbf{B}^{(i)})^{-1}$, is called the mixing matrix. Because $\mathbf{B}^{(i)}$ could be permuted to a strictly lower triangular matrix, each $\mathbf{A}^{(i)}$ could be permuted to a lower triangular matrix with all non-zero elements along its diagonal. We denote the inverse of $\mathbf{A}^{(i)}$ as $\mathbf{W}^{(i)} = (\mathbf{A}^{(i)})^{-1} = \mathbf{I} - \mathbf{B}^{(i)}$, which could also be permuted to lower triangular matrix with non-zero elements on the main diagonal.

Interestingly, we find that the form of Eq. (3) is similar to the Shared Independent Component Analysis (ShICA) model [Richard *et al.*, 2021], where \mathbf{s} can be viewed as shared components, and $\mathbf{e}^{(i)}$ can be viewed as noises components for i -th individual. The mixing matrices $\mathbf{A}^{(1)}, \mathbf{A}^{(2)}, \dots, \mathbf{A}^{(m)}$ have been proven to be identifiable up to sign and permutation under mild assumptions on the distributions of \mathbf{s} and $\mathbf{e}^{(i)}$ [Richard *et al.*, 2021].

In light of the ShICA, the distribution of environment variables \mathbf{s} can be non-Gaussian, or there may be partial Gaussian in the identification of the model (3) with mild assumptions. In the first case, we assume that at most one environment variable follows a Gaussian distribution. In the second case, we assume that there are two or more environment variables following Gaussian distributions.

Although the mixing matrix in the model (3) is identifiable, the identification of ILAM model also requires addressing the unique transformation from the mixing matrix to the causal strength matrix. Regarding permutation indeterminacy, it pertains to the relationship between observed variables and their corresponding components. That is, the true $\mathbf{W}^{(i)}$ is a matrix with non-zero diagonals because our model is acyclic. Intuitively, we can find the correspondence between the latent components and the observed variables by a row permutation. The lemma proved in the work [Shimizu *et al.*, 2006] shows that the row permutation is unique and eliminates the indeterminacy of permutation. With this connection, we can provide the theorems for the identification of the proposed model.

First, we consider the case where at most one environment variable follows a Gaussian distribution. The identifiability of ILAM is guaranteed by the following theorem.

Theorem 4.1 (Identifiability with at most one Gaussian component). *Suppose there is at most one Gaussian component in shared environment variables \mathbf{s} and the number of individuals m is at least 3. Given enough observed data $\mathbf{x}^{(1)}, \mathbf{x}^{(2)}, \dots, \mathbf{x}^{(m)}$ generated from model (2), then the model (2) is identifiable up to sign.*

Theorem 4.3 shows that with the data of different individuals, it can provide the pseudo-supervised information to estimate our model (2). It can relax the non-Gaussian assumption, different from the existing model for one individual. Inspired by the Multiset Canonical Correlation Analysis (CCA) [Kettenring, 1971] and the ShICA technique, if there are more than two Gaussian components in \mathbf{s} , we require an additional assumption on the variances of $\mathbf{e}^{(i)}$ within each individual. Let \mathfrak{N} denote the set of Gaussian components in \mathbf{s} , i.e., $\forall j \in \mathfrak{N}$, s_j follows the Gaussian distribution.

Assumption 4.2 (Noise diversity in Gaussian variances). For $i \in \{1, 2, \dots, m\}$, $\forall j, j' \in \mathfrak{N}, j \neq j'$, the sequences $(\Sigma_{j,j}^{(i)})_{i=1, \dots, m}$ and $(\Sigma_{j',j'}^{(i)})_{i=1, \dots, m}$ are different.

In this case, the identifiability of model can be guaranteed by the following theorem:

Theorem 4.3 (Identifiability with noise diversity). *Suppose Assumption 4.2 holds and the number of individuals m is at least 3. Given enough observed data $\mathbf{x}^{(1)}, \mathbf{x}^{(2)}, \dots, \mathbf{x}^{(m)}$ generated from model (2), then model (2) is identifiable up to sign.*

For the sign indeterminacy, while both Theorem 4.1 and Theorem 4.3 only guarantee the identifiability of model (2) up to sign, we can demonstrate that this is tolerable. Since model (2) is identified up to sign and $\mathbf{B}^{(i)}$ corresponds to a DAG, we can transform the estimated $\widehat{\mathbf{B}}^{(i)}$ into a lower triangular matrix by the equal permutation. The permuted order of observed variables reflects the causal order. Based on the causal order, we can re-estimated the coefficients of $\mathbf{B}^{(i)}$ using regression methods, like Adaptive Lasso, which helps us discover the sign of the coefficients of $\mathbf{B}^{(i)}$. For the condition of enough observed data, note that in practice, the data is enough when the error of the objective function is small, which will also be shown in the experimental results.

5 ICSL Algorithm

In this section, we provide an Individual Causal Structure Learning (ICSL) algorithm, using the observed data $\mathbf{X} = \{\mathbf{x}^{(1)}, \mathbf{x}^{(2)}, \dots, \mathbf{x}^{(m)}\}$ generated from model (2).

Based on the identifiability of model (3), we can obtain the estimated $\widehat{\mathbf{A}}^{(i)}$ for all individuals by estimating the mixing matrix on the population data. Then, we can calculate the inverse of the mixing matrices $\widehat{\mathbf{W}}^{(i)}$ for every individual. In the first stage, we can use the approaches provided in ShICA [Richard *et al.*, 2021] to estimate the mixing matrix. Because the ShICA is identifiable up to sign and permutation, our next step is to find the correct sorting for the components, which corresponds to the combination of \mathbf{s} and $\mathbf{e}^{(i)}$. Fortunately, we know that the true $\mathbf{W}^{(i)}$ is a matrix with all-ones on the main diagonal. Therefore, for each individual, we can reorder the components by finding a row permutation $\mathbf{P}^{(i)}$ of $\widehat{\mathbf{W}}^{(i)}$ such that $\widetilde{\mathbf{W}}^{(i)} = \mathbf{P}^{(i)}\widehat{\mathbf{W}}^{(i)}$ have non-zero on the diagonal. Subsequently, to obtain the causal strength matrix, we divide each row of $\widetilde{\mathbf{W}}^{(i)}$ by its corresponding diagonal element to obtain $\overline{\mathbf{W}}^{(i)}$ with all-one elements on the diagonal. And then we can obtain the estimated causal matrix $\widehat{\mathbf{B}}^{(i)} = \mathbf{I} - \overline{\mathbf{W}}^{(i)}$. Finally, for each individual, we try to find an equal permutation $\widetilde{\mathbf{P}}^{(i)}$ on the row and column of $\widehat{\mathbf{B}}^{(i)}$, such that $\widetilde{\mathbf{B}}^{(i)} = \widetilde{\mathbf{P}}^{(i)}\widehat{\mathbf{B}}^{(i)}(\widetilde{\mathbf{P}}^{(i)})^\top$ is close to a strictly lower triangular. Every permutation matrix $\widetilde{\mathbf{P}}^{(i)}$ reflects the causal order of the observed variables for the i -th individual.

In the above steps, we can use two approaches to estimate the mixing matrices. The first approach involves estimating the mixing matrices via multiset CCA, followed by the joint-diagonalization algorithm for improved estimation, denoted as ICSL-J. This method can be applied to the data that

are Gaussian or non-Gaussian. The second approach entails maximizing the likelihood of components using the EM algorithm to estimate parameters, denoted as ICSL-ML. This method can be used when at most one dataset is Gaussian. In practice, we can first test whether the dataset is Gaussian or non-Gaussian to choose the proper method.

6 Experiments

In this section, we conduct experiments on synthetic and real-world data to evaluate the performance of our method.

6.1 Synthetic Data

We randomly generate the synthetic data according to our model (2). To show the efficacy of the proposed method, we generate data with two kinds of distributions for environment variables: Gaussian and non-Gaussian distributions. For the Gaussian distribution, we sample environment variables from a standard Gaussian distribution $\mathbf{s} \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$. For the non-Gaussian distribution, we sample environment variables from the Laplace distribution $\mathbf{s} \sim \text{Laplace}(\mathbf{0}, \mathbf{I})$. In each setting, we synthesize data with fixed parameters while traversing the target parameter. In detail, we vary the number of nodes with $n = 6, 8, 10, 12$, the sample size per individual with $l = 50, 100, 500, 1000$, the number of individuals $m = 3, 5, 7, 9$ and the number of different causal structures with $d = 1, 2, 4, 8$. The default parameters are marked as bold. The causal strength from one observed variable to another is randomly generated with the range of $[0.5, 1.2]$.

In these experiments, we use PC [Spirtes *et al.*, 2000], ICA-LiNGAM [Shimizu *et al.*, 2006], Direct-LiNGAM [Shimizu *et al.*, 2011], CD-NOD [Huang *et al.*, 2020b], SSCM [Huang *et al.*, 2019], MSS [Perry *et al.*, 2022] and J-PCMCI+ [Günther *et al.*, 2023] as the baseline methods. Among these methods, PC, ICA-LiNGAM, and Direct-LiNGAM are typical methods for causal discovery from each individual data. The CD-NOD method uses conditional independence tests for causal discovery from non-stationary or heterogeneous data. The SSCM method aims to recover the share and specific causal relation among observed variables. The MSS method identifies the causal graph based on the mechanism shift hypothesis. The J-PCMCI+ method is a constraint-based method that takes into account the influence of temporal and spatial context information on multiple datasets, and recovers their common causal structure.

To evaluate the correctness of causal structure learning, we use the F1 score and Structural Hamming Distance (SHD) as evaluation metrics. Furthermore, to evaluate the correctness of the estimated causal strength, we also use the mean squared error (MSE) between the true $\mathbf{B}^{(i)}$ and the estimated one, as a metric. Each setting was conducted 10 times and the average of the evaluation metrics was calculated as the final evaluation metric. The results on SHD for all methods are provided in the appendix due to the space limit.

Experimental results for non-Gaussian noises. Figure 2 illustrates the experimental results from synthetic data with noises generated by Laplace distributions. In the experimental result for sensitivity to number of nodes, with the increase in the number of nodes, fluctuations or decreases

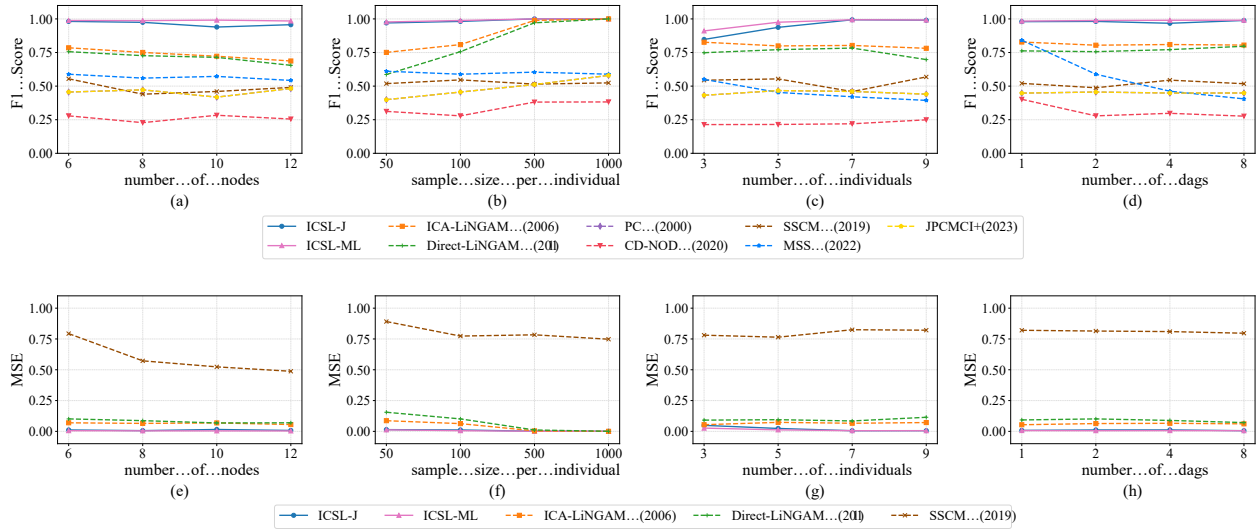


Figure 2: F1 scores of the recovered causal structure and MSE of the recovered causal strength in the setting with Laplace noises.

in F1 score are observed across all methods. The F1 score of our method consistently maintains a high level and outperform other methods. It is worth noting that F1 scores of ICA-LiNGAM and Direct-LiNGAM remain lower than our method. Considering that, under a single individual, the generated data follows the LiNGAM model assumption, we attribute this phenomenon to the issue of sample size. The dataset with a sample size of 100 may not be sufficient for ICA-LiNGAM and Direct-LiNGAM to estimate accurate causal structure. Our method, on the other hand, aggregates generated data from different individuals, thereby compensating for the inadequacy in sample size.

In the experimental result for sensitivity to sample size, F1 scores for methods other than ours, ICA-LiNGAM, and Direct-LiNGAM consistently remain lower. As the sample size increases, F1 scores for our method, ICA-LiNGAM, and Direct-LiNGAM also steadily rise. When the sample size is 500 or 1000, F1 scores of our method, Direct-LiNGAM, and ICA-LiNGAM are close to 1. However, as the sample size gradually decreases, the gap between these two baseline methods and ours becomes more pronounced. When the sample size is 10, ICA-LiNGAM’s F1 score are below 0.8, Direct-LiNGAM’s F1 score is below 0.6, while our method maintains the F1 score above 0.9. This also corroborates the earlier conjecture: our method, by aggregating generated data from different individuals, is capable of compensating for the problem of a small sample size per individual.

In the experimental result for sensitivity to the number of individuals, we observe a clear increase in our method’s F1 score as the number of individuals grows. When the number of individuals is sufficiently high, our F1 score approaches 1. Even in the case of a smaller number of individuals ($m = 3$), our method performs well. In contrast, there is no significant improvement in the F1 scores of other methods as the number as the number of individuals grows. The F1 score

of MSS shows a notable decline. This is attributed to the MSS method relying on the sparsity of mechanism changes to estimate causal structures. In our assumed data generation process, mechanism changes across different individuals lack sparsity. These findings demonstrate that our method can better leverage shared information among different individuals to enhance the accuracy of causal structure estimation.

In the experimental result for sensitivity to the number of different dags, for varying d , our method consistently maintains stable F1 scores close to 1. In contrast, as d increases, both CD-NOD and MSS experience a significant drop in their F1 scores. This is because CD-NOD and MSS assume that different individuals share the same underlying causal structure and utilize the principle of minimal changes in causal mechanisms for estimation. In our setting, individual causal structures vary widely, and different individuals just share common information from the environment, which is the foundation for our method’s estimation.

Furthermore, for all the experimental results mentioned above, the MSE of our method is close to zero, while the MSE of SSCM is generally greater than 0.5. ICA-LiNGAM and Direct-LiNGAM only achieve close-to-zero MSE when the sample size is 500 or 1000. This indicates that using our method, our model is identifiable up to sign.

Experimental results for Gaussian noises. In these experiments, we randomly sample noises from Gaussian distribution and yield the synthetic data based on the randomly generated DAG. We make the variances of individual noises to be different. Figure 3 illustrates the experimental results.

In the experimental result for sensitivity to the number of nodes, compared to the case without Gaussian noises, F1 scores of Direct-LiNGAM decreases significantly, because Gaussian noises violate the assumptions of the LiNGAM model. In contrast, our F1 score approaches 1. This indicates that with the diverse individual noises, our method can

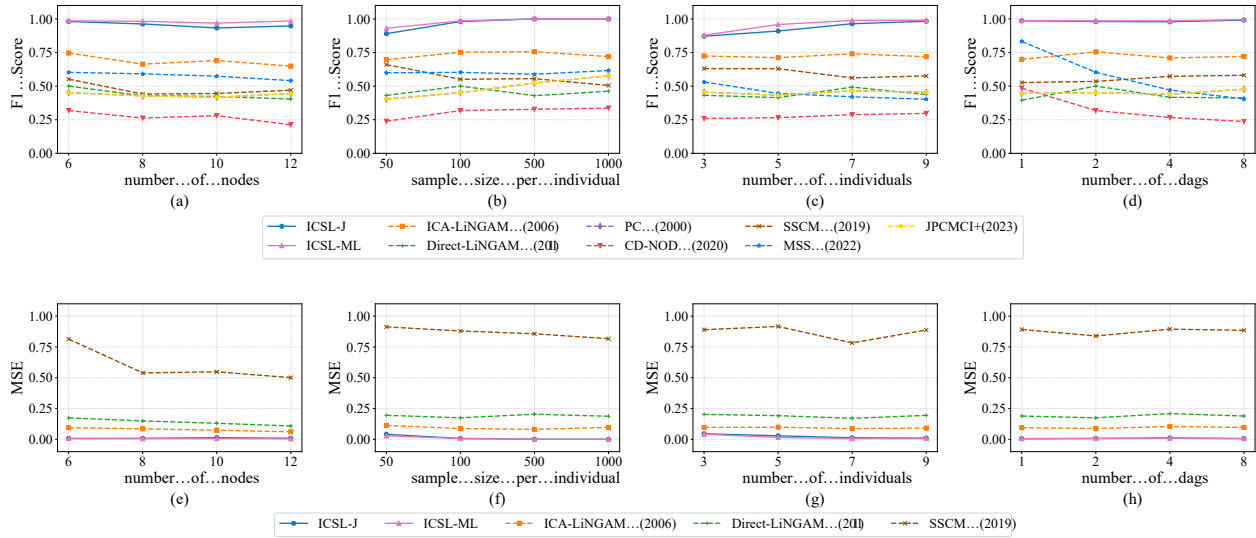


Figure 3: F1 scores of the recovered causal structure and MSE of the recovered causal strength in the setting with Gaussian noises.

identify the causal structure.

In the experimental result for sensitivity to sample size, the F1 score of our method is significantly higher than that of baseline models. As the sample size increases, our F1 score increases and eventually approaches 1. However, F1 scores of ICA-LiNGAM and Direct-LiNGAM are relatively low, indicating that Gaussian noise makes them challenging to accurately estimate the causal structure.

Similar to the previous experiments, in the experimental result for sensitivity to number of individuals, F1 scores of ICA-LiNGAM and Direct-LiNGAM exhibit a noticeable decrease. While the F1 score of our method increases with the growing number of individuals. In the experimental result for the number of different dags, CD-NOD and MSS show a noticeable decrease as d increases. From all experimental results, the MSE of our method is close to zero. Unlike the previous experiments, the MSE of ICA-LiNGAM and Direct-LiNGAM no longer approaches zero, even when we increase the sample size.

6.2 Real World Data

FMRI Data

To test the performance of our method in a real-world problem, we applied the algorithm to real functional magnetic resonance imaging (fMRI) task data [Ramsey *et al.*, 2010]. This fMRI dataset was acquired by a 3T scanner with TR= 2 s, resulting in a sample size of 160 [Sanchez-Romero *et al.*, 2019] per subject. The raw data can be obtained from the OpenfMRI project¹. Our experiment uses the preprocessed data². The dataset contains data for 9 individuals, each consisting of nine variables that were judged to rhyme with or without a pair of visual stimuli. It includes one input variable (Input) and eight regions of interest (ROIs). The input

variables were created by combining the rhyming task’s box-car model with the standard hemodynamic response function, which reflects how the brain’s blood flow changes in response to neural activity. The eight ROIs include the left and right occipital cortex (LOCC, ROCC), left and right anterior cingulate cortex (LACC, RACC), and left and right inferior frontal gyrus (LIFG, RIFG); as well as the left inferior and right inferior parietal lobule (LIPL, RIPL). We treat the fMRI data of all subjects as population data and apply our method to discover the causal relations for every subject. Subsequently, we aggregate the results from each individual to obtain the final causal graph. For each edge, if more than 50% of individuals believe it exists, we retain the edge in the aggregated causal graph; otherwise, we remove the edge from the aggregated causal graph.

For fMRI data, a common point of view is that the stimulus input is expected to traverse the left occipital cortex and propagate from the left to the right. The result of the fMRI data

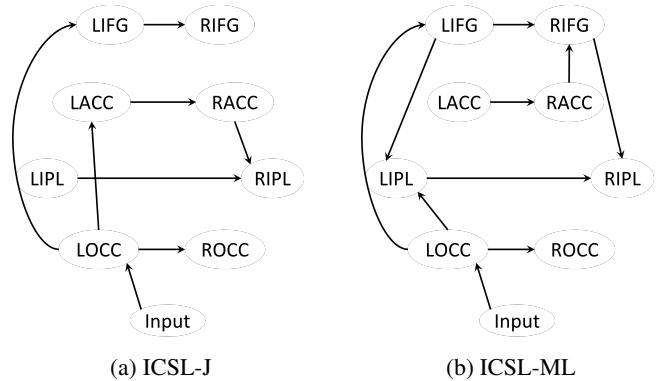


Figure 4: Causal graphs learned from fMRI task data.

¹<https://openfmri.org/dataset/ds000003/>

²<https://github.com/cabal-cmu/Feedback-Discovery>

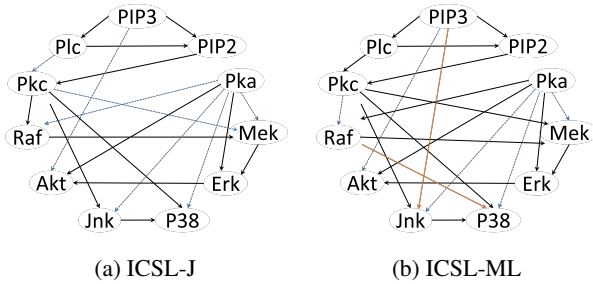


Figure 5: Causal graphs learned from sachs data. Black lines: correct edges. Blue dashed lines: missing edges. Orange lines: wrong edges.

is shown in Figure 4. It can be seen that the only edge connecting to Input is from it to the left occipital cortex, which is consistent with the common viewpoint. Also, both the left and right brain connected edges are from the left half of the brain to the right half of the brain, which is also consistent with the actual situation. Although there are some additional edges in our results, we can observe that some of the findings are consistent with those in [Ramsey *et al.*, 2010]. Specifically, in the results ICSL-J, the edges LOCC → LIFG and LOCC → LACC maintain a consistent causal order with the outcomes in [Ramsey *et al.*, 2010]. For the result of ICSL-ML, the feedforward edges LOCC → LIFG → LIPL are also consistent with the results in [Ramsey *et al.*, 2010].

Sachs Data

We also applied our methods to Sachs data [Sachs *et al.*, 2005]. The data consists of a collection of data sets, where each data set corresponds with a different experiment in which a perturbation was applied to sets of individual cells. The dataset we used contains 9 experimental observed data. We treat the data of each experiment as individual observed data. We aim to discover the causal relationship of different types of cellular protein. We use obtain the aggregated causal graph by the same way as the experiment for fMRI Data.

The learned causal structures by ICSL-J and ICSL-ML are shown in Figure 5. It is shown that our estimated structures contain many edges in ground truth. We observed that all estimated edges from the ICSL method are contained within the ground truth causal structure. For the structure estimated by ICSL-ML, all edges except for edges Raf → P38 and PIP3 → Jnk are included in the ground truth causal structure.

Model	Precision	Recall	F1 score	SHD
ICSL-J	0.92	0.55	0.69	10
ICSL-ML	0.76	0.65	0.7	11
PC (2000)	0.53	0.50	0.51	18
ICA-LiNGAM (2006)	0.60	0.60	0.60	16
Direct-LiNGAM (2011)	0.59	0.50	0.54	17
CD-NOD (2020)	0.45	0.45	0.45	31
SSCM (2019)	0.17	0.40	0.24	50
MSS (2022)	0.29	0.10	0.15	21
J-PCMCI+(2023)	0.53	0.5	0.51	18

Table 1: Evaluation results on sachs data

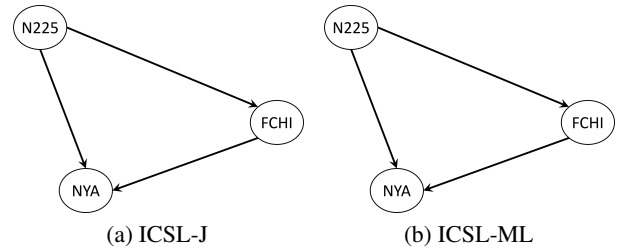


Figure 6: Causal graphs learned from Yahoo stock indices data.

We calculated the Precision, Recall, F1 score and SHD between the ground truth and the estimated causal structures of all methods. In the table 1, we can see that our method outperforms all the other methods. Specifically, our F1 score is higher than that of ICA-LiNGAM and Direct-LiNGAM, and our SHD is lower than theirs. We attribute this to the aggregation of information from multiple datasets, providing more abundant information for causal structure recovery.

Yahoo Stock Indices Data

We also apply our algorithm to stock indices data that is collected from the Yahoo finance database for 5 years (from 2015 to 2019). We use the adjusted closing prices for the stocks. This data contains 3 stock indices, which are N225 from Japan, FCHI from Europe and NYA from the United States. We treat the data of each year as the observed data of each individual. Then, we aim to find the causal structure between 3 stork indices with our method. Due to the different time zones, it is expected the causal order of ground truth is N225 → FCHI → NYA. The results of ICSL-J and ICSL-ML are shown in Figure 6, which are consistent with expectations.

7 Conclusion

In this paper, we introduce the Individual Linear Acyclic Model (ILAM) to describe the data generation process for each individual where it shares common environment information with other individuals. Additionally, we propose a novel method named Individual Causal Structure Learning (ICSL) to uncover causal structures for each individual. ICSL estimates the mixing matrix first and then determines a row permutation for the inverse of the mixing matrix to establish the correspondence between noise terms and observed variables. We demonstrate that our model is identifiable up to sign when at most one component of the shared noises follows a Gaussian distribution. Even with more than two components following a Gaussian distribution, our model remains identifiable under additional mild assumptions. We experimentally demonstrate that our method performs well when dealing with shared environment variables following a non-Gaussian or Gaussian distribution. Even in scenarios with limited sample size, a common challenge in real-world applications, our method consistently outperforms other baseline approaches. It is noted that the linear assumption of ILAM can be relaxed, and future work will focus on the more flexible causal relationships.

Ethical Statement

There are no ethical issues.

Acknowledgments

This research was supported in part by the National Science and Technology Major Project (2021ZD0111501), the National Science Fund for Excellent Young Scholars (62122022), the Natural Science Foundation of China (62206064), the Guangzhou Basic and Applied Basic Research Foundation (2024A04J4384), the major key project of PCL (PCL2021A12), the Guangdong Basic and Applied Basic Research Foundation (2023B1515120020), and the Jihua laboratory scientific project (X210101UZ210).

References

- [Cai *et al.*, 2024] Ruichu Cai, Yunjin Wu, Xiaokai Huang, Wei Chen, Tom ZJ Fu, and Zhifeng Hao. Granger causal representation learning for groups of time series. *Science China Information Sciences*, 67(5):152103, 2024.
- [Chen *et al.*, 2021] Wei Chen, Ruichu Cai, Kun Zhang, and Zhifeng Hao. Causal discovery in linear non-gaussian acyclic model with multiple latent confounders. *IEEE Transactions on Neural Networks and Learning Systems*, 33(7):2816–2827, 2021.
- [Chen *et al.*, 2024] Wei Chen, Zhiyi Huang, Ruichu Cai, Zhifeng Hao, and Kun Zhang. Identification of causal structure with latent variables based on higher order cumulants. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 38, pages 20353–20361, 2024.
- [Gates and Molenaar, 2012] Kathleen M Gates and Peter CM Molenaar. Group search algorithm recovers effective connectivity maps for individuals in homogeneous and heterogeneous samples. *NeuroImage*, 63(1):310–319, 2012.
- [Gates *et al.*, 2010] Kathleen M Gates, Peter CM Molenaar, Frank G Hillary, Nilam Ram, and Michael J Rovine. Automatic search for fmri connectivity mapping: An alternative to granger causality testing using formal equivalences among sem path modeling, var, and unified sem. *NeuroImage*, 50(3):1118–1125, 2010.
- [Ghassami *et al.*, 2018] AmirEmad Ghassami, Negar Kiyavash, Biwei Huang, and Kun Zhang. Multi-domain causal structure learning in linear systems. *Advances in Neural Information Processing Systems*, 31, 2018.
- [Glasser *et al.*, 2016] Matthew F Glasser, Stephen M Smith, Daniel S Marcus, Jesper LR Andersson, Edward J Auerbach, Timothy EJ Behrens, Timothy S Coalson, Michael P Harms, Mark Jenkinson, Steen Moeller, et al. The human connectome project’s neuroimaging approach. *Nature Neuroscience*, 19(9):1175–1187, 2016.
- [Günther *et al.*, 2023] Wiebke Günther, Urmi Ninad, and Jakob Runge. Causal discovery for time series from multiple datasets with latent contexts. In *Uncertainty in Artificial Intelligence*, pages 766–776. PMLR, 2023.
- [Hu *et al.*, 2018] Shoubo Hu, Zhitang Chen, Vahid Par-tovi Nia, Laiwan Chan, and Yanhui Geng. Causal inference and mechanism clustering of a mixture of additive noise models. *Advances in Neural Information Processing Systems*, 31, 2018.
- [Huang *et al.*, 2019] Biwei Huang, Kun Zhang, Pengtao Xie, Mingming Gong, Eric P Xing, and Clark Glymour. Specific and shared causal relation modeling and mechanism-based clustering. *Advances in Neural Information Processing Systems*, 32, 2019.
- [Huang *et al.*, 2020a] Biwei Huang, Kun Zhang, Mingming Gong, and Clark Glymour. Causal discovery from multiple data sets with non-identical variable sets. In *Proceedings of the AAAI conference on artificial intelligence*, volume 34, pages 10153–10161, 2020.
- [Huang *et al.*, 2020b] Biwei Huang, Kun Zhang, Jiji Zhang, Joseph Ramsey, Ruben Sanchez-Romero, Clark Glymour, and Bernhard Schölkopf. Causal discovery from heterogeneous/nonstationary data. *The Journal of Machine Learning Research*, 21(1):3482–3534, 2020.
- [Kettenring, 1971] Jon R Kettenring. Canonical analysis of several sets of variables. *Biometrika*, 58(3):433–451, 1971.
- [Miller *et al.*, 2016] Karla L Miller, Fidel Alfaro-Almagro, Neal K Bangerter, David L Thomas, Essa Yacoub, Junqian Xu, Andreas J Bartsch, Saad Jbabdi, Stamatios N Sotiropoulos, Jesper LR Andersson, et al. Multimodal population brain imaging in the uk biobank prospective epidemiological study. *Nature Neuroscience*, 19(11):1523–1536, 2016.
- [Pashami *et al.*, 2018] Sepideh Pashami, Anders Holst, Juhee Bae, and Sławomir Nowaczyk. Causal discovery using clusters from observational data. In *FAIM’18 Workshop on CausalML, Stockholm, Sweden, July 15, 2018*, 2018.
- [Perry *et al.*, 2022] Ronan Perry, Julius Von Kügelgen, and Bernhard Schölkopf. Causal discovery in heterogeneous environments under the sparse mechanism shift hypothesis. *Advances in Neural Information Processing Systems*, 35:10904–10917, 2022.
- [Ramsey *et al.*, 2010] Joseph D Ramsey, Stephen José Hanson, Catherine Hanson, Yaroslav O Halchenko, Russell A Poldrack, and Clark Glymour. Six problems for causal inference from fmri. *NeuroImage*, 49(2):1545–1558, 2010.
- [Richard *et al.*, 2021] Hugo Richard, Pierre Ablin, Bertrand Thirion, Alexandre Gramfort, and Aapo Hyvarinen. Shared independent component analysis for multi-subject neuroimaging. *Advances in Neural Information Processing Systems*, 34, 2021.
- [Sachs *et al.*, 2005] Karen Sachs, Omar Perez, Dana Pe’er, Douglas A Lauffenburger, and Garry P Nolan. Causal protein-signaling networks derived from multiparameter single-cell data. *Science*, 308(5721):523–529, 2005.
- [Saeed *et al.*, 2020] Basil Saeed, Snigdha Panigrahi, and Caroline Uhler. Causal structure discovery from distribu-

- tions arising from mixtures of dags. In *International Conference on Machine Learning*, pages 8336–8345. PMLR, 2020.
- [Sanchez-Romero *et al.*, 2019] Ruben Sanchez-Romero, Joseph D Ramsey, Kun Zhang, Madelyn RK Glymour, Biwei Huang, and Clark Glymour. Estimating feedforward and feedback effective connections from fmri time series: Assessments of statistical methods. *Network Neuroscience*, 3(2):274–306, 2019.
- [Shimizu *et al.*, 2006] Shohei Shimizu, Patrik O Hoyer, Aapo Hyvärinen, and Antti Kerminen. A linear non-gaussian acyclic model for causal discovery. *Journal of Machine Learning Research*, 7(Oct):2003–2030, 2006.
- [Shimizu *et al.*, 2011] Shohei Shimizu, Takanori Inazumi, Yasuhiro Sogawa, Aapo Hyvärinen, Yoshinobu Kawahara, Takashi Washio, Patrik O Hoyer, and Kenneth Bollen. Directlingam: A direct method for learning a linear non-gaussian structural equation model. *The Journal of Machine Learning Research*, 12:1225–1248, 2011.
- [Smith *et al.*, 2011] Stephen M Smith, Karla L Miller, Gholamreza Salimi-Khorshidi, Matthew Webster, Christian F Beckmann, Thomas E Nichols, Joseph D Ramsey, and Mark W Woolrich. Network modelling methods for fmri. *NeuroImage*, 54(2):875–891, 2011.
- [Spirtes *et al.*, 2000] Peter Spirtes, Clark N Glymour, Richard Scheines, and David Heckerman. *Causation, prediction, and search*. MIT press, 2000.
- [Wang and Drton, 2023] Y Samuel Wang and Mathias Drton. Causal discovery with unobserved confounding and non-gaussian data. *Journal of Machine Learning Research*, 24(271):1–61, 2023.
- [Zhang *et al.*, 2017] Kun Zhang, Biwei Huang, Jiji Zhang, Clark Glymour, and Bernhard Schölkopf. Causal discovery from nonstationary/heterogeneous data: Skeleton estimation and orientation determination. In *IJCAI: Proceedings of the Conference*, volume 2017, page 1347. NIH Public Access, 2017.
- [Zhang *et al.*, 2023] Xinhe Zhang, Lin Han, Chenxuan Lu, Roger S McIntyre, Kayla M Teopiz, Yiyi Wang, Hong Chen, and Bing Cao. Brain structural and functional alterations in individuals with combined overweight/obesity and mood disorders: A systematic review of neuroimaging studies. *Journal of Affective Disorders*, 2023.