PCFL CIF Algorithm: Uncovering Causal Connections in Unemployment, Inflation, and GDP

Team 13
Xiao Zhijia 120090046
Meng Shengnan 120090505
Yao Yuan 119010396
Wang Yue 120090778

1 Introduction

The project of our group aims to explore the causal relationships [1] between unemployment rate, inflation value, and GDP using Pragmatic Causal Feature Learning (PCFL) algorithm and confounder identification-free (CIF) method.

First, we collected the unemployment rate, inflation value and GDP of five countries, including China, France, USA, Denmark, and Germany. Then, we analyzed the data by calculating their correlation coefficient. We also determined the causal relationship using the directed acyclic graph and Additive Noise Model. After that, we use two methods, CIF method and PCFL method, to further explain the casual relationship. In CIF method, we built a model which not only can forecast the GDP when given the inflation value and unemployment rate, but also can ignore the effect from confounders. PCFL method shows how much effect each factor has on the GDP by identifying coarsened variables.

Ultimately, we concluded that the unemployment rate has a causal effect on inflation rate and GDP. Also, inflation rate has a causal effect on GDP. In addition, two models were set up to explain how unemployment rate and inflation rate affect the GDP.

2 Data Processing

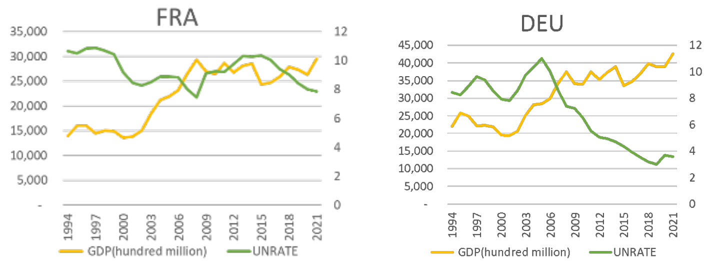
We collect GDP, inflation rate and unemployment rate data for some countries from 1994 to 2021. The included countries are China, the United States, Germany, Denmark and France. In our project, GDP measures the monetary value of final goods and services produced in a country in a given period of time. Inflation rate refers to an overall increase in the Consumer Price Index (CPI), which is a weighted average of prices for different goods. Unemployment rate is measured in numbers of unemployed people as a percentage of the labor force.

To visualize the relationship between unemployment rate, inflation value and GDP, we first drew the line charts of each country and calculated the correlation coefficient. According to the graphs and coefficient, it is easy to find out that unemployment rate and inflation value have negative correlation with GDP, though the correlation is not that strong. Due to the page limits, we only display the line charts which typically show the relationship.

3 Methodology and Results

3.1 DAG

To get the first view of the causal relationship between features [2], we introduce directed acyclic graphs (DAG) and additive noise model (ANM)[5]. To normalize the overall data, we regard Germany, Denmark and France as a whole country. We calculate the combined GDP of these three countries. We also average the inflation rates and unemployment rates in the three countries. In order to distinguish the data of different countries, we label China as 0, the United States as 1 and European countries as 2. We use Pandas in python to merge data and



(a) (b)

Figure 1: The line chart between GDP and Unemployment Rate (with correlation coefficients equal to -0.40 in France and -0.76 in German)

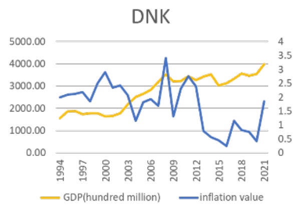
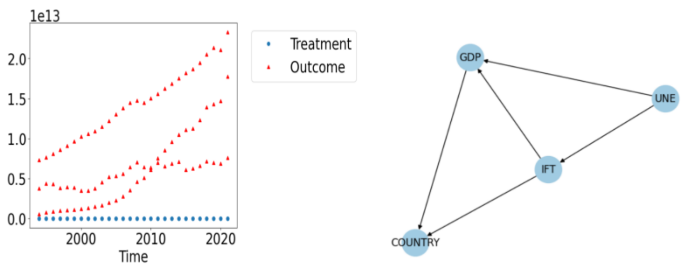


Figure 2: The line chart between GDP and Inflation value (with correlation coefficient equals to -0.43 in Denmark)

finally we get a 112*4 data table. In a DAG, each vertex represents an object or event, and each directed edge represents a causal relationship between two objects or events. In specific, we firstly estimate the sparse inverse covariance matrix between variables to discover conditional independence relationships between variables and construct the skeleton structure. Next, we decompose each variable into its causal and non-causal components. ANM represents each variable as a linear combination of its parent nodes plus a random noise term. Then, by performing regression analysis on these variables, it estimates the causal and non-causal parts of each variable and uses the relationships between these parts to infer causal structure. Finally, we use network library in python to visualize the causal graph. From Fig 3, we can see that from 1994 to 2021, the correlation between GDP and inflation rate and the correlation between GDP and unemployment rate are weak, but unemployment rate has a causal effect on inflation rate and GDP and inflation rate has a causal effect on GDP. Therefore, the correlation cannot correctly show the real relationships between features. We should research deeply by the following models.



(a) (b)

Figure 3: DAG method

3.2 Confounder Identification-free (CIF) Method

3.2.1 Methodology

One of the methods we implemented is the confounder identification-free method introduced by Li et al. [4]. This method eschews the identifications of confounders by applying front-door criterion instead of back-door criterion.

Back-door criterion is typically used to achieve the *do* operation, as the formula displays:

$$P(Y \mid do(X=x)) = \sum_{c} P(Y \mid X=x, C=c)P(C=c).$$

While the implementation of this criterion acquires the access to distributions of all confounders $c \in C$, front-door criterion avoid that by introducing an intermediate variable Z, as the Fig 4(a) shows. Then, the effect of X on Y can be formulated as:

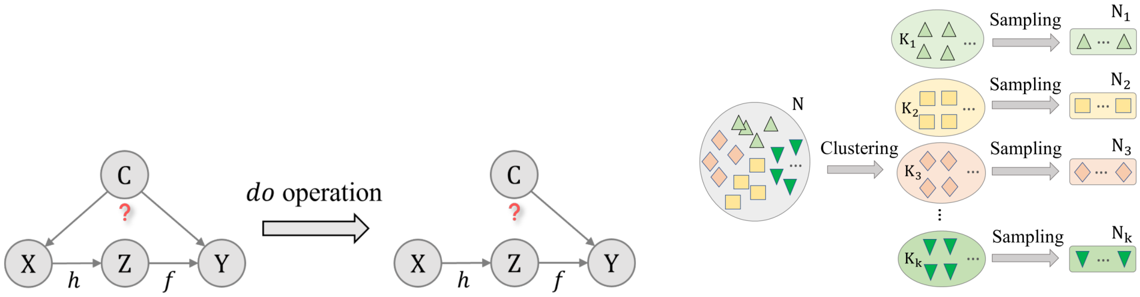
$$\begin{aligned}
P(Y \mid do(X)) &= \sum_{z} P(Z=z \mid X) \sum_{\tilde{x}\in X} P(Y \mid Z=z, \tilde{x})P(\tilde{x}) \\
&= \sum_{\tilde{x}\in X} P(Y \mid Z=h(x), \tilde{x})P(\tilde{x}) \\
&= \sum_{\tilde{x}\in X} f_{\theta_{\tilde{x}}}(Z=h(x))P(\tilde{x})
\end{aligned}$$

The intermediate variable Z is determined by X, i.e. $Z = h(x)$ where model h is fixed. Model f is what we trained later, and the parameter $\theta_{\tilde{x}}$ of f is updated with the gradient $g_{\tilde{x}}$ of sample $\tilde{x}$ as $\theta_{\tilde{x}} = \theta - \alpha g_{\tilde{x}}$.

Applying the first-order Taylor's expansion, the effect of X on Y can be approximated as the following:

$$\begin{aligned}
P(Y \mid do(X)) &\approx f_{\theta}(h(x)) - \alpha\Big(\sum_{\tilde{x}\in X} g_{\tilde{x}}P(\tilde{x})\Big)\nabla_{\theta} f_{\theta}(h(x)) \\
&\approx f_{\theta}(h(x)) - \alpha g_{\dagger}\nabla_{\theta} f_{\theta}(h(x)) \\
&\approx f_{\theta_{\dagger}}(h(x)),
\end{aligned}$$

where $\sum_{\tilde{x}\in X} g_{\tilde{x}}P(\tilde{x})$ is approximated to $g_{\dagger} = \frac{1}{M}\sum_{k=1}^{K}\sum_{j=1}^{N_k} g_{\tilde{x}_{j,k}}$ with *clustering-then-sampling* algorithm displayed in Fig 4(b). Then, the effect of X on Y without global intervention can be viewed as model f updated with $\theta_{\dagger} = \theta - \alpha g_{\dagger}$. Thus, we can train the model f with the cross entropy loss between $f_{\theta_{\dagger}}(h(x))$ and the ground-truth y.



(a) Front-door criterion

(b) Clustering-then-sampling algorithm

Figure 4: Confounder identification-free method

3.2.2 Results

The result of CIF method is a neural network model, whose input is all the related information and output is the predicted conditional probability $P(Y \mid X)$. In our project, the input of model are the unemployment rate and inflation value of China, and the output is the label of China's GDP level. In practice, the input data can involve much more information than only unemployment rate and inflation value. The method is so general that we can even add the previous GDP of the selected country to further increase the prediction value. Nevertheless, this will be irrelevant to the causality among unemployment rate, inflation and GDP.

3.3 Pragmatic Causal Feature Learning (PCFL)

This section focuses on the coarsened causal relationships between several features and GDP throughout five countries by Pragmatic Causal Feature Learning (PCFL). Although the previous model can achieve great performance with its confounder-free method, we also need to the real cause and effect relationships in reality instead of a black box.

Kinney and Watson [3] have proved that PCFL avoids certain counter-intuitive and undesirable consequences of the original CFL algorithm without sacrificing its attractive measure theoretic properties. Our approach is also more computationally efficient if agents rely on utility functions less complex than the KNN algorithm.

3.3.1 Utility Function

$$U(unemployment, inflation) = labor\ constant \times productivity^{1-unemployment} \times (1 + inflation)$$

In PCFL, the utility function is to guide the model to identify important causal features related to the target variable (such as GDP). It can ensure that the causal relationship extracted from the data is more consistent with the application scenario. The above utility function is learned from Cobb–Douglas production function, and is transferred to extract causal features (i.e. inflation and unemployment rate). We collect the labor constant and productivity of different countries from World Bank.

3.3.2 Process

We use a polynomial regression model to capture the complex relationships in the data and predict the target variable, GDP. After that, we used the K-means to group similar observations on the causal features provide useful information for subsequent pragmatic interventions. Next, we fitted a linear regression model with polynomial features to estimate the GDP values based on the coarse causal features. After obtaining the causal features, we computed the conditional probability tables for the clustered data and identified the coarsened variables.

3.3.3 Results

Our research shows that the PCFL method provides a robust framework for analyzing causal relationships. The coarsened causal variables crs_C0, crs_C1, and crs_C2 are clusters of the original input variables and $U(unemployment, inflation)$. Similarly, crs_E0, crs_E1, and crs_E2 are the coarsened effect variables of GDP. For example, given crs_C0, there is about 34% to observe crs_E0 and 66% of observe crs_E2

	crs_E0	crs_E1	crs_E2
crs_C0	0.341	0.0	0.659
crs_C1	0.0	1.0	0.0
crs_C2	0.905	0.095	0.0

Table 1: Results of PCFL

4 Significance

In this project, the two models we use owns certain degree of innovations. For PCFL method, it is the developed method of causal feature learning (CFL). It helps to find out more targeted feature learning while preserving the basic information of causality. For CIF method, it is first applied on picture feature learning. However, through our efforts, it can be used on the concrete data and gives a confounder-free method. The results of our project help people get comprehensive understanding about the economic phenomenon and provide some provide guidance on macroeconomic policy making. To improve the GDP, government should pay more attention on reducing the unemployment rate and inflation value.

5 Conclusion

In conclusion, our project showed that there are causal relationships between inflation rate and GDP. Also, unemployment rate has a causal effect on both GDP and inflation value. Two models were set up in our project to explain how unemployment rate and inflation rate affect the GDP. The CIF model can predict GDP while provide the unemployment rate and inflation rate. The PCFL model can show which factor is the main causal factor.

References

[1] Krzysztof Chalupka, Frederick Eberhardt, and Pietro Perona. Causal feature learning: an overview. *Behaviormetrika*, 44:137–164, 2017.

[2] Krzysztof Chalupka, Pietro Perona, and Frederick Eberhardt. Visual causal feature learning. *arXiv preprint arXiv:1412.2309*, 2014.

[3] David Kinney and David Watson. Causal feature learning for utility-maximizing agents. In *International conference on probabilistic graphical models*, pages 257–268. PMLR, 2020.

[4] Xin Li, Zhizheng Zhang, Guoqiang Wei, Cuiling Lan, Wenjun Zeng, Xin Jin, and Zhibo Chen. Confounder identification-free causal visual feature learning. *arXiv preprint arXiv:2111.13420*, 2021.

[5] Pramod Kumar Parida, Tshilidzi Marwala, and Snehashish Chakraverty. A multivariate additive noise model for complete causal discovery. *Neural Networks*, 103:44–54, 2018.