Analyzing the Influence of Family Background on Students' Learning Motivation Based on Structural Causal Inference Approach

--Taking CEPS data as an example

Yige Li*

School of Business and Management ,Shanghai International Studies University, Shanghai, China

*Corresponding author: lyg2335407262@qq.com

Abstract. Analysing the influence of external factors on students' learning motivation is a hot issue in educational research. This paper takes CEPS data as an example and analyses the influence of various external factors on learning motivation based on the causal inference method, and the results of the analysis and the proposed method can support the optimization of educational decision-making.

Keywords: Family background, parental education, learning motivation, Structural Causal Inference, CEPS.

1. Introduction

Learning motivation is often affected by multiple factors, including social factors, school factors, family factors and other objective factors, as well as subjective factors such as interest in learning, learning goals, learning styles and methods, which are interconnected and affect each other. This paper mainly focuses on the influence of family background on learning motivation, in which the concept of family background mainly focuses on the education level of parents.

This paper adopts the structural causal inference method, which can effectively improve the blind model problem caused by machine learning, and through the causal analysis from the surface to the inside, gradually deepen the causal analysis, to dig out the root causes affecting learning motivation, which is of practical significance for improving the family education environment.

2. Related Work

Li Chongzhao and Wang Fayu used the cluster analysis machine learning method to preprocess and cluster analysis of the collected data of students' behavioral trajectories, and finally came up with the GRU-Attention model that can predict students' motivation [1].

Cheng Shuaiqiang proposed a causal analysis process based on knowledge push, and established a causal analysis architecture based on knowledge management, driven by knowledge demand and knowledge active push, which provides new ideas and directions for the subsequent research on how to use knowledge to assist causal analysis [2].

Lu Shanting, Chen Yang, and Zhao Chunfeng explore the application of causal analysis in teaching specific disciplines-electrical engineering internship. It is proposed that the idea of causal analysis is to evaluate, analyse and classify the possible or important factors that have an effect on the outcome [3].

In summary, a number of scholars use machine learning method to research and explore the problems in the field of education, while the research of structural causal inference method in the field of education is still in the growth stage.This paper attempts to break through the shortcomings of the existing research, based on the student's perspective, and with the help of the structural causal inference method of the establishment of the causal model (Structural Causal Model, SCM) based on students' perspectives to explore the influence of family background.

3. Method

3.1. Structural causal inference method

The structural causal inference method originates from the NIPS 2017 report of Judea Pearl, the father of Bayesian networks [4].Pearl believes that the seven pillars of causal inference theory can solve the problems that cannot be solved by the existing machine learning methods.

One of the major applications of SCM is to combine data for analysis, and the data used in this paper are from the 2013-2014 school year follow-up data of the China Education Panel Survey (CEPS).The part of the analysis focusing on "family background" and "learning motivation" of the research participants is the part of data from the questionnaires of Grade 7 and 9 students and parents after the data processing.

3.2. Description of Variables

3.2.1 General description

This paper combines the common sense of life and questionnaire data subjective and objective two ideas of analysis, to derive the variables of this paper. The main subject of the descriptive object is the family's cause variables: parents' education level (E), economic level (M), educational expectations (W), material input (A), time investment (T), and the main subject of the descriptive object is the student's outcome variables: learning environment (O), physical and mental environment (S), and the final explanatory variable: motivation to learn (P).

3.2.2 Definition of individual variables

Educational background (E): family background in this paper refers to the parents' level of education, as follows Table 1 Options are listed in ascending order by degree of education.

Table 1. Educational background

Score	Corresponding level of education
1	No education at all
2	Primary school
3	Junior high school
4	Secondary/technical school
5	Vocational high school
6	High school
7	University speciality
8	Undergraduate
9	Postgraduate and above

Economic level (M): the subjectively chosen relative economic situation was chosen to be scored (as in Table 2) as an indication of the level of economic conditions of the family.

Table 2. Economic Level

Score	Corresponding Occupation
1	Very difficult
2	Comparatively difficult
3	Moderate
4	Relatively well-off
5	Very affluent

Expectation of education (W): In this paper, the composition of expectation of education is related to the expectation of achievement, degree and career in the questionnaire, and the scores of relative expectation of education are obtained through equal weighting, of which the attitudes of "1.Don't study now", "10. The attitudes of "1. Don't study now" and "10. Doesn't matter" in part B31 are all

regarded as indifference to education, so they are scored as zero, and the rest of the scores are scored in order (as shown in Table 3).

Table 3. Educational Background

Score	Corresponding education level
0	Don't study now, don't care
2	Primary school
3	Junior high school
4	Junior college/technical school
5	Vocational high school
6	High school
7	University speciality
8	Undergraduate
9	Postgraduate and above

Material input (A):choose"the cost of children attending out-of-school tutoring classes or interest classes" in the parent questionnaire as the data for A.

Time investment (T): Study guidance and life accompaniment are equally important in family education, so this paper chooses "checking homework" and "guiding homework" and "reading" and "doing homework" as data.

Learning environment (O): the learning environment in this paper mainly refers to the learning environment formed by the interaction between students in school. Some data are less than 10% (see Figures 1 and 2), which can be treated as invalid data or abnormal data.

		Frequency	Percent	Valid Percent	Cumulative Percent
valid	None	17571	90.2	91.9	91.9
	One or two	1253	6.4	6.6	98.4
	A lot	299	1.5	1.6	100.0
	Total	19123	98.1	100.0	
Missing	System	364	1.9		
Total		19487	100.0		

Fig. 1 Statistics on smoking and alcohol consumption of participants' friends

		Frequency	Percent	Valid Percent	Cumulative Percent
valid	None	17887	91.8	93.6	93.6
	One or two	1026	5.3	5.4	99.0
	A lot	199	1.0	1.0	100.0
	Total	19112	98.1	100.0	
Missing	System	375	1.9		
Total		19487	100.0		

Fig. 2 Statistics on withdrawal of participants' friends from school

Learning environment (O) = Positive factors 1+...+ Positive factors 5 - Negative factors 1-...- Negative factors 5

Physical and mental environment (S): Physical and mental environment in this paper includes two categories: physical health status and psychological status.

Learning motivation (P): Since motivation is a psychological concept that is difficult to measure with specific data, this paper chooses the questionnaire C22, "What level do you want to read to?" to approximate learning motivation.

3.3. Cause and Effect Diagram

The causal analysis method, is a comprehensive representation of the causes and consequences of the graph is intuitive and clear, and can be more objective assessment of the development of events [5].

In this paper, we introduce the idea and concept of causal analysis, starting from the parents' education level E to study whether the intermediate variables have a causal relationship with each other, which ultimately affects the motivation to study P. This process is shown in Fig. 3, where arrows with direction are connected to indicate that there is a causal relationship between two variables.

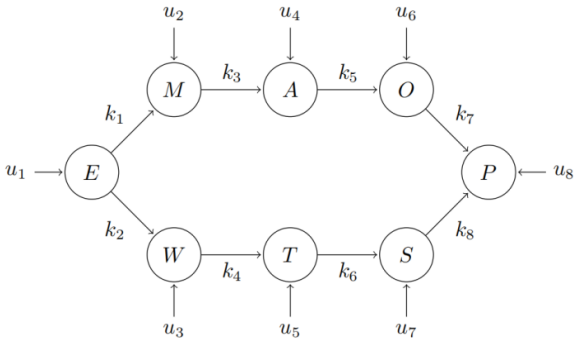


Fig. 3 Cause and effect diagram

The meaning of the causal diagram shown in Figure 6 is as follows:

In particular, the causal effect of parents' education level E on economic level M and educational expectations W: from the perspective of family education, data from different countries or regions show that indicators of socio-economic status, such as parents' occupations and education levels, can have an intergenerational effect on their children's access to education and their academic achievements. For example, parents with higher education levels tend to have higher socio-economic status and are able to create better family environments and educational opportunities for their children, making material inputs very rich; parents with higher education levels tend to have higher cultural capital and better educational concepts and behaviors as well as family cultural atmosphere [6].

Economic level M and educational expectations W are causal for material inputs A and time inputs T: childcare inputs consist of two aspects, the quantity and quality of inputs on the material side and the quantity and quality of inputs on time [7]. Empirically, the level of economic level affects the amount of material inputs and the level of educational expectation affects the amount of time inputs.

The causal relationship between material inputs A and time inputs T for learning environment O and physical and mental environment S: material inputs determine the impact of material level factors on learning motivation, for example, the school private learning tuition is expensive and the learning environment is better, which enhances students' motivation to learn at the material level. And time input determines the influence of mental level factors on learning motivation, for example, parents guide students' homework and enhance parent-child relationship, which enhances students' learning motivation in the mental aspect.

3.4. Construction of structural equations

From Figure 6, the equation variables are parental education (E), economic level (M), educational expectation (W), material input (A), time input (T), learning environment (O), physical and mental environment (S), and the final dependent variable: motivation to study (P), and the introduction of confounding variable $u(x) = u_n$ (n taken as 2-8) as the other influencing factors.

The causal relationship between the variables is derived from Figure 1, and in order to quantitatively describe the causal relationship and simplify the solution, the functional relationship is assumed to be in the form of a simple primary function, i.e., $Y=F(x)=k_mx+u(x)$, where m is taken as 1-7 to obtain:

$$\begin{aligned} E &= u_1 \\ M &= F(E, u_2) = k_1 u_1 + u_2 \\ W &= F(E, u_3) = k_2 u_1 + u_3 \\ A &= F(M, u_4) = k_3 M + u_4 \\ T &= F(W, u_5) = k_4 W + u_5 \\ O &= F(A, u_6) = k_5 A + u_6 \\ S &= F(T, u_7) = k_6 T + u_7 \\ P &= F(O, S, u_8) = k_7 O + k_8 S + u_8 \end{aligned} \tag{1}$$

The equation for the final variable P is obtained by substituting the coefficients from the equation:

$$P=(k_1k_3k_5k_7+k_2k_4k_6k_8)u_1+k_3k_5k_7u_2+k_4k_6k_8u_3+k_5k_7u_4+k_6k_8u_5+k_7u_6+k_8u_7+u_8 \tag{2}$$

The equations were solved by combining the data and the following results were obtained：

$$\begin{aligned} M &= 0.043E + 2.470 \\ W &= 0.031E + 14.563 \\ A &= 737.055M - 820.641 \\ T &= -0.187W + 18.300 \\ O &= 0.0000241A + 7.985 \\ S &= 0.183T - 5.983 \\ P &= 0.202O + 0.022S + 5.056 \end{aligned} \tag{3}$$

The final substitution of each intermediate variable yields the equation for P and E as:

$$P = 0.00013235\ E + 6.605096 \tag{4}$$

From the above equation, it is obtained that parents' education level E has a positive relationship with family income level M and educational expectation W, which is in line with the causal analysis in the earlier part of this paper; while educational expectation W has an inverse relationship with time payment T, i.e., in the case of higher expectations, the time payment is rather less, which reflects the overall situation where parents do not make a high investment of time due to high educational expectations. Ultimately, the range of fluctuation of P with the change of E is [0,9], based on the observation that the coefficient of the functional relationship that exists between P and E is 0.00013235, the value of which indicates that the effect of the value of E on P is almost negligible, which means that the functional correlation between E and P is not significant.

4. Conclusion

This paper uses the structural causal inference method, based on CEPS data, to derive an equation for the impact of family educational background on student motivation. The method can be extended to the analysis of similar problems in the field of education to provide support for educational decision-making.

References

[1] LI Chongzhao &WANG Fayu. (2022),Prediction model of students' learning enthusiasm based on gated recurrent unit-attention

[2] Cheng Shuaiqiang. (2021). Research on Causal Analysis Method Based on Knowledge Push(006), 037.

[3] LU Shan-ting &CHEN Yang &ZHAO Chun-feng.(2020),Application of Cause and Effect Analysis Mathod in Practice Teaching

[4] Judea Pearl.(2017),Theoretical Impediments to Machine Learning With Seven Sparks from the Causal Revolution

[5] ZHENG Zhao-ning & WU Yu-ying & BAO Han-ling.(2001),The Combined Application of Fishbone Diag gram and AHP for Corporation Diagnosis

[6] Liu tianyu.(2021),Research on the Influence of Parents' Occupations on Children's Participation in Commercial Medical Insurance.

[7] Sampasa-Kanyinga, H., Chaput, J. P., & Hamilton, H. A. (2019). Social media use, school connectedness, and academic performance among adolescents. The journal of primary prevention, 40, 189-211.