Article

Increased scientific research policies and the sustainable development of the education chain—Taking China as an example

Heng-Guo Zhang*, Tailong Li

Center for Economic Research, Shandong University, Jinan 250100, China
*Corresponding author: Heng-Guo Zhang, zszhg@126.com

Abstract: This paper models 54,559 Chinese news items about education industry and scientific industry by machine learning during the COVID-19 epidemic to build China’s increased scientific research policy (ISRP) index. The result of interrupted time series analysis indicates that, the ISRP has an emphatic positive causality on the education industry advancement and promotes the development of the education industry. The ISRP also has a remarkable positive causality on the development of the scientific industry. Moreover, the result of causal network indicates that, a virtuous circle within the ISRP, the education industry and the scientific industry has been formed, which has promoted the sustainable development of the education chain.

Keywords: COVID-19; scientific industry policy; education industry; scientific industry; causal network; sustainable development

1. Introduction

The purpose of this paper is to investigate the influential mechanism of China’s increased scientific research policies (ISRP) to facilitate the sustainable development of the education industry during the coronavirus disease 2019 (COVID-19) epidemic. Khan et al. (2022) have studied supply chains, this paper wants to study education chain. Therefore, it is to test whether a virtuous circle can be formed among ISRP policy, the education industry and the scientific industry, so as to promote the sustainable development of education chain. Sustainable development (SD) has attracted the attention of many researchers, and government policy support is regarded as one of the keys to the successful implementation of sustainable development (Filho et al., 2019). Human society needs to find a sustainable development path to cope with increasing pressures such as public health risks (Didham and Ofei Manu, 2020). Individuals with COVID-19 are now confirmed in almost every country. The current spread of COVID-19 is a global pandemic and has urged affected countries to adopt containment and suppression measures, such as travel restrictions, social distancing, treatment of patients, and increased scientific industry, as methods to slow the virus (Anderson et al., 2020; Hellewell et al., 2020; Shim et al., 2020).

Chinese government has proposed policies to increase scientific industry, organize interdisciplinary and cross-disciplinary scientific industry teams, and promote cooperation with other countries in scientific industry, education, and clinical work. Positive progress has been made in just over one month, providing strong scientific and technological support for epidemic prevention and control. In addition, as an important environmental problem, air pollution has become increasingly serious worldwide, and air pollution is an important factor restricting the sustainable...
development of the urban environment. Epidemiological studies have shown that air pollution will increase the death and morbidity rates associated with certain diseases (Cheng et al., 2007; Liu et al., 2018). The government needs to develop concrete ways of implementing policies and solutions and to carefully evaluate the link between efforts to implement these policies and the sustainability of their results.

There are three problems studied in this paper. First, the existing literature pays little attention to the relationship between scientific research policy and the sustainable development of education chain during the epidemic period. This lack of attention is that the time of such public emergencies is too short to get macro data, which are prone to the “dimension curse” (Marcellino and Sivec, 2016). Therefore, it is often difficult to conduct a comprehensive analysis of economic shocks using traditional methods because of data limitations (Galariotis et al., 2018). How to use machine learning to construct macro data through news texts is the first problem studied in this paper.

Second, the interrupted time series analysis (ITSA) can compare the results of the indicators before and after the intervention to evaluate the effects of policy interventions (Wagner et al., 2002). Therefore, whether China’s increased scientific research policy (ISRP) has promoted the development of the education industry and the scientific industry is the second question studied in this paper, which is done using the ITSA method.

Third, when analyzing the public emergencies in the existing literature, most studies use intervention models, case studies, event studies, natural experiments, etc. They only conduct comparative analyses before and after public emergencies, not to study the economic effect of the public emergencies themselves (Boehm et al., 2019). How to analyze the impact of China’s increased scientific research policy on the sustainable development of education chain during the epidemic is the third issue studied in this paper.

In response to the above problems, there are three innovations in this paper. First, to our knowledge, there is no scientific research policy index research in the existing literature, such as Tejedor et al. (2020). This paper analyses the daily news related to education and scientific industry by latent Dirichlet allocation (LDA) model under the COVID-19 epidemic, which are used to build China’s increased scientific research policy index. The economic significance of the index lies in the quantitative measurement of ISRP itself in the face of the COVID-19 epidemic. In addition to news text data, real-time data such as video data can be studied in the future to construct policy evaluation index.

Second, existing literature often analyzes local aspects of the education industry, such as online classes (Khan et al., 2022). This paper tries to study the entire education industry. The interrupted time series analysis (ITSA) method is a good method to evaluate the longitudinal effects of policy interventions (Wagner et al., 2002). This paper analyzes the economic impact of ISRP policy on the education industry before and after its implementation by ITSA method. This study finds that ISRP not only promotes the sustainable development of the education industry but also promotes the sustainable development of scientific industry.

Third, as far as we know, there is no research on the influential mechanism of emergency regulatory policies to the education chain in existing literature, such as online courses (Hodges et al., 2020). In this paper, PCMCI model is to estimate the
education chain causal network and quantify its strength. Compared with the traditional two-variable causal model, this method simultaneously evaluates the degree of mutual influence among one variable and other variables (Runge et al., 2019; Zhang et al., 2021). The use of causal network models compensates for the shortcomings of existing research methods (such as intervention models, case studies, event studies, and natural experiments). Therefore, this paper can analyze the influential mechanism of ISRP on the sustainable development of education chain. The results show that after the implementation of the ISRP policy, a virtuous circle within the ISRP policy, the education industry and the scientific industry has also been formed, which has promoted the sound development of education chain.

2. Literature review

Over the past two decades, the role of education in promoting sustainable development has aroused the interest of many researchers and scholars around the world (Alvarezgarcia et al., 2018; Merrittet al., 2019). Education chain has the ability to develop human capital in a more sustainable way by improving individuals’ knowledge, skills and abilities (Nousheen et al., 2020). Higher education institutions need to participate in sustainable development practices, with education, research, internal management and community participation as the main areas for research and development (Filho et al., 2019). Some researchers have proposed multiple methods to incorporate sustainability into education (Rusinko, 2010). For example, one method is by proposing new courses (Pappas et al., 2013) and evaluation models (Savelyeva and McKenna, 2011). Another method is to incorporate a manual on sustainable development and curriculum integration (Ceulemans and De Prins, 2010). Other researchers suggest a comprehensive reorganization of the university to meet this challenge (Aktas et al., 2015; Filho et al., 2017; Isenmann et al., 2020).

The education has been affected by COVID-19 (Tejedor et al., 2020). Aguilera-Hermida (2020) found that education motivation, self-efficacy, and education technology played important roles in students’ cognitive engagement. Cao et al. (2020) showed that anxiety symptoms among college students were generally positively correlated with economic impact and delayed academic activities. Wilde and Hsu (2019) found that compared to other individuals with higher overall self-efficacy, individuals with lower overall self-efficacy in searching for alternative experiential information have significantly smaller benefits for their self-efficacy in completing established tasks. Hodges et al. (2020) indicated that a carefully planned online learning experience was significantly different from online courses that responded to crises or disasters. Yakubu and Dasuki (2019) analyzed in detail various factors in developing countries that affected the adoption and use of educational technology by students in higher education institutions. Khan et al. (2022) discussed college students’ views on the acceptance of online learning during COVID-19.

Public emergencies have aroused widespread concern among scholars. Guan et al. (2020) quantitatively evaluated the short-term supply chain impact of different containment strategies between countries and industries, revealing how the economic losses associated with COVID-19 will be distributed in the global supply chain. Mirza et al. (2020) found that social entrepreneurship funds were limited by their flexibility...
during the COVID-19. Goodell and Huynh (2020) examined the effect of COVID-19 news on United States industries. Hsiang et al. (2020) found that anti-infection policies can reduce the spread of COVID-19. The differences in the stringency of these policies and the speed of policy relaxation reflect different assessments of the public health risks of COVID-19 and the social and economic impacts of different policies (Wells et al., 2020). Many scholars have studied Chinese air pollution. Shao et al. (2016) investigated the air pollution control capacity from 1998 to 2012. Li (2016) conducted an empirical study on the factors for Beijing’s air pollution. Leng and Du (2016) found that energy price has a positive effect on air pollution in China from 2001 to 2010. Huang (2017) investigated whether fiscal decentralization had an impact on air pollution from 2001 to 2010. Shao et al. (2019) measured China’s air control capabilities from 1998 to 2013. Zhang and Chen (2020) found that air pollution control is still relatively low in China.

In summary, there are several shortcomings in the evaluation of emergency regulatory policies to the education chain in existing literature. Firstly, due to the short duration of emergencies, it is often difficult to obtain relevant real-time data. Secondly, after relevant economic data is available, evaluating the effectiveness of emergency regulatory policies has a strong lag in policy evaluation. Thirdly, the current method of evaluating emergency regulatory policies, such as intervention models, case studies, event studies, and natural experiments, cannot analyze the influential mechanism of emergency regulatory policies.

The remaining arrangements are as follows. Section 3 presents the method. Section 4 displays the data. Section 5 shows the empirical results. Section 6 exhibits the conclusion.

3. Experimental methods

This paper uses a variety of methods to study the impact of scientific industry policy on the education industry, the scientific industry, COVID-19 transmission and air quality from multiple perspectives.

3.1. Causal network model

Consider a causal network model (PCMCI) \( Z_t = (Z_t^1, \ldots, Z_t^N) \) with variable independence (Runge et al., 2019):

\[
Z_t^i = f_j(P(Z_t^j), \xi_t^j)
\]

(1)

where \( f_j \) is nonlinear function, and \( \xi_t^j \) is noise. \( P(Z_t^j) \subset Z_t^\neg = (Z_{t-1}, Z_{t-2}, \ldots) \) is the causal parents of the time variables \( Z_t^j \) with lag. The PCMCI model is also on account of a conditionally independent framework \( Z_t^i = \hat{P}(Z_t^i) \beta \) and contains two phases.

To uncover the relevant conditions \( \hat{P}(Z_t^i) \) of all time series variables \( Z_t^i \in \{Z_t^1, \ldots, Z_t^N\} \) by the PC1 algorithm is the first phase. The second phase shows the conditional independence test (MCI) to test whether \( Z_{t-\tau} \rightarrow Z_t^i \) has the following relationship:

\[
\text{MCI: } Z_{t-\tau} \forall Z_t^i | \hat{P}(Z_T^i) \setminus \{Z_t^i | Z_{t-\tau}\}, \hat{P}(Z_{t-\tau})
\]

(2)
So, the MCI test involves the time condition changes for the parents of $Z_t^1$ and $Z_{t-\tau}$.

### 3.2. Interrupted time series analysis method

The standard interrupted time series analysis (ITSA) model takes the following formula (Linden, 2015):

\[
Y_t = \gamma_0 + \gamma_1 T_t + \gamma_2 Z_t + \gamma_3 Z_t T_t + \zeta_t
\]

\[
\zeta_t = \varphi \zeta_{t-1} + \sigma_t
\]

where $Y_t$ is the summary result variable at time $t$, $Z_t$ is an indicator variable, and $Z_t T_t$ is an interaction term. $\gamma_0$ represents the starting variable. $\gamma_1$ stands for the slope of the outcome variable before the intervention is introduced. $\gamma_2$ indicates the change after the intervention occurs. $\gamma_3$ denotes the difference of the slope before and after the intervention. $\varphi$ reveals the correlation coefficient with $|\varphi| < 1$, and $\sigma_t$ are independent with distribution $N(0, \sigma^2)$.

### 4. Experimental data

The selected experimental data measure the return on equity (ROE) for 48 listed companies from 17 January 2020 to 2 July 2020 related to the education chain from the CSMAR database. These 48 listed companies are split across two industries related to the education chain. Specifically, they are split into the education industry (industry code P82), from which 21 listed companies are selected, and the scientific industry (industry code M73), from which 27 listed companies are selected. The company profit is measured by return on equity (ROE). The industry’s ROE (ROE73 or REO82) is to evaluate the profit of the industry. In addition, to study the effect of China’s increased scientific industry policies on air pollution control, the Air Quality Index (AQI) from 17 January 2020 to 2 July 2020, is selected from the CSMAR database. The bigger the AQI value is, the higher the level of air pollution.

Due to the short time span of emergencies, relevant data is often difficult to obtain. Existing literature is not possible to conduct in-depth analysis of emergencies (Galariotis et al., 2018). In addition, existing literature needs to wait for relevant economic data to be available before evaluating policy effectiveness, which often has a strong lag and cannot be evaluated in real-time. Traditional economic models cannot handle text data, but machine learning can handle news text data. Therefore, this paper adopts machine learning (LDA model) for processing text data.

We use a LDA model (Blei et al., 2003) to get China’s increased scientific research policy index (ISRP). 54,559 Chinese news items related to education and scientific industry from 17 January 2020 to 2 July 2020, were extracted from the China InfoBank Database. The LDA model was used to construct the ISRP index to represent the real-time implementation effect of policies. Figure 1 shows the daily changes of the index.
In addition, the total number of COVID-19 cases in China from 17 January 2020 to 2 July 2020, is gathered from the Chinese Center for Disease Control and Prevention to study the effect of the ISRP on preventing the spread of COVID-19.

5. Experimental results

In this section, the influence of the ISRP on the education industry and the scientific industry is studied from multiple perspectives.

5.1. Examining the scientific industry

The ITSA method is to examine whether the ISRP can facilitate the advancement of the scientific industry (ROE73). Table 1 displays the results of scientific industry.

Table 1. The results of the scientific industry.

<table>
<thead>
<tr>
<th>Test A</th>
<th>( \beta_0 )</th>
<th>( \beta_1 )</th>
<th>Value</th>
<th>( \beta_2 )</th>
<th>( \beta_3 )</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Test A</td>
<td>0.0925***(0.000)</td>
<td>−0.0001***(0.000)</td>
<td></td>
<td>0.0037***(0.000)</td>
<td>0.0002***(0.000)</td>
<td></td>
</tr>
<tr>
<td>Test B</td>
<td>Trend</td>
<td>Value 0.0001***(0.000)</td>
<td>Test C</td>
<td>Value 91.028***(0.000)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Note: ***, ** and * indicate significance at the 1%, 5% and 10% levels, respectively. Test A shows the Newey-West Test. Test B is Post-intervention Linear Trend Test. Test C displays Cumby-Huizinga Test. () is \( p \)-values.

Figure 2 clearly shows the impact of the ISRP on the return on equity of the scientific industry (ROE73). The time on the horizontal axis represents the number of days. To our knowledge, there is no scientific research policy index research in the existing literature, such as Tejedor et al. (2020). This paper tries to study the impact of ISRP on the entire scientific industry. Due to limited epidemic data, we took the middle of the sample period as the starting point for the policy intervention. As displayed in Table 1, the results of the test A indicate that the initial return for ROE73 is 0.0925. Before the implementation of the ISRP, ROE73 had a downward trend of −0.0001. On the first day after the policy intervention, ROE73 increased dramatically by 0.0037. The trend after the intervention was larger by 0.0002 than 0.0925, which is consistent with the results of test B in Table 1. The results of test B show that ROE73
has increased at a rate of 0.0001 per day after the implementation of the ISRP. The result of test C in Table 1 displays that autocorrelation can be obtained if the lag is 1. So, the lag (1) command can correct for the autocorrelation. In Figure 2, after the implementation of the ISRP, ROE73 has a positive growth trend, no longer a negative growth trend. It means that the ISRP have a significant positive causality on the development of the scientific industry.

Figure 2. The results of the scientific industry.

5.2. Examining the education industry

The ITSA method is to examine whether the ISRP can upgrade the advancement of the education industry (ROE82). Table 2 shows the results of education industry.

Table 2. The results of education industry.

<table>
<thead>
<tr>
<th>Test</th>
<th>$\beta_0$</th>
<th>Value</th>
<th>$\beta_1$</th>
<th>Value</th>
<th>$\beta_2$</th>
<th>Value</th>
<th>$\beta_3$</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td></td>
<td>0.1855*** (0.000)</td>
<td></td>
<td>0.0013*** (0.000)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>B</td>
<td></td>
<td>0.0033 (0.396)</td>
<td></td>
<td>-0.0009*** (0.000)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>C</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Test B: Post-intervention Linear Trend Test. Test C displays Cumby-Huizinga Test. () is $p$-values.

Figure 3 reveals the impact of the ISRP on the return on equity of the education industry (ROE82). The time on the horizontal axis represents the number of days. Existing literature often analyzes local aspects of the education industry, such as online classes (Khan et al., 2022). This paper tries to study the impact of ISRP on the entire education industry. Due to limited epidemic data, we took the middle of the sample period as the starting point for the intervention. The results of test A in Table 2 reveal that the initial return for ROE82 is 0.1855. Before the implementation of the ISRP,
ROE82 had an upward trend of 0.0013. After the policy intervention, the return on net assets rose by 0.0033 on the first day. The trend decreases by −0.0009 from a baseline of 0.1855, which is consistent with the results of test B. The results of test B indicate that ROE82 has increased at a rate of 0.0005 per day after the implementation of the ISRP. The result of test C is similar to Table 1. In Figure 3, after the implementation of the ISRP, ROE82 is further improved. This result shows that the ISRP has a positive causality impact on the advancement of the education industry.

![Figure 3. The results of education industry.](image)

5.3. Causal network test on the education chain

In this paper, the education chain includes all companies in scientific industry (ROE73) and education industry (ROE82). The causal network (MCI) test is to analyze the influential mechanism of the ISRP on the sustainable development of the ROE73 and ROE82. Table 3 shows the results of MCI test.

<table>
<thead>
<tr>
<th>Table 3. The result of the MCI test.</th>
</tr>
</thead>
<tbody>
<tr>
<td>ROE73</td>
</tr>
<tr>
<td>ROE82</td>
</tr>
<tr>
<td>ROE82</td>
</tr>
</tbody>
</table>

Note: ***, ** and * indicate significance at the 1%, 5% and 10% levels, respectively. () is p-values.

In general, there are three criteria for determining p-values in literature: 0.01, 0.05, and 0.1 (Su et al., 2016; Zhang and Li, 2023). If the p-value is less than 0.01, it
is considered significant at the 0.01 level. If the p-value is less than 0.05 (and greater than or equal to 0.01), it is considered significant at the 0.05 level. If the p-value is less than 0.1, it is considered significant at the 0.1 level. Therefore, all three criteria are statistically significant, with only differences in accuracy. It is acceptable to take a p-value less than 0.1 in this paper.

Figure 4 exhibits the MCI test of the ISRP on the education chain. As far as we know, there is no research on the influential mechanism of scientific research policies to the education chain in existing literature, such as online courses (Hodges et al., 2020). The use of causal network models compensates for the shortcomings of existing research methods (such as intervention models, case studies, event studies, and natural experiments). In this paper, the ISRP promotes a virtuous circle within the education chain. Specifically, ISRP has a bidirectional causality with ROE73. The positive causal strength is 0.301 in Table 3. ISRP has a bidirectional causality with ROE82. The positive causal strength is 0.213 in Table 3. ROE82 has a bidirectional causality with ROE73, and its positive causal strength is 0.194 in Table 3. Therefore, the ISRP has promoted the virtuous circle of the education chain (education industry and scientific industry).

![Figure 4. The result of MCI test.](image)

6. Conclusion

The objective of this paper is to investigate the influential mechanism of emergency regulatory policies to the education chain during the COVID-19. Due to the short time span of emergencies, relevant data is often difficult to obtain. Existing literature is not possible to conduct in-depth analysis of emergencies (Galariotis et al., 2018). In addition, existing literature needs to wait for relevant economic data to be available before evaluating policy effectiveness, which often has a strong lag and cannot be evaluated in real-time. Existing research methods, such as intervention models, only conduct comparative analyses before and after public emergencies, not to study the economic effect of the public emergencies themselves (Boehm et al., 2019). And they also cannot analyze the influential mechanisms of emergency regulatory policies.

In order to make up for the gap in evaluating emergency regulatory policies to
the education chain in existing literature, this paper models 54,559 Chinese news items related to scientific industry and education industry during the COVID-19 epidemic, and constructs China’s increased scientific research policy (ISRP) index. This paper uses machine learning methods to construct the ISRP policy index to represent the real-time implementation effect of policies. In addition to using the interruption time series analysis method to evaluate the effectiveness of policy intervention, this paper also uses a causal network model to analyze the mechanism of emergency regulatory policies to the education chain. This is a supplement to existing literature as a policy evaluation method.

Existing literature often analyzes local aspects of the education chain, such as online classes (Khan et al., 2022). This paper studies the entire education industry chain, covering a wider and more comprehensive scope. The results show that after the implementation of the ISRP, the return on equity in the scientific industry changed, from a negative growth trend to a positive growth trend. This means that the ISRP has a positive causality on the scientific industry. The ISRP also has a positive causality on the education industry. The return on equity in the education industry has been further improved. Moreover, a virtuous circle within the ISRP, the education industry and the scientific industry has also been formed, which has promoted the sound development of education chain. Therefore, the ISRP promotes the sustainable development of the education chain.

**Author contributions:** Conceptualization, HGZ and TL; methodology, HGZ; software, HGZ; validation, HGZ and TL; formal analysis, HGZ; investigation, HGZ; resources, HGZ; data curation, HGZ; writing—original draft preparation, HGZ; writing—review and editing, HGZ; visualization, HGZ; supervision, HGZ; project administration, HGZ; funding acquisition, TL. All authors have read and agreed to the published version of the manuscript.

**Funding:** This document is the results of the research project funded by [Heng-Guo Zhang of the Fundamental Research Funds of Shandong University] grant number [2020GN009] and by [Tailong Li of the National Natural Science Foundation of China] grant number [72173072].

**Conflict of interest:** The authors declare no conflict of interest.

**Notes**

1. For the specific technical details of the PCMCI model, please refer to Runge et al. (2019) and Zhang et al. (2021) in the references.
2. For the specific technical details of the ITSA model, please refer to Linden (2015) and Zhang et al. (2021) in the references.
3. For the specific technical details of the LDA model, please refer to Blei et al. (2003) and Zhang et al. (2021) in the references.

**References**


