Analysis of Chinese Bayh-Dole Policies Using Natural Language Processing Tools and RDD Model

Ruoya Wang(✉) and Yaodong Zhou

School of Economics and Management, Beijing Jiaotong University, Beijing, China
zoeywry@126.com

Abstract. Based on the text data of BD policies of colleges and universities compiled by hand, this paper uses jieba tool to analyze the characteristics of science and technology achievement transformation policies of colleges and universities, to investigate the attention of colleges and universities to patent transformation activities. Empirical Analysis of the Incentive Impact of Science and Technology Achievement Conversion Policies on Patent Conversion Activities in Universities Using Exact Breakpoint Regression Models. The research finds that the BD policies of universities reflect that Chinese universities pay a high level of attention to scientific and technological achievements, especially patent transfer, licensing and price-setting, which is conducive to the development of social innovation. The local average treatment effect of universities' attention to conversion activities on conversion results is positive, indicating that university BD-type policies have a significant positive contribution to patent conversion.

Keywords: University Patents · Patent Transformation · Policy Analysis · Attention Studies · Incentive Effects

1 Introduction

Under the vigorous promotion of the innovation-driven development strategy, China's patent market has grown by leaps and bounds over more than a decade since it became the world's top patent filing country in 2011, and its share of the global patent market has risen from 25% to 46% (by 2020). Universities are an important force in the patent application in China.

To promote the application and dissemination of patents in universities, the government has issued a series of policies and regulations, which are in common with the content of the Bayh-Dole Act promulgated by the United States in 1980. In response to the call of the national policy, universities have formulated their school-level regulations such as management measures and implementation rules to promote the transformation of scientific and technological achievements. In recent years, universities have generally paid attention to issues related to the transformation of patents, and most of them have introduced relevant incentive policies.

D. Qiu et al. (Eds.): ICBBEM 2022, AHIS 5, pp. 1158–1167, 2023.
https://doi.org/10.2991/978-94-6463-030-5_116

How much attention do universities pay to the transformation of patents? Through natural language processing, this study conducts a textual analysis of universities' Baidoo-type policies. The innovation of this study is to measure the policy intensity based on the concept of policy makers' attention, to extract word frequency indicators from policy texts using natural language processing to characterize universities' attention to patent conversion; to analyze the impact effect of BD-type policies in universities using the exact breakpoint regression method.

Attention is originally a psychological concept that refers to mental concentration on a particular piece of information [3]. Simon (1947) [2] first introduced the study of attention into the field of management and proposed the doctrine of Attention management, which argues that attention is a key scarce resource in the decision-making process. As decision makers are constrained by time, energy and cost to deal with multiple tasks at the same time, attention is reflected in the process of making limited decisions by selectively focusing on certain information and ignoring others.

2 University Attention Analysis

2.1 Index Construction and Data

2.1.1 The Jieba Tool

There is currently no consensus among scholars on how to quantify attention; Whorf et. al (2012) [1] suggest that language determines behaviour, and Wade et. al (1997) [4] argue that the frequency of word occurrences in a text can reflect the writer's focus of attention. Some scholars have suggested that textual analysis is the most effective way to quantitatively measure text-based information. In policy analysis, the mental processes of policy makers can often be reflected through verbal words, and textual analysis can be used to make better use of the unstructured data implicit in policy texts to more objectively measure the attention of policy makers on specific matters. In concrete practice, scholars generally analyse the frequency of keyword words that are directional in policy texts. The importance of a particular word increases proportionally with its number of occurrences in the policy text, indicating that the policy maker pays more attention to that aspect. There is currently little research applying word frequency analysis to the field of economics.

The TF-IDF algorithm is implemented specifically through the jieba word splitting tool in python software. The general steps are to first build a data dictionary, followed by word frequency statistics.

2.1.2 The Sample Texts

In order to capture the attention (focus) of universities on patent conversion activities, the texts of BD-type policies of 72 ministerial institutions were collected. The scope of the collection was the BD-type policies of Chinese universities that have been promulgated and implemented up to 2020, specifically collected through the web search route. Firstly, we searched for relevant policy documents on the official websites of the science and technology departments, scientific research departments or other technology transfer

offices of universities; secondly, we used the internet search engines Baidu and Bing to search for policies and documents by "name of university" plus "management of intellectual property" and "transformation of technological achievements". Secondly, we searched for relevant policy documents on the websites of Baidu and Bing, using "name of university" plus "intellectual property management" and "transformation of scientific and technological achievements" as keywords. The policy documents obtained through the above two channels were collated to obtain a text database of BD policies of universities, according to which statistics on the implementation time and distribution of benefits of each university were compiled to provide a basis for text analysis.

As the data of language, finance and art colleges were seriously missing, and these colleges were less active in applying for patent protection and implementing patent transformation, 11 language, finance and art colleges were excluded according to the classification of colleges and universities by the Ministry of Education's Sunshine College Entrance Examination Information Platform, and the final sample contained 72 colleges and universities.

Descriptive statistics of sample data are shown in Table 2.

2.2 Analysis of Sample Texts

The text analysis method can be implemented with the help of software such as Rostcm, Nvivo and Python, and the degree of attention can be measured by keyword frequency, frequency or TF- IDF value of keywords. In this paper, the TF-IDF algorithm (term frequency-inverse document frequency), a common weighting technique for information retrieval and text mining, was specifically used in constructing the term frequency metric to assess the importance of a word to a document set or a document in a corpus.

A lexicon suitable for measuring policy information was built by analysing 72 university texts on BD-type policies and hand-selecting words that fit the relevant descriptions. Referring to the construction idea of the LM dictionary, the lexicon was selected based on a careful reading of the policy texts, following professionalism and precision. Considering that the epithet of Chinese words is strongly influenced by context, the vocabulary was streamlined using the principle of preferring shortage to overuse in order to improve the accuracy of word selection, specifically by removing repetitive words with consistent ideograms and by merging and collating them. Based on the results of the jieba lexical classification, the word frequency counts of the words within the lexicon in all policy texts were counted, and the frequency counts of individual words were measured as a proportion of the total word frequency counts of all words in all texts, and the results were sorted in descending order to distinguish high-frequency words in the lexicon. Table 1 lists the top 50 words in terms of word frequency, giving the word frequency of each word and its proportion of the total word frequency in the lexicon.

The results are shown in Table 1. It can be found that the words "scientific and technological achievements", "transformation" and "transfer" have the highest word frequency, which directly reflects the key contents of BD-type policies in universities. The words "intellectual property rights", "contracts" and "assets" also have a high word frequency share, indicating the main objects and connotations of the policies; "promoting The importance of words such as "promote", "service" and "benefit distribution" is also high in the lexicon, indicating the purpose, function and mechanism of BD policies.

Table 1. Descriptive statistics of the sample universities.

word	Frequency of the words	percentage
Scientific and technological achievements	4357	7.32%
conversion	3423	5.75%
The school	3096	5.20%
technology	1553	2.61%
management	925	1.55%
transfer	916	1.54%
The transfer	753	1.26%
Intellectual property rights	630	1.06%
The implementation of	630	1.06%
reward	582	0.98%
position	514	0.86%
The license	507	0.85%
evaluate	469	0.79%
The contract	469	0.79%
earnings	442	0.74%
investment	422	0.71%
assets	400	0.67%
assessment	386	0.65%
patent	342	0.57%
equity	340	0.57%
agreement	329	0.55%
enterprise	622	1.04%
The scientific research	311	0.52%
cooperation	289	0.49%
To promote	283	0.48%
perform	273	0.46%
Take a stake in	269	0.45%
organization	267	0.45%
convention	255	0.43%
The proportion	255	0.43%
use	247	0.41%
institutions	247	0.41%

(*continued*)

Table 1. (*continued*)

word	Frequency of the words	percentage
pricing	230	0.39%
To apply for	227	0.38%
activity	222	0.37%
team	217	0.36%
service	215	0.36%
The development of	210	0.35%
personal	199	0.33%
business	193	0.32%
audit	191	0.32%
trading	190	0.32%
Science and technology personnel	189	0.32%
entrepreneurship	186	0.31%
To undertake	182	0.31%
Income distribution	449	0.75%
contribution	175	0.29%
The price	171	0.29%
college	167	0.28%
To handle the	167	0.28%

Table 2. Descriptive statistics of sample data.

indicators	frequency
Mean	50.30
maximum value	114.00
Minimum value	9.00
standard deviation	25.45
number of samples	72.00

Combining the essence of BD-type policies and the performance of patent transformation in universities at the present stage, the study concludes that the words listed in the table can better reflect the positive messages released by the sample universities on the implementation of patent transformation. The 50 words in Table 1 represent 2.32% of the total number of words in the lexicon, while the total number of word frequency accounts for 48% of the total word frequency of all words in the total text sample, implying that

the results of the text analysis can better indicate the content characteristics of the sample. In selecting the key words, considering that the word "scientific and technological achievements", which is the first word in terms of frequency, only indicates the subject of transformation and cannot characterize the essence of patent transformation in universities; the word "transformation" indicates the key initiative of patent transformation in universities, so the word "transformation" was selected.

3 Empirical Analysis

3.1 Hypothesis

Universities, as decision makers, are the subjects that exert attention on patent conversion activities, and BD-type policies are the carriers of such attention. It is generally believed that the stronger the attention of a policy-making subject on a certain matter, the richer the corresponding results achieved are likely to be. Accordingly, hypothesis H1 is proposed.

H1: The attention of universities has influence on patent conversion activities.
H1a: The attention of universities has a positive contribution to patent conversion activities.
H1b: The attention of universities has a negative inhibitory effect on patent conversion activities.

3.2 Models and Variables

In this paper, we choose the clear breakpoint regression method, and in order to assess the effect of the BD-type policy, we need to estimate the local average treatment effect of the policy at the breakpoint. The expression of the local average treatment effect estimator LATE (Local Average Treatment Effect) is as follows.

$$LATE = \frac{\lim\limits_{x_i \uparrow c} E[Y|x_i] - \lim\limits_{x_i \downarrow c} E[Y|x_i]}{\lim\limits_{x_i \uparrow c} E[D_i|x_i] - \lim\limits_{x_i \downarrow c} E[D_i|x_i]}$$

$$D_i = \begin{cases} 0, x_i < c \\ 1, x_i \gg c \end{cases} \tag{1}$$

In Eq. (1), Y denotes the explanatory variable, i.e., the results of patent conversion activities of universities; x_i is the grouping variable, i represents each university individually, c represents the breakpoint value of each grouping variable, and D_i is a dummy variable determined according to the corresponding grouping variable. $x_i \uparrow c$ represents the grouping variable on the left side of the breakpoint, and $x_i \downarrow c$ represents the grouping variable on the right side of the breakpoint. According to the research hypothesis, the grouping variable x_i package is an indicator reflecting the attention or focus of universities on the transformation of scientific and technological achievements. One of the keys to clear breakpoint regression design is to determine the breakpoint, if the attention of universities to patent transformation is greater than c, the dummy variable takes the value of 1, and vice versa takes the value of 0.

If LATE takes a value greater than 0, it means that the BD-type policy has a positive effect on patent conversion in universities, and vice versa, it shows a negative effect.

The mainstream analysis methods in the breakpoint regression design include nonparametric and parametric estimation. Among them, the nonparametric method does not need to set the functional form, while the parametric estimation method uses specific estimating equations.

The model settings for parametric estimation are as follows.

$$\begin{aligned} &Y_i = \alpha + \beta_1(x_i - c) + \delta D_i + \beta_2(x_i - \mathrm{c})D_i + \gamma X_i + \varepsilon_i \\ &(\mathrm{c} - \mathrm{h} < x_i < \mathrm{c} + \mathrm{h}) \end{aligned} \tag{2}$$

(2) in the parameter estimation model with the primary term in Eq. (1), where Xi denotes a set of covariates affecting patent conversion activities in universities and h is the symmetric bandwidth. In the parameter estimation equation, δ is the LATE estimate at x = c. The nonparametric estimation method then generally uses kernel density estimation to calculate the specific value of the LATE estimates.

The explanatory variables are the number and value of contracts for the sale of patents in universities. There are two specific indicators, the number of patent sale contracts in 2020 (items) contract and the total amount of patent sale in 2020 (yuan) con-value, and the data are obtained from the Ministry of Education Data Collection. These two variables characterize the patent conversion results of each university in the observation year in terms of the number of patent conversions and the amount of revenue, respectively.

The subgroup variables are word frequency indicators reflecting each university's attention to the transformation of scientific and technological achievements. The word frequency indicator reflects the attention of universities to the transformation of scientific and technological achievements, frequency, specifically means the absolute frequency of the word "transformation" in the policy text.

3.3 Breakpoint Regression Results

3.3.1 Graphical Analysis

The presence of breakpoints is visually examined through linear and multinomial fit graphs, and the baseline regression is performed using the parameter estimating equations.

The relationship between the explanatory and grouping variables can be visually judged by the 4 times polynomial fit graphs. As shown in Fig. 1, when keyword frequency is used as the grouping variable, the conversion results jump upward to the left of the "0" point. According to the results of the graphical analysis, the degree of attention of universities to patent conversion has a positive promotion effect on patent conversion (Fig. 2).

3.3.2 Parameter Estimation Results

Based on the graphical analysis, the baseline breakpoint regression was performed using the parameter estimation method and robustness testing was performed using second,

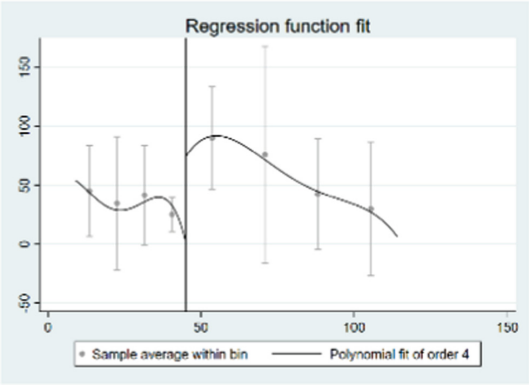

Fig. 1. The effect of RDD.The explanatory variable of the graph is the number of patent sale contracts.

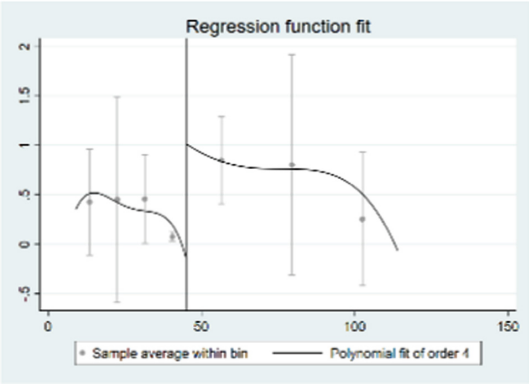

Fig. 2. The effect of RDD. The explanatory variable of the graph is the amount of patent sale contracts.

third, and fourth-order polynomial methods. The polynomial was set to address the possible multicollinearity problem in the model. After local linear regression and polynomial regression of model (2), covariates are added for estimation.

According to Table 3, the attention of universities to patent conversion showed a facilitating effect on conversion outcomes at the breakpoint, but the correlation between the incentive of revenue distribution to inventors and conversion outcomes was not fully consistent with the graphical results. According to the results of full-sample parameter estimation, equity gain distribution has a negative effect on patent sale contracts at the breakpoint, but the effect is positive in the 4th-order polynomial estimation of the same number of local effects; assignment and license fee sharing is basically positively correlated with transformation outcomes (all polynomial estimation results are positive).

Table 3. Parameter estimation results.

Explained variables	*contract*	*con-value*
Subgroup variables	*frequency*	*frequency*
No covariates	(1)	(2)
LATE estimates	58.1545** (2.05)	0.7075** (2.15)
Local 2nd order polynomial	31.1810 (0.82)	0.5672 (0.97)
Local 3rd order polynomial	8.0121 (0.16)	0.7150 (1.00)
Local 4th order polynomial	−17.9459 (−0.29)	0.6987 (0.82)
Adding covariates	(3)	(4)
LATE estimate	57.1385** (2.22)	0.5641* (1.78)
Local 2nd order polynomial	39.5338 (1.31)	0.3462 (0.58)
Local 3rd order polynomial	32.7042 (0.83)	0.6999 (1.07)
Local 4th order polynomial	−12.2732 (−0.25)	0.7504 (0.99)

Table 4. Robustness tests for explanatory variables with replacement variables.

	Triangular Nucleus			Rectangular nucleus		
	Optimal bandwidth	0.5x bandwidth	2x bandwidth	Optimal bandwidth	0.5x bandwidth	2x bandwidth
LATE	(1)	(2)	(3)	(4)	(5)	(6)
frequency	9.7744* (1.95)	9.1544 (1.17)	13.4022*** (3.16)	15.6077*** (2.64)	9.0414 (1.20)	15.6077*** (2.64)

3.4 Robustness Test

In this section, the explanatory variables are replaced with the amount of technology transfer contracts tech-contract in 2020 of the sample universities (in thousands of RMB), and the results of Sect. 4.2 are tested for robustness. For universities, technology transfer is not limited to patents, but involves a larger scope. The results of the robust type test are generally consistent with Sect. 4.2. Universities' attention to patent conversion positively contributes to the increase in contract amount (Table 4).

4 Conclusions

The word "transformation" is the key initiative of patent transformation in universities, and the frequency of the word "transformation" can reflect the degree of attention of universities to patent transformation. As can be seen from Table 1, the word frequency of "conversion" accounts for 5.75% of the overall frequency of the lexicon, and the word frequency is in the second place, which confirms the validity of the indicator selection.

Based on the results of textual analysis, the relationship between BD-type policies and university patent conversion is examined by applying the exact breakpoint regression method, and it is found that the higher the attention of universities to patent conversion, the more fruitful the conversion results will be, on the one hand, the number of patent sale contracts will become larger, and on the other hand, the value of the contracts will rise accordingly.

References

1. B. L. WHORF, J. B. CARROLL, S. C. LEVINSON, et al. The MIT Press, (2012).
2. H. A. SIMON. New York: Macmillan, (1947).
3. H. T. DAVENPORT, C. B. JOHN. Harvard Business School, (2002).
4. J. B. WADE, J. F. PORAC, T. G. POLLOCK. J. ORGAN. BEHAV. 18,7(1997).