International Journal of Innovative Computing, Information and Control
Volume **20**, Number **2**, April **2024**

ICIC International ©2024 ISSN 1349-4198
pp. **509–524**

THE IMPACT OF CARBON EMISSION TRADING PILOT POLICY ON DIGITAL TRANSFORMATION OF ENTERPRISES BASED ON TEXT ANALYSIS METHOD

Wenyan Song* and Wenjun Xiao

School of Economics
Dongbei University of Finance and Economics
No. 217, Jianshan Street, Dalian 116025, P. R. China
xiaowenjun7@163.com
*Corresponding author: songwydufe@dufe.edu.cn

Received June 2023; revised September 2023

Abstract. *As an emerging driver force, the role of digitalization in the carbon reduction may be underestimated. This paper aims at investigating the impact of carbon emissions trading pilot policy on enterprise digital transformation in China. To construct a comprehensive index representing the degree of digital transformation in enterprises, text analysis and factor analysis methods are considered together in this study. Initially, a substantial amount of text is gathered from relevant official documents and research reports to establish a specialized corpus. Subsequently, the Word2Vec Skip-Gram model is applied to training the corpus. Additionally, text similarity analysis is conducted to expand the initial lexicon. Following this, a Chinese word splitting method is employed to determine the frequency of keywords from enterprise annual reports. Then, utilizing the factor analysis, a comprehensive indicator is constructed. Finally, based on 144 listed enterprises from 2008 to 2020, the digital transformation effect of the carbon pilot policy is analyzed using the staggered Difference-in-Difference (DID) method. The results show that the implementation of carbon pilot policy significantly improves the digital transformation of enterprises, and this conclusion still holds after a series of robustness tests.*
Keywords: Text analysis, Word2Vec, Factor analysis, Digital transformation, Carbon trading

1. **Introduction.** With the rapid development of a new generation of digital technologies, the integration of digital technology and real economy is deepening. As an emerging driving factor, the impact of digitalization on the carbon reduction may be underestimated. At the regional level, the digital economy has made significant contributions to low-carbon development since the implementation of the pilot policy on carbon emissions trading [1]. At the enterprise level, the environmental regulations have compensation effects, which can stimulate enterprises to make digital transformation [2].

As we know, the Porter effect is one of the important theoretical bases for the impact of carbon market policies on digital transformation. The Porter effect suggests that stringent but well-designed environmental regulations can incentivize innovation in businesses, partially or completely offsetting the costs of compliance and thereby enhancing the core competitiveness of enterprises. The implementation of carbon markets can raise environmental awareness among enterprises, making them realize the inefficiency of current resource utilization, reducing uncertainty in investments valuable to the environment, and changing the traditional competitive landscape, imposing continuous pressure on enterprises for innovation and progress [3]. If environmental regulations stimulate innovation

DOI: 10.24507/ijicic.20.02.509

in enterprises, but there is uncertainty about whether this innovation will lead to an improvement in the core competitiveness of the enterprise, it is referred to as "weak Porter effect". If the implementation of market-oriented environmental regulations not only offsets compliance costs but also enhances the competitiveness of enterprises, achieving a win-win situation for environmental protection and economic benefits, it is termed as "strong Porter effect" [4]. Carbon emission trading markets are a typical form of market-oriented environmental regulation, and some studies have demonstrated the existence of the Porter effects in China's carbon market. At the provincial level, the implementation of carbon markets significantly increases the rate of technical progress in various pilot provinces, supporting the existence of the "weak Porter effect" [5]. At the enterprise level, the "weak Porter effect" still holds. The implementation of carbon markets stimulates the continuous improvement of the level of innovation in enterprises [6], and this stimulation is influenced by the liquidity of the carbon market. In carbon markets with higher carbon quota trading price, larger daily trading volume, and higher annual trading activity, businesses are more compelled to engage in independent innovation [7].

Furthermore, due to the need for carbon footprint management, enterprises are motivated to undergo digital transformation. Carbon market policies require businesses to track their carbon footprint and accurately quantify carbon emissions, which heavily relies on the application of digital technologies. To achieve carbon neutrality, businesses need sufficient understanding of their own carbon footprint and precise quantification of carbon emissions throughout the product life cycle. The MRV (Monitoring, Reporting and Verification) system refers to the process of quantifying and ensuring the quality of carbon emission data and it is crucial for supporting the sustainable development of carbon markets [8]. Introducing digital technologies into the MRV system can enhance data accuracy and reduce the burden of supervision. Based on the Porter effect and carbon footprint management, this paper puts forward a core hypothesis that the implementation of the carbon emissions trading pilot policy can significantly enhance the digital transformation of pilot enterprises.

The measurement of the degree of digital transformation in enterprises poses an academic challenge. Many existing studies have employed word frequency counts in annual reports as a proxy variable for digital transformation, yet the selection of keywords remains subjective. To mitigate this subjectivity, some scholars have begun to expand the lexicon using some text similarity analysis methods. Certain research endeavors utilized the BERT (Bidirectional Encoder Representations from Transformers) model [9] or the Word2Vec's CBOW (Continuous Bag-Of-Words) model [10] and the Skip-Gram model [11] for text similarity analysis. Both the BERT and the Word2Vec methods are unsupervised, allowing model training by simply feeding the corpus, which avoids the need for extensive data annotation. The BERT model, without fine-tuning, exhibits poorer performance in text similarity analysis [12], while the Skip-Gram model of the Word2Vec, despite having a longer training time, boasts higher semantic accuracy than the CBOW model of the Word2Vec [13]. In light of these considerations, this paper opts for the Skip-Gram model of the Word2Vec to train the corpus.

Considering that enterprise digital transformation is a multi-dimensional comprehensive indicator, using some specific variables as proxy variables has limitations. In the past, scholars often used annual report keyword frequency as a proxy variable for enterprise digital transformation. Nowadays, some scholars are focusing on the construction of comprehensive indicators of enterprise digital transformation. Analytic hierarchy process [14], principal component analysis [15] and fuzzy rough set method [16] have been used to construct comprehensive indicators for digital transformation of enterprises. Hence, we choose the factor analysis method to construct comprehensive indicators.

In this paper, the Skip-Gram model of the Word2Vec is considered to expand the keyword and the factor analysis method is used to construct a comprehensive index of enterprise digital transformation, which not only avoids the subjectivity of keyword selection, but also avoids the problem that the estimates obtained by adding one to the word frequency data and then estimating log-linear regressions have no meaningful explanation [17]. In addition, most previous studies have considered digital transformation as an independent variable to examine the benefits of digital transformation. However, we regard digital transformation as a dependent variable and explore paths to improve the degree of digital transformation, providing reference for governments and enterprises.

The remainder of the paper is organized as follows. In Section 2, the model and the data source are introduced. In Section 3, the processing of text analysis and indicator construction are described. In Section 4, the empirical results and robustness tests are presented. The conclusions are presented in the last section.

2. **Empirical Model and Data Sources.**

2.1. **Staggered DID model.**

The formal trading time of the Emission Trading Scheme (ETS) varies in different regions of China, so the staggered DID model is selected for modeling and identification. The keys of the DID method are to find the control group with the same development trend as the treatment group, and to exclude other disturbances to get a clean effect of the policy. As shown in Equation (1), a three-way fixed effects model with individual, time and industry fixed effects is set up for the benchmark regression:

$$Digital\_index_{itd} = \beta_0 + \beta_1 ETS_{it} + \sum \theta_{it} X_{it} + \mu_i + \pi_t + \eta_d + \varepsilon_{it} \tag{1}$$

where i denotes firm, t denotes year, and d denotes industry. The dependent variable $Digital\_index_{itd}$ refers to the degree of digital transformation of firm i in industry d at year t. A larger $Digital\_index_{itd}$ indicates a higher degree of digital transformation. β_0 is the constant term, representing the expected value of $Digital\_index_{itd}$ when all independent variables are zero. ETS_{it} is the core independent variable which is recorded as 1 after the pilot firm is affected by the policy, and 0 in all other cases. The coefficient β_1 represents policy effects of the ETS. If β_1 is significantly positive, it indicates that the carbon market does promote the degree of digital transformation of the pilot companies. X_{it} represents a series of control variables, and θ_{it} is the respective coefficient for each variable. μ_i is the firm fixed effect to control for individual unobservable factor that does not change over time. π_t is the year fixed effect to control for national shock in each year. η_d is the industry fixed effect to control for industrial unobservable factor that does not change over time. ε_{it} is the random disturbance term.

2.2. **Data sources.**

2.2.1. *Treatment and control groups.* The samples used in this paper include a total of 144 firms within the treatment and control groups, and the sample period is 2008-2020. The treatment group consists of enterprises in the pilot areas that are included in the key emission list, which is obtained manually through the public documents of the development and reform commission and the ministry of environmental resources in China The filtering process for the treatment group is as follows. 1) Due to the presence of new enterprises entering and old enterprises exiting, only enterprises that have not exited since they were included in the ETS are left in the treatment group to obtain clean policy effects. 2) Considering the availability of enterprise information, only enterprises that have been listed on the A-share before the implementation of the policy are included. 3)

The key emission list of Beijing included a number of listed companies in the financial sector, which were excluded due to the special financial indicators of the financial sector. Finally, 82 enterprises are included in the treatment group, including 23 in Beijing, 27 in Shenzhen, 8 in Hubei, 14 in Shanghai, 4 in Fujian, 1 in Guangzhou, and 5 in Chongqing.

Furthermore, considering that when selecting pilot enterprises, the governments of the pilot regions will make reference to indicators such as energy consumption, annual carbon dioxide emissions and building footprint, and enterprises meeting the criteria will be included in the pilot list, it will be difficult to find matching control group within the pilot regions. This paper looks for a control group in a non-pilot area. The filtering process for the control group is as follows. 1) We manually compiled the list of enterprises in non-pilot regions that was included in the national carbon emission trading pilot (which officially traded in 2021) list. 2) Those enterprises that were listed on the A-share before 2014 were screened out. 3) We excluded companies in the financial sector. Finally, a total of 62 eligible enterprises are identified as the control group.

The distribution of the treatment and control groups is shown in Figure 1. The color distinguishes the treatment group from the control group, and the size of the circle represents the number of enterprises. Figure 2 shows the processing time point of each research object. The horizontal axis represents year, and the vertical axis represents company. As illustrated in Figure 2, the light-colored portion represents the control group, the dark-colored portion represents the treatment group, and the white portion represents missing data. Due to companies not being listed or being in an ST (Special Treatment) status, it results in data deletion. Figure 2 reflects the following two facts. 1) There is no group that

Figure 1. Heat map of the treatment and control groups

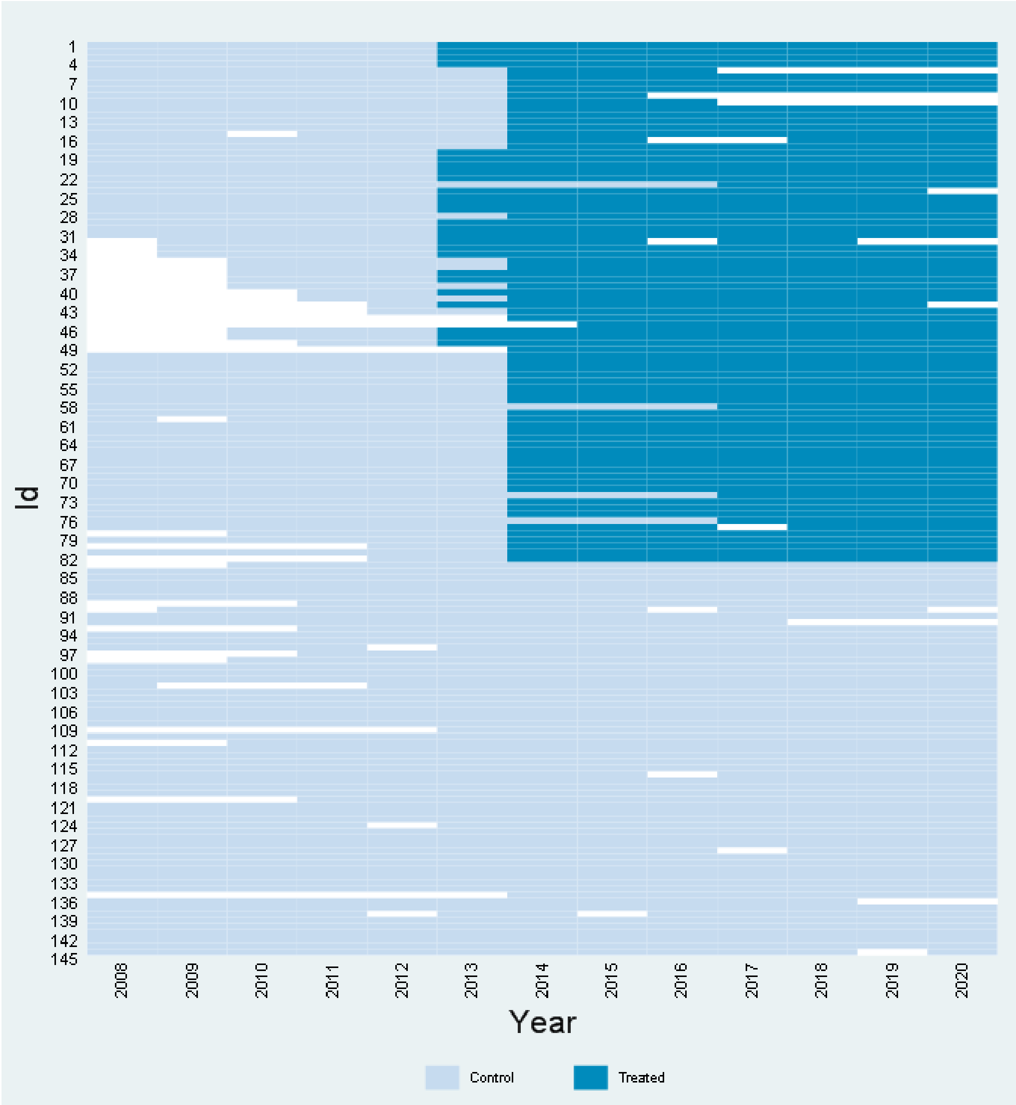


Figure 2. Processing time point diagram

was already treated before the year 2008. 2) Many individual samples have never been treated. These facts indicate that even if there is a potential bias in the Two-Way Fixed Effect (TWFE) estimator, the bias is not significant. This suggests that the selection of research years and samples in this study is reasonable.

2.2.2. *Variable definition and measurement.* The data sources and definitions, as well as the measurement methods, are shown in Table 1. The control variables are obtained from the CSMAR database. The top and bottom 1% data are removed to avoid the effect of extreme values. The word frequency data are obtained by performing word frequency count on the annual reports.

2.2.3. *Descriptive statistics and correlation tests.* Table 2 shows the descriptive statistics analysis results of variables. We also test the variance inflation factor, and the maximum value of VIF is 1.81, which is well below the standard of 10.00, indicating that there is no significant covariance between the independent variables.

Figure 3 displays the results of the correlation analysis. It can be observed that the policy variable *ETS* is positively correlated with each indicator of digital transformation

Table 1. Variable definition and measurement

Type	Definition	Variable name	Measure	Source
Dependent variables	Comprehensive indicators of enterprise digital transformation	*Digital_index*	Multiple variables were downscaled using the factor analysis to obtain '*Digital_index*'	
	Digital technology	*Digital_tech*	Word frequency count	Annual report
	Digital applications	*Digital_apply*	Word frequency count	Annual report
	Digital facilities	*Digital_fac*	Word frequency count	Annual report
	Number of digital patents	*Digitalpat*1	Number of corporate digital economy invention patent applications	CNRDS
		*Digitalpat*2	The number of enterprise digital economy invention patents granted	CNRDS
		*Digitalpat*3	The number of enterprise digital economy utility model patent applications	CNRDS
		*Digitalpat*4	The number of enterprise digital economy utility model patents granted	CNRDS
	Digital patent quality	*Citednum*	Number of corporate digital economy patents cited (excluding self-citations)	CNRDS
Independent variable	Policy variable	*ETS*	$ETS = 1$ after the carbon market at the pilot company's location is operational Other cases $ETS = 0$	Official documents
Control variables	Enterprise size	*Size*	Logarithm of total assets	CSMAR
	Corporate credit indicators	*Lev*	Ratio of total liabilities to total assets	CSMAR
	Corporate operating indicators	*Cashflow*	Net cash flow from operating activities divided by total assets	CSMAR
		Roe	Return on net assets	CSMAR
		Ato	Total assets turnover ratio	CSMAR
	Nature of business ownership	*Soe*	State-owned enterprises take the value of 1, others are 0	CSMAR
	Corporate social wealth creativity	*Tobinq*	Tobin's Q value	CSMAR
	Degree of corporate centralization	*Top*1	Percentage of shareholding of the largest shareholder	CSMAR
	Business maturity indicators	*Firmage*	The number of years the company has been established is taken as a logarithm	CSMAR

TABLE 2. Descriptive statistics

Variables	N	Mean	SD	Min	Max	VIF
Digital_index	1,739	0.0003	0.585	−0.213	6.253	
Digital_tech	1,739	4.015	17.05	0	269	
Digital_apply	1,739	2.361	6.565	0	80	
Digital_fac	1,739	4.727	10.31	0	124	
*Digitalpat*1	1,739	50.41	391.2	0	6,727	
*Digitalpat*2	1,739	22.88	192.1	0	3,067	
*Digitalpat*3	1,739	12.41	41.50	0	460	
*Digitalpat*4	1,739	11.27	38.79	0	586	
Citednum	1,739	223.6	1,409	0	17,592	
ETS	1,739	0.325	0.469	0	1	1.19
Size	1,739	4.535	1.356	1.308	8.582	1.81
Lev	1,739	0.491	0.192	0.0436	0.925	1.55
Cashflow	1,739	0.0671	0.0620	−0.180	0.283	1.30
Roe	1,739	0.0683	0.116	−0.912	0.409	1.37
Ato	1,739	0.727	0.438	0.0589	2.918	1.11
Soe	1,739	0.618	0.486	0	1	1.26
*Top*1	1,739	0.390	0.161	0.0832	0.758	1.23
Tobinq	1,739	1.649	0.955	0	8.989	1.33
Firmage	1,739	2.817	0.373	0.693	3.555	1.37

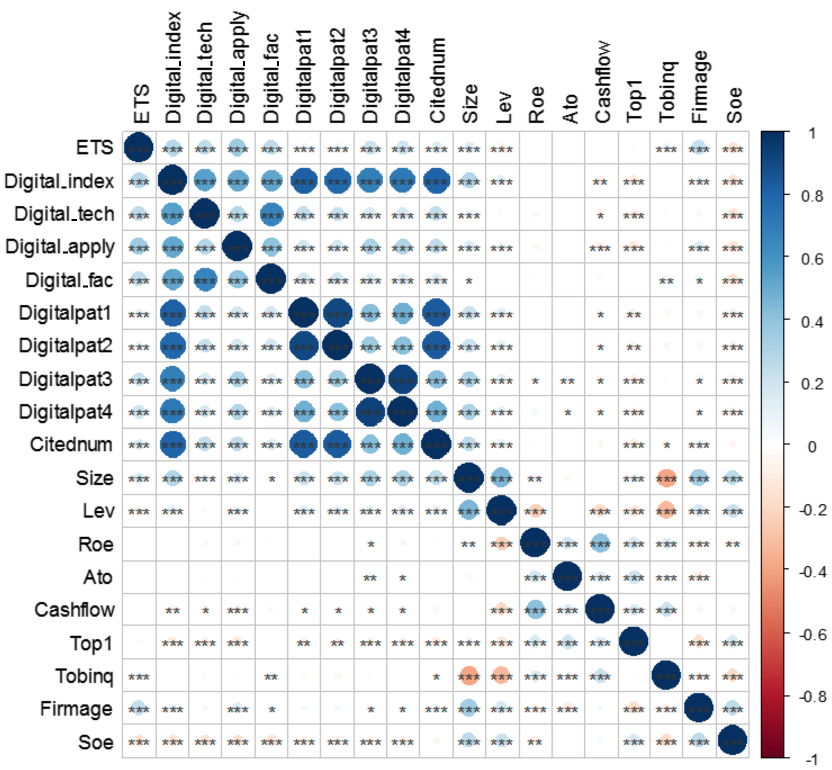


FIGURE 3. (color online) Correlation matrix

at the 1% significance level. Overall, after the implementation of the carbon market, the comprehensive digital transformation index of the treatment group enterprises has shown improvement compared to the control group. Furthermore, specific indicators such as digital technology, digital applications, digital facilities, quantity of digital-related patents and number of citations of digital-related patents have all experienced enhancement in the treatment group. This provides preliminary evidence supporting the assertion that the carbon market indeed has a certain facilitating effect on the digital transformation of enterprises. At the same time, the majority of control variables also exhibit a positive correlation with *Digital_index*, indicating that the selection of control variables in this study is reasonable.

3. Indicator Construction.

The primary innovative contributions of this section are as follows. 1) Construct a specialized corpus for digital transformation; 2) Divide the lexicon into three dimensions to analyze different dimensions of digital transformation; 3) Employ the factor analysis method to construct a comprehensive indicator.

3.1. Text analysis methodology.

The process of text analysis is shown in Figure 4. Firstly, build a professional digital transformation corpus. Secondly, put the corpus into the Word2Vec model for the word vector construction. Then find the digital transformation words with 80% or more similarities to the keywords by calculating the cosine similarity and incorporate them into the initial lexicon. Finally, split words and count keywords frequency on annual reports based on the final lexicon.

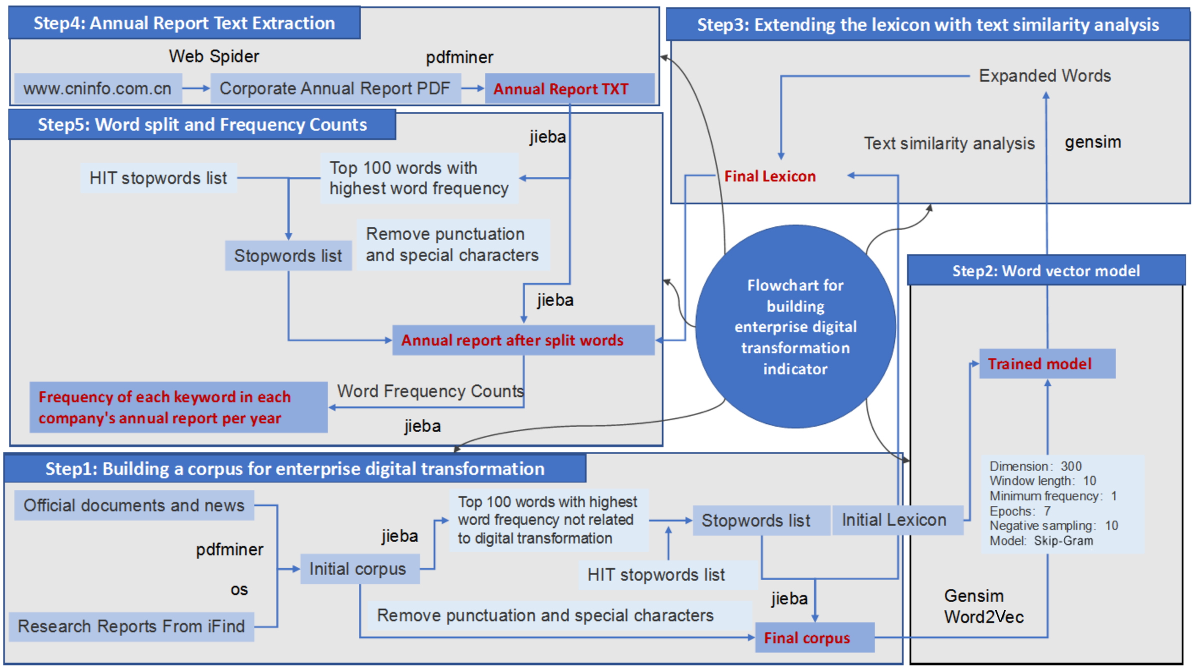


Figure 4. Flowchart for text analysis

3.1.1. *Building a corpus for enterprise digital transformation.*

1) Specialized corpus construction

Previous studies used enterprise annual reports as a corpus, but enterprise annual reports are not a corpus specifically for digitalization. To address the above problem, this paper constructs a specialized corpus for digital transformation to prevent irrelevant corpus from reducing word vector effects. The process of corpus construction is shown as Step 1 in Figure 4. Search digital transformation keywords from government official website and iFinD database to download all news and reports. Use the Python's pdfminer library to convert PDF files to TXT format, and merge multiple TXT files into one large TXT file as a professional corpus for digital transformation.

2) Corpus cleaning and word segmentation

As depicted in Step 1, a list of stopwords is initially established for cleaning the corpus. The stopwords list comprises two parts. The first part includes words with the top 100 word frequencies in the corpus that are unrelated to digitization. The second part is the stopwords list developed by the Harbin Institute of Technology. Importing the stopwords list helps eliminate interference from noisy words, punctuation and special symbols in the corpus. Following this, an initial lexicon is established, and the jieba library is employed for word segmentation. The purpose of creating an initial vocabulary is to prevent important digital-related terms from being separated during word segmentation. The initial lexicon is constructed by referring to previous studies and industry research reports from brokerage firms, consulting and other institutions. To analyze different dimensions of digital transformation, this paper categorizes the lexicon into three dimensions, i.e., digital technology, digital applications and digital facilities. A total of 151 initial words are obtained in the initial lexicon.

3.1.2. *Training the Word2Vec model.* As illustrated in Step 2 of Figure 4, this paper utilizes the Gensim library to train the cleaned and segmented corpus, constructing word vectors for each word. The Skip-Gram model is used with the following parameters. The word vector dimension is set to 300. The window length is set to 10. The minimum word frequency is set to 1. The number of epochs is set to 7. And the number of negative samplings is set to 10.

3.1.3. *Extending the lexicon with text similarity analysis.* As shown in Step 3 of Figure 4, utilizing the cosine similarity algorithm of the Gensim library, the initial words are analyzed for similarity. Words related to digital transformation are selected based on a similarity of 80% or more to the initial words. Additionally, the vocabulary for enterprise digital transformation technologies, applications and facilities is expanded separately. Eventually, 68 words are added to the initial lexicon, resulting in a final lexicon that incorporates 219 key words. The final lexicon is presented in Figure 5.

3.1.4. *Extracting annual report text.* As shown in Step 4 of Figure 4, utilizing the Python spider technology, this paper crawls all the PDF files of enterprise annual reports from 2008 to 2020 in the treatment group and the control group from the Juchao.com. Subsequently, these files are converted into TXT format using the pdfminer library and merged into a single large TXT file for later word segmentation preparation.

3.1.5. *Splitting words and conducting frequency counts.* As shown in Step 5 of Figure 4, this paper utilizes the jieba library to count the words in the top 100 word frequencies in annual reports and integrates them with the HIT stopwords list to form the final stopwords list. This list is then imported into the final lexicon, and the word frequency counts are conducted to determine the frequency of each keyword for each company every year. Figure 6 displays word clouds for digital technology, applications and facilities, respectively. In these word clouds, the larger the font size and the darker the color, the higher the frequency of the word.

3.2. **Factor analysis methodology.** Most previous studies only used word frequency as a proxy variable, but the word frequency index could not fully reflect the degree of digital transformation. In order to synthesize several dimensions of annual report word frequency and digital patent quantity and quality, a factor analysis method is used in this section to construct comprehensive indicators. First, the KMO test and the Bartlett test are used to determine whether the data were suitable for the factor analysis. As shown in Table 3, the KMO value is 0.728 which is higher than the standard of 0.5, and the p-value

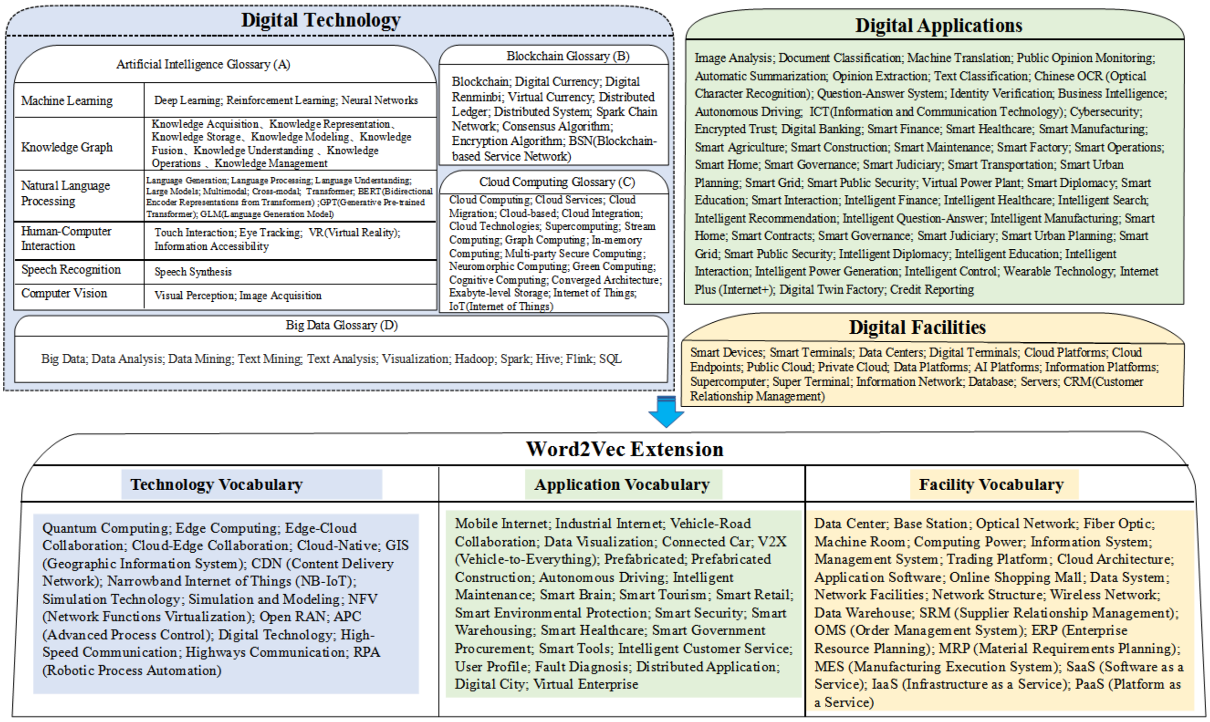
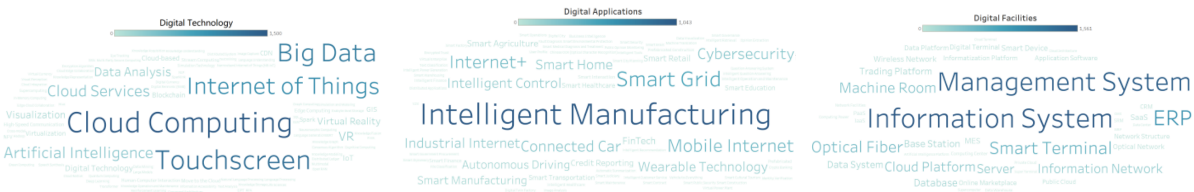


FIGURE 5. Final lexicon



(a) Digital technology word cloud (b) Digital application word cloud (c) Digital facility word cloud

FIGURE 6. Digital words clouds

TABLE 3. KMO and bartlett tests

KMO	Bartlett		
	Chi-square	Degrees of freedom	p-value
0.728	10840.21	28	0.00

is 0, indicating that it is suitable for the factor analysis. Then as shown in Table 4, three factors with eigenvalues greater than 1 are selected as common factors. The cumulative variance contribution of these three factors is 83.12%, which well reflects the majority of the information contained in the original eight variables.

The rotated component matrix is shown in Table 5. *Digitalpat*1, *Digitalpat*2, and *Citednum* can be categorized as Factor1. These three variables represent the results of digital R&D of enterprises, which have strong inventive, innovative and advanced implication, so the factor is named as the Digital Innovation Factor. *Digitalpat*3, *Digitalpat*4 and *Digital_apply* can be categorized as Factor2. These three variables represent the degree of digital application of enterprise, so the factor is the Digital Utility Factor. *Digital_tech* and *Digital_fac* can be categorized as Factor3. These two variables represent the importance companies place on digital technology and facilities, which require significant investment

Table 4. Variance interpretation rate

Factor	Initial eigenvalue			Explanation of variance before rotation			Explanation of variance after rotation		
	Eigenvalue	Proportion	Cumulative	Eigenvalue	Proportion	Cumulative	Eigenvalue	Proportion	Cumulative
Factor1	3.79352	47.42%	47.42%	3.79352	47.42%	47.42%	2.71612	33.95%	33.95%
Factor2	1.61568	20.20%	67.61%	1.61568	20.20%	67.61%	2.05616	25.70%	59.65%
Factor3	1.24063	15.51%	83.12%	1.24063	15.51%	83.12%	1.87754	23.47%	83.12%
Factor4	0.69312	8.66%	91.79%						
Factor5	0.31896	3.99%	95.77%						
Factor6	0.17231	2.15%	97.93%						
Factor7	0.09439	1.18%	99.11%						
Factor8	0.07139	0.89%	100%						

Table 5. Rotated component matrix

Variable	Factor1	Factor2	Factor3
Digital_tech	0.01665	−0.10787	**0.47265**
Digital_apply	−0.13317	**0.21109**	0.28436
Digital_fac	−0.04380	−0.06833	**0.50988**
*Digitalpat*1	**0.38199**	−0.08073	−0.03963
*Digitalpat*2	**0.40743**	−0.13257	−0.03401
*Digitalpat*3	−0.12095	**0.53703**	−0.06747
*Digitalpat*4	−0.07894	**0.50580**	−0.07725
Citednum	**0.35413**	−0.05157	−0.03362

in technology introduction and facilities construction, so this factor is named as the Digital Input Factor.

According to the following Equation (2), the percentage of variance contribution of each factor is used as the weight to get the comprehensive index of digital transformation of each company every year. If *Digital_index* is greater than zero, it indicates that the enterprise digital transformation is above the mean value. If it is less than zero, it indicates that the enterprise digital transformation is below the mean.

$$Digital\_index = (33.95\% * Factor1 + 25.70\% * Factor2 + 23.47\% * Factor3)/83.12\% \quad (2)$$

4. **Empirical Results and Robustness Checks.**

4.1. **Empirical results.** Table 6 shows the benchmark regression results of the implementation of carbon policies on the enterprise digital transformation. Both Model_1 and Model_2 control for individual, time and industry fixed effects. Model_1 shows that the coefficient of *ETS* is significantly positive at the 1% level. Model_2 adds control variables on the basis of Model_1, and the policy effect is partially absorbed by the control variables, but still significant at the 1% level. The coefficient of *ETS* in Model_2 is significantly positive, indicating that the difference in *Digital* between firms in the treatment and control groups before and after the implementation of the pilot policy is significantly higher by 0.1112. So the core hypothesis of this paper is validated.

4.2. **Robustness tests.**

4.2.1. *Parallel trend test and dynamic effect test.* The event study method is considered to construct the dynamic effect model as Equation (3). β_s^{pre} is the difference in effect of the s-period before carbon market trading relative to the base period. β_s^{post} is the difference in effect of the s-period after carbon market trading relative to the base period. D_i is the grouping variable, with a value of 1 for enterprises in the pilot area and 0 for others. T_D is the year when the carbon market started trading. $\mathbf{1}(t - T_D = s)$ is an indicative function, which takes 1 when the conditions in parentheses are met, and 0 in other cases.

TABLE 6. Benchmark regression

	Model_1 *Digital_index*	Model_2 *Digital_index*
ETS	0.1490*** (0.0472)	0.1112*** (0.0318)
Control variables	No	Yes
Id FE	Yes	Yes
Year FE	Yes	Yes
Ind. FE	Yes	Yes
N	1742	1739
adj. R^2	0.8440	0.8634

Note: 1) Clustering robust standard errors at the provincial level are reported in parentheses; 2) *, **, *** indicate significance at the 1%, 5%, and 10% levels, respectively.

Other parameters in Equation (3) have the same meaning as in Equation (1). The policy shock point of this paper is defined as the year when each carbon market starts trading. In fact, each region releases the list of companies to be included in the carbon market approximately one year before the carbon market starts trading. Considering the above issues, it is appropriate to use pre_2 as the base period.

$$Digital\_index_{it} = \beta_0 + \sum_{s=-8}^{-3} \beta_s^{pre}[D_i \times \mathbf{1}(t - T_D = s)] + \sum_{s=-1}^{7} \beta_s^{post}[D_i \times \mathbf{1}(t - T_D = s)] + \sum \theta X_{it} + \mu_i + \pi_t + \eta_d + \varepsilon_{it} \tag{3}$$

As shown in Figure 7, the policy effects in each period before the base period are significantly zero, which means that the years before the policy implementation pass the parallel trend test and the policy coefficient of the benchmark regression represents a relatively clean policy effect. All periods after the base period are significantly different from the base period, and the difference is widening, indicating that the impact of digital transformation brought by the carbon pilot policy to the enterprises in the treatment group is continuous, and the enterprises insist on digital transformation to cope with the adverse impact of limiting carbon emissions.

4.2.2. *Placebo test.* In order to exclude the interference of other random factors, the interaction term *ETS* is sampled 500 times to conduct the individual placebo test. Figure 8 shows that the coefficients obtained from 500 regressions are all centralized around 0. No coefficient falls to the right of the actual coefficient, and most of the p-values are larger than the actual p-value, indicating that the impact of the carbon market on the digital transformation of enterprises is basically not disturbed by other random factors.

4.2.3. *Bacon decomposition and heterogeneity robust estimator.* For the staggered DID model, the Two-Way Fixed Effects (TWFE) require that the treatment effects are constant over time and cannot exit dynamic effect [18]. According to Goodman's approach, we decompose the TWFE estimates into four categories of estimates by using the Bacon decomposition method. Here "Earlier T vs. Later C" refers to a regression where companies included in the carbon market earlier as the treatment group and those included later as the control group. "Later T vs. Earlier C" refers to a regression where companies included in the carbon market later as the treatment group and those included earlier as the control group. "T vs. Never treated" refers to a regression where newly added

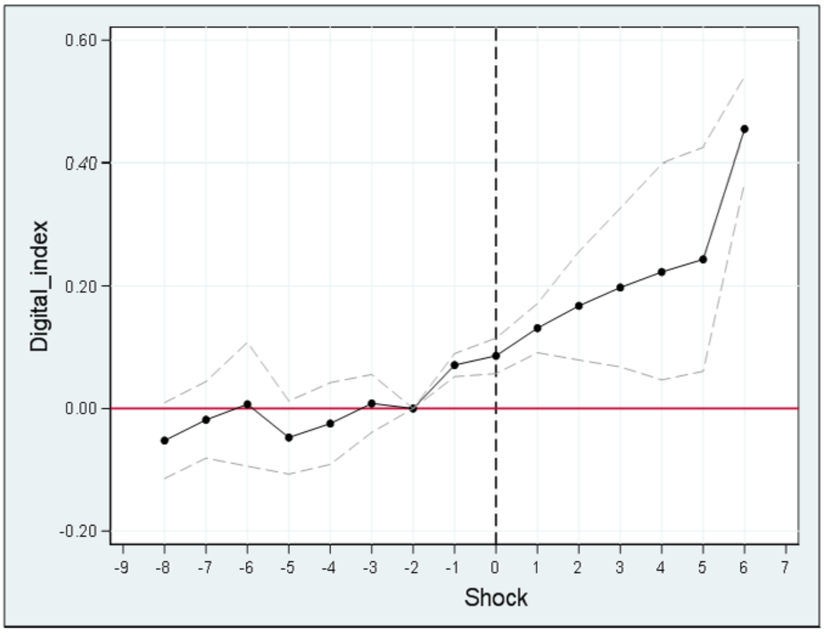


Figure 7. Parallel trend test

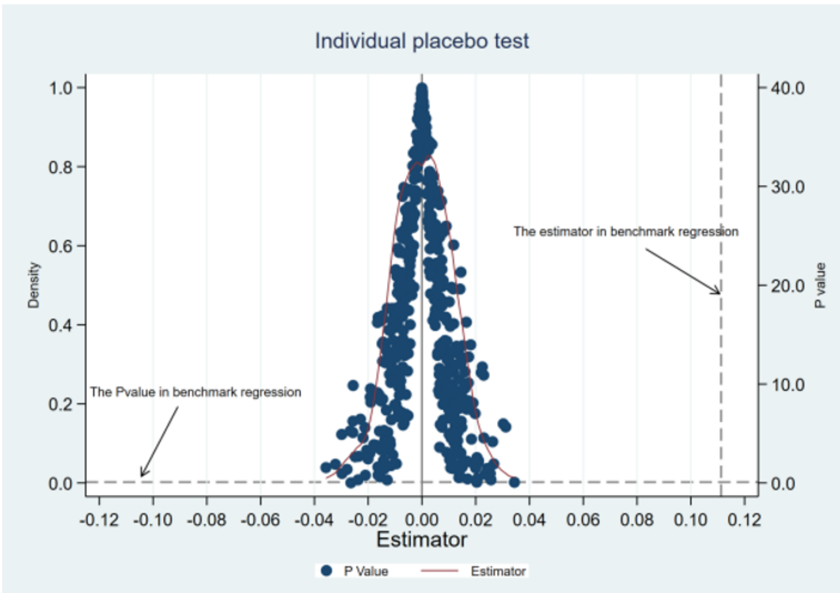


Figure 8. Placebo test

companies to the carbon market as the treatment group and those have never joined the carbon market as the control group. “T vs. Already treated” refers to a regression where newly added companies to the carbon market as the treatment group and those have been included all the time as the control group. “Earlier T vs. Later C” and “T vs. Never treated” can get clean policy effects; however, “Later T vs. Earlier C” and “T vs. Already treated” are influenced by the heterogeneity of treatment effects, resulting in a misestimation of policy effects. As shown in Table 7, the dashed lines represent a 95% confidence interval, while the solid line represents the average treatment effect of policy for each year. “Later T vs. Earlier C” and “T vs. Already treated” do not have much weight, so there is a slight heterogeneity in the treatment effects of benchmark regression.

TABLE 7. Bacon decomposition

DD comparison	Weight	Avg DD Est
Earlier T vs. Later C	0.027	0.058
Later T vs. Earlier C	0.070	−0.249
T vs. Never treated	0.849	0.187
T vs. Already treated	0.022	−0.015

Next, the CSDID (a DID model proposed by Callaway and Sant'Anna) method is adopted to obtain heterogeneity robust estimators and to highlight treatment effect heterogeneity across different dimensions as well as to summarize the overall effect of participating in the treatment [19]. As shown in Table 8, both Model_3 and Model_4 control for individual, time and industry fixed effects. Model_4 adds control variables on the basis of Model_3. The estimators aggregated by all available group-time average treatment effects (Simple weighted average), by group (Group-specific effects), by length of exposure to carbon market policy (Event study) and by calendar time, are consistent with the idea that carbon market policy had a positive effect on enterprise digital transformation.

TABLE 8. CSDID results

	Model_3 *Digital*	Model_4 *Digital*
Simple weighted average	0.1428*** (0.0386)	0.1494*** (0.0415)
Group-specific effects	0.1311*** (0.0318)	0.1382*** (0.0354)
Event study	0.1832*** (0.0498)	0.1895*** (0.0519)
Calendar time effects	0.1331*** (0.0386)	0.1396*** (0.0417)
Control variables	No	Yes
Id FE	Yes	Yes
Year FE	Yes	Yes
Ind. FE	Yes	Yes

Note: 1) Clustering robust standard errors at the provincial level are reported in parentheses; 2) *, **, *** indicate significance at the 1%, 5%, and 10% levels, respectively.

5. **Conclusion and Discussion.** In this paper, a text analysis method and a factor analysis method are used to construct an indicator of enterprise digital transformation. Furthermore, the staggered DID method is implemented to investigate the impact of carbon trading on enterprise digital transformation. The results show that in China the carbon trading policy has significantly improved the enterprise digital transformation in pilot regions. However, the corpus built in this paper is not large enough. In the future, we can use some big data methods to crawl digital transformation related news and documents, the bigger the corpus the more accurate the text similarity analysis. Additionally, further exploration will be conducted on the mechanisms and channels of carbon markets to facilitate enterprises' digital transformation.

Acknowledgment. This work is partially supported by the Humanities and Social Science project for Ministry of Education of China (21YJA630079). The authors also gratefully acknowledge the helpful comments and suggestions of the reviewers, which have improved the presentation.

REFERENCES

[1] J. Zhang, Y. Lyu, Y. Li and Y. Geng, Digital economy: An innovation driving factor for low-carbon development, *Environmental Impact Assessment Review*, vol.96, DOI: 10.1016/j.eiar.2022.106821, 2022.

[2] W. Chen, Can low-carbon development force enterprises to make digital transformation?, *Business Strategy and the Environment*, DOI: 10.1007/s11365-023-00852-7, 2022.

[3] M. E. Poter and C. V. D. Linde, Toward a new conception of the environment-competitiveness relationship, *Journal of Economic Perspectives*, vol.9, no.4, pp.97-118, 1995.

[4] W. Yu and Q. Chen, 20 Years of Porter Hypothesis – A literature review on the relationship among environmental regulation, innovation and competitiveness, *Science Research Management*, vol.36, no.5, pp.65-71, 2015.

[5] D. Fan and W. G. Wang, Analysis of the performance of carbon emissions trading right in China – The evaluation based on the difference-in-difference model, *China Environmental Science*, vol.37, no.6, pp.2383-2392, 2017.

[6] H. Chen, Z. H. Xie and X. Y. Wang, The impact of carbon trading policy implementation on corporate research and development – An empirical study based on multi-period differences-in-differences model, *Social Scientist*, no.9, pp.75-82, 2022.

[7] N. Huang, *Carbon Emission Trading Scheme, Innovation and Firm Performance*, Ph.D. Thesis, Jinan University, 2018.

[8] X. Z. Liu, Q. K. Zhu, X. Sun, J. Y. Mao, J. Liu and Y. T. Shang, Study on EU ETS MRV system and the enlightenments for China, *Forum on Science and Technology in China*, no.8, pp.164-173, 2018.

[9] Z. G. Li, Y. R. Wu and Y. K. Li, Technical founders, digital transformation and corporate technological innovation: Empirical evidence from listed companies in China's STAR market, *International Entrepreneurship and Management Journal*, DOI: 10.1007/s11365-023-00852-7, 2023.

[10] W. Tu and J. He, Can digital transformation facilitate firms' M&A: Empirical discovery based on machine learning, *Emerging Markets Finance and Trade*, vol.59, no.1, pp.113-128, 2023.

[11] A. Ohno, Acquisition of synonyms for colloquial expressions in scientific reports using Word2Vec, *ICIC Express Letters, Part B: Applications*, vol.14, no.6, pp.657-662, DOI: 10.24507/icicelb.14.06.657, 2023.

[12] B. Li, H. Zhou, J. He, M. Wang, Y. Yang and L. Li, On the sentence embeddings from pre-trained language models, *Proc. of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pp.9119-9130, 2020.

[13] T. Mikolov, K. Chen, G. Corrado and J. J. C. S. Dean, Efficient estimation of word representations in vector space, *arXiv Preprint*, arXiv: 1301.3781, 2013.

[14] Q. M. Chen, W. Zhang, N. S. Jin, X. C. Wang and P. R. Dai, Digital transformation evaluation for small- and medium-sized manufacturing enterprises using the fuzzy synthetic method DEMATEL-ANP, *Sustainability*, vol.14, no.20, 2022.

[15] R. Li, J. Rao and L. Y. Wan, The digital economy, enterprise digital transformation, and enterprise innovation, *Managerial and Decision Economics*, vol.43, no.7, pp.2875-2886, 2022.

[16] X. Tang, Y. Miao, Y. Sun and L. Dong, Research on digital maturity measurement and influencing factors of advanced equipment manufacturing enterprises, *Science Research Management*, vol.43, no.9, pp.10-19, 2022.

[17] J. B. Cohn, Z. C. Liu and M. I. Wardlaw, Count (and count-like) data in finance, *Journal of Financial Economics*, vol.146, no.2, pp.529-551, 2022.

[18] A. Goodman-Bacon, Difference-in-differences with variation in treatment timing, *Journal of Econometrics*, DOI: 10.1016/j.jeconom.2021.03.014, 2021.

[19] B. Callaway and P. H. C. Sant'Anna, Difference-in-Differences with multiple time periods, *Journal of Econometrics*, DOI: 10.1016/j.jeconom.2020.12.001, 2020.

Author Biography

Wenyan Song received the B.Sc. degree in Mathematics from Beijing Normal University, China, 2002; the Ph.D. degree in Applied Mathematics from Beijing Normal University, China, 2007.

Dr. Song is currently a full-time associate professor at the School of Economics, Dongbei University of Finance and Economics, China. Her research interests include fuzzy system modelling, neural network and management optimization.

Wenjun Xiao received the B.Sc. degree in Industrial Engineering from Beijing Information Science and Technology University, China, 2019. She is currently pursuing the M.Sc. degree at the School of Economics, Dongbei University of Finance and Economics, China. Her research interests include text analysis and carbon neutrality.