Green Energy, Environment and Sustainable Development
C. Wang et al. (Eds.)

doi:10.3233/ATDE230302

A Hybrid Detection Approach for Carbon Emission Intensity Reduction Mechanism Under Environmental Regulations

Jing WANG and Xiang YI[1]
Business School, Dongguan City University, Dongguan 523419, Guangdong, China

Abstract. In this paper, an action mechanism of the carbon emission intensity (CE) reduction relevant to the level of industrial agglomeration (IA), scientific and technological progress (TECH), foreign direct investment (F), and environmental regulation (ER) is studied by applying a hybrid framework driven by the K-means clustering algorithm and the stepwise multiple linear regression (SMLR)-based models. To concisely clarify the relationship between the objected variable and impact factors, three echelons of urban agglomerations are summarized in this study from the different industry characteristics of the whole 21 cities in Guangdong, which exhibits a strong growth of the secondary industry in China. Stepwise multiple linear regression (SMLR) analysis is proposed to reveal that strict environmental regulations for the urban agglomeration with low overall development level and weak industrial foundation are more likely to stimulate the "innovation compensation" effect. For the urban agglomeration in the growth stage of industrial economic development or distributed in the developed region, strengthening the level of IA and highlighting the core guiding role of environmental regulation seems to make more sense to reduce the carbon emission intensity and promote the regional industrial green upgrading.

Keywords. Environmental regulation, carbon emission intensity reduction, SMLR-based model, K-means clustering algorithm

1. Introduction

As the climate crisis has become the focus of the international governance agenda, there are increasingly more discussions about the impacts of industrial factors on green development [1]. Chen et al. established panel data and GMM model and found a positive correlation between foreign direct investment (FDI) and energy efficiency. Their study stated that giving full play to the driving role of FDI in regional technological innovation is the premise of boosting high-quality economic development, helping enterprises to promote technological innovation and adopt new technologies [2]. By building a SAC model using the panel data of China's thirty provinces, Zhong et al. revealed a positive relationship between high-technology industry agglomeration and green total factor productivity. The study noticed that when industrial agglomeration develops to a relatively mature stage, industrial agglomeration can promote industrial ecological efficiency to improve economic benefits [3]. In addition, some scholars used the

[1] Corresponding author, Xiang YI, Business School, Dongguan City University, Dongguan 523419, Guangdong, China; Email: xiang.yi@outlook.com.

generalized linear regression model to consider the impacts of environmental regulation intensity on the development of green finance from the perspective of enterprises. The related study [4] demonstrated that environmental regulation would positively influence green finance through short-term or long-term external financing and positively affect high-quality economic development.

There is no denying that promoting industrial green upgrading vigorously and enhancing the efficiency of environmental regulation is the breakthrough to practice new measures of green development [5]. The famous “Porter hypothesis” supports an appropriate environmental regulation that will stimulate industrial and technological innovation, ensure the positive role of science and technology in formal environmental regulation and technological innovation, and improve the innovation performance of enterprises, to offset the loss of production costs caused by environmental regulation [6]. Ramanathan et al. demonstrated the existence of the “port hypothesis”. They pointed out that environmental regulation is a crucial factor effectively motivating green innovation technology upgrading, the government can achieve the balance between emissions targets and economic targets, and enterprises can improve the ability of sustainable development by adopting positive innovation in response to environmental regulation [7]. However, the academic attitude toward the impact of environmental regulation on the industry is inconsistent. The difference in the strength of the environmental regulation, the low pollution enterprise area is challenging to produce innovation effect [8]. Hart et al. believed that strong environmental regulation would lower enterprises’ economic performance. The increasing capital and technology investment would offset the cost reduction of eliminating residual emissions [9]. Wu et al. argued that if a region implements strict environmental regulation, polluting enterprises have to shift to the region with weak environmental regulation, where they take as “shelter”. Otherwise, some scholars believe no conflict exists between the “pollution shelter” and the “Porter hypothesis” [10]. During the government’s participation in governance, it should pay attention to playing its due role to avoid excessive intervention caused by market failure. The win-win goals can be achieved through the appropriate design of the policy in terms of the simplified impacts of enterprises’ response to environmental regulation.

Note worthily, there are three views on the trend of industrial agglomeration under environmental regulation. First, technological innovation can promote the development of positive environmental externalities of industrial agglomeration and positively affect the relationship between industrial agglomeration and environmental performance [11]. Environmental regulation accelerates the dissemination of green knowledge and technology, stimulates the willingness and enthusiasm of enterprises for green innovation, and forces industrial optimization and upgrading by agglomeration technology spillover effect. Technological innovation and upgrading make up for the adverse impacts of environmental regulation on industry. Second, environmental regulation forces industrial clusters to shift location. The strict environmental regulation will increase production costs and lower profit margins. By studying the spatial interactive spillover effect between environmental regulation and the transfer of polluting industries, Jang et al. verified strict environmental regulation would increase the transfer of polluting industries [12]. Third, environmental regulation does not significantly impact industrial agglomeration. To determine whether the change of environmental regulation intensity is the determinant of industrial cluster transition, Cole et al. constructed an econometric model of influencing factors of pollution-intensive industry transfer based on the panel data of industrial zones in Guangdong, China using the input-output function principle. They stated that the spatial transfer behavior of pollution-intensive industries is not significantly different from that of other industries, which can greatly affect the traditional "pollution haven" hypothesis [13].

Exploring the impacts of environmental regulation and industrial factors on regional emission reduction shows vital theoretical and practical significance to realize the "win-win" results of ecological protection and economic development. This study focuses on the 21 prefecture-level cities in Guangdong, China. To concisely clarify the relationship between the objected variable and impact factors, the K-means clustering algorithm is applied in our research to form three echelons of urban agglomerations according to the time series of the annual carbon emission intensity (CE) data from these cities. Furthermore, the Stepwise Multiple Linear Regression (SMLR) analysis was employed to analyze the correlation between the average carbon emission intensity (CE) corresponding to the output of the secondary industry in different urban agglomerations, and the industrial agglomeration level (IA), the level of technological progress (TECH), the level of opening to the outside world (CF), and the level of environmental regulation (ER). In addition to clarifying the differences in the average CE corresponding to the output of the secondary industry in the three classes of urban agglomerations with the improvement of environmental regulation intensity, this work focuses on the mechanism of adjusting the environmental regulation, which can make its effect on the green economic development of urban agglomerations optimal. The remaining contents of this work are arranged as follows. The second part explains the source of the original data, analysis method, and model setting. The third part analyzes the correlation between variables and the significant impacts and explores the interference factors of the difference in impact degree of environmental regulation in three classes of urban agglomerations according to the model results. Finally, a summary reviews the basic conclusions of this work based on the model results, aiming to provide suggestions for optimizing the regulation according to the regional situation, industrial characteristics, industrial structure, and economic level. Meanwhile, it provides some ideas for further research.

2. Data Description and Model Setting

2.1. Data Description

According to the data from the "General Administration of Customs of China", as one of the provinces with the most prominent export-oriented economy characteristics in China, Guangdong reached a total foreign trade volume of 8.27 trillion yuan in 2021, with a year-on-year growth of 16.70%, ranking first in China for 36 consecutive years. After years of construction, the region has gradually formed a coordinated development and management system for service trade in the Greater Bay Area, promoting the integration of service trade in the Greater Bay Area [14]. However, along with economic development, the ecological pressure in the Greater Bay Area is still prominent. According to the latest data from the Guangdong Provincial Department of Ecology and Environment, in the second quarter of 2022, the provincial average monthly Air Quality Index (AQI) compliance rate was 93.60% the second quarter of the same year, down 0.32% year on year. Ozone and NO_2 ranked high in the proportion of pollutants. The environmental regulation measures adopted in this region are necessary to promote the organic integration of industrial development and ecological protection, to give full play to the leading role of Guangdong Province in upgrading industries, and to promote the

Guangdong-Hong Kong and Macao Bay Area into a world-class economic belt.

This work selects 21 prefecture-level cities in Guangdong Province as research samples, and their statistical data from 1997 to 2018 are collected. The original data are from the "Guangdong Statistical Yearbook", China Carbon Accounting Database, China Economic Network Database, EPS database, and the statistical yearbook of each prefecture-level city. The representation and descriptive statistics of the explained variables are shown in Table 1, and those of the explanatory variables are shown in Table 2. Due to the missing data and the difference in data magnitude among variables, the gray prediction method is employed to supplement the missing data, and the data is standardized.

Table 1. Descriptive statistics of the explained variables.

Variable	Definition	Observed value	Mean value	Standard deviation	Min	Max
C	Carbon dioxide emissions (ton)	462	17.89	13.85	2.34	77.00
G	Secondary industry added value (thousand yuan)	462	92.76	146.52	0.64	999.59
CE	Carbon emission intensity (One thousand yuan per ton)	462	0.50	0.51	0.04	4.27

Table 2. Descriptive statistics of the explanatory variables.

Variable	Definition	Observed value	Mean value	Min	Max
ER	Industrial waste gas emissions (billion cubic meters)	462	982.85	2.00	4676.77
IA_1	Location entropy of employment in secondary industry (%)	462	0.98	0.21	1.60
IA_2	Location entropy of added value of secondary industry (%)	462	1.04	0.14	4.56
TECH	Internal expenditures on R&D funding (million yuan)	462	3502.15	3.36	96674.82
F	Amount of foreign capital used (million yuan)	462	774.68	1.83	8203.01

For the first time in human history, the Kyoto Protocol, which restricts emissions of greenhouse gases (including carbon dioxide (CO_2), methane (CH_4), nitrous oxide (N_2O), hydrofluorocarbons (CFCs), and perfluorocarbons (CF_4)) with regulations, came into effect in 2005. The most important factor causing the climate warming effect is CO_2, which accounts for up to 66% of the total warming effect factors [15]. The emission reduction effect analyzed in this work mainly refers to CO_2 emission reduction, and the CE concerned mainly refers to the research results of Lv et al. [16], that is, the ratio of CO_2 emission and the added value of the secondary industry in this region. According to the "China Energy Statistical Yearbook" data, the total CO_2 emission of the secondary industry has maintained a proportion of about 83%-86% among the three major industries in the past two decades. The changing trend of the carbon emission curve in China is almost consistent with that of the total carbon emission curve in the secondary industry [17]. Since the secondary industry is the main factor driving the total regional carbon emissions, the carbon emission level of cities in this work mainly focuses on the secondary industry. In Table 1, based on the CO_2 emissions and the added value of the secondary industry in 21 cities of Guangdong Province from 1997 to 2018, the explained variable is determined to be CE, that is, the ratio of CO_2 emissions to the added value of the secondary industry.

In this study, a single index of industrial wastewater discharge is adopted to represent the ER. The less industrial wastewater discharged, the stricter the environmental regulation. In addition, this work refers to the setting of SV Lall in the study [18] and uses location entropy to measure industrial agglomeration. Relevant indicators are set from two perspectives in this work to distinguish different classes of agglomeration produced by labor-intensive and regional industrial advantages. The first Angle is based on the number of employments in the secondary industry, and the other one is based on the operation of the added value of the secondary industry (equations (1) and (2), respectively).

$$\mathrm{IA}_{1_{\mathrm{ij}}} = \frac{a_{1_{ij}}/a_{1_j}}{a_{1_i}/a_1} \tag{1}$$

$$\mathrm{IA}_{2_{ij}} = \frac{a_{2_{ij}}/a_{2_j}}{a_{2_i}/a_2} \tag{2}$$

In equation (1), $\mathrm{IA}_{1_{ij}}$ represents the location entropy of employment number in the industry i in region j. This work will focus on case $i = 2$. Where j is the integer within [1, 21], a_{ij} represents the employment number of the i-th industry in region j, a_{1_j} is the total employment number of the region j, $a_{1\mathrm{i}}$ refers to the total employment number of the secondary industry in the country, a_1 is the total employment number of the country. If the value is greater than 1, the specialization degree of the regional primary industry is higher. On the other hand, $\mathrm{IA}_{2_{ij}}$ represents the locational entropy of the added value of the i-th industry in region j. The only difference from equation (1) is that the industry's added value replaces the essential data. The higher the locational entropy, the higher the industrial agglomeration. The greater the locational entropy of the number of employments in the secondary industry is, the greater the locational entropy of the industrial added value in the secondary industry is, the stronger the production capacity of enterprises in the region, and the more significant the agglomeration effect. In addition, other explanatory variables include TECH and F. TECH refers to the research content of Sanchez-Sellero et al. [19]. In some studies, internal expenditure of research and development (R&D) funds is undertaken to measure the TECH. This data refers to the expenditure of enterprises for internal scientific research. The more internal expenditure of R&D funds, the sufficient capital and technological foundation an enterprise has to promote its innovation and development. In this work, F is described by the actual utilization of foreign capital amount, which usually refers to the amount of foreign money obtained after signing a contract with foreign enterprises [20].

2.2. K-means Clustering Dimensionality Reduction Processing

K-means algorithm is a typical clustering method with fast convergence and effective processing of large data sets, so it has been successfully applied to document clustering, market segmentation, image segmentation, and feature learning [21-22]. In this work, CE is selected as the classification index of urban agglomeration for cluster analysis. The process follows the following steps. First, the K points from the data set are randomly selected as the center points of our clustering, and the classification index corresponding to the population unit is assigned to the nearest class center point based on the Euclidean distance. After K classes are formed, the center point of each class is recalculated, and the above steps are repeated. The process is iterated, and the algorithm converges until the class no longer changes.

The optimal number of clusters K is defined according to this work's "elbow method" [23]. It is not difficult to understand that the aggregation of various categories and the sum of the sample's square error (SSE) will decrease accordingly when the number of clusters increases. Before the optimal number of clusters appears, the magnitude of SSE reduction corresponding to each unit of K value is significant. This will continue until the K value reaches the optimal number of clusters. At this time, when the K value continues to increase, the decline range of SSE will decrease significantly, and the changing trend of the SSE curve will flatten out.

The case discussed in this work is shown in Figure 1, with obvious inflection point features at K =3. Therefore, the below relations are determined: the optimal number of clusters based on the K-means clustering algorithm is 3. Therefore, the 21 cities in Guangdong Province are divided into three categories (as shown in Figure 2a). The carbon emission intensity of the first, second, and third-class cities is represented by CE_1, CE_2, CE_3, respectively, and different colors in Figure 2a mean the spatial distribution of the cities after classification. The first-class (Heyuan) and the second-class urban agglomeration (Zhaoqing, Shanwei, Yunfu, Shaoguan, Qingyuan, and Meizhou), except Shanwei, are inland cities. The third-class urban agglomerations (Guangzhou, Shenzhen, Foshan, Dongguan, Zhongshan, Zhuhai, Jiangmen, Huizhou, Shantou, Chaozhou, Jieyang, Zhanjiang, Maoming, and Yangjiang) are all coastal. Some of them represent the central regions of China participating in economic globalization (Pearl River Delta urban agglomeration), and include the emerging economic belt of Eastern Guangdong. In Figure 2b, the ratio of CO_2 emissions to the industrial-added value of the secondary industry of the three classes of cities from 1997 to 2018 is defined as CE. As illustrated in the figure, CEs of the three classes of urban agglomerations all shows a fluctuating and declining trend over time. Among them, Heyuan is conditioned by the not perfect industrial structure, the level of economic development, and the most crucial carbon intensity. The third-class urban agglomeration, represented by Guangzhou, shows an excellent geographical location and sound economic development, giving full play to the degree of regional environmental consciousness and environmental protection efforts.

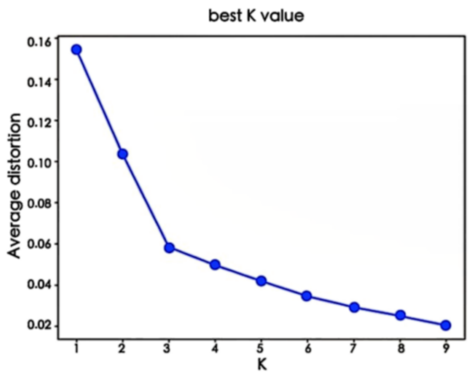
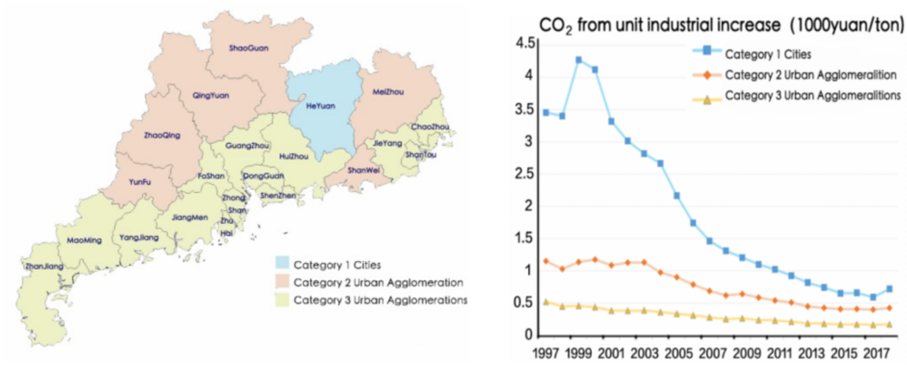


Figure 1. Optimal clustering results.

2.3. The SMLR-Based Model

SMLR is an essential method in multivariate statistical analysis, which is widely used in the social, economic, and natural sciences [24-26]. In the SMLR-based model, the dependent and the independent variables can not only be associated but this association can also be quantified. The preexisting variable significance will be tested after each newly accepted variable entry when the optimal model is selected in stepwise regression. It will be removed if the primary variable becomes no longer significant due to introducing a new variable. Such selection and exclusion procedures will be repeated until the model cannot be improved by introducing more new explanatory variables and all variables included in the regression model are significant.



(a) Cluster distribution diagram (b) CO_2 generated by a unit of industrial added value

Figure 2. The distribution of clustered urban agglomerations and the CO_2 trend generated by unit industrial added value.

In this work, the CE values corresponding to the clustering centers of three classes of urban agglomerations are selected as the input data of the explained variables (dependent variables). The related IA, TECH, F, and ER of them on the time dimension are undertaken as input data of explanatory variables (independent variables) input data. The multiple linear regression model is established by using the stepwise regression method (as shown in equation (3)):

$$\begin{cases} CE = a + b_1\, IA_1 + b_2\, IA_2 + b_3 \mathrm{TECH} + b_4 \mathrm{F} + b_5 \mathrm{ER} + \varepsilon_i \\ \varepsilon_i \sim N(0, \sigma^2), i = 1,2, \cdots\cdots, n \end{cases} \tag{3}$$

In the above equation, a is the constant term, $b_1, b_2 \cdots$ are the coefficients of the explanatory variables (independent variables), and ε_i is the error term conforming to the Gaussian distribution. The optimal equation can be constructed based on the highest multiple determination coefficient values, $\mathrm{R}^2 = \sum (\widehat{\mathrm{CE}_\mathrm{i}} - \overline{\mathrm{CE}})^2 / \sum (\mathrm{CE}_\mathrm{i} - \overline{\mathrm{CE}})^2$. Here, CE_i and $\overline{\mathrm{CE}}$ are the actual observed value and corresponding average value, respectively, $\widehat{\mathrm{CE}_\mathrm{i}}$ is the model estimate. In addition, the values corresponding to the overall test will be used to evaluate the model performance. The model prediction is better when the value is small and close to 1.

3. Empirical Results and Analysis

3.1. Correlation Analysis

Except for the first class, the other two classes of urban agglomerations include multiple cities. If the historical data of all these cities are analyzed one by one, it is evident that the characteristics of regional industries are challenging to summarize quantificationally due to excessive specific differentiation. Such an operation does not promote the subsequent regression model establishment and practical analysis.

Therefore, the median at each time point is undertaken as the expected value in the original variable matrix to define the time sequence vector of each variable in various urban agglomerations to maintain the pertinence of the discussion (as shown in Table 3). Before the multiple progressive regression model analysis, the linear correlation among the various variables is analyzed to avoid the interference caused by excessive variables and reduce the difficulty of the optimal model selection.

Table 3. Description of the characteristic data of each urban agglomeration.

Urban agglomerations	Variables	Unit	Mean value	Median level	Min	Max
1	ER	Billion cubic meters	261.68	201.11	24.81	632.08
	IA_1	%	84.04	89.87	23.13	127.66
	IA_2	%	95.14	94.18	18.36	341.34
	TECH	Million yuan	124.39	88.95	9.42	323.77
	F	Million yuan	166.34	170.97	69.73	321.66
2	ER	Billion cubic meters	802.91	778.96	169.30	1662.80
	IA_1	%	94.54	93.95	76.94	107.27
	IA_2	%	74.12	79.49	56.96	88.23
	TECH	Million yuan	272.32	182.30	30.39	862.68
	F	Million yuan	78.81	70.15	32.69	150.84
3	ER	Billion cubic meters	890.30	807.92	202.30	2399.33
	IA_1	%	97.34	100.61	87.11	105.51
	IA_2	%	106.01	105.89	102.52	111.62
	TECH	Million yuan	1974.07	1198.86	306.06	5881.40
	F	Million yuan	460.94	498.64	234.72	643.16

$$\rho_{X,Y} = \frac{\sum XY - \frac{\sum X \sum Y}{N}}{\sqrt{\left(\sum X^2 - \frac{(\sum X)^2}{N}\right)}\sqrt{\left(\sum Y^2 - \frac{(\sum Y)^2}{N}\right)}} \tag{4}$$

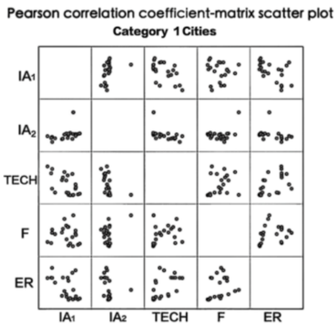
The Pearson correlation coefficient is a standard measure of the linear correlation among different variables (as shown in equation (4)). When the value is (-1, 0), there is a negative correlation, while (0, 1) indicates a positive correlation, and there is no correlation if it is 0. The closer the correlation coefficient is to 1, the stronger the correlation among the variables. When the absolute value of the correlation coefficient is (0.1, 0.3) and (0.3, 0.7), when the absolute value of the correlation coefficient is more significant than 0.7, it is generally believed that there is a strong correlation among the variables [27].

According to the above standards, the correlation analysis of variables of the three classes of urban agglomeration is shown in Table 4. In first-class urban agglomeration, the linear correlation between IA_1 and two variables (TECH and ER) are significant. The correlation coefficients are -0.568 and -0.528, respectively, indicating that based on the current situation of Heyuan city, the increase of ER and TECH can affect the original characteristics of labor-intensive industries. In addition, guiding the technological transformation of local advanced and traditional industries may bring pain points to the industry in the short term, such as cost increase (development uncertainty and long landing cycle), insignificant income, difficulties in expanding new business, professional ability requirements, and mismatch of local resources. Therefore, such an operation will not enhance the location role of the local secondary industry in the short term. This is shown by the significant relationship between IA_2 and TECH in Table 4 and the correlation coefficient values of -0.474. Overall, the correlation coefficient is less than 0.7 for all the variables. The spatial scatter distribution of Figure 3a shows no highly linear correlation among all variables.

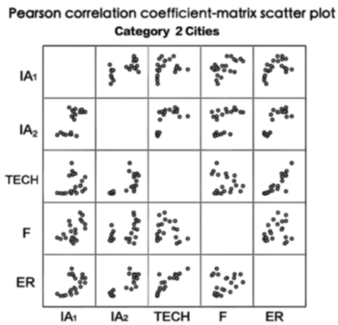
Table 4. Correlation of Category I urban agglomeration.

Correlation	IA_1	IA_2	TECH	F	ER
IA_1	1	0.395	-0.568^{**}	-0.045	-0.528^{*}
IA_2	0.395	1	-0.474^{*}	0.354	-0.316
TECH	-0.568^{**}	-0.474^{*}	1	0.254	0.543^{**}
F	-0.045	0.354	0.254	1	0.295
ER	-0.528^{*}	-0.316	0.543^{**}	0.295	1

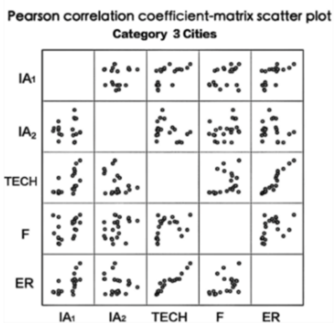
Note: *, ** represent significance at the 5% and 1% significance levels, respectively.



(a) Scatter plot of correlation of variables in correlation of variables of first-class urban agglomeration



(b) Scatter plot of second-class urban agglomeration



(c) Scatter plot of the variable correlation in the third-class urban agglomeration

Figure 3. Diagram of the correlation scatter between the variables in each urban agglomeration.

For the second-class urban agglomeration, the correlation coefficients between IA_1 and three variables (F, IA_2 and ER), as given in Table 5, are 0.66, 0.546, and 0.485, respectively. Moreover, the correlation coefficient values between IA_2 and TECH, F and ER are 0.679, 0.473, and 0.826, respectively, and the exact relationship is significant. This shows that the guidance of FDI in the local secondary industry, industrial technology upgrading, and environmental regulation can improve the local industrial location. Notably, $\mathrm{IA_2}$ is highly linearly associated with ER (as shown in Table 5 and Figure 3b). Meanwhile, the correlation coefficients of both $\mathrm{IA_1}$ and $\mathrm{IA_2}$ with TECH are close to 0.7. To avoid multicollinearity, they will all be cautiously treated in the subsequent optimal model explanatory variable selection.

Table 5. Correlation of the second-class urban agglomeration.

Correlation	$\mathrm{IA_1}$	$\mathrm{IA_2}$	TECH	F	ER
$\mathrm{IA_1}$	1	0.660**	0.233	0.546**	0.485*
$\mathrm{IA_2}$	0.660**	1	0.679**	0.473*	0.826**
TECH	0.233	0.679**	1	-0.074	0.087**
F	0.546**	0.473*	-0.074	1	0.233
ER	0.485*	0.826**	0.087**	0.233	1

Note: * and ** represent significance at the 5% and 1% significance levels, respectively.

For the third-class urban agglomeration, the correlations between $\mathrm{IA_1}$ and the two variables (TECH and ER) given in Table 6 are 0.62 and 0.693, respectively, and the linear relationship is significant at the 1% level. This indicates that the guidance of investment in industrial technology upgrading and environmental regulation for the local secondary industry is of positive significance for the agglomeration of professional talents. However, it is worth noting that ER is highly linearly correlated with TECH and $\mathrm{IA_1}$ (as shown in Table 6 and Figure 3c).

Table 6. Correlation of the third-class urban agglomeration.

Correlation	$\mathrm{IA_1}$	$\mathrm{IA_2}$	TECH	F	ER
$\mathrm{IA_1}$	1	0.040	0.620**	0.408	0.693**
$\mathrm{IA_2}$	0.040	1	-0.337	0.363	-0.159
TECH	0.620**	-0.337	1	0.375	0.957**
F	0.408	0.363	0.375	1	0.483*
ER	0.693**	-0.159	0.957**	0.483*	1

Note: * and ** represent significance at the 5% and 1% significance levels, respectively.

3.2. Analysis of the Multiple Stepwise Regression Models

Stepwise linear regression is carried out for three classes of urban agglomerations, and the variables that could pass the F-test and the t-test are otherwise excluded. Finally, the stepwise regression analysis determines that the indicators significantly impacting CE in urban cluster I are IA, TECH, F, and ER.

$$\mathrm{CE_1} = 3.332 + 0.825\mathrm{IA_1} - 0.002\mathrm{TECH} - 0.003\mathrm{ER} - 0.006\mathrm{F} \tag{5}$$

Based on the regression results (Table 7), the R^2 value in the model is 0.922, and the corresponding value p is 5.553×10-10. From this perspective, the combined effect of the CE caused by these explanatory variables is significant. The correlation analysis reminds us to consider the collinearity of explanatory variables. Therefore, the variance inflation factor tests the collinearity of explanatory variables. Here, the variance expansion factor of the first explanatory variable is defined below:

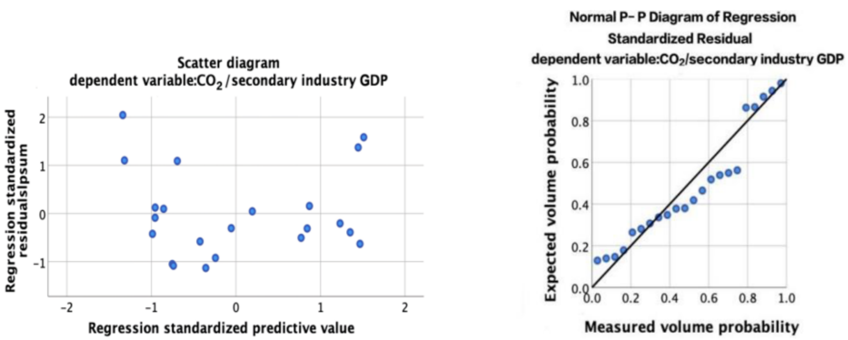
$$\mathrm{VIF_i} = \frac{1}{1-\mathrm{r_i^2}} \tag{6}$$

In equation (6), r_i is x_i the determination coefficient of the regression of other independent variables x_j, which takes x_i as the independent variable [28]. For the third-class urban agglomeration, the corresponding variance expansion factors are 1.679, 1.156, 1.708, and 1.739, respectively. According to the above results, there is no multicollinearity in the model.

Table 7. Results of the stepwise regression process of urban cluster.

Regression model	Regression coefficient B	*t*-Statistic	*t*-Statistics probability value	R^2	Adjusted R^2	F Statistics probability value	Multicollinearity test	
							Allowance	VIF
Constant	3.332	8.784	0.000					
ER	-0.003	-7.207	0.000	0.937	0.922	5.553E-10	0.596	1.679
F	-0.006	-4.984	0.000				0.865	1.156
IA_2	0.825	2.569	0.020				0.585	1.708
TECH	-0.002	-2.361	0.030				0.575	1.739

The graph on Figure 4 shows the relationship between the regression residuals and the predicted values of the CE. The abscissa is the predicted value, and the ordinate represents the residual value. Obviously, most of the residues are basically distributed in the moderately wide region of the level band of residue 0 values. Thus, the residuals of the regression equations essentially satisfy the assumption of homogeneity of variance. Figure 4b shows the normal probability diagram of the residual. Obviously, except for a few outliers, most of the residuals are near the reference line. It suggests that residuals of the regression equation satisfy the assumption of normality.



(a) Residual analysis (b) Normal probability plot of the residuals

Figure 4. The relationship between the regression residual and the fitted value of first-class urban agglomeration and the normal probability diagram corresponding to the residual.

The first-class urban agglomeration (Heyuan) is in the mountainous area of northern Guangdong. Its overall industrial development is low, its leading industry is single, its industrial foundation is weak, and its economic development is relatively backward. As shown in equation (5), the ER, IA, F, and TECH can significantly affect the change of carbon emissions in this region. The positive coefficient of the constant value and the IA exhibit that CE of the second industry in this region is high in Guangdong Province, and its regional agglomeration characteristics formed by the original enterprises can effectively promote CE. For such regions with industrial characteristics, the green development of the local economy can be effectively promoted by strengthening the guidance of environmental regulation, increasing the green technology upgrading of the industry, and promoting the FDI in the local secondary industry. In this model, it is worth emphasizing that environmental regulation is the key to curbing the FDI based on the industrial transfer motivation of the "pollution shelter" and realize the regional industrial upgrading for the relatively backward regions.

$$\mathrm{CE}_2 = 1.853 - 1.100\mathrm{IA}_1 - 3.550 \times 10^{-4}\mathrm{ER} \tag{7}$$

$$\mathrm{CE}_3 = 0.434 - 1.570 \times 10^{-4}\mathrm{ER} \tag{8}$$

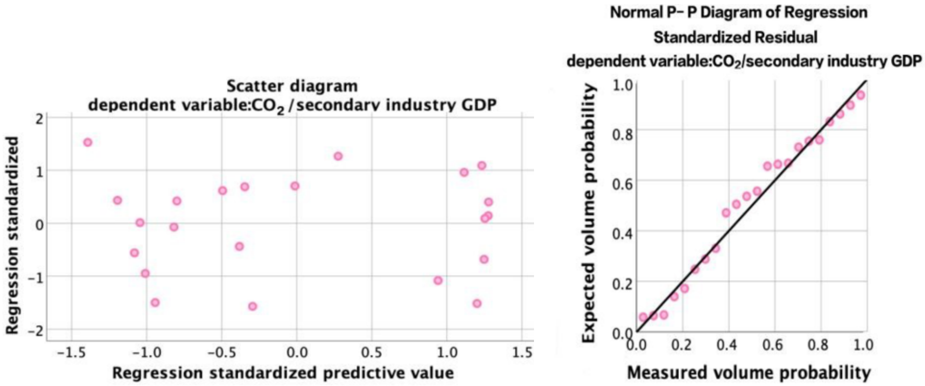
Next, the regression equations of the second-class and third-class urban agglomerations are constructed (as shown in equations (7) and (8), respectively). Zhaoqing and other cities included in the second-class urban agglomeration are mainly distributed in the east and west wings of Guangdong, where the secondary industry accounts for a prominent proportion, but the integrity of the industrial chain is poor, and the development of the regional industrial economy is still in the growth stage [29]. Regarding spatial distribution, the third-class urban agglomeration is mainly concentrated in the Pearl River Delta coastal region, where the net worth industry is high, the industrial infrastructure and supporting services are perfect, the economic structure is relatively perfect, and the overall economic level is better than the other two. Therefore, it has the best economic vitality in Guangdong Province.

Table 8 presents the stepwise regression results of the third-class urban agglomeration. The location quotient and ER have become the core factors affecting the CE of the secondary industry in this region. The value of R^2 and the adjusted R^2 are 0.954 and 0.950 respectively, which are very close. Moreover, the probability value p of the corresponding statistic F is 1.786×10-13. Therefore, it can be concluded that the combined effect of these explanatory variables on CE is significant. The variance expansion factor of the two explanatory variables is 3.150, so the model has no multicollinearity.

Table 8. Results of the stepwise regression of the second-class urban agglomeration.

Regression model	**Regression coefficient B**	***t* Statistic**	***t* Statistics probability value**	**R^2**	**Adjusted R^2**	**F Statistics probability value**	**Multicollinearity test**	
							Allowance	**VIF**
Constant	1.853	14.665	8.186E-12					
Ia_2	-1.100	-6.625	2.000 E-6	0.954	0.950	1.786E-13	0.317	3.150
Er	-3.550E-4	-6.625	5.900 E-5				0.317	3.150

On the other hand, the regression-normalized residual correlation data is distributed in the equal-width interval [-2, 2] (Figure 5). In addition, the data points in Figure 5b are all near the reference line, so the residuals of the regression equation satisfy the assumptions of normality and homogeneity of variance. In the model, the regression coefficient of the explanatory variables is all negative. It indicates that further improving the integration of the industrial chain and strengthening the IA is essential to lower CE of the secondary industry in the region and to promote the completion of regional green industrial upgrading in addition to strengthening the constraints of environmental regulation on the industry.



(a) Residual analysis (b) Normal probability plot of the residuals

Figure 5. The relationship between the regression residual and the fitted value of second-class urban agglomeration and the normal probability diagram corresponding to the residual.

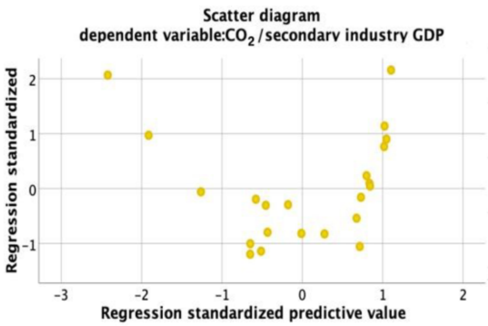
Table 9 presents the gradual regression results of the third-class urban agglomeration. The R^2 in the model and the adjusted R^2 are 0.778 and 0.766, respectively, and the corresponding p-value is 5.895×10-8, respectively. Therefore, it can be concluded that the combined effect of these explanatory variables on CE is significant. On the other hand, the residual difference of the regression equation satisfies the assumptions of normality and homogeneity of variances (Figure 6).

Table 9. Stepwise regression results of the third-class urban agglomeration.

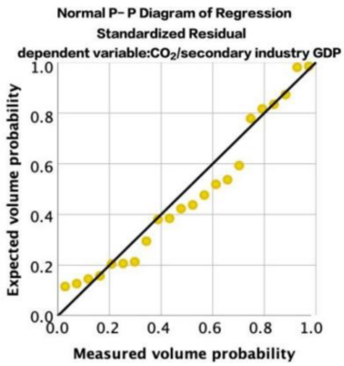
Regression model	Regression coefficient B	t Statistic	t Statistics probability value	R^2	Adjusted R^2	F Statistics probability value	Multicollinearity test	
							Allowance	VIF
Constant	0.434	21.509	2.683E-15	0.778	0.766	5.859E-8		
Er	-1.570E-4	-8.361	5.859E-8				1.000	1.000

Unlike the unit CEs of the first-class and second-class urban agglomerations, which are affected by multiple factors, ER mainly restricts that of the third-class urban agglomeration. This also indicates that with the improvement of the economic level, in the region where the industrial structure tends to be perfect, the factors interfering with the unit CE of the secondary industry are gradually decreasing. At this time, the core guiding role of environmental regulation will be prominent.

According to equations (5), (7), and (8), the impact degree of environmental regulation in the three classes of urban agglomerations varies significantly, and the coefficients are -0.003, -3.550×10^{-4} and -1.570×10^{-4}, respectively. It is worth noting that lowering the impact degree of environmental regulation does not mean that the constraint degree of environmental regulation can be reduced by improving urban specifications and economic development levels. For the first-class urban agglomeration (Heyuan), driven by local industrial upgrading needs, the strict environmental regulation is more likely to stimulate the “innovation compensation” effect. Meanwhile, it can force some enterprises that fail to meet the environmental regulatory standards to exit the market to improve the regional green total factor productivity comprehensively. For the second-class urban agglomeration, implementing environmental regulation should improve the industrial chain integration and strengthen the IA, to promote the green upgrading of regional industries more efficiently. Therefore, the targeted introduction of "incentive environmental regulation" (such as "green innovation subsidy" and "emission trading") to cooperate with the "command environmental regulation" policy and appropriately reduce the intensity of regulation helps local industries reduce the cost of green transformation, realize industrial integration, and encourage innovation. In addition, the third-class urban agglomeration, including the Pearl River Delta and the coastal areas of Guangdong Province, has always been the critical location of green and low-carbon industry projects in Guangdong Province. The third-class urban agglomeration features a clear division of labor, a low degree of industrial overlap, and a significant innovation-leading effect. This also makes the lowest unit-added value of the secondary industry in the region among the three classes of urban agglomerations. However, when the core guiding role of environmental regulation in this region is discussed, Lei mentioned that the improvement of ER had an excellent inhibitory effect on industrial emission reduction only in the short term, and its marginal effect was gradually decreased [20]. The unique economic pattern has curbed the upgrading of the green industry in the third-class urban agglomeration. In this case, further strengthening the central polarization of the core region and expanding the economy can not only promote the environmental regulation to exert the effect of inhibiting CE by affecting the economic pattern but also promote the realization of the win-win situation of new economic agglomeration and environmental governance.



(a) Residual analysis



(b) Normal probability of the residuals

Figure 6. The relationship between the regression residual and the fitted value of the third-class urban agglomeration and the normal probability diagram corresponding to the residual.

4. Conclusion

The increasing importance of environmental issues in economic research and environmental regulation in economics make environmental regulation a current research hotspot. The traditional theory insists that the strict environmental regulation will increase equipment investment and support the enterprise's technical and human capital investment. To adapt to the environmental regulation, the cost increase caused by intensity enhancement is much more significant than the profit growth brought by the technological innovation driven by environmental regulation, and labor productivity will be significantly reduced. The emergence of "Porter Hypothesis" has dramatically challenged the traditional concept. According to "Porter Hypothesis", the choice of reasonable ER can encourage enterprises to carry out adaptive technological innovation, form the technology diffusion effect in reducing enterprise costs, and promote upgrading and optimizing industrial structure, improving production efficiency and market competitiveness enterprises, and economic growth. Undeniably, for any region in the world facing an energy resource gap and ecological environment pressure, it is essential to use direct or indirect regulation tools of the government to solve the externality and optimize the allocation of market resources to promote the internalization of environmental costs and then realize the harmonious development of industry and ecology. For China, which permanently adheres to green development, improving the efficiency of environmental regulation is a crucial way to achieve carbon peaking and carbon neutrality. In this context, discussing the relationship of the environmental regulation contract with industrial factors and green development is conducive to analyzing and understanding the balance factors of environment and industrial upgrading and improving the theoretical framework of environmental regulation design strategy.

Based on the panel data from 21 cities in Guangdong province from 1997 to 2018, this work discusses the correlation between the secondary industry elements and CE by building a generalized linear model. Due to the large volume of historical data composed by the explanatory variable matrix of each city, the K-means algorithm is adopted, and the CE is undertaken as the classification index of urban agglomeration for cluster analysis to improve the analysis efficiency. After iterative operations, the 21 cities in Guangdong province are divided into three classes. Heyuan is only included in the first class. The second-class urban agglomeration includes six cities: Zhaoqing, Shanwei, Yunfu, Shaoguan, Qingyuan, and Meizhou. Among them, except for Shanwei, the rest are inland cities. There are 14 cities in third-class urban agglomerations, including Guangzhou, Shenzhen, Foshan, Dongguan, Zhongshan, Zhuhai, Jiangmen, Huizhou, Shantou, Chaozhou, Jieyang, Zhanjiang, Maoming, and Yangjiang. Except for the first class, the second-class and third-class urban agglomerations include multiple cities. In order to clarify the characteristics of regional secondary industry elements and maintain the pertinence of the discussion, the median values of related IA, TECH, F, and ER in three classes of urban agglomerations of 21 prefectural cities in Guangdong Province are undertaken as representative values in the time dimension for correlation analysis. It is found that for first-class urban agglomeration, the improvement of ER or TECH significantly impacts the characteristics of the original labor-intensive industries. In addition, guiding the green transformation of traditional industries with local advantages may bring some pain points to the industry in the short term, such as cost increase (development uncertainty and long landing cycle), nominal income growth, difficulty in new business expansion, professional capacity requirements, and mismatch of local resources. For the second-class urban agglomeration, the guidance of FDI in the local secondary industry, industrial technology upgrading, and environmental regulation will improve the local industrial location. For the third-class urban agglomeration, strengthening the investment in industrial green technology upgrading and the guidance of environmental regulation will be of positive significance to the agglomeration of professional talents.

Then, CE values corresponding to the clustering centers of three classes of urban agglomerations are selected as the input data of the explained variables (dependent variables). The median values of the secondary industry elements-related variables in the time dimension of 21 cities are undertaken as the input data of the explanatory variables. In addition, the stepwise regression method establishes three groups of multiple linear regression models. The determination coefficients and tests of each group support the significance of the model, and the expansion factor test excludes the multicollinearity. In addition, the residual analysis of the model concludes the normality and homogeneity of variance. On the other hand, the analysis points out that for the first-class urban agglomeration with a low overall development level, single leading industry, and weak industrial foundation, strengthening the guidance of environmental regulation, increasing industry green technology upgrading, and promoting the local secondary industry of FDI to promote the green development of the local economy can exert a positive role. What is more worth emphasizing is that for such regions, taking environmental regulation as the guidance is the key to curbing the behavior of FDI based on the motivation of "pollution haven" industry transfer and realizing regional industrial upgrading. Different from the first-class urban agglomeration, the leading industrial characteristics of the second-class urban agglomeration are that the secondary industry accounts for a prominent proportion, but the integrity of the industrial chain is poor, and the development of the regional industrial economy is still in the growth stage. The variable coefficient feature in the model suggests that in this region, further improving the industrial chain integration and strengthening the IA are essential methods to reduce the CE of the secondary industry and promote green industrial upgrading in addition to strengthening the constraints of environmental regulation on the industry. Regarding spatial distribution, the third-class urban agglomeration is mainly concentrated in the Pearl River Delta and the coastal areas, dominated by high-net-worth industries and relatively excellent industrial infrastructure and support services. Different from the situation where the unit CE of the first two classes of urban agglomerations is affected by multiple factors, the core guiding role of environmental regulation will become more prominent in the region where the industrial structure tends to be perfect with the improvement of the economic level.

Meanwhile, the model points out significant differences in environmental regulation impacts in the three classes of urban agglomerations. With the improvement of the regional and economic status of urban agglomeration, the impacts of environmental regulation are gradually weakened, which does not mean that the degree of environmental regulation constraint should be decreased. For regions with relatively backward economic development in Guangdong Province, strict environmental regulation is more likely to stimulate the "innovation compensation" effect by promoting the demand for industrial upgrading. For the second-class urban agglomeration located in the east and west wings of Guangdong province, the targeted introduction of "incentive environmental regulation" to cooperate with the "command environmental regulation" policy to appropriately reduce the intensity of regulation will help improve the local IA. Further, it will help the region to complete the industrial chain integration and promote industrial green upgrading. For the third-class urban agglomeration, further strengthening the central polarization degree of the core region can not only promote the environmental regulation to exert the effect of inhibiting CE by affecting the economic pattern but also promote the realization of the win-win situation of new economic agglomeration and environmental governance.

The research method used in this work can be extended for follow-up studies to explore the correlation between other industrial factors and the emission reduction effect. Similarly, suitable means can also be employed to explore further the green transformation and upgrading mechanism of environmental regulation in highly

Polluting subsectors of the secondary industry (such as papermaking and glass manufacturing). In addition, some other relevant issues (such as the timing of implementation decisions in environmental regulation design) are not discussed in this work. The subsequent exploration based on these contents will help to explore the further optimization direction of the environmental regulation design and provide more theoretical and data support.

Acknowledgments

This research is supported by National College Students' Innovation and Entrepreneurship Training Program (No. 202113844006), Major research and cultivation project 2020 (No. 2020YZDYB06R), and Soft Science Research Project (No.2020A1010020060). The authors would like to thank the referees for their helpful comments and suggestions on the manuscript.

References

[1] Mikulčić H, Baleta J, Wang X, et al. Sustainable development in period of climate crisis. Journal of Environmental Management. 2022; 303:114271.

[2] Chen J, Tan H, Ma Y. Distinguishing the complex effects of foreign direct investment on environmental pollution: Evidence from China. The Energy Journal. 2022;43(4).

[3] Zhong C, Hamzah H Z, Yin J, et al. Impacts of Industrial Agglomeration and Energy Intensity on Industrial Eco-Efficiency-Analysis Based on Spatial Correlation and Mediating Effect. Frontiers in Environmental Science. 2022:1077.

[4] Xu Y, Li S, Zhou X, et al. How environmental regulations affect the development of green finance: Recent evidence from polluting firms in China. Renewable Energy. 2022;189:917-926.

[5] Song M, Tao W, Shen Z. Improving high-quality development with environmental regulation and industrial structure in China. Journal of Cleaner Production. 2022;366:132997.

[6] Lanoie P, Laurent-Lucchetti J, Johnstone N, et al. Environmental policy, innovation and performance: New insights on the Porter hypothesis. Journal of Economics & Management Strategy. 2011; 20(3):803-842.

[7] Ramanathan R, He Q, Black A, et al. Environmental regulations, innovation and firm performance: A revisit of the Porter hypothesis. Journal of Cleaner Production. 2017; 155:79-92.

[8] Wu S, Zhang H. The existence and mechanism of the domestic pollution haven hypothesis: evidence from 265 cities in China. Letters in Spatial and Resource Sciences. 2021:1-24.

[9] Hart S L, Ahuja G. Does it pay to be green? An empirical examination of the relationship between emission reduction and firm performance. Business strategy and the Environment. 1996;5(1):30-37.

[10] Ranocchia C, Lambertini L. Porter hypothesis vs pollution haven hypothesis: Can there be environmental policies getting two eggs in one basket?. Environmental and Resource Economics. 2021;78(1):177-199.

[11] Peng H, Shen N, Ying H, et al. Can environmental regulation directly promote green innovation behavior?——based on situation of industrial agglomeration. Journal of Cleaner Production. 2021;314:128044.

[12] Jiang Z, Lyu P. Stimulate or inhibit? Multiple environmental regulations and pollution-intensive Industries' Transfer in China. Journal of Cleaner Production. 2021; 328:129528.

[13] Cole M A. Trade, the pollution haven hypothesis and the environmental Kuznets curve: examining the linkages. Ecological Economics. 2004; 48(1):71-81.

[14] Xu X, Arshad M A, Mahmood A. Analysis on international competitiveness of service trade in the Guangdong-Hong Kong-Macao Greater Bay Area based on using the entropy and gray correlation methods. Entropy. 2021; 23(10):1253.

[15] Nullis C, Greenhouse gas bulletin: Another year another record. World Meteorological Organization. 2021; 25102021.

[16] Lv KJ, He YX. Economic agglomeration, technological progress and carbon emission intensity in yangtze river delta urban agglomeration: An empirical study based on spatial econometric and intermediary effect. Ecological Economy. 2021;37(01):13-20. (in Chinese)

[17] He YD, Wen H, Sun CW. Forecasting China's total carbon emission and its structure in the 14th Five-Year Plan: Based on mixed-frequency ADL-MIDAS model. On Economic Problems, 2021;4:31-40. (in Chinese)
[18] Lall SV, Shalizi Z, Deichmann U. Agglomeration economies and productivity in Indian industry. Journal of Development Economics. 2004;73(2):643-673.
[19] Sánchez-Sellero P, Bataineh MJ. How R&D cooperation, R&D expenditures, public funds and R&D intensity affect green innovation?. Technology Analysis & Strategic Management. 2021;1-14.
[20] Lei YT, Sun JJ, Huang ZX. Urban agglomeration economy, environmental regulation and haze reduction effect: An empirical study based on China's three urban agglomerations. Macroeconomics. 2021;(01):131-149. (in Chinese)
[21] Coates A, Ng A Y. Learning Feature Representations with k-means in NEURAL NETWORKS: Tricks of the Trade. Springer, Berlin, Heidelberg. 2012; 561-580.
[22] Likas A, Vlassis N, Verbeek JJ. The global k-means clustering algorithm. Pattern Recognition. 2003; 36(2): 451-461.
[23] Kodinariya TM, Makwana PR. Review on determining number of cluster in K-Means clustering. International Journal. 2013; 1(6):90-95.
[24] Wang H, Meng J. Predictive modeling on multivariate linear regression. Journal of Beijing University of Aeronautics and Astronautics. 2007; 33(04):500.
[25] Uyanık GK, Güler N. A study on multiple linear regression analysis. Procedia-Social and Behavioral Sciences. 2013;106:234-240.
[26] Yi X, Liu S. Impact of environmental factors on pulmonary tuberculosis in multi-levels industrial upgrading area of China. Environmental Research. 2021;195:110768.
[27] Baak M, Koopman R, Snoek H, et al. A new correlation coefficient between categorical, ordinal and interval variables with Pearson characteristics. Computational Statistics & Data Analysis. 2020;152:107043.
[28] Fop M, Murphy TB. Variable selection methods for model-based clustering. Statistics Surveys. 2018;12:18-65.
[29] Bui T, Yang D, Jones W, Li J. China's Economic Powerhouse: Economic Reform in Guangdong Province. Springer. 2002.