Research on the Model of "Effects of Land Use Projects on Ecosystem Services"

Haowen Jing[a], Rui Yang[b]

School of electrical engineering, North China Electric Power University, Baoding 071000, China;

[a]2920671885@qq.com, [b]2627751949@qq.com

Keywords: land use projects, ecosystem services, ecosystem assessment model

Abstract: In order to account for the cost-benefit ratio of environmental costs to the land use projects, we establish an ecosystem services assessment model to predict the final environmental governance costs based on the pollution level of existing projects. According to the principal component analysis method, four indicators with the weakest correlation and the largest weight were selected as the measuring tools of environmental pollution degree and then the regression relationship between environmental governance cost and each index were bulit. To further eliminate the correlation between indicator variables, we used principal component regression analysis method. After obtaining the regression equation, R^2 test, F test, and T test were performed successively and the model is proved to be feasible. Finally, we use the grey prediction method to predict the coefficients of the regression model to make the model more applicable.

1. Introduction

For a long time, most land use projects have not considered the environmental costs they generated when planning. However they would lower the value of ecosystem services greatly when added up. Therefore, it's an urgency to work out an ecosystem services evaluation model of land use projects to help land planner and manager better estimate the economic costs in the development process, so as to achieve the balance between economic and sustainable ecological benefits.

2. General Assumptions

Ecosystem services and land project planning are comprehensive issues with multiple indicators. It is impossible to quantify them one by one. Based on this, we have made some simplifications and bring forward the following assumptions when building the model.

- We assume that all sources of data collected are authentic and reliable.
- During the evaluation phase, the project will not be affected by strong external conditions.
- We take into account the environmental degradation posed during project construction only.

3. Establishment of evaluation system

After preliminary processing and analysis of the collected data, we find that there is a certain relationship between pollution treatment costs and pollution emissions. Through expert review, it is conceivable to make a prediction on possible pollution during the project planning period. Then we can get the suppositional cost of environmental governance by combining predicted results with the deterministic relationships. For project planners, they can forecast the environmental treatment cost based on the information provided by our model, and then make corresponding adjustments to the project in advance.

3.1 Index selection

According to Costanza's global ecosystem services valuation model, the direct and indirect values of ecosystems can be segmented into 17 categories: gas regulation, climate regulation, erosion control, sediment retention, soil formation, waste disposal, biological control, food production, raw materials and so on [1]. Employing principal component analysis, we selected four indicators from four aspects that are convenient for quantification to measure the economic losses caused by environmental degradation. It should be noted that due to the different project sizes, we use the pollutant emission per unit of gross domestic product.

In the model, we take the environmental protection expenditure (unit: ten thousand yuan) to represent the economic loss. In the meanwhile, we give four quantitative indicators of environmental pollution as the following table shows:

Table.1. Evaluation indicators

Target	Quantitative index	Expression	Units
Air contamination	SO_2 emissions per unit area	x_1	mg/m^3
Waste disposal	Solid emissions per unit area per year	x_2	$t/\left(a \cdot km^2\right)$
Water regulation	Chemical oxygen demand (COD)	x_3	mg/L
Interference regulation	Random detection (db)	x_4	db

3.2 Model: Principal Component Regression (PCR)

According to regression analysis, we can obtain a linear relationship between dependent variable (y) and independent variables (x_1, x_2, x_3, x_4). Under each index, we select nine projects and denote them as n_i ($i = 1, 2, \ldots, 9$). However, in the ecosystem service evaluation system, there are mutual constraints among different independent variables. The coefficients won't pass the significance test with general regression analysis only. Therefore, we use the Principal Component Regression (PCR) analysis method instead.

PCR analysis is based on normal regression model. Firstly, it transforms the original regression variables into component variables, then, selects the principal component that has vital functions as a new independent variable. This not only separates the variables that affect each other, but also decreases the dimension of the data, making regression analysis more accurate.

3.2.1 Data preprocessing

(1) Establish the raw data matrix

Firstly, we define the $m \times n$ matrix A:

$$A = \left(a_{ij}\right)_{m\times n}, \quad i = 1, 2, \ldots, m; \quad j = 1, 2, \ldots, n$$

Where a_{ij} is used to denote the i-th project indicator under the j-th indicator.

(2)Vector normalization

As Table1 shows, the four indexes selected have different dimensions. It's infeasible to directly determine their shares in total environmental pollution. We obviously need to standardize matrix A.

Therefore, we develop matrix B

$$B = \left(b_{ij}\right)_{m\times n}, \quad i = 1, 2, \ldots, m; \quad j = 1, 2, \ldots, n$$

And we denote element b_{ij} as follows:

$$b_{ij} = \frac{a_{ij} - \mu_j}{S_j}, \quad i = 1, 2, \ldots, m; \quad j = 1, 2, \ldots, n$$

Through this process, we can analyze the relationship between indicators and their relationship with environmental pollution index with little hindrance.

3.2.2 Quantification of indicators

Firstly, we set up the correlation coefficient matrix C and we let

$$C = \left(c_{ij}\right)_{m\times m}, \quad i, j = 1, 2, \ldots, m;$$

Where $c_{ij} = \frac{1}{n-1}\sum_{i=1}^{n} b_{ki} b_{kj}$, $i, j = 1, 2, \ldots, m$;

We select data from nine projects of different sizes [2] and apply the method above to calculate the specific correlation coefficient matrix. The result is shown below:

$$\begin{pmatrix} 1 & -0.2937 & -0.2078 & -0.3060 \\ -0.2937 & 1 & -0.1200 & -0.3128 \\ -0.2078 & -0.1200 & 1 & 0.4282 \\ -0.3060 & -0.3128 & 0.4282 & 1 \end{pmatrix}$$

Then, we use Matlab to solve the eigenvalues and eigenvectors of the matrix above and gain the following consequences:

$$\lambda_1 = 1.6693 \qquad U_1 = [-0.3777, -0.6540, 0.4021, 0.5176]$$

$$\lambda_2 = 1.2950 \qquad U_2 = [-0.2521, 0.7500, 0.2898, 0.5385]$$

$$\lambda_3 = 0.6547 \qquad U_3 = [-0.5897, 0.0129, 0.7900, -0.1671]$$

$$\lambda_4 = 0.3810 \qquad U_4 = [0.6678, -0.0982, -0.3608, -0.6436]$$

If adding λ_1, λ_2 and λ_3 up, we can clearly find that the cumulative contribution rate of λ_1, λ_2 and λ_3 reaches 90.47%. Consequently, we pick up λ_1, λ_2 and λ_3 as the principal component of the index.

Ultimately, we obtain the multiple regression equation:

$$\hat{y} = 55123.8 + 25303.7x_1 + 164.4x_2 + 5.1x_3 - 531.1x_4 \qquad (1)$$

3.2.3 Testing of Residuals and Confidence Intervals

The magnitude of the residual reflects the extent to which the original data deviates from the regression equation. The confidence interval reflects the accuracy of the parameter estimates, which are given by the following formula:

$$S_i^2 = \frac{\left(y_i - \hat{y}_i\right)^2}{n-k-1}$$

And at confidence level $1-\alpha$, the confidence interval for y is:

$$\left[\hat{y}_i - t_{\alpha/2}(n-k-1)S\sqrt{x_{jj}}, \hat{y}_i + t_{\alpha/2}(n-k-1)S\sqrt{x_{jj}}\right]$$

Using Matlab, we obtain the following residual graph:

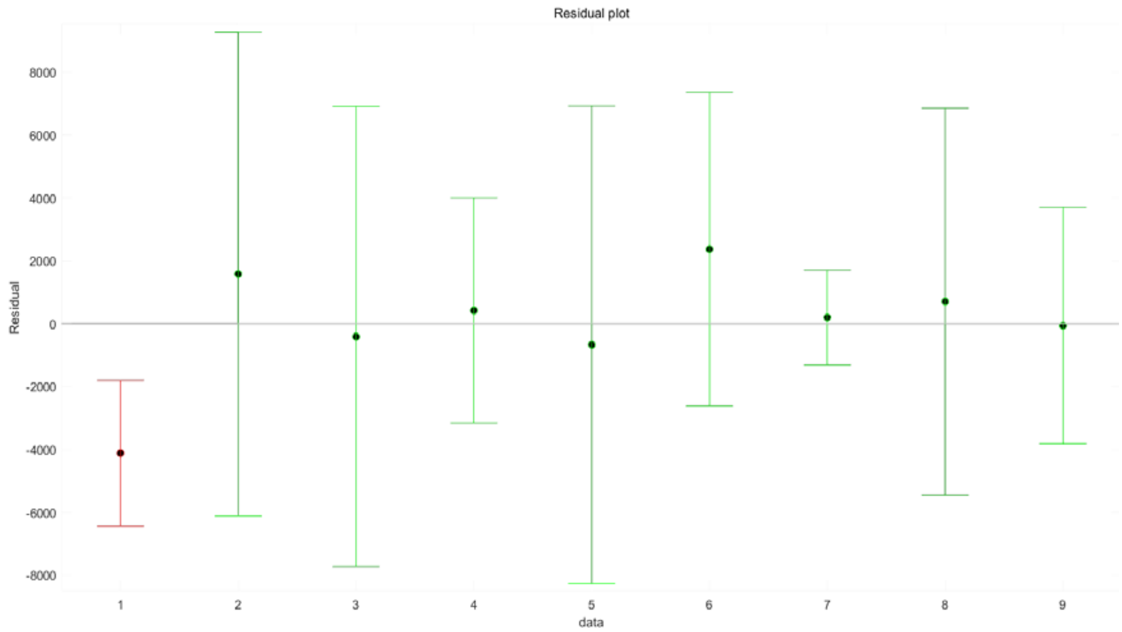

Figure 1. Residual graph

4. Model validity analysis

To elucidate the three-W: whether there is a strong correlation between variables, whether the regression equation is reliable and whether the discipline is trial, we need to perform model testing. Here we select three general and also convincing methods of calibration: Goodness of Fit Test (R^2 Test), Equation Significance Test (F Test) and Variable Significance Test (T Test).

● R^2 Test [3]

Goodness of Fit Test is to test the fitting degree of regression equation to sample observation value. We use R^2 to mirror the degree of fitting as follows:

$$R^2 = 1 - \frac{S_2}{S_1}$$

where $n = 9,\ k = 4$.

From the formula, we can draw a conclusion that the closer R^2 is to 1, the higher the goodness of fit of the regression equation is. In our multivariate regression model , if $R^2 = 0.99$, we then, affirm that there is a high correlation between explanatory variable and explained variable.

●F Test [4]

The Equation's Significance Test is usually used to analyze a statistical model applying multiple parameters to determine whether all or part of the parameters in the model are suitable for estimation.

The null hypothesis H_0 of the test and the alternative opposite hypothesis H_1 are defined as follows:

$H_0 : a_0 = a_1 = a_2 = a_3 = a_4$ H_1 : There exists a not equal to 0

It can be proved that S_2 and S_3 are independent of each other. Furthermore, S_2 and S_3 obey

$$F = \frac{S_3 / k}{S_2 / (n - k - 1)} \sim F(k, n - k - 1)$$

If $F > F_\alpha(k, n - k - 1)$, then we refuse H_0.

In the content, we have $n = 9,\ k = 4$; After computing, we get F=216.1454. At significance level

α=0.05, the range was beyond 0.1041 ~ 9.6045. Hence, we accept H_1. That means the regression equation has a strong linear relation at significance level α=0.05.

● T Test.

For multiple regression models, the overall significance of the equation does not imply that the effect of each explanatory variable on the explained variable y is significant. If an explanatory variable is not significant, it should be removed from the equation to reestablish a simpler equation.

The null hypothesis $H_0(j)$ and the alternative hypothesis $H_1(j)$ of the test are respectively :

$H_0 : a_j = 0$; $H_1 : a_j$ is not equal to 0 for $j = 1,2,3,4,5$

When H_0 is satisfied, we affirm:

$$t_j = \frac{\hat{a}_j / \sqrt{x_{ij}}}{\sqrt{S_2 / (n-k-1)}} \sim t(n-k-1)$$

If $|t_j| > t_{\alpha/2}(n-k-1)$, then, we reject H_0.

In the contest, we have $n-k-1=4$. And $t_{0.025}(4) = 2.776$.

Testing each coefficient separately, we get the consequences below:

$$t_1 = 16.0565 \quad t_2 = -6.1702 \quad t_3 = 19.1732 \quad t_4 = -7.5879$$

It's apparent that when the significance level α is 0.05, the values of all t are out of the range from -2.776 to 2.776. This indicates that the variables selected can significantly affect the value of y.

In summary, our model has passed R^2 Test, F Test and T Test. It strongly suggests that our model is reasonable and has a good imitative effect. And the model can explain the relationship between variables precisely. Thereby, applying our evaluation model, we can help project planners make rational adjustments to land project planning by predicting environmental governance costs in advance.

5. Analysis of typical projects

To verify the generality of the models above, we pick out the specific enterprise project - Binheng Minsheng Cogeneration Project - and make a detailed analysis on its environmental governance costs based on our model.

According to the information found, the thermal power plant is located in the north of Hanting District, Weifang City. According to the collected data, the building area covers 2480 square kilometer, and the total investment is 6731.18 million yuan. Statistically, the plant can supply 5929,000 GJ of heat and 3326,648,000 kWh of power. And the total expected revenue is 7838.76 million yuan [5].

In order to evaluate this project, we first collect relevant information of other land projects, classify them by clustering algorithm, and obtain the pollution level range at each level. Then, after experts evaluate the grade of each indicator in our selected project, the corresponding average pollution level is substituted into formula (1) as approximate estimation value, and get the virtual environment management cost of the project, approximately 122.78 million yuan. According to the environmental assessment report of Binheng Minsheng Cogeneration Project, their estimate of environmental protection investment is 126.44 million yuan [5], which is not much different from the environmental cost value calculated by our model.

6. Grey system prediction model

Mastering a large number of land use project data in recent years, we can make predictions on the parameters of our multiple linear regression model, and then predict the trend of parameter changes in a short period of time. In this way, the cost of projects can be predicted over the next few years.

The prediction methods that can be used are gray prediction, time series prediction, regression prediction, and so on. Here, we use the gray prediction method to predict the trend of the parameters.

We use GM (1, 1) to predict the values of a_0, a_1, a_2, a_3, a_4. Firstly, we define

$$a^{(0)}=\left(a^{(0)}(1), a^{(0)}(2), \ ...\ , a^{(0)}(7)\right)$$

Similarly, we have

$$a^{(1)}=\left(a^{(1)}(1), a^{(1)}(2), \ ...\ , a^{(1)}(n)\right)$$

Where

$$a^{(1)}(k)=\sum_{i=1}^{k} a^{(0)}(i), \quad k=1,2,...,n$$

The mean generation sequence of $a^{(1)}$ is:

$$z^{(1)}=\left(z^{(1)}(1), z^{(1)}(2), \ ...\ , z^{(1)}(n)\right)$$

Among them

$$z^{(1)}(k)=0.5\, a^{(1)}(k)+0.5a^{(1)}(k-1),\ k=2,3,...,n$$

Then we establish a white differential equation:

$$\frac{da^{(1)}}{dt}+aa^{(1)}(t)=b \tag{2}$$

Let $u=[\alpha,\beta]^T$, $Y=\left[\left(a^{(0)}(2), a^{(0)}(3), \ ...\ , a^{(0)}(n)\right)\right]^T$ and $B=\begin{bmatrix} -z^{(1)}(2) & -z^{(1)}(3) & ... & -z^{(1)}(n) \\ 1 & 1 & ... & 1 \end{bmatrix}^T$

Then we use the Least squares, we can get the estimate of u:

$$\hat{u}=\left[\hat{a},\hat{b}\right]^T=\left(B^T B\right)^{-1} B^T Y$$

Bring it in equation (2), we obtain:

$$\hat{a}^{(1)}(k+1)=\left(a^{(0)}-\frac{\hat{\beta}}{\hat{\alpha}}\right)e^{-\hat{\alpha}k}+\frac{\hat{\beta}}{\hat{\alpha}}, \quad k=0,1,...,n-1,...$$

We select the coefficient of x_3 as the representative, use the above formula to predict its value, and get the results shown in the following picture. Where $xp3$ represents the value of the coefficient calculated from the existing data over time, and $x3$ represents the value of the coefficient obtained by the gray prediction over time.

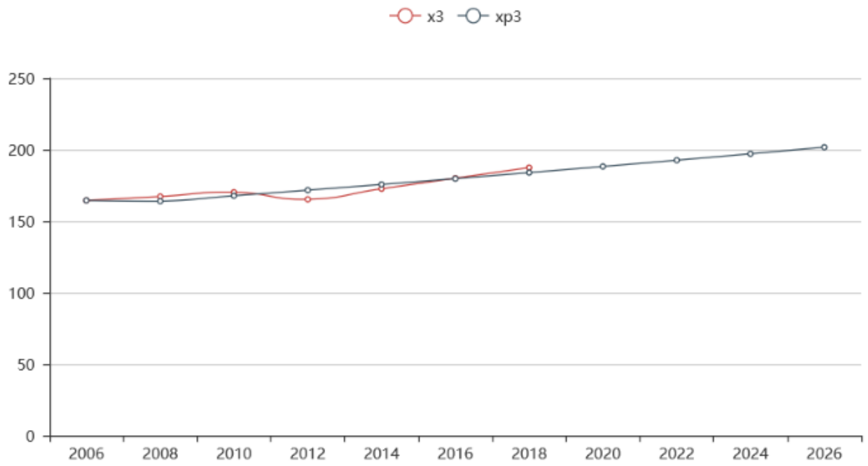

Figure 2. Coefficient Three

It can be seen from the original data that although the coefficient has the tendency to fluctuate, the overall trend is obvious, which is consistent with the predicted results.

7. Conclusion and discussion.

Our model adopts the method of multiple linear regression, which can estimate the cost of environmental governance to a certain extent and provide guidance for project planners. However, the influencing factors in the actual project are variable, and there are many factors influencing the prediction results:

(1) If the influence factor is not considered comprehensively, the constant term of the regression equation will be a variable. It may affect the final prediction.

(2) The influence factors we choose are all linear. If they are non-linear, they must be linearized first and then substituted into the calculation. Otherwise, there will be a great deviation.

(3) An inevitable correlation exists between the chosen indicators inevitably. It can be reduced by using principal component analysis, but can not be completely eliminated.

References

[1] Costanza, Robert. The value of the world's ecosystem services and natural capital [J]. Nature, 1997, (387): 253-260.

[2] Data collected from: http://www.mee.gov.cn/home/rdq/hjyxpj/jsxmhjyx\\pj/xmslqk/

[3] Information about Panyu Mine and Coal Preparation Project in Nalinhe Mining Area, Inner Mongolia, Retrieved from http://hps.mee.gov.cn/jsxm/xmsl/201805/t20180511_439252.shtml

[4] Jingjuan Liu. Multivariate Linear Regression Model Test Method [J]. Journal of Hunan Tax College, 2005, (05): 49-50+60.

[5] Information about Binheng Minsheng Cogeneration Project, Retrieved from: http://www.sdein.gov.cn/zwgk/spq/jsxml_1442/201804/t20180419_1269802.html