Study On the Current Status of Illegal Wildlife Trade Based on ARIMA And Logistic Regression Modelling

Zishuo Liu †, *, Senyuan Ma † and Yanli Zhang †

School of International Education, Hebei University of Technology, Tianjin, China

* Corresponding Author Email: lzs1326709002@163.com

† These authors also contributed equally to this work

Abstract. The purpose of this paper is to explore effective programs to address illegal wildlife trade. With increasing environmental problems, illegal wildlife trade threatens biodiversity and social stability. The government is identified as the client through hierarchical analysis and a comprehensive project program is proposed for a period of five years. The necessity of the project is analyzed through international cooperation, government capacity, human health and biodiversity, and the possible impacts of not implementing the project are demonstrated using ARIMA time series model. Further modeling of expected goals based on logistic regression methods to assess the success of the project. Promote the implementation of the project through cooperation and contribute to solving the problem of illegal wildlife trade. This study provides useful reference and practical guidance for taking effective measures to deal with IWT.

Keywords: Hierarchical analysis; ARIMA; Logistic regression.

1. Introduction

Illegal wildlife trade (IWT) is one of the pressing challenges facing the world today, posing a serious threat to biodiversity and social stability [1]. Despite international treaties such as the Convention on International Trade in Endangered Species of Wild Fauna and Flora (CITES), which were developed to regulate wildlife trade, IWT remains rampant, leading to a dramatic acceleration in the rate of species extinction. It is estimated that up to 25% of tropical forest wildlife species could become extinct by 2020, with nearly one-fifth of existing species facing extinction [2].

Against this backdrop, this paper aims to present a comprehensive 5-year project targeting the IWT problems faced. The government is identified as the client through hierarchical analysis and the need for international cooperation, government capacity, human health and biodiversity is analyzed in depth to show the urgency of the project. The possible impacts of non-implementation of the project are demonstrated through ARIMA time-series modeling to emphasize the need for the government to actively implement the multifaceted 5-year project. Further, the success of the project is assessed by modeling the expected goal assessment through logistic regression method. The purpose of this paper is to provide theoretical support and practical guidance for the development of effective IWT countermeasures, and to promote the government's cooperation with all parties globally to work together to solve the problem of illegal wildlife trade. Under the current unstable global situation, mutual cooperation will be a key factor in promoting the implementation of the project with a view to achieving greater success.

2. Clients Selection Model

2.1. Calculation of Weights – AHP

To build a model to evaluate client choice, we use power (P), resources (R), and interest (I) three indicators. There are several methods to select the weights of indexes. We use AHP method to calculate the weight of each indicator, and by constructing judgement matrix to see the degree of importance [3]. Then we conduct consistency test to ensure the reliability and stability at different conditions.

We use three hierarchies, which are destination layer, criterion layer and project layer re-spectively. Destination layer aims to choose the most suitable client. Criterion layer contains three indicators to evaluate their ability of executing the project. Project layer includes options we can choose. The hierarchy chart is drawn in Fig. 1.

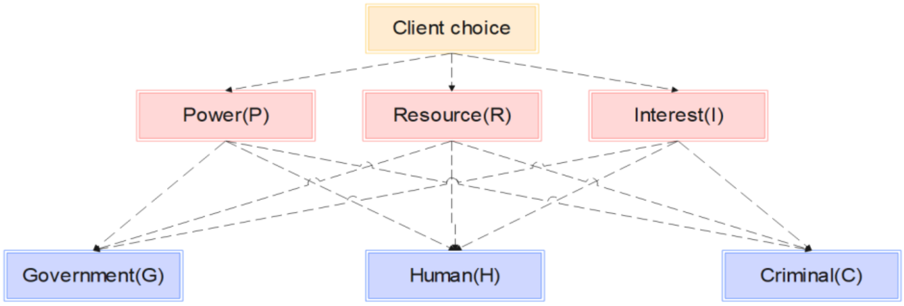

Fig. 1 Hierarchy chart

We construct a 3×3 judgement matrix in Table 1 named A ($A=a_{ij}$), to stand for how important i is compared with j. We also list the importance chart on the right as Table 2 shows.

Table 1. Judgement matrix

	P	R	I
P	1	1	3
R	1	1	2
I	1/3	1/2	1

Table 2. Importance chart

Scale	Meaning
1	Same important
3	Slightly important
5	Obvious important
7	Strong important
9	Extreme important
2,4,6,8	Mid-value

Through the contrast above, we can see that this comparison is a little bit subjective, so we managed to do consistency text to get objective results.

2.2. Consistency Text

Step 1: Calculate consistency index CI.

We first calculate the maximum eigenvalue ($\lambda_{\max}$) of the matrix, then we conduct the consistency text using the following formula:

$$CI=\frac{\lambda_{\max}-n}{n-1} \tag{1}$$

In this formula, CI denotes consistency index, n denotes the order of the matrix.

The table of RI values corresponding to different quantity indicators is shown in Table 3.

Step 2: Search for the corresponding average random consistency index RI.

Table 3. RI corresponding to different quantity indicators

$\boldsymbol{n}$	1	2	3	4	5	6
$\boldsymbol{RI}$	0	0	0.52	0.89	1.12	1.26

Step 3: Calculate consistency ratio CR.

The formula is as follows:

$$CR = \frac{CI}{RI} \begin{cases} CR < 0.1 \\ CR \geqslant 0.1 \end{cases} \tag{2}$$

When CR < 0.1 is set up, we can regard this judgement matrix is consistent. The reason why we choose 0.1 as the judgement condition is according to Monte Carlo simulation, where CR < 0.1, we can get the optimal results.

From the first judgement matrix, we can get CR is 0.0176, for table 5, we can calculate that CR is 0.0516, for table 6, CR is 0.0088, for the table 7, the result of CR is 0.0516, we can know that all CR<0.1.

2.3. Eigenvalue Method

Through the calculations above, we use Eigenvalue Method to get this weight char shown in Table 4, which shows the score each index has.

Table 4. Weights chart

	Weights	***G***	***H***	***C***
P	0.4434	0.5936	0.1571	0.2493
R	0.3874	0.5396	0.2970	0.1634
I	0.1692	0.3108	0.4934	0.1958

We use this formula to get final score:

$$G_j = \sum_{i=1}^{n} \omega_i a_{ij} \tag{3}$$

Where G_j means the j-th clients synthesis score, ω_i denotes the i-th index criterion weight. Finally, according to the formula we can calculate the final score of G is 0.5248, H is 0.2682, C is 0.2070. The score of G is the highest one.

The government have economic management ability, which can provide healthy environment for market, such as public facility and basic support. They have social capacity, providing social support, protecting ecological environment. In addition, they have political ability, aiming to ensure social harmony and maintain stability of policy. So, we select government as our client.

3. Trend Prediction Model

We will estimate if we don't persuade our clients to do the project, then what will happen in the next few years. We are managed to use time series prediction model. We use Auto-regressive Integrated Moving Average (ARIMA) to predict the trend if the government doesn't execute this project what will happen [4]. This model requires the sequence satisfies sta-tionarity. We first choose three variables, including living planet index, global trade value and trade volume of ivory. We select living planet index (LPI) as example to analyze in detailed. Because the living planet index is an authoritative assessment, which can better describe the situation we faced. The other two index use the same method, so we won't do more explana-tion.

3.1. Judge Smooth

Smooth means the mean, variance, and expectations of the sequence do not change sig nificantly. We use Augmented Dickey-Fuller (ADF) test to get the table 6 below.

The critical value 1%, 5%, 10% are the different degrees to refuse the previous assumption compared with ADF tests. When ADF less than 1%, 5%, 10% simultaneously, it shows good rejection of the hypothesis. The essence of differencing order is the latter value subtracts the former value, aiming to eliminate fluctuations of data to get smooth.

Table 5. ADF test table

ADF test table						
Variable	Differencing order	t	p	Critical Value		
				1%	5%	10%
LPI	0	-4.176	0.001	-4.332	-3.233	-2.749
	1	-0.557	0.880	-4.012	-3.104	-2.691

We are aimed to get smooth time series data by analyzing the value of t. If the significance ($P < 0.05$), the previous hypothesis is false, this sequence is a stable time series. If not, that means it is not smooth.

This sequence results show that, based on the variable LPI, when the differencing order is 0, the significance value p is 0.001. the level showed significant. We reject the previous hypothesis, and this sequence is a stable time series. On the contrary, when the differencing order is 1, we can't refuse the previous hypothesis. Through ADF test verifies that our model has passed the stationarity test, and then we build a formula:

$$x_{(t)} = \mu + \sum_{i=1}^{p} \gamma_i x_{t-i} + \sum_{i=1}^{q} \beta_i \varepsilon_{t-i} \tag{4}$$

Where μ means constant, variable γ_i means arL_1 and variable β_i stands for arL_2, $x_{(t)}$ means the future predicted data of LPI, t means time period .

If it is not smooth, we will use Difference Method to make it smooth. The formula is as follows:

$$\Delta x_t = x_{t+1} - x_t \tag{5}$$

Where x_t stands for data of every year, Δx_t means the difference of some certain time corresponding data.

3.2. Judge white noise sequence

We need to calculate the autocorrelation coefficient (ACF) and partial autocorrelation coefficient (PACF) of the series respectively based on the obtained smoothed time series data and check whether it decays to 0 when it is close to zero or whether it decays to 0 at a lag greater than 0 when it decays to zero.

If the ACF of the series is close to zero at non-zero lags and the PACF is close to zero at lags greater than zero, then is a stationary white noise sequence. If not, then it is a stationary non-white noise sequence. We should do other calculations below:

Calculate Autocorrelation Coefficient (ACF):

$$(\alpha_k) = ACF = \frac{\sum_{t=1}^{n=k} (x_t - \overline{x})(x_{t+k} - \overline{x})}{\sum_{t=1}^{n} (x_t - \overline{x})^2} \tag{6}$$

The formula we used to calculate ACF is shown as above, where α_κ means autocorrelation coefficient, k means the interval is k years. Autocorrelation plot of the final differenced data is showed below in Fig.2.

In Fig.2, x-axis stands for number of delays, y-axis means autocorrelation. ACF truncate at order q in the plot, with the PACF exhibiting a tail. ARMA model can be simplified to MA(q) model.

Calculate partial autocorrelation coefficient (PACF):

$$(\beta_k) = PACF = \frac{Cov\,[x_t - \overline{x}_t,\ x_{t-k} - \overline{x}_{t-k}]}{\sqrt{Var(x_t - \overline{x}_t)} \cdot \sqrt{Var(x_{t-k} - \overline{x}_{t-k})}} \tag{7}$$

Then we draw a partial autocorrelation plot of the final differenced data in Fig.3 below.

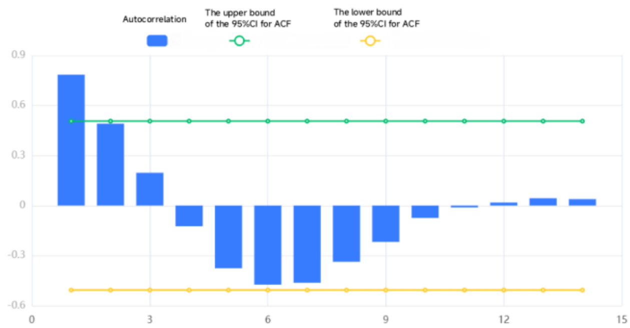

Fig. 2 Autocorrelation of the final differential data

Fig. 3 Partial autocorrelation of the final differenced data

Truncating at order p at PACF figure, with a decaying ACF, the ARMA model can be simplified to an AR(p) model. ARIMA model requires that there is no autocorrelation in residuals, meaning that the model residuals are white noise.

We search the Model Parameter table. We choose the value of information criteria Akaike Information Criterion (AIC) and Bayesian Information Criterion (BIC), used for multiple model analysis and comparison. The formula is as below, θ stands of maximum likelihood estimator, N means number of samples, R^2 means degree of linear fit.

$$AIC = 2N - 2\ln(\theta) \tag{8}$$

$$BIC = \ln(N) - 2\ln(\theta) \tag{9}$$

$$R^2 = 1 - \frac{\sum_{i=1}^{N} (x_i - \hat{x}_i)^2}{\sum_{i=1}^{N} (x_i - \overline{x}_t)^2} \tag{10}$$

Model parameter test, based on the variable LPI and the value of AIC (13.577) and BIC (16.39) and Q-statistic, automatically searching for the optimal parameters, the result of this model is ARIMA (2,1,0). The formula of this model is as follows:

$$x_t = -0.702 + 1.288x_{(t-1)} - 0.051x_{(t-2)} \tag{11}$$

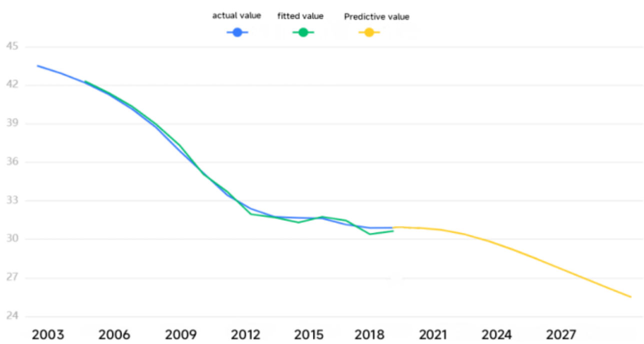

Fig. 4 Living plant index

Fig.4 presenting the data situation of the time series model over the past five years. That implies that the amount of biology is decreasing. After our calculation, LPI is decreased by 10.26%.

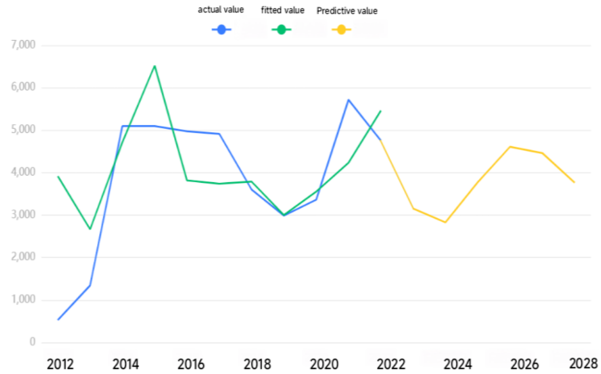

Fig. 5 Global trade value

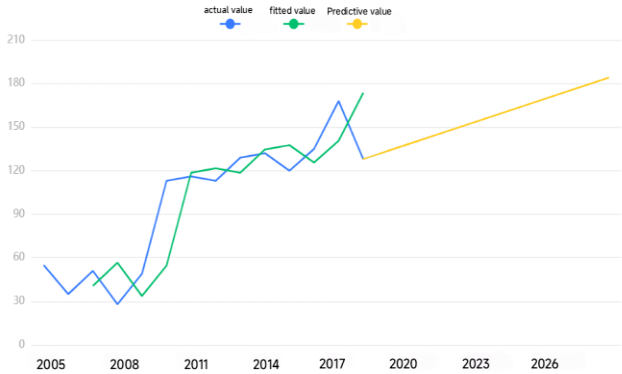

Fig. 6 Trade volume of ivory

Through Fig.5 and Fig. 6, we can know that TV will be increased by 15.16%, TVI will be increased by 19.78%. Based on the variable trade value and trade volume of ivory, we can get the formula as below respectively, which the second formula is constant coefficient formula.

$$x_t = 3917.923 + 0.635x_{(t-1)} - 0.72x_{(t-2)} \tag{12}$$

Among analyzing these figures above, we can know that it is extremely urgent for us to take measures now. LPI is decreasing year by year, some wildlife has no good conditions to live. It affects biodiversity significantly. Trade value is at an increasing speed, especially the volume amount of ivory, reveling more criminal behavior is happening. In a nutshell, the government needs to take actions now to regulate this sever situation and build a bridge between humanity and nature to have a harmonious coexistence.

4. Expected Targets Assessment Model

We use logistic regression prediction method to assess the impact in the future after executing our project [5]. We use this formula (sigmoid function) to calculate the probability according to our project.

Step 1: Calculate probability.

$$h_{\theta(x)} = \frac{1}{1+e^{-(\theta_0+\theta_1 F_0+\theta_2 C_0+\theta_3 A_0)}} \qquad h_{\theta(x)} \epsilon [0,1] \tag{13}$$

Step 2: Model evaluation.

From the formula we know that the success rate is 1, failure rate is 0, and we set a threshold $P(M)$=0.5. Then we conduct likelihood ratio chi-square test (LRCT), we can get the value AIC =52.213, BIC = 62.634. The results of LRCT for the model show a significant P-value of 0.031, likelihood ratio chi-square value is 44.431. It shows high significance, satisfying principles of AIC and BIC and demonstrating the effectiveness of our model.

Based on the results of binary logistic regression, the significance P-value for A_0 is 0.005, the significance P-value for C_0 is 0.016, the significance P-value of F_0 is 0.011, constant is 0.036, which both less than 0.5 and have significant influence on LPI. They all can be the key indicators. With each additional unit, the success rate of LPI will increase greatly. Regression coefficient for A_0, C_0 and F_0 is -1.565, 0.012, 1.576 respectively. Significance is observed at the confidence level, rejecting the previous hypothesis. Therefore, it will have a significant impact on the LPI. OR (odds ratio) value means the ratio of the probability of success to the probability of failure.

Step 3: Getting classification result confusion matrix plot.

True positive (TP) stands for the number of positive cases, false negative (FN) means the number of negative cases. Fig.7 shows the frequency of positive and negative cases of LPI.

Step 4: Evaluate classification result.

We combine true positive rate (TPR) with false positive rate (FPR) to analyze the stability and classification level of our model. We use the following formula to calculate:

$$TPR = \frac{TP}{TP+FN} \tag{14}$$

$$FPR = \frac{FP}{FP+TN} \tag{15}$$

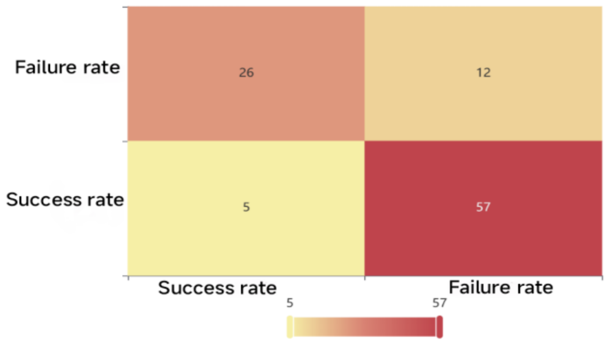

Fig. 7 Classification result confusion matrix plot

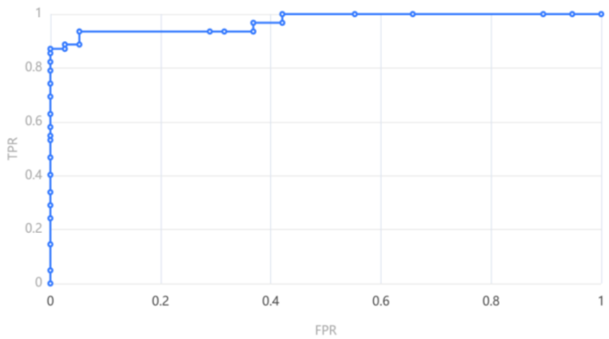

Fig. 8 Receiver operating characteristic (ROC)

We can see from Fig.8 that, when TPR approaches to 1, FPR approaches to 0, the stability and performance of the model is better.

Step 5: Model test.

To reduce the error between sample prediction and real value, we use the following formula (Cross-Entropy Loss Function) to measure the magnitude of errors made by the model, aiming to reduce the discrepancy between predicted values and actual values in the sample.

$$J_\theta = -\frac{1}{m}\left[\sum_{i=1}^{m} y_i \ln\left(h_\theta x_i\right)\right] + (1-y_i)\ln\left(1-h_\theta x_i\right) \tag{16}$$

Where m means sample number, y_i means the true probability of the i-th sample (0 or 1).

Then we use Mini-Batch Gradient Descent method:

$$\Delta\frac{\alpha J_0}{\alpha\theta} = \frac{1}{m}\sum_{i=1}^{m}\left[h_\theta x_i - y_i\right]\cdot x_i \tag{17}$$

We get result that the parameter values of $\theta_0 - \theta_3$ corresponding to the minimum value of J_θ, we can conclude that the model is well-constructed.

Step 6: Draw the conclusion.

$$h_{\theta(x)} = \frac{1}{1+e^{-(-116.895-7.565F_0+0.012C_0+1.574A_0)}} \tag{18}$$

After the calculation, we get the result of probability. From the classification evaluation metrics, we know that Area Under the ROC Curve (AUC) is 0.972 (the closer to 1, the better the classification performance), the success rate of LPI is 85.0%. The success rate of TV, TVI is 71.4%, 78.6% respectively. (Because the calculation method is the same, so we won't represent detail process)

The current international situation is turbulent, and global climate is unstable. We are facing numerous challenges. That will influence the probability and ability of realizing our project. However,

ever country is committed to improving technique level, innovative development, which is an effective basis of executing our project. And there are lots of development projects between each country, which can foster prosperity of development, also be the driver for improving the success rate. The awareness of individuals is increasing as the development of economy. Community can play a significant role on educating and managing them. Each country worldwide has consensus on climate change and calling for global governance. This provides a conducive international environment consensus, contributing to reach our expected goal.

5. Summary

In the face of the daunting challenge of Illegal Wildlife Trade (IWT), our research proposes a comprehensive five-year project to initiate action on this global problem. By identifying governments as clients through a hierarchical approach and analyzing in depth the need for international cooperation, government capacity, human health and biodiversity, we call on governments to act quickly. Our research suggests that if governments fail to act, tropical forest wildlife species may be at greater risk of extinction, while the scale of the illegal trade grows further. Within the framework of this study, we emphasized the key role of the government in implementing the 5-year project and developed a model for evaluating the expected goals through logistic regression methods to assess the success of the project. The results show that the probability of achieving the desired objectives is quite high, which provides a solid basis for the viability of our project. However, we also realize that the turbulence of the global situation may have an impact on the implementation of the project, so mutual cooperation will be one of the key factors to promote the success of the project. In summary, our study provides important theoretical support and practical guidance for solving the problem of illegal wildlife trade.

References

[1] Margulies J D, Bullough L A, Hinsley A, et al. Illegal wildlife trade and the persistence of “plant blindness” [J]. Plants, People, Planet, 2019, 1(3): 173-182.

[2] Ceballos G, Ehrlich P R, Raven P H. Vertebrates on the brink as indicators of biological annihilation and the sixth mass extinction[J]. Proceedings of the National Academy of Sciences, 2020, 117(24): 13596-13602.

[3] Parekh H, Yadav K, Yadav S, et al. Identification and assigning weight of indicator influencing performance of municipal solid waste management using AHP[J]. KSCE Journal of Civil Engineering, 2015, 19(1): 36-45.

[4] Noor T H, Almars A M, Alwateer M, et al. Sarima: a seasonal autoregressive integrated moving average model for crime analysis in Saudi Arabia[J]. Electronics, 2022, 11(23): 3986.

[5] Mohanty R P, Sahoo G, Dasgupta J. Identification of risk factors in globally outsourced software projects using logistic regression and ANN[J]. International Journal of Sup. Chain. Mgt, 2012, 1: 2-11.