Research On the Strategy of Combating Illegal Wildlife Trade Based on Logistic Regression-ARIMA Model

Lanqi Guo 1, #, *, Ruiqi Mao 2, #, Mingrui Li 2, *

1 School of Business Administration, Southwestern University of Finance and Economics, Chengdu, China, 611130
2 School of Accounting, Southwestern University of Finance and Economics, Chengdu, China, 611130
* Corresponding Author Email: 18804379955@163.com
# These authors contributed equally.

Abstract. In view of the continuous growth of the global illegal wildlife trade, this paper proposes a series of methods and strategies to combat the illegal wildlife trade and explores the feasibility and effectiveness of the implementation of the strategy by the United Nations Environment Programme (UNEP) based on logistic regression and ARIMA prediction model. The results of this paper show that the success rate of the strategy implemented by the UNEP is 71.429%, and it will achieve the effect of reducing the global illegal wildlife trade by nearly 5% per year after the implementation of the strategy. The data show that the strategy proposed in this paper can help the UNEP and other environmental regulators to combat illegal wildlife trade and provide evidence for the promotion of the strategy in the future.

Keywords: Illegal Wildlife Trade, Logistic Regression Model, Autoregressive Integrated Moving Average Model.

1. Introduction

Wildlife is a bridge between mankind and the planet and is also a friend of mankind. However, statistics show that the trade in illegal wildlife is increasing year by year, involving as much as $26.5 billion annually, and is ranked as the fourth largest illegal trade in the world. Studies show that the global illegal trade in wildlife may pose a threat to more than 2,200 species of plants and animals, including more than 900 species of amphibians, birds, reptiles, fish, flowering plants and other plants and animals that are not included in the protection of globally endangered species [1]. This will seriously disrupt the ecological balance and food chain, reduce the resilience and self-regulation of the whole system, and have a serious impact on ecosystem service functions such as climate regulation, soil conservation, water purification, etc., and ultimately jeopardize the safety of human beings themselves. In addition to this, about 44 trillion dollars of global economic value output depends on nature and its services [2], and once the natural environment deteriorates, economic growth will be unsustainable. Therefore, research on strategies to combat illegal wildlife trade is an issue that scholars need to pay attention to. Currently, scholars’ research on combating illegal wildlife trade mostly focuses on methods and means but does not assume what expected results will be achieved if the strategies are applied to specific environmental protection organizations. In this paper, we will explore the success rate of the strategy applied to the UNEP and the effect of combating illegal wildlife trade based on the logistic regression model and ARIMA prediction model. The simulation results show that the prediction model is effective and has good reference value for the research of strategies to combat illegal wildlife trade.

2. Strategy

2.1. Strategies to reduce the volume of illegal wildlife trade

1) Cooperate to crack down on illegal trade.
2) Legislative guarantee
3) Sustainable development path
4) technical support
5) publicity and education
The implementation plans:
  • The first year: Invest in research and development; organize the training of relevant law enforcement personnel; Combating illegal wildlife trade (mainly in Asia) through international cooperation, reducing the export value of illegal wildlife by 10% and the import value by 12%.
  • The second year: Send researchers to areas with wildlife trade as the economic pillar to explore other sustainable development routes. Invest in the pharmaceutical industry and start looking for substitutes for wild animals used as medicine; Organize training activities for relevant law enforcement personnel in Asia for 4 times; Expand the area to crack down on illegal wildlife trade to North America, to reduce illegal wildlife exports by 15% and imports by 18%.
  • The third year: Law enforcement officers in all parts of the country are equipped with advanced scientific and technological devices to improve supervision efficiency; Organize 20 community publicity activities in Asia to involve the community widely; We will expand the area to crack down on illegal wildlife trade to the whole world, and the export value of illegal wildlife will be reduced by 20% and the import value will be reduced by 20%.
  • The fourth year: Evaluate the results of the project, improve the imperfections, deepen cooperation among countries in the form of international forums, and reach a consensus on cracking down on illegal wildlife trade. The global import and export of illegal wildlife is expected to decrease by 25%.
  • The fifth year: Summarize the experience and lessons of the project and plan the next five-year plan, which is expected to reduce the global import and export of illegal wildlife by 28%.

Expected results:
Achieve a sharp decline in the trade volume of illegal wildlife within five years; Global species diversity, gene diversity and ecological environment diversity have been effectively protected; The public's awareness of wildlife protection has improved significantly; The contradiction between traditional medicine and biological protection has been alleviated. Internationally, some countries at the border have reached a consensus, and wildlife protection laws are converging, forming a strategic system to combat illegal wildlife trade.

2.2. Data collecting and cleaning
Collecting sufficient data is the basis of developing a complete index system. Data used in this article are mainly collected from the CITES trade database, UNEP website, Our work in data website. The data sources are summarized in Table.1.

<table>
<thead>
<tr>
<th>Database</th>
<th>Website</th>
</tr>
</thead>
<tbody>
<tr>
<td>Our World in Data</td>
<td><a href="https://ourworldindata.org/sdgs/life-on-land">https://ourworldindata.org/sdgs/life-on-land</a></td>
</tr>
<tr>
<td>CITES trade data</td>
<td><a href="https://tradeview.cites.org/en/overview">https://tradeview.cites.org/en/overview</a></td>
</tr>
<tr>
<td>CITES Trade Database</td>
<td><a href="https://trade.cites.org">https://trade.cites.org</a></td>
</tr>
<tr>
<td>UNEP</td>
<td><a href="https://www.unep.org">https://www.unep.org</a></td>
</tr>
<tr>
<td>IUCN</td>
<td><a href="https://www.iucnredlist.org/resources/summary-statistics#Summary%20Tables">https://www.iucnredlist.org/resources/summary-statistics#Summary%20Tables</a></td>
</tr>
</tbody>
</table>

Data pre-processing is divided into two steps: data filling, handling outliers.
• Data Filling: If there is a relatively strong correlation with other years, we use the regression interpolation method, or we use the average value of other years to fill in the missing ones.
• Handling Outliers: We analyzed each indicator and deviated from the abnormal data that may damage the accuracy and efficacy of our models.

2.3. Use Logistic Regression Model to calculate the success probability of the strategy

To estimate the possibility of success of the strategy proposed, this paper uses the logistic regression model to analyze the data of the land scope, investment funds and the number of global patent applications of the United Nations from 1986-2020, so as to reasonably predict the future and obtain the success rate.

2.3.1 Model overview

Logistic Regression is a generalized linear regression analysis model, often used in data mining, automatic disease diagnosis, economic prediction, and other fields[3]. Logical regression estimates the probability of occurrence of events based on a given independent variable data set and, since the result is a probability, the dependent variable ranges between 0 and 1[4]. The basis of the model form is wx + b, where w and b are the parameters to be found. logistic regression corresponds to wx + b to a hidden state p, p =L (wx + b), and then determines the value of the dependent variable according to the size of p and 1-p.

In this paper, the result of the implementation policy is defined as "success" (by 1) or "failure" (by 0) probability greater than 0.5[5]. On the other hand, this model assumes that there is no perfect or high multicollinearity between explanatory variables, because highly correlated explanatory variables may lead to instability in model estimates and interpretation difficulties[6].

2.3.2 Select explanatory variables

The explanatory variable is the main facilitator of UNEP in the project. The larger the explanatory variable, the stronger the UNEP ability, and the higher the probability of successfully completing the project. According to the construction and selection of variables, their principles are scientific, comprehensive, fair, pertinence, legitimacy, operability, etc., and the corresponding evaluation system is constructed accordingly. There are four indicators of the evaluation system, and their description are as Table.2.

<table>
<thead>
<tr>
<th>Variable quantity</th>
<th>Reference point</th>
<th>Characterization</th>
</tr>
</thead>
<tbody>
<tr>
<td>UNEP self-capability indicator (X1)</td>
<td>Number of projects successfully completed in the fight against illegal wildlife trade</td>
<td>The larger the X1, the higher the probability of success</td>
</tr>
<tr>
<td>Technical innovation index (X2)</td>
<td>There is the number of technological achievements on the fight against illegal wildlife trade</td>
<td>Additional manpower allocated for similar projects</td>
</tr>
<tr>
<td>Human resource index (X3)</td>
<td>Additional manpower allocated for similar projects</td>
<td>Additional manpower allocated for similar projects</td>
</tr>
<tr>
<td>Financial support indicators (X4)</td>
<td>Additional manpower allocated for similar projects</td>
<td>Additional manpower allocated for similar projects</td>
</tr>
</tbody>
</table>

2.3.3 Model formula (estimate model parameters using MLE)

The logistic regression model was established through the data from 1986-2020, and the success probability of project implementation is Pi and the probability of failure is 1-Pi, where i represents the first year i (i=1986,1987... 2020). The correlation coefficient of the index in the regression equation is fitted by the variables, and the model is:

$$
\log \left( \frac{P(Y=1|x)}{1-P(Y=1|x)} \right) = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_4 X_4
$$
Symbol description:
• Y binary dependent variable
• X_1, X_2... X_k is a set of explanatory variables
• P (Y=1|X) The probability of project success given the explanatory variable X.

\[ P(Y = 1 \mid X) = \frac{e^{\beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_4 X_4}}{1 + e^{\beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_4 X_4}} \]  

Symbol description:
• Logit (P) Log odds of success probability
• \( \beta_0, \beta_1 \ldots \beta_k \) model parameters

2.3.4 model analysis

Using the above formula, the success rate of the UNEP implementation plan in 1986-2020 is calculated, and then mapped between 0-1 by mapping. We can conclude that the closer the P value is to 1, the greater the success probability of the implementation plan in that year; the closer the P value is to 0, the smaller the success probability of the implementation plan in that year. The results are shown in the Table.3. and Figure 1.

Table.3. Schematic diagram of the success probability
dependent variable| option| frequency| percentage (%) |
---|---|---|---|
Success / failure| 1.0| 15| 71.429 |
| 0.0| 6| 28.571 |
Sum| 21| 100 |

Figure 1. Success probability map

2.3.5 model evaluation

The model was tested by likelihood ratio chi-squared test and analyzed for likelihood test ratio significance. In the Table.4.

Table.4. Significance analysis of the likelihood-ratio test

<table>
<thead>
<tr>
<th>Likelihood ratio chi square values</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td>10.227</td>
<td>0.005***</td>
</tr>
</tbody>
</table>

Note: ***, ** and * represent the significance levels of 1%, 5% and 10%, respectively.
• The results of the likelihood ratio chi-square test of the model show that the significance P-value is 0.005 ** *, showing significance at the level, rejecting the null hypothesis, and thus the model is valid.

2.3.6 Model conclusion

It is predicted that if the UNEP implements our program in the future, the probability of success is 71.429%. Given the unpredictable future changes in international politics, economic environment and national relations, this leads to a low probability of success. However, as can be seen from the figure above, with the growth of time, the probability of success keeps rising, with the average annual average stable at 1 from 2018-2020. The reason may be continued economic development, growing awareness of biodiversity conservation, and increasing capacity of UNEP.

3. Autoregressive Integrated Moving Average Model

3.1. Model introduction

ARIMA (p, d, q) model is called the differential autoregressive moving average model, which is a famous time series prediction method proposed by Box and Jenkins in the early 1970s[7]. AR is autoregressive, p is the autoregressive term, MA is the average moving, q is the number of moving average terms, and d is the difference times done when the time series becomes smooth[8].

The basic idea of the model is: the data sequence formed by the predicted object over time is regarded as a random sequence, and a certain mathematical model is used to approximately describe the sequence. Once the model is identified, it can predict the future value from the past value and current value of the time series.

3.2. Model assumptions

1. Stationarity assumption: the statistical properties of the time series model do not change with time.[9]
2. Linear assumption: ARIMA model can be predicted by linear equations, which means that the current value and future value in the time series can be predicted by a linear combination of the past error terms.
3. Independent and identical distribution of error terms: it is assumed that the error terms in the time series are independent and identical, and there is no autocorrelation between the error terms at each time point.
4. Model parameter stability: it is assumed that the model parameters are constant during the data observation period.[10]

3.3. Stationarity test of time series

The data we chose come from the World Illegal Wildlife Seizure Database of the United Nations Office on Drugs and Crime from 1999 to 2021. Draw the time series graph of the sequence, as shown in the figure, and judge its stationarity preliminarily. It can be seen from the Figure 2 that although the number of illegal wildlife seizures fluctuates, the overall trend is still rising. Then use ADF to test the stationarity of the original sequence, and the result is p = 0.270. Therefore, the original assumption of non-stationarity of the sequence cannot be rejected, so the sequence can be determined to be non-stationary.
3.4. Data preprocessing

The original data is non-stationary. The ADF test is used to test the stationarity of the first order and second-order difference sequences. The test results are shown in the table below. The results of the sequence test show that, based on the variable UN Office on Drugs and Crime World Illegal Wildlife Seizure Database, when the difference is first-order, the significance P value is 0.000***, which is significant at the level, rejecting the null hypothesis, and the sequence is a stationary time series (as shown in the Figure 3 and 4). When the difference is second-order, the null hypothesis can also be rejected, as shown in the Table 5.

<table>
<thead>
<tr>
<th>Variable</th>
<th>Difference Order</th>
<th>t</th>
<th>P</th>
<th>AIC</th>
<th>Cut-off Value</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1%</td>
</tr>
<tr>
<td>Global wildlife trade</td>
<td>0</td>
<td>-2.038</td>
<td>0.270</td>
<td>941.854</td>
<td>-3.621</td>
</tr>
<tr>
<td></td>
<td>1</td>
<td>-7.107</td>
<td>0.000***</td>
<td>914.133</td>
<td>-3.627</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>-3.441</td>
<td>0.010***</td>
<td>881.122</td>
<td>-3.7</td>
</tr>
</tbody>
</table>

Note: ***, **, and * represent the significance levels of 1%, 5%, and 10%, respectively.

3.5. Selection and fitting of ARIMA model.

To establish a correct time series model, the order of the model is first known. The method is to judge through the autocorrelation and partial autocorrelation graphs after the first-order difference method. The test results of autocorrelation and partial autocorrelation: the autocorrelation coefficients are shown in the Figure 5 and 6.
Figure 5. Sample Autocorrelation Function

Figure 6. Sample Partial Autocorrelation Function

Note: the green line represents the upper bound of the ACF95% confidence interval, and the yellow line represents the lower bound of the ACF95% confidence interval.

According to the ACF and PACF graphs, the system automatically finds the optimal parameters based on the AIC information criterion (the lower the better), and determines that the ARIMA model is ARIMA (0,1,0). The Table. 6. of model parameters is shown below.

Table 6. ARIMA model (0,1,0) test table

<table>
<thead>
<tr>
<th>item</th>
<th>symbol</th>
<th>value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Df Residuals</td>
<td>N</td>
<td>36</td>
</tr>
<tr>
<td>Q statistic</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Q6</td>
<td>1.671(0.196)</td>
<td></td>
</tr>
<tr>
<td>Q12</td>
<td>3.64(0.725)</td>
<td></td>
</tr>
<tr>
<td>Q18</td>
<td>15.232(0.229)</td>
<td></td>
</tr>
<tr>
<td>Q24</td>
<td>20.684(0.296)</td>
<td></td>
</tr>
<tr>
<td>Q30</td>
<td>23.074(0.515)</td>
<td></td>
</tr>
<tr>
<td>Information guideline</td>
<td></td>
<td></td>
</tr>
<tr>
<td>AIC</td>
<td>1289.179</td>
<td></td>
</tr>
<tr>
<td>BIC</td>
<td>1292.4</td>
<td></td>
</tr>
<tr>
<td>Goodness of fit</td>
<td>R²</td>
<td>0.686</td>
</tr>
</tbody>
</table>

Note: ***, **, and * represent the significance levels of 1%, 5%, and 10%, respectively

- ARIMA model requires that the residual error of the model does not have autocorrelation, that is, the residual error of the model is white noise. Check the model test table, and test the white noise of the model according to the P value of Q statistic (P value greater than 0.1 is white noise). The model Q6 does not show significance at the horizontal level, and cannot reject the hypothesis that the residual error of the model is a white noise sequence.
- $R^2$ represents the degree of fitting of the time series, and the closer it is to 1, the better the effect. The model $R^2$ is 0.686, and the model performs well, basically meeting the requirements.
3.6. Significance test of the model

The residual autocorrelation coefficient and partial autocorrelation coefficient of the model are calculated as follows:

As shown in the Figure 7, the correlation coefficients are all within the dotted line, and the residual error of the autoregressive model (AR) is a white noise sequence, meeting the requirements. As shown in the Figure 8, the correlation coefficients are basically within the dotted line, and the residual error of the moving average model (MA) is a white noise sequence, meeting the requirements of the time series.

The system automatically finds the optimal parameters according to the AIC information standard based on the variables of global wildlife trade volumes. The results are obtained through the test table of ARIMA model (0,1,0): \( y(t)=483812.703 \).

The model formula is as follows

\[
(1 - \sum_{i=1}^{p} \alpha_i L^i)(1 - L)^d y_t = \alpha_0 + (1 + \sum_{j=1}^{q} \beta_j L^j)\varepsilon_t
\]  

3.7. Model prediction

Scenario 1 (no intervention): Assuming project isn’t implemented in the next five years, the current trend continues, and illegal wildlife trade volume in the next five years is showed in Figure 9.

Scenario 2 (intervention): Based on the expected project impact, the model parameters are adjusted to reflect the expected intervention effect.

The intervention analysis model is introduced, and the intervention variable \( D_t \) is a dummy variable, which represents the implementation of our project by the United Nations Environment Program. The ARIMA model is extended to predict the illegal wildlife trade volume in the next ten years.

The extended model is:

\[
(1 - \sum_{i=1}^{p} \varphi_i L^i)(1 - L)^d Y_t = \left(1 + \sum_{j=1}^{q} \theta_j L^j\right)\varepsilon_t + \delta D_t
\]  

\( Y_t \) is the time series data that takes into account the intervention effect, and \( \delta \) is the intervention effect coefficient we want to estimate. Then, the model parameters and the intervention effect coefficient are estimated by the MLE method. The estimated value of the intervention effect
coefficient $\delta$ represents the immediate change of the time series after the implementation of the intervention. (If $\delta$ is significantly different from 0, it can be considered that the intervention has a significant impact. A positive value indicates that the intervention leads to an increase in the time series, and a negative value indicates a decrease. The result is $\delta$: $7.6882 \times 10^{-10}$.

The amount of illegal wildlife trade is shown in the following Table 7. Besides, the impact of implementing strategies on the volume of illegal wildlife trade is shown in Figure 10.

Table 7. Prediction of illegal wildlife trade volume

<table>
<thead>
<tr>
<th>Time</th>
<th>2022</th>
<th>2023</th>
<th>2024</th>
<th>2025</th>
<th>2026</th>
</tr>
</thead>
<tbody>
<tr>
<td>Trade</td>
<td>43802993</td>
<td>42177423</td>
<td>40551853</td>
<td>39739068</td>
<td>38926283</td>
</tr>
</tbody>
</table>

Figure 10. The impact of the implementation of the project on the volume of trade

According to the above figure, it can be seen that for illegal wildlife trade, the implemented project is always lower than the unimplemented project, and with the change of time, the gap between the two is becoming larger and larger. The amount of illegal wildlife trade is decreasing year by year. Our project has a significant effect.

4. Conclusions

This paper establishes a prediction model based on logistic regression and ARIMA for the problem of illegal wildlife trade and carries out prediction simulation using existing historical data. Firstly, this paper comprehensively considers the possibility of UNEP implementing the strategy and achieving the expected goal and calculates the success rate of strategy implementation as 71.429% by establishing a logistic regression model. Secondly, this paper establishes an ARIMA model to predict the expected impact of UNEP's implementation of the strategy, and the calculated data shows that after the implementation of the strategy, the global trade in illegal wildlife will shrink by nearly 5%. Therefore, based on the results of this paper, it is feasible for UNEP to implement the strategy proposed in this paper and the impact of the strategy on illegal wildlife trade in the future is significant.

However, there is still room for improvement of the model in this paper. The accuracy of the model's predictions depends on the quality of the input data, and errors and incompleteness in the historical data may interfere with the accuracy of the output results. In addition, the influencing variables in the model in this paper also have certain limitation. If we want to conduct a more in-depth exploration of the implementation effect of the strategy, we need to take into account the future changes in the international situation, policy changes, economic direction and other factors, so as to make the prediction results more accurate, thus making the prediction results more informative.
References


