# The Exploration and Establishment of an Economic Operation Monitoring and Early Warning Model Driven by Big Data

Xiaoping Zhang [1, *], Yang Lei [2], Wancai He [3], Liu Chang [1], Yang Peng [2], Yingying Jiang [2], ZhengMing Kou [4]

[1] Yinchuan City Company of Ningxia Hui Autonomous Region Tobacco Company

[2] China National Tobacco Corporation Ningxia Hui Autonomous Region Company

[3] Wuzhong City Company of Ningxia Hui Autonomous Region Tobacco Company

[4] Shizuishan Branch of Ningxia Hui Autonomous Region Tobacco Company

* Corresponding Author Email: klnh92@126.com

**Abstract.** The tobacco industry is a crucial component of the national economy. Enhancing the foresight, relevance, and effectiveness of macroeconomic regulation to ensure that the cigarette economy operates within a reasonable range is vital for improving macroeconomic alignment and stabilizing its contribution. This study proposes a monitoring index system for the cigarette economy. The SARIMA model is applied to forecast key indicators scientifically, while algorithms like the BP neural network and the prosperity index establish a big data-driven early warning system for cigarette economic operations. This system identifies real-time trends, detects abnormal fluctuations, shifts from reactive regulation to proactive prevention, and provides data-driven support for management decisions.

**Keywords:** Economic operation of cigarettes; Forecast and early warning; SArima model; Prosperity analysis; back propagation.

## 1. Introduction

As the world's largest tobacco producer and consumer, China's tobacco industry's development is closely tied to macroeconomic trends. The industry must accurately position itself within the broader economic landscape, skillfully manage the dialectical relationship between "stability and progress," and continually enhance the stability and sustainability of its economic operations. Ensuring the economy operates within a reasonable range is both a fundamental requirement for high-quality industrial development and a reflection of its responsibility to stabilize and contribute comprehensively. At present, China's economic recovery remains at a critical stage, with positive macroeconomic factors increasing, and the national economy continuing its upward trajectory. However, the ongoing impact of insufficient macroeconomic demand and weak consumption poses significant challenges to the tobacco industry. Local markets exhibiting "two highs" (high prices and high inventory) remain sluggish, making it increasingly difficult to balance short-term growth stabilization with long-term sustainable development. Ensuring the stable operation of the industry faces considerable pressure. This study aims to establish a big data-driven economic monitoring and early warning model to conduct pre-analysis, forecasting, and judgment of cigarette economic operations. The model will enable real-time monitoring of economic trends and abnormal fluctuations during operations, providing data-driven insights and theoretical guidance for accurately assessing market conditions and effectively managing emerging, directional, and trend-related challenges.

## 2. Research Background and Significance

### 2.1. Research background

Following the "total control, slightly tight balance, reasonable growth rate, and sustained development" policy outlined by the State Administration, tobacco enterprises must take proactive steps to ensure the economy operates within a reasonable range. These efforts should emphasize the continuous advancement of brand-driven strategies and steady progression toward high-quality enterprise development. This project aims to establish a cigarette economic operation monitoring and early warning analysis model driven by big data. The model is expected to digitally represent trends across dimensions such as macroeconomics, cigarette operations, and economic development. It will enable comprehensive evaluations of economic trends over the coming years at the regional, municipal, and local levels, achieving the goal of effective monitoring and early warning. Ultimately, it will assist decision-makers in formulating strategies for the future development of cigarette operations.

### 2.2. Overview of domestic and foreign research

Economic operation early warning systems encompass macroeconomic and microeconomic dimensions. Macroeconomic monitoring and early warning systems serve as "barometers" or "alarms" for economic trends by leveraging a series of economic indicators. As early as the 20th century, renowned American economic statistician Moore proposed the "Diffusion Index" theory, which significantly advanced the field of macroeconomic monitoring and early warning. International research on corporate early warning systems primarily adopts empirical methods, with a particular focus on financial risk prediction. In China, research on macroeconomic early warning systems began in the mid-1980s. Early studies often built upon foreign models, refining them to better suit domestic contexts.

As modernization progresses, the tobacco industry has gradually conducted related research on monitoring and early warning systems.

Despite these advancements, most tobacco industry research on monitoring and early warning remains focused on proactive prevention, such as combating illegal cigarette outflow. However, there is a notable lack of studies addressing the monitoring and early warning of trends in cigarette economic operations and development. This gap underscores the need for comprehensive research and innovation in this area.

## 3. Main research content and methods

This study closely adheres to the "16-character policy." It focuses on achieving the goal of modernized tobacco management by identifying key indicators that influence the economic dynamics of the cigarette market across dimensions of macroeconomics, cigarette business, and economic development. The study employs scientific classification methods, such as cluster analysis and factor loading, to categorize these indicators. It calculates their correlations and degrees of influence on economic operations. By combining seasonal adjustment coefficients with the adaptive ARIMA model, it eliminates the impact of seasonal factors, enabling more accurate forecasting of all indicators to align with real-world scenarios. On this foundation, algorithms like BP neural networks and the prosperity index are applied to determine leading, coincident, and lagging indicators. A big data-driven economic operation monitoring and early warning model is constructed to monitor and detect abnormalities in cigarette economic operations. The model also analyzes and predicts changes in the macro market and Ningxia's tobacco economic operations. This system provides a mathematical foundation and reference model for cigarette business decision-making, ensuring the positive development of cigarette economic operations and enhancing stability and growth in the market.

### 3.1. Establishing a monitoring and early warning indicator system for the operation of the cigarette economy

This paper based on the importance, scientificity and rationality of the construction of the indicator system, focuses on achieving the goal of modern tobacco management, and selects statistical indicators that complement and relate to each other from the dimensions of macroeconomics, cigarette operations, and economic development, and can fully characterize the economic operation of the cigarette market, reflecting the economic operation of cigarettes from different aspects and angles.

### 3.1.1. Identify indicators.

In light of the characteristics of the economic operation of the tobacco industry, focused on the three dimensions of macroeconomics, cigarette operations, and economic development, according to the three categories of total indicators, ratio indicators, and speed indicators, and following the principles of accessible indicator data, consistent data cycles, and standardization, 15 indicators that affect the economic operation of the cigarette market were selected from the 25 indicators.

**Table 1.** Index system of the forecast and early warning model for the economic operation of the cigarette market.

| Serial number | Dimensions | category | index |
|---|---|---|---|
| 1 | Macroeconomics | Total indicators | Resident population |
| 2 | | | Gross Regional Product |
| 3 | | | Total retail sales of consumer goods |
| 4 | | | Consumer Price Index |
| 5 | | | Per capita disposable income |
| 6 | Cigarette business | Total indicators | Cigarette sales |
| 7 | | | Cigarette sales revenue |
| 8 | | | Cigarette box sales |
| 9 | | | Social Inventory |
| 10 | | Ratio Indicator | Price Index |
| 11 | | | Inventory-to-sales ratio |
| 12 | Economic Development | Speed Index | Tax profit growth rate |
| 13 | | | Cigarette sales growth rate |
| 14 | | | Cigarette sales revenue growth rate |
| 15 | | | Cigarette box sales growth rate |

### 3.1.2. Data collection.

Combining 15 indicators from three dimensions, data are collected from four aspects: ① Through querying websites or contacting the Statistics Bureau, collect macroeconomic indicators of Ningxia, local companies and county-level companies in the past ten years; ② Through the industry business decision-making system, collect cigarette sales indicators of Ningxia, local companies and county-level companies in the past ten years (months); ③ Through the collection of consumer data, stratified sampling surveys and statistics on smoking rate, per capita smoking volume and other data, and finally calculate the interval value of the cigarette market capacity in the whole region. ④ through the market information collection system, collect economic development indicators of Ningxia, local companies and county-level companies in the past ten years (months).

### 3.1.3. Indicator classification.

The 11 - year data from 2012 to 2022 were used for exploratory factor analysis (EFA), and the complex variables were integrated into a few core factors through dimensionality reduction technology.

**Data preprocessing**

(a) Data conversion. Convert annual growth indicators such as tax and profit growth rate, cigarette sales growth rate, and cigarette single sales growth rate into monthly growth rates.

Annual growth rate = (1 + monthly growth rate) ^12-1

Monthly growth rate = (1 + annual growth rate) ^ (1/12) - 1

(b) Data standardization: Perform Z-standardization on the 14 columns of indicator data to convert data of different values into Z-Scores of unified measurement.

After treatment X = (before treatment X - mean) / standard deviation

**Exploratory factor analysis**

The factor analysis automatically extracted 4 factors. The factor loading coefficients after rotation are shown in Table 2.

**Table 2.** Factor loading coefficient table after rotation.

| name | Factor loading coefficient | | | | Common factor variance |
|---|---|---|---|---|---|
| | Factor 1 | Factor 2 | Factor 3 | Factor 4 | |
| S_Total population | 0.963 | -0.160 | 0.067 | 0.005 | 0.957 |
| S_Gross Regional Product | 0.977 | 0.065 | 0.075 | 0.133 | 0.981 |
| S_Total retail sales of consumer goods | 0.899 | -0.301 | 0.065 | -0.155 | 0.927 |
| S_Per capita disposable income of urban residents | 0.988 | -0.036 | 0.072 | 0.015 | 0.982 |
| S_Stock | 0.610 | -0.374 | 0.399 | -0.202 | 0.712 |
| S_Stock -to-sales ratio | 0.519 | -0.532 | 0.202 | -0.398 | 0.752 |
| S_Price Index | 0.369 | 0.653 | -0.087 | 0.350 | 0.693 |
| S_Tax profit growth rate | -0.402 | 0.866 | 0.019 | -0.118 | 0.925 |
| S_Single box sales growth rate | -0.177 | 0.946 | 0.038 | 0.027 | 0.928 |
| S_Single box sales | 0.391 | -0.135 | 0.759 | -0.331 | 0.857 |
| S_Sales | -0.095 | 0.023 | 0.966 | 0.076 | 0.949 |
| S_Sales amount | 0.131 | 0.024 | 0.968 | 0.031 | 0.956 |
| S_Consumer Price Index | -0.233 | -0.125 | -0.026 | 0.868 | 0.824 |
| S_Sales Growth Rate | 0.240 | 0.272 | -0.004 | 0.792 | 0.758 |

Among them, the research items inventory and macroeconomic indicators correspond to factor 1, and the consumer price index and sales growth rate correspond to factor 4, which is inconsistent with the research expectations. After deleting the inventory and consumer price index, the factor analysis was repeated. Three factors were automatically extracted. The factor loading coefficients after rotation are shown in Table 3. The commonality values (common factor variance) corresponding to all research items are higher than 0.4, indicating that there is a strong correlation between the research items and the factors, and the factors can effectively extract information.

**Table 3.** Factor loading coefficient table after rotation.

| name | Factor loading coefficient | | | Common factor variance |
|---|---|---|---|---|
| | Factor 1 | Factor 2 | Factor 3 | |
| S_Total population | 0.977 | -0.030 | 0.086 | 0.964 |
| S_Gross Regional Product | 0.961 | 0.229 | 0.092 | 0.983 |
| S_Total retail sales of consumer goods | 0.929 | -0.234 | 0.095 | 0.927 |
| S_Per capita disposable income of urban residents | 0.980 | 0.091 | 0.097 | 0.978 |
| S_Stock -to-sales ratio | 0.562 | -0.585 | 0.219 | 0.705 |
| S_Price Index | 0.272 | 0.785 | -0.082 | 0.696 |
| S_Tax profit growth rate | -0.549 | 0.681 | 0.083 | 0.772 |
| S_Sales Growth Rate | 0.246 | 0.611 | -0.080 | 0.441 |
| S_Single box sales growth rate | -0.325 | 0.836 | 0.101 | 0.815 |
| S_Single box sales | 0.384 | -0.234 | 0.802 | 0.845 |
| S_Sales | -0.101 | 0.031 | 0.941 | 0.896 |
| S_Sales amount | 0.125 | 0.034 | 0.963 | 0.944 |

According to the results of factor analysis, the 12 indicators are divided into three dimensions, which can be summarized as macroeconomics, cigarette business and economic development. A cigarette economic operation forecast and early warning indicator system is constructed with cigarette economic operation forecast and early warning as the target layer, 3 first-level indicators and 12 second-level indicators, as shown in Table 4.

**Table 4.** Cigarette economic operation forecast and early warning indicator system.

| Target level | First level indicator | Secondary indicators |
|---|---|---|
| Cigarette economic operation forecast and early warning indicator system | Macroeconomics | Total population |
| | | Gross Regional Product |
| | | Total retail sales of consumer goods |
| | | Per capita disposable income of urban residents |
| | Economic Development | Inventory-to-sales ratio |
| | | Price Index |
| | | Tax profit growth rate |
| | | Sales growth rate |
| | | Sales growth rate per box |
| | Cigarette business | Sales per box |
| | | Sales |
| | | Sales Amount |

### 3.1.4. Determine the indicator weights.

The weights are calculated using factor loading coefficient information, as shown in Table 5.

Linear combination coefficient = loading matrix / Sqrt (eigen)

Load factor divided by the square root of the corresponding characteristic root

Comprehensive score coefficient = cumulative (linear combination coefficient * variance explanation rate) / cumulative variance explanation rate

That is, the linear combination coefficients are multiplied by the variance explanation rate, added together, and then divided by the cumulative variance explanation rate; the comprehensive score coefficients are normalized to obtain the weight values of each indicator .

**Table 5.** Linear combination coefficients and weight results.

| name | Factor 1 | Factor 2 | Factor 3 | Comprehensive score coefficient | Weight |
|---|---|---|---|---|---|
| Characteristic root (after rotation) | 4.732 | 2.667 | 2.569 | | |
| Variance Explained | 39.43% | 22.23% | 21.41% | | |
| S_Total population | 0.4494 | -0.0182 | 0.0535 | 0.2222 | 11.89% |
| S_Gross Regional Product | 0.4416 | 0.1400 | 0.0574 | 0.2619 | 14.01% |
| S_Total retail sales of consumer goods | 0.4272 | -0.1433 | 0.0593 | 0.1797 | 9.61% |
| S_Per capita disposable income of urban residents | 0.4505 | 0.0556 | 0.0604 | 0.2443 | 13.07% |
| S_Stock -to-sales ratio | 0.2582 | -0.3579 | 0.1369 | 0.0621 | 3.32% |
| S_Price Index | 0.1249 | 0.4805 | -0.0511 | 0.1747 | 9.35% |
| S_Single box sales | 0.1764 | -0.1432 | 0.5004 | 0.1744 | 9.33% |
| S_Tax profit growth rate | -0.2523 | 0.4169 | 0.0517 | 0.0051 | 0.27% |
| S_Sales Growth Rate | 0.1130 | 0.3744 | -0.0501 | 0.1409 | 7.54% |
| S_Single box sales growth rate | -0.1495 | 0.5120 | 0.0629 | 0.0823 | 4.40% |
| S_Sales | -0.0465 | 0.0187 | 0.5870 | 0.1342 | 7.18% |
| S_Sales amount | 0.0573 | 0.0205 | 0.6010 | 0.1876 | 10.03% |

### 3.2. Eliminating the influence of special factors

Almost all monthly or quarterly economic time series contain seasonal factors, which will affect the accuracy of measuring economic cycle fluctuations, this paper uses multiple methods of cluster analysis and the seasonal autoregressive difference moving average model (SARIMA) method to predict 15 key indicator data, and conducts comparative analysis based on the prediction error values. It is finally concluded that the SARIMA model is more suitable for the prediction of the key indicator data of this project and has higher accuracy.

#### 3.2.1. SArima model analysis method

The SARIMA model explicitly supports univariate time series with seasonal components, and can model and predict time series with seasonal cyclical changes.

Taking cigarette sales as an example, based on the historical data of cigarette sales from 2011 to 2020, the SARIMA model is used to predict the sales volume in 2021 and compared with the actual value.

After comparing the prediction results with the cluster analysis method, the prediction results directly analyzed by the SArima model have smaller errors. Through analysis, it is found that the SArima model has already taken into account the trend, periodicity and irregularity of the data, and it is finally determined to use the SArima model to predict and analyze future data to eliminate the influence of special factors.

#### 3.2.2. Apply the SARIMA model to realize single indicator data prediction.

Focusing on the sales volume data of 11 years from January 2011 to December 2021, the first 10 years of data are imported as simulated time series data, the SARIMA model is generated on the data training using the optimal parameters finally determined, and the generated forecast values are compared with the actual reset values. Iterative methods such as gradient descent or Newton-Raphson method are used to gradually adjust the parameter values to find the optimal parameter combination.

(1) Find the best parameters. By continuously running the data to find the best parameter values, the annual sales volume, monthly sales volume, and regional sales volume data errors are required to be relatively small, that is, to meet the needs of different time periods and different regions, and determine the optimal model with the smallest error.

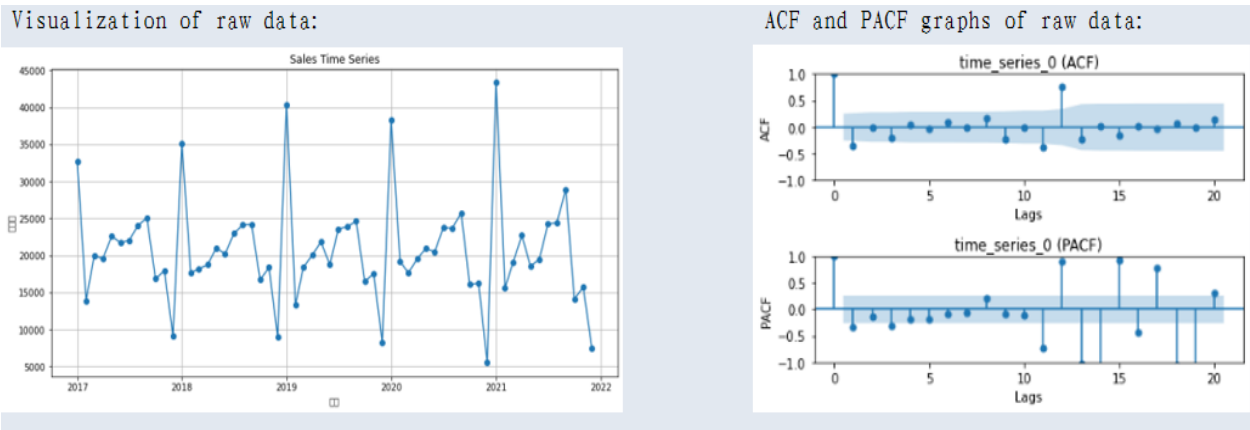



**Figure 1.** Visualization of raw data.

It can be seen from the graph analysis that: the original data has seasonal changes, and the period is S=12; the overall periodicity of the original data shows a slight increasing trend, and it is initially judged that the original data shows weak stationarity; it is suitable for short-term forecasting, but not suitable for long-term forecasting; the SARIMA model is suitable Forecast demand.

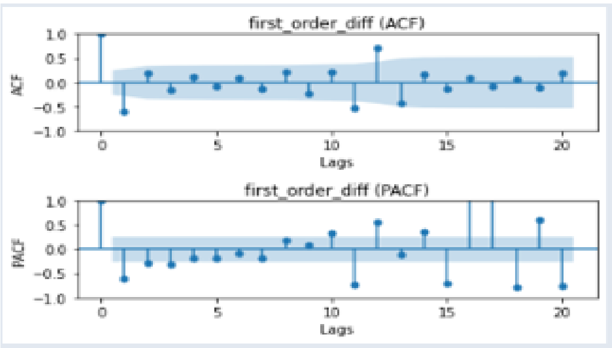
(2) Stability detection. Use the PDF test method in the unit root test to detect stationarity. The test statistic ADF Statistic value is greater than the critical value of the confidence level of 1% and 5%.It can be seen that the original data tends to be weakly stationary in trend and seasonality, and does not meet the assumptions of the SARIMA model, requiring further differentiation.

**Table 6.** PDF verification method.

| Portable Document Format | |
|---|---|
| Critical Values | |
| 1% | −3. 5745892596209488 |
| 5% | −2. 9239543084490744 |
| 10% | −2. 6000391840277777 |
| p−value | 0. 05 |
| raw data | |
| ADF Statistic | −2. 352922148228983 |
| p−value | 0. 15546681682849312 |

(3) First order difference


**Figure 2.** First_order difference visualization chart.

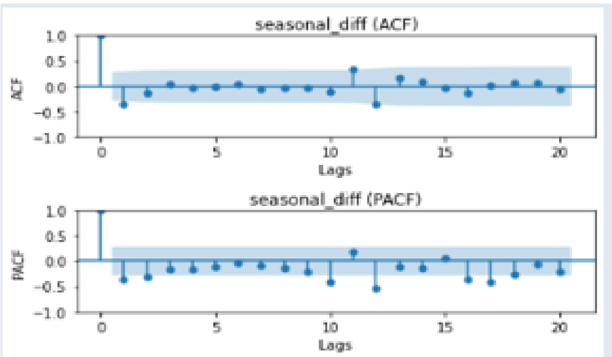



| PDF Test method | |
|---|---|
| Critical Values | |
| 1% | -3.5745892596209488 |
| 5% | -2.9239543084490744 |
| 10% | -2.6000391840277777 |
| P-Value | 0.05 |
| First-order difference data | |
| ADF Statistic | -7.048342555589031 |
| P-Value | 5.610357977576743e-10 |

**Figure 3.** First order differential ACF and PACF graphs.

From the first-order difference visualization chart, it can be seen that the data after the first-order difference are concentrated up and down a line, with no trend change, and the data is stable, which satisfies the SARIMA model assumption.

(4) Seasonal difference




| PDF Test method | |
|---|---|
| Critical Values | |
| 1% | -3.5745892596209488 |
| 5% | -2.9239543084490744 |
| 10% | -2.6000391840277777 |
| P-Value | 0.05 |
| Seasonal differencing data | |
| ADF Statistic | -6.145344938840606 |
| P-Value | 7.791591054183054e-08 |

**Figure 4.** Seasonal difference ACF and PACF graphs.


**Figure 5.** Seasonal difference visualization chart.

It can be seen that the data after seasonal difference show strong stationarity and meet the assumptions of the SARIMA model.

(5) Observe the ACF and PACF graphs to determine the initial parameters.

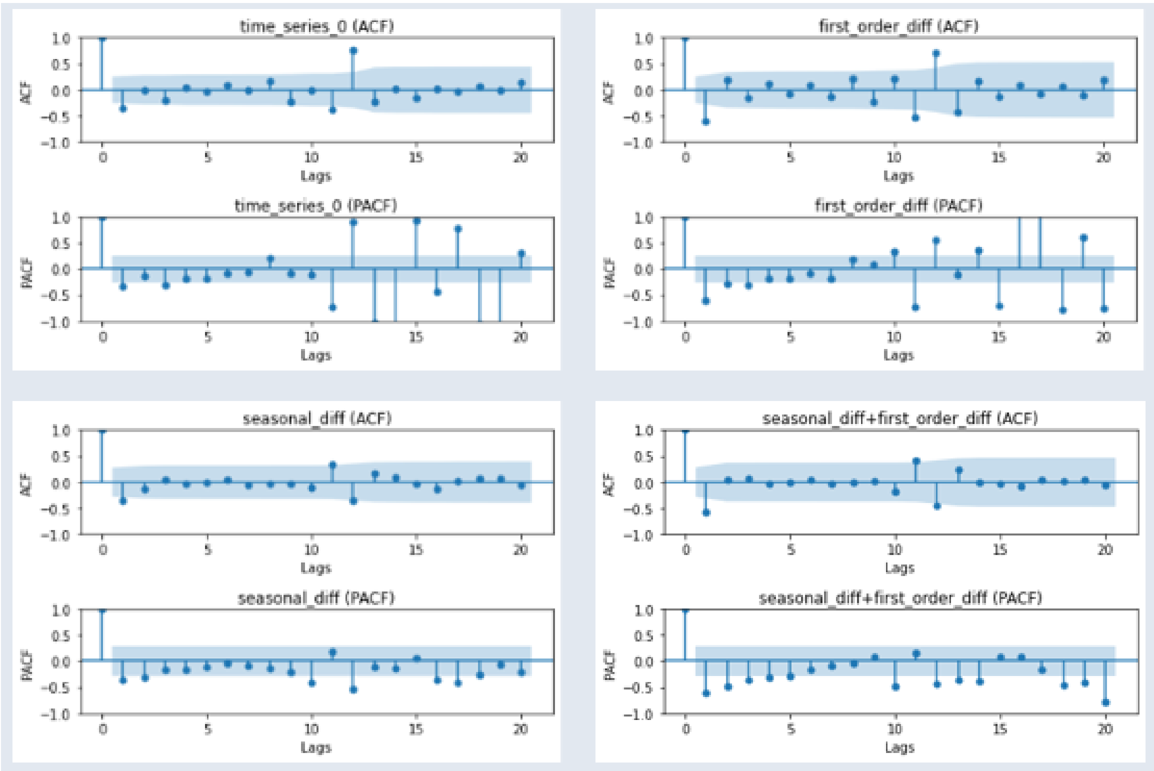



**Figure 6.** ACF and PACF graphs of dates.

Through the stationarity test, it can be seen that the original data are strongly stationary after one first-order difference or one seasonal difference, which meets the assumptions of the SARIMA model. Observe the ACF graph of the original data and we can see that truncation occurs every 12th lag, so S=12.

(6) Grid search + information criterion. According to the analysis of ACF and PACF graphs, the range of p, q, P, and Q is within 3, and the range of d and D is 0 or 1. Iterating within this range, it is finally concluded that SARIMAX (2, 1, and 2) x (0, 1, 1, and 12) produces the lowest AIC value of 12.0. Combined with the observations of ACF and PACF graphs, it can be considered that this is the best choice among all the models considered.

(7) Model parameter evaluation.

**Table 7.** SARIMA coefficient table.

SARIMAX Results

| | | | |
|---|---|---|---|
| Dep. Variable: | sales volume | No. Observations: | 60 |
| Model: | SARIMAX(2, 1, 2)x(0, 1, [1], 12) | Log Likelihood | -429.817 |
| Date: | Wed, 23 Aug 2023 | AIC | 871.634 |
| Time: | 10:42:50 | BIC | 882.735 |
| Sample: | 01-01-2017 - 12-01-2021 | HQIC | 875.812 |
| Covariance Type: | opg | | |

The maximum likelihood estimation was used to evaluate the optimal parameters that were finally determined, indicating that the model fitted by this group of parameters had a general fit quality for the data, and the model parameters needed to be further optimized.

(8) Model diagnosis

**Table 8.** Model diagnosis results.

| | coef | std err | z | P>\|z\| | [0.025 | 0.975] |
|---|---|---|---|---|---|---|
| ar.L1 | 0.4500 | 0.138 | 3.267 | 0.001 | 0.180 | 0.720 |
| ar.L2 | -0.2485 | 0.151 | -1.642 | 0.101 | -0.545 | 0.048 |
| ma.L1 | -1.4457 | 0.156 | -9.274 | 0.000 | -1.751 | -1.140 |
| ma.L2 | 1.0000 | 0.185 | 5.407 | 0.000 | 0.638 | 1.362 |
| ma.S.L12 | -0.0922 | 0.125 | -0.740 | 0.459 | -0.336 | 0.152 |
| sigma2 | 3.939e+06 | 7.37e-08 | 5.35e+13 | 0.000 | 3.94e+06 | 3.94e+06 |

By comparing the Ljung -Box (L1) (Q) value, Jarque-Bera (JB) value, Prob (Q) and Prob (JB) value, Heteroskedasticity (H) value, Skew (skewness), Kurtosis (kurtosis) value After comprehensive analysis of residual characteristics and fitness, the residuals of the model are basically consistent with the model assumptions, making the model suitable for fitting the data to a certain extent.

(9) Model prediction. Through the evaluation of model parameters and the results of model diagnosis, it can be seen that SARIMAX (2, 1, and 2) x (0, 1, 1, and 12) has a general degree of fit with historical data.

Through model parameter evaluation and model diagnostic results analysis, the impact of ar.L2 (autoregressive coefficient 2) and ma.S.L12 (seasonal moving average coefficient) is not significant, indicating that p=2 and Q=1 are not optimal. The parameters need to be optimized and adjusted. Combined with the previous data analysis, the maximum likelihood estimation is further used to optimize the parameters. Finally, the best model is SARIMA(0, 1, 2)x(0, 1, 2, 12).Observing the forecast results, we can see that the forecast results in the first few months are close to the actual values, with small errors, and the errors in the full-year forecast results are also small.

Based on the above analysis, based on the sales data for 11 years from January 2011 to December 2021, the model suitable for the sales forecast of the whole region is finally concluded as SARIMAX(0, 1, 2)x(0, 1, 2, 12).

### 3.2.3. Predict future data.

The indicator data in the three dimensions of macroeconomics, cigarette business, and economic development are gradually substituted into the SArima analysis method to find the best parameters, build the best model, and predict future data values one by one.

### 3.2.4. Parameter model verification effect.

Through verification and analysis of several years of actual data and prediction model data, the error value is very small, indicating that the single indicator prediction model has a high accuracy.

### 3.3. Constructing an economic operation monitoring and early warning model

In the early stage, three pre-selected economic operation early warning models were selected by searching literature materials: BP neural network model, ARIMA model, and prosperity index model. After literature analysis, discussion and research, two pre-selected models were determined: one is the BP neural network model that uses the prosperity index as the output indicator data, and the other is the ARIMA model that uses the prosperity index as the forecast indicator data.

#### 3.3.1. Analyze the impact of macro indicators on cigarette economic operation indicators.

Annual and monthly data are used to analyze the impact of macro indicators on cigarette economic operating indicators.

(1) Use annual data. Perform the Pearson correlation test to find out the indicators that have the greatest impact on the two indicators of sales amount (10,000 yuan) and sales volume (boxes). Using the stepwise regression method, we get:

Sales amount = -63369.970 + 904.827*Per capita disposable income of urban residents - 11290.694*Gross regional product_tertiary industry. The independent variables in the model have no significant impact.

(2) Use monthly data. Perform a Pearson correlation test to find the indicator with the greatest impact on the two indicators of sales amount (10,000 yuan) and sales volume (boxes). Use the stepwise regression method to obtain:

Sales amount = 18599.113 + 15.920*per capita disposable income of urban residents , which means that for every 1 yuan increase in per capita disposable income of urban residents per month, the sales amount of the entire district will increase by an average of 18,615.033 yuan per month.

#### 3.3.2. Analysis of the economic operation of cigarettes.

This paper adopts the internationally accepted prosperity index method to conduct a prosperity analysis on the economic operation of the entire Ningxia region, and comprehensively evaluates the status of the cigarette economic operation by calculating the diffusion index and the composite index.

In order to construct a prosperity index that describes the operation effect of the cigarette economy, we first used the grey correlation method to analyze the correlation between 15 key indicators and the price index, and divided them into leading indicators, consistent indicators and lagging indicators. Then, we used the structural equation model (SEM) method to estimate the influence coefficients of these indicators on the prosperity index. Finally, through annual and monthly data, we applied these coefficients to the corresponding indicators and calculated the annual and monthly prosperity indexes to describe the operation status and development trend of the cigarette economy in the corresponding period.

(1) Annual data is used.

a. Based on GDP, the leading, consistent and lagging indicators are divided. The division results are shown in Table 6:

**Table 9.** Statistics of leading indicators, coincident indicators, and lagging indicators.

| index | K=-1 | K=0 | K=1 | type |
|---|---|---|---|---|
| sales amount | 0. 968 | 0. 951 | 0. 923 | advance |
| Single box sales | 0. 932 | 0. 926 | 0. 911 | advance |
| price index | 0. 626 | 0. 566 | 0. 303 | advance |
| Per capita disposable income of urban residents | 0. 982 | 0. 985 | 0. 984 | consistent |
| sales volume | 0. 183 | 0. 344 | 0. 138 | consistent |
| Total number of retail stores | 0. 954 | 0. 965 | 0. 962 | consistent |
| Inventory-to-sales ratio | 0. 720 | 0. 752 | 0. 716 | consistent |
| tax profit growth rate | -0. 480 | -0. 54 | -0. 6 | consistent |
| Gross Regional Product_Tertiary Industry | 0. 981 | 0. 986 | 0. 986 | consistent |
| sales growth rate | 0. 6 | 0. 6 | 0. 51 | consistent |
| total population | 0. 913 | 0. 937 | 0. 952 | lag |
| Total retail sales of consumer goods | -0. 136 | 0. 046 | 0. 868 | lag |
| consumer price index | -0. 004 | -0. 346 | -0. 429 | lag |
| Single box sales growth rate | -0. 225 | -0. 364 | -0. 484 | lag |
| Growth rate of total number of retail households | 0. 239 | 0. 171 | 0. 287 | lag |

Benchmark: Gross Regional Product

After the indicators are standardized, they are put into the structural equation SEM model to find the fitting relationship:

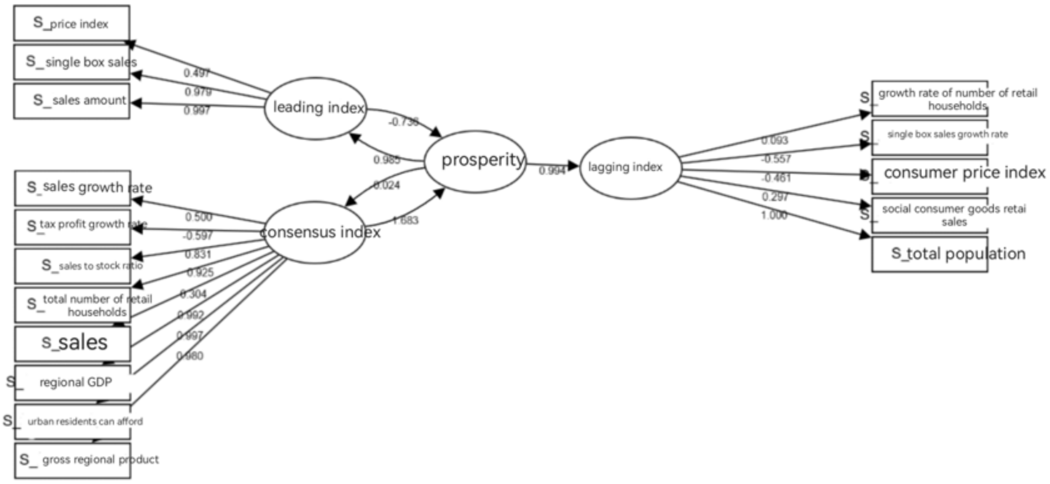



**Figure 7.** SEM model coefficient calculation results.

Calculate the annual prosperity index:

**Table 10.** The range of prosperity index from 2011 to 2023.

| | | | |
|---|---|---|---|
| prosperity index | expect | | 414433. 9948 |
| | standard deviation | | 17002. 8943 |
| | | | |
| normal | 397431. 1006 | | 431436. 8891 |
| Hot | 431436. 8891 | | 448439. 7834 |
| Cold | 380428. 2063 | | 397431. 1006 |

**Table 11.** Calculation of the prosperity index values from 2011 to 2023

| years | Leading index | consensus index | lagging index | prosperity index | degree of prosperity |
|---|---|---|---|---|---|
| 2011 | 30075. 8923 | 3812. 969032 | 409655. 172 | 391478. 6112 | Cold |
| 2012 | 36732. 74 | 4052. 363772 | 415608. 2018 | 392899. 3842 | Cold |
| 2013 | 40335. 94938 | 4145. 441277 | 421164. 2913 | 395926. 8244 | Cold |
| 2014 | 44294. 95736 | 4263. 099985 | 426830. 1619 | 398842. 8896 | normal |
| 2015 | 48488. 84863 | 4163. 696946 | 431671. 4989 | 400401. 1792 | normal |
| 2016 | 46047. 55095 | 4000. 088358 | 437046. 2252 | 407265. 099 | normal |
| 2017 | 48381. 59557 | 4032. 041759 | 457352. 3275 | 425785. 2854 | normal |
| 2018 | 49765. 18976 | 4058. 736825 | 460962. 9555 | 428400. 8522 | normal |
| 2019 | 51025. 46492 | 4088. 988361 | 466349. 9666 | 432878. 8921 | Hot |
| 2020 | 52164. 61201 | 4099. 750204 | 467840. 8438 | 433540. 5239 | Hot |
| 2021 | 57468. 72096 | 4218. 569935 | 470757. 7605 | 432736. 0885 | Hot |
| 2022 | 62528. 4869 | 4299. 012222 | 472682. 5898 | 431060. 7654 | normal |
| 2023 | 64414. 38216 | 4350. 40392 | 459268. 4034 | 416425. 5375 | normal |

By calculating the prosperity index values from 2012 to 2021, it can be clearly seen that the cigarette economy has been steadily growing every year. The prosperity index from 2011 to 2013 was relatively low, and the cigarette economy was in a state of undercooling. The economic prosperity index was relatively high from 2019 to 2021, and the cigarette economy was in a relatively hot state, consistent with the trend of enterprise economic development, indicating that the monitoring and warning model is consistent with reality.

b. based on the price index, the indicators are divided into leading, coincident and lagging indicators.

After the indicators are standardized, they are put into the structural equation SEM model to find the fitting relationship:

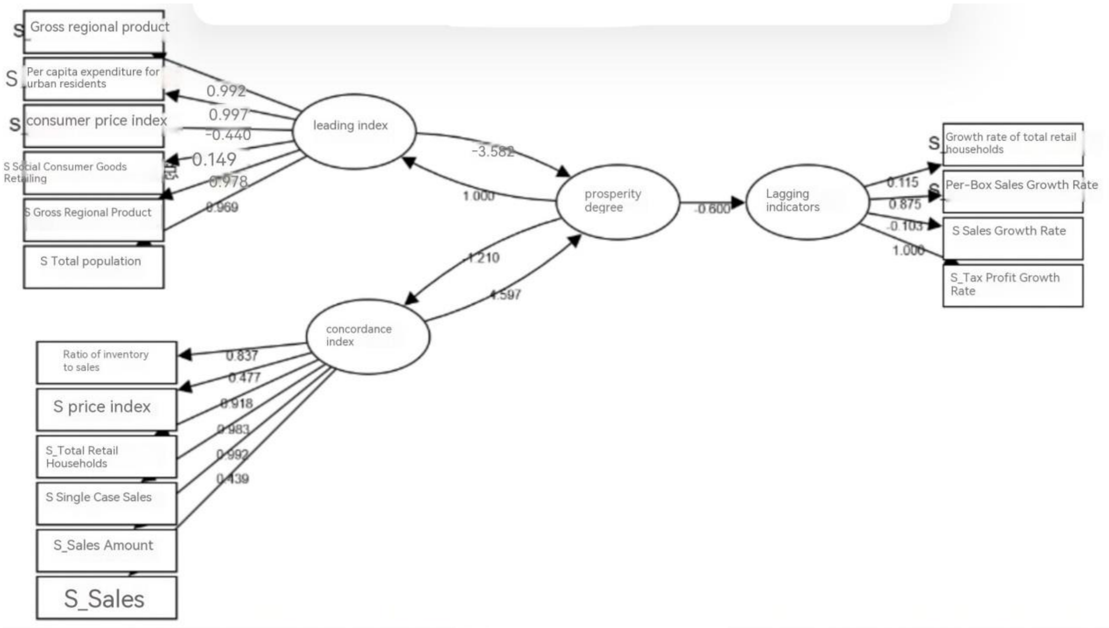



**Figure 8.** SEM model coefficient calculation results.

**Table 12.** The range of prosperity index from 2011 to 2023.

| | | |
|---|---|---|
| prosperity index | expect | 359516.5854 |
| | standard deviation | 12177.0568 |
| | | |
| normal | 347339.5285 | 371693.6422 |
| Hot | 371693.6422 | 383870.6991 |
| Cold | 335162.4717 | 347339.5285 |

**Table 13.** Calculation of the prosperity index values from 2011 to 2023

| years | Leading index | consensus index | lagging index | prosperity index | Prosperity range |
|---|---|---|---|---|---|
| 2011 | 393888.1622 | 37005.33314 | 0.001591555 | 349111.7081 | normal |
| 2012 | 399166.2283 | 44035.35384 | 0.0018417 | 345883.449 | Cold |
| 2013 | 404027.9094 | 47705.45917 | 0.001073911 | 346304.3032 | Cold |
| 2014 | 409043.7483 | 51815.41497 | 0.001115371 | 346347.0955 | Cold |
| 2015 | 413350.7051 | 55888.96485 | 0.00203325 | 345725.0564 | Cold |
| 2016 | 418125.9652 | 53258.95908 | -6.01679E-05 | 353682.6248 | normal |
| 2017 | 437215.4402 | 55607.39957 | 0.000535514 | 369930.4864 | normal |
| 2018 | 440355.2929 | 57029.67554 | 0.000282966 | 371349.3853 | normal |
| 2019 | 445251.0992 | 58289.18502 | 0.000337875 | 374721.1851 | Hot |
| 2020 | 447185.9286 | 59496.30234 | 0.000337685 | 375195.4025 | Hot |
| 2021 | 449865.1969 | 65124.3329 | 0.000764004 | 371064.7537 | normal |
| 2022 | 451728.1121 | 70304.70887 | 0.000268608 | 366659.4143 | normal |
| 2023 | 445307.3844 | 72369.12266 | 0.000601644 | 357740.7456 | normal |

By calculating the prosperity index values from 2012 to 2021, it can be clearly seen that the cigarette economy has been steadily growing every year. The prosperity index from 2012 to 2015 was relatively low, and the cigarette economy was in a state of undercooling. The economic prosperity index was relatively high from 2019 to 2020, and the cigarette economy was in a relatively hot state, consistent with the trend of enterprise economic development, indicating that the monitoring and warning model is consistent with reality.

### 3.3.3. Divide the cigarette economic operation warning range.

(1) Single indicator warning intervals.

In order to achieve a comprehensive evaluation of the economic development trends of Ningxia, municipal bureaus (companies) and regions, it is necessary to quantify the five economic interval values of "economic overheating, hot, general, cold, and too cold", and clarify the warning interval and balance interval. This paper adopts the quartile method to divide the historical data of 15 single indicators and the predicted data for the next three years, a total of 15 years (2011-2025), and obtains the five economic warning intervals of "economic overheating, hot, general, cold, and too cold", and then judges the economic development trend by analyzing the interval of the predicted data results. The specific quartile function division formula is:

Lower limit=Q1-1.5(Q3-Q1), Q1= value (0.25position), Q3=value (0.75position)

Upper limit = Q3 + 1.5 (Q3 - Q1)

(2) Single indicator warning intervals.

This paper adopts the normal distribution principle. When the prosperity index value is far from the expected value, $\mu$the closer it is, the greater its probability. On the contrary , when the prosperity index is farther from the expected value , $\mu$the smaller the probability., Combined with the actual situation of Ningxia tobacco, the standard deviation area within 2 times is the normal range. It can be seen that the five state intervals of Ningxia cigarette economic operation warning are "overcooled", "cold", "normal", "hot", and "overheated".

**Table 14.** The state intervals of Ningxia cigarette economic operation warning of 2012-2021.

| Warning range | Overcooling | cold | generally | hot | overheat |
|---|---|---|---|---|---|
| 2012 | | | √ | | |
| 2013 | | √ | | | |
| 2014 | | √ | | | |
| 2015 | | √ | | | |
| 2016 | √ | | | | |
| 2017 | | | √ | | |
| 2018 | | | | | √ |
| 2019 | | | | √ | |
| 2020 | | | | √ | |
| 2021 | | | | √ | |

## 4. Conclusion

### 4.1. Main conclusions

This article builds a cigarette economic operation monitoring and early warning model and verifies it based on actual data from 2011 to 2021. It can understand the operating status of the cigarette economy in real time, correctly grasp the development and operation trend of the cigarette economy, and provide timely early warning and reminders for abnormal fluctuations in the operation of the cigarette economy, thereby realizing the transformation from post-control to pre-prevention, providing a mathematical basis and reference model for cigarette business decision-making, and ensuring the stable, healthy and sustainable development of the cigarette economy.

### 4.2. Outlook for deficiencies

#### 4.2.1. Data collection is not comprehensive and accurate enough.

The cigarette market is changing rapidly, and there are many factors that affect the operation and development of the cigarette economy. How to comprehensively find these influencing factors and divide them by dimension is a difficult point. At the same time, facing the 15 early warning indicators we have sorted out, some of the related data need to be obtained by coordinating with relevant units and departments, some need to invite statistical technicians to collect statistics based on social reality, and some need to be obtained with the help of databases. How to collect data one by one accurately is a major challenge.

#### 4.2.2. The conditions for realizing information-based and intelligent operation are not yet sufficient.

The development of future systems will involve research and development technologies such as front-end development, back-end development, database development, algorithm compilation, and web page development, which will also involve certain difficulties. In the next step, if the data of the integrated platform cannot be connected, other digital platforms such as Yida, Qingliu, and QUICK BI will be used to develop an economic operation monitoring and early warning system to realize online data collection, evaluation and analysis, real-time monitoring, system early warning, improvement and promotion, etc., steadily improve the efficiency of economic operation statistics, forecasting, monitoring, and early warning management, and gradually improve the level of economic operation management.

## References

[1] Liu Xiao. Research on Urban Logistics Prosperity Monitoring and Early Warning System [D]. Beijing Jiaotong University, 2010.

[2] Zheng Guihuan, Zhang Xun, et al. Economic Prosperity Analysis Methods [M]. Beijing: Science Press. 2011: 4-5.

[3] Xie Xianfen, Model construction and monitoring and early warning research of economic operation index [D]. Jinan University, 2016.

[4] Ma Xiaoyu, Zhang Huifen, Liu Zongjie, et al. Temperature prediction model of electrical equipment based on variable weight combination [J]. Control Engineering, 2023.

[5] Wang Taiji, Liu Weiyi, Li Zhuyu. Comparison and empirical test of dynamic regression models for financial interval data [J]. Statistics and Decision, 2011.

[6] Yue Li, GAO Xincai, Zhang Qinzhi. Evaluation of regional circular economy development based on entropy method: A case study of Gansu Province [J]. Soft Science, 2011, 25(10): 74-77

[7] Xu Bingbing, Cen Keting, Huang Junjie, et al. A review of graph convolutional neural networks [J]. Chinese Journal of Computers, 2020, 43(5):26. DOI: 10.11897 /SP.J.1016.2020.00755.

[8] Xiao Bin Ding. Study on the Forecast Model of Water Quantity on the Basis of BP Artificial Neural Network [J]. Applied Mechanics and Materials, 2014, 2911 (475).

[9] Zhu Weidong, Sun Yibo, Wu Yong, et al. Research on early warning model of cigarette retailers selling counterfeit cigarettes based on neural network [J]. Journal of Tobacco Science, 2014, 20 (6): 127-132.

[10] XIAO Y. Anomaly detection of cigarette sales using ARI-MA on lunar calendar [J]. Journal of Chmical&Pharma-ceutical Research, 2014, 6 (5): 266-270.

[11] Automatic and Generic Periodicity Adaptation for KPI Anomaly Detection [J].Zhao Nengwen,Zhu Jing, Wang Yao,Network and Service Management, IEEE Transactions on.2019, Issue 3

[12] Xiao Xiao, Feng Pengcheng, Liu Luni, et al. Early warning of illegal tobacco sales based on time series prediction and anomaly detection [J]. Journal of Guizhou Normal University (Natural Science Edition), 2023, 41(3): 119-124.