

Time Series Analysis of Air Quality Index (AQI) in Beijing from 2014 to 2021

Yuqi Zhang

School of Statistics and Mathematics, Central University of Finance and Economics, Beijing, China
2019310900@email.cufe.edu.cn

Abstract. Air problems seriously affect people's physical health, quality of life and happiness in life. Beijing as the capital of China, air quality problems should be paid more attention to. This paper analyzes the monthly data (AQI values) of air pollution in Beijing from 2014 to 2021, discusses the trend effects and seasonal effects of AQI data, constructs the ARIMA multiplication model with the highest degree of matching, and predicts the future air quality of Beijing. The results show that the air quality in Beijing from 2014 to 2021 will improve year by year, and the air quality will be better in the next three years. This paper also analyzes whether the improvement of air quality in Beijing is relevant to the document promulgated in 2018 whose name is "Beijing's Three-Year Action Plan to Win the Blue Sky Defence War", and whether the impact is sustainable. The results show that the promulgation of the document has effectively improved the air quality in Beijing without delay effect, and we can reasonably predict that the air quality will improve further on the basis of this impact in 2022-2024.

Keywords: ARIMA multiplication model, On-stationary series, Intervention analysis.

1. Introduction

Beijing, as the capital of China, attracts more and more people to live here, and young people regard staying in Beijing to work and live as the embodiment of their ability and value. Meanwhile, with the increasing resident population, Beijing bears growing pressure on its air quality. Therefore, air pollution control of great significance cannot be ignored.

Intertwined with people's health, life quality, and happiness, air pollution is indispensable to the construction of ecological civilization in China. Only by solving environmental problems can we improve the environment and people's livelihood as well as develop an economy characterized by high quality. In addition, we can enhance the image and credibility of the government and lay the foundation for China's prosperity.

On September 12, 2013, the State Council issued the Action Plan on Air Pollution Prevention and Control, which was positively echoed by Beijing Clean Air Action Plan from 2013 to 2017 released at the same time. In September 2018, the Thirteenth Five-Year Plan was put forward by the Three-Year Action Plan for Winning the War to Protect Blue Skies, which learned and summarized the experience of the Beijing Clean Air Action Plan from 2013 to 2017 to highly improve the air quality. With the support of this series of policies, Beijing has achieved a lot in air pollution control.

2. Description and Analysis of AQI in Beijing

Firstly, the data on air pollution in Beijing from 2014 to 2021 sourced from the China Air Quality Online Monitoring Platform (<https://www.aqistudy.cn/>) are analyzed by time series mapping. The data includes daily AQI values from 2014 to 2021 and air pollutants indicators such as PM_{2.5}, PM₁₀, SO₂, CO, and NO₂. In this data analysis, the daily AQI values are taken as monthly average AQI values, and the monthly data of AQI values from 2014 to 2021 is adopted as an analysis sequence, a relatively classical time series.

Draw the annual calendar charts of the AQI value of Beijing in 2014 and 2021 respectively, as shown in Figures 1 and 2.

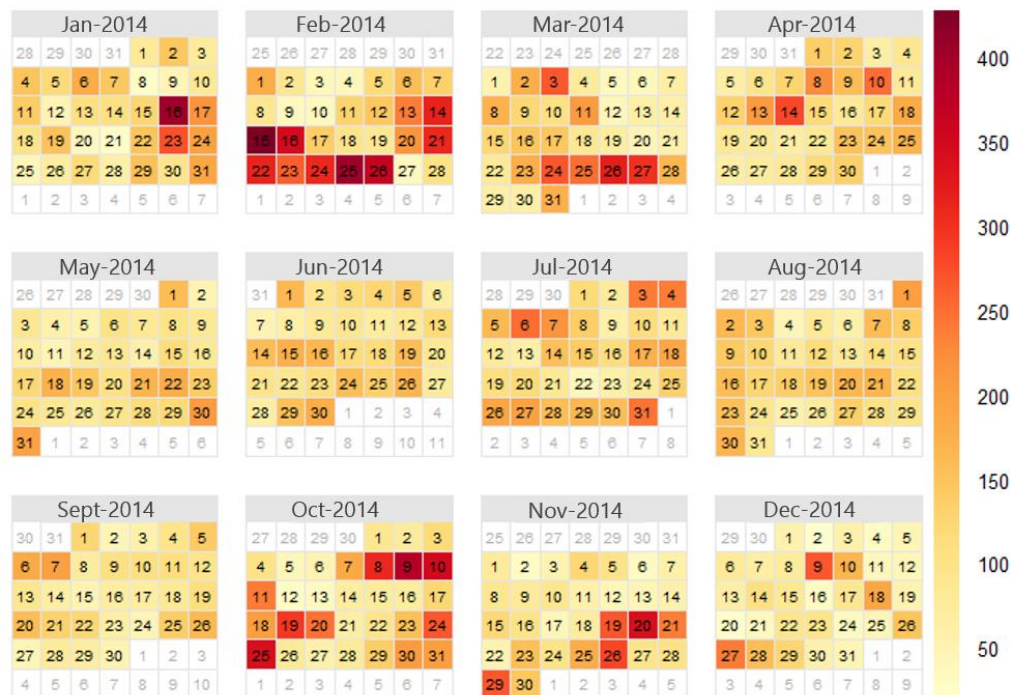


Figure 1. Annual Calendar Chart of AQI Value of Beijing in 2014

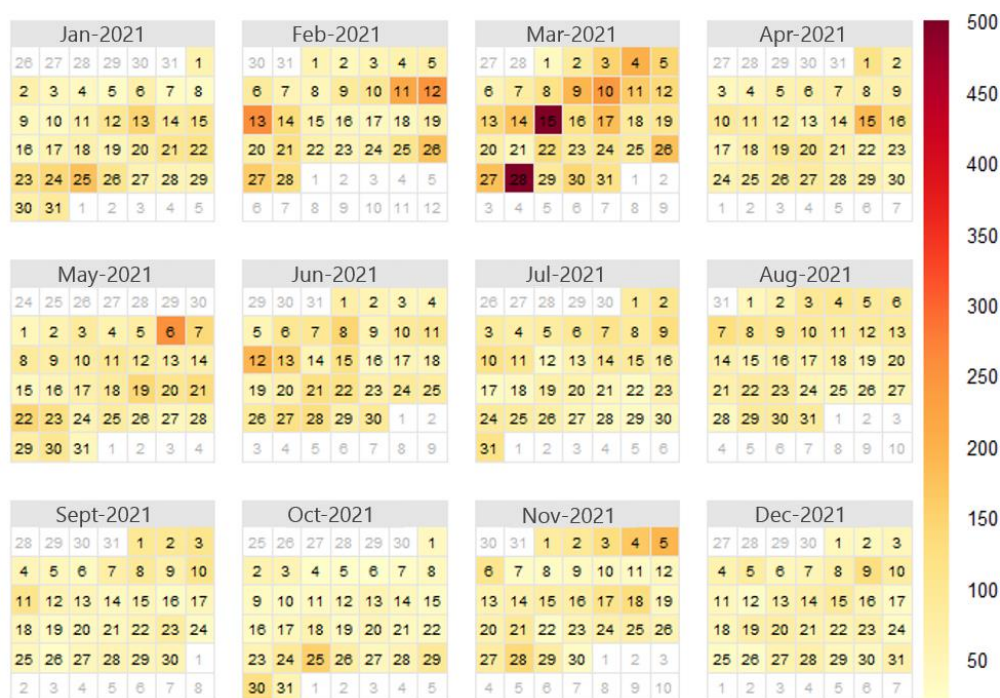


Figure 2. Annual Calendar Chart of AQI Value of Beijing in 2021

It can be seen from the figure that the air quality of Beijing is poor in February and March but better in May and August. Its air quality in 2021 is much higher than that in 2014. In recent years, good results have been achieved in air pollution control.

3. Recognition and Order Determination of ARIMA Model

3.1 Draw a Sequence Diagram

First of all, we draw a sequence diagram. As can be seen from the figure, the monthly time series of the AQI value of Beijing is non-stationary. In the trend of the long term, the series has a downward

trend year by year, that is, the air quality in Beijing has an upward trend year by year. As for its seasonal change, the series has a large fluctuation annually, that is, the average value of AQI in Beijing is monthly different. Therefore, it is necessary to extract the degree of trend and seasonality in subsequent modeling.

In the pure randomness test, the P value of the delayed 6-order Q_{LB} statistics is $4.452 \times 10^{-14} < 0.05$ and the P value of the delayed 12-order Q_{LB} statistic $< 2.2 \times 10^{-16} < 0.05$, so the original hypothesis that the sequence is pure random is significantly rejected. Therefore, this sequence is non-pure random, which is necessary to model it.

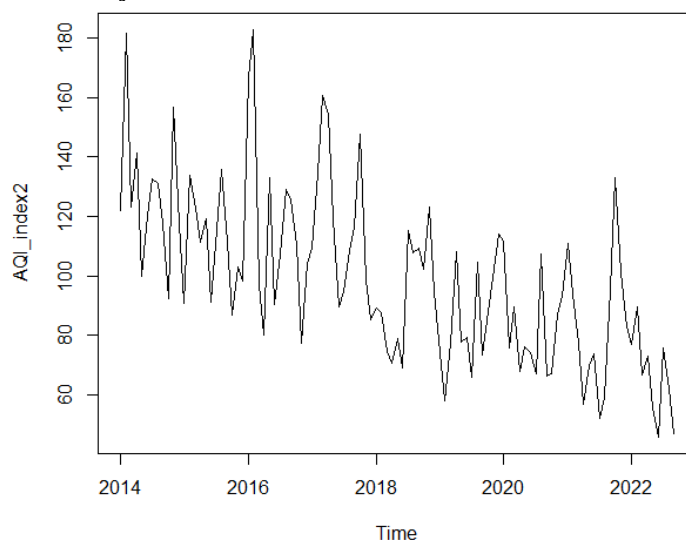


Figure 3. Monthly Time Series of AQI Value of Beijing

3.2 Decomposition of Sequence Trend, Seasonality, and Pure Random Factor

The trend, seasonality, and pure random factors of the monthly time series of AQI value of Beijing are decomposed by a simple multiplication model. The decomposition results are shown in the following figure.

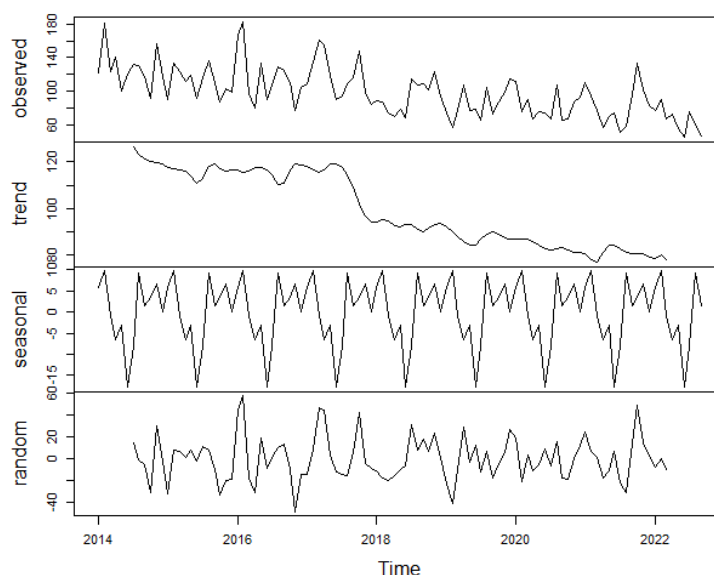


Figure 4. Decomposition Diagram of Trend, Seasonality, and Pure Random Factors of Monthly Time Series of AQI Value of Beijing

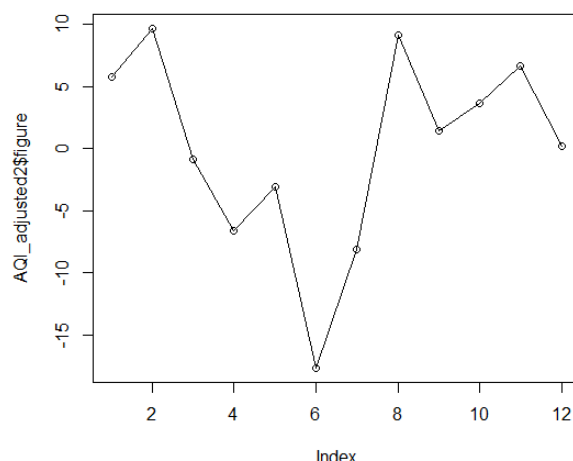


Figure 5. Seasonal Index Effect Diagram of Monthly Time Series of AQI Value of Beijing

As can be seen from sequence decomposition, the trend and seasonal effect of this time series are obvious. According to Figure 4, the AQI value of Beijing has a downward trend year by year with a largely downward trend from the end of 2017 to 2018. Based on the seasonal index renderings, the AQI value of Beijing is lower from March to July but higher in other months.

3.3 Model Order Determination

Due to the non-stationary sequence being affected by both long-term trend and seasonal factors, we should first achieve the first order series with 12-step differences and test their stationary behaviors.

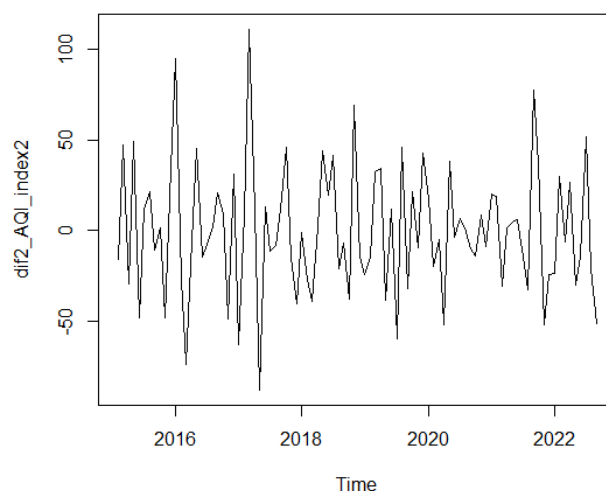


Figure 6. Monthly Time Series of AQI Value of Beijing after First Order 12-Step Differences

Table 1. Stationarity Test Results of Time Series after Difference

	Lag	ADF	P value
Type 1: No drift no trend	0	-11.30	< 0.01
	1	-10.15	< 0.01
	2	-9.48	< 0.01
	Lag	ADF	P value
Type 2: With drift no trend	0	-11.24	< 0.01
	1	-10.09	< 0.01
	2	-9.43	< 0.01
	Lag	ADF	P value
Type 3: With drift and trend	0	-11.19	< 0.01
	1	-10.03	< 0.01
	2	-9.39	< 0.01

The time series after the difference is shown in the above figure. It can be seen from the time sequence diagram that the time series after the difference is roughly stationary. In order to make a more rigorous conclusion, ADF stationarity and pure randomness tests are carried out on the post-difference sequence.

Table 2. Test Results of Pure Randomness of Time Series after Difference

X-squared	Df	P value
14.346	6	0.026
33.227	12	0.001

The results of the stationarity test show that the P values of the three types of τ statistics are all less than the significance level (0.05), so the sequence after the difference is significantly stationary.

According to the results of the pure randomness test, the P value of the delayed 6 order Q_{LB} statistic is $0.026 < 0.05$ and that of the delayed 12 order Q_{LB} statistic is $0.001 < 0.05$. Thus, the differential sequence is a non-white noise sequence.

The above results show that a stationary time series can be obtained by first order 12-step difference.

3.4 Modeling

The monthly AQI series of Beijing is non-stationary with the long-term trend effect, seasonal effect, and random fluctuation effect interacting at the same time. Therefore, the ARIMA multiplication model is used to model this series.

Assuming that the d-order trend difference and d-order seasonal difference with period S as its length are carried out on the sequence, the ARIMA multiplication model is established as follows:

$$\nabla^d \nabla_S^D x_t = \frac{\Theta(B)\Theta_S(B)}{\Phi(B)\Phi_S(B)} \varepsilon_t$$

where,

$$\begin{aligned}\Theta(B) &= 1 - \theta_1 B - \dots - \theta_q B^q \\ \Phi(B) &= 1 - \phi_1 B - \dots - \phi_p B^p \\ \Theta_S(B) &= 1 - \widetilde{\theta}_{12} B^S - \dots - \widetilde{\theta}_{12Q} B^{QS} \\ \Phi_S(B) &= 1 - \widetilde{\phi}_{12} B^S - \dots - \widetilde{\phi}_{12P} B^{PS}\end{aligned}$$

That is, the ARIMA multiplication model is abbreviated as: $ARIMA(p, d, q) \times (P, D, Q)_S$

First, draw the autocorrelogram and partial autocorrelation of the first order 12-step difference sequence. Then, determine the parameters of the ARIMA multiplication model.

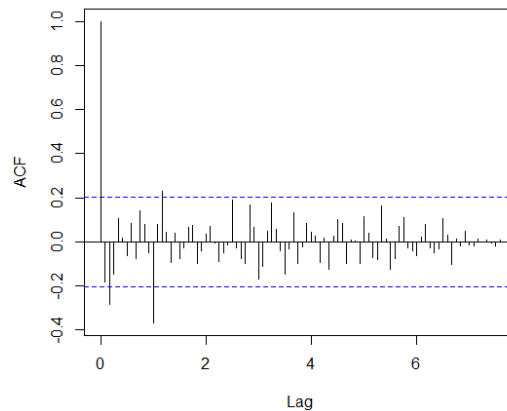


Figure 7. Autocorrelogram of Time Series after First Order 12-Step Difference

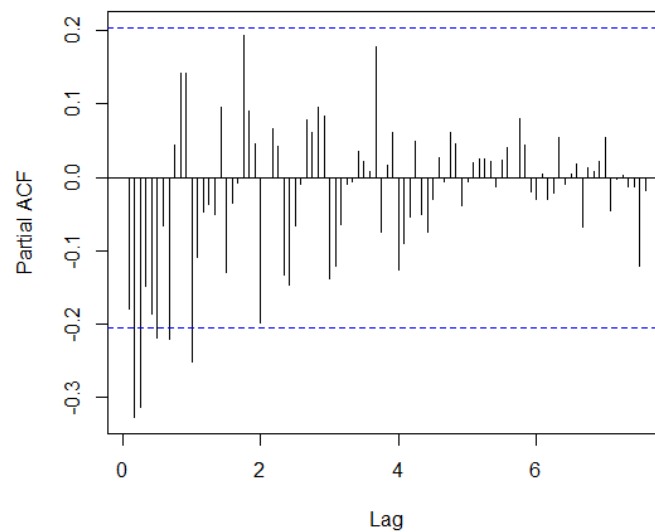


Figure 8. Partial Autocorrelogram of Time Series after First Order 12-Step Difference

After observing the coefficient of the autocorrelogram and partial autocorrelogram in one period (delay less than 12 orders) of the differential sequence, the ACF graph is truncated in order 2, the PACF graph is trailed, and the short-term autocorrelation information of the differential sequence is extracted by fitting MA (2) model, that is, $p=0$, $q=2$. Next, analyzing the seasonal autocorrelation characteristics after observing the coefficient characteristics of autocorrelogram and partial autocorrelogram in terms of period length, such as delays of 12 orders, 24 orders, and 36 orders after difference. ACF diagram shows that the autocorrelation coefficient of delay 12 (1 period) is significantly non-zero, but the autocorrelation coefficients of delay 24 and 36 (2 periods and 3 periods) fall within 2 times the standard deviation, while the partial autocorrelogram shows that delay 12 and 24 orders are significantly non-zero. Thus, the characteristics of seasonal autocorrelation are truncated by 1 period of autocorrelation coefficient and trailed by partial autocorrelation coefficient, that is, $P=0$, $Q=1$, and $S=12$.

At the same time, according to the above analysis, the time series is stable after the first-order 12-step difference with d and D as 1. So the fitted ARIMA multiplication model is $ARIMA(0,1,2) \times (0,1,1)_{12}$ and the model expression is:

$$\nabla \nabla_{12} x_t = (1 - \theta_1 B)(1 - \theta_2 B^2)(1 - \tilde{\theta}_{12} B^{12}) \varepsilon_t$$

Using the mixed method of conditional least squares and maximum likelihood estimation, the parameter estimation results of the ARIMA multiplication model are obtained. According to the regression results, we fit the ARIMA multiplication model as follows:

$$x_t = x_{t-1} + x_{t-12} - x_{t-13} + (1 - 0.6764B)(1 - 0.3236B^2)(1 - 0.7824B^{12}) \varepsilon_t$$

$$Var(\varepsilon_t) = 513.2$$

After fitting, the significance of the fitting model is tested. The following figure shows the significance test chart of the monthly series fitting model of the AQI value of Beijing.

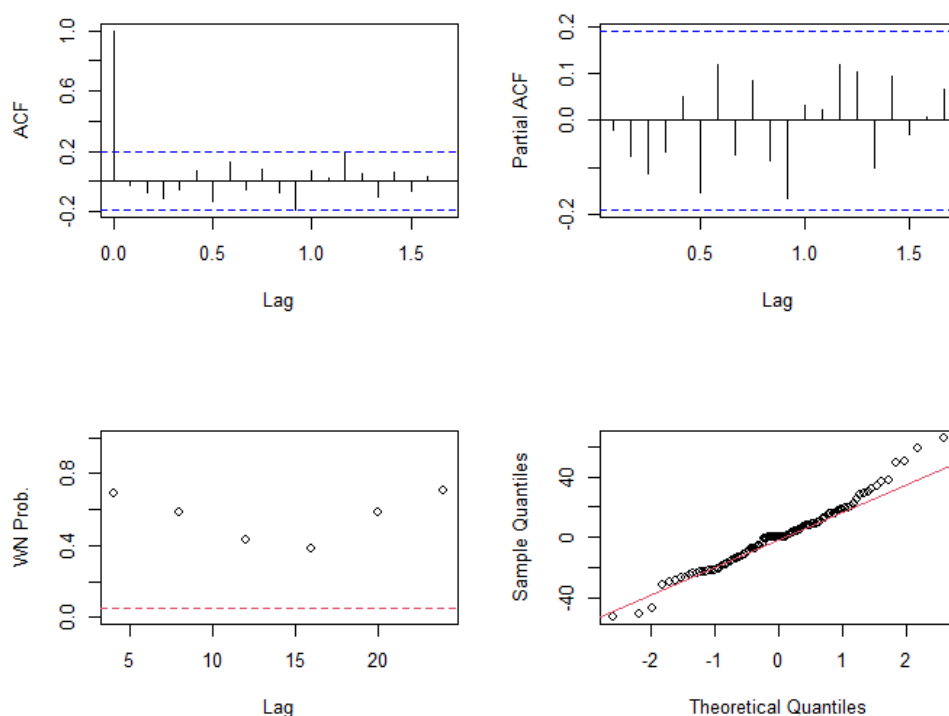


Figure 9. Significance Test Chart of Monthly Series Fitting Model of AQI value of Beijing

In the lower left chart of the pure randomness test of residuals, the P values of pure randomness tests (Q statistics) delayed within 20 orders are all greater than 0.05. The original hypothesis cannot be rejected under the significance level of 0.05, that is, the residuals are white noise sequences. The ARIMA multiplication model was better used to fit the monthly AQI value of Beijing.

3.5 Sequence Prediction

According to the principle of minimum prediction variance, the monthly AQI value of Beijing in the next two years is predicted and the prediction result by using the forecast function in R modeling language is shown in the following figure.

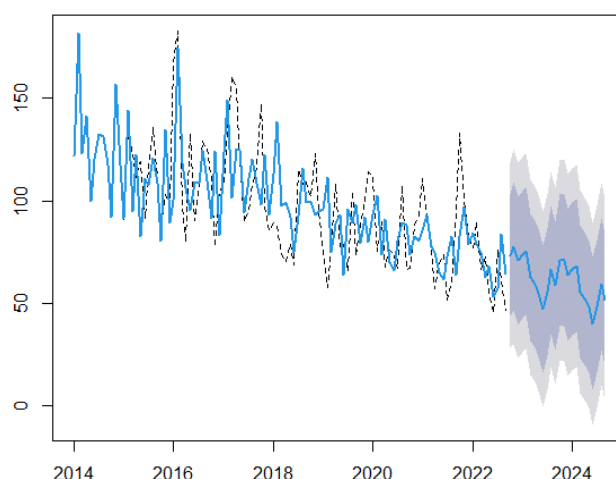


Figure 10. Forecast Results of Monthly AQI of Beijing in the Next Two Years

In the figure, the black dotted line is the observed value of the series, and the blue solid line is fitted and predicted values of the series under model fitting $ARIMA(0,1,2) \times (0,1,1)_{12}$. Besides, the dark shaded part is the 80% confidence interval of the predicted value and the light shaded part is the 95% confidence interval of the predicted value. It can be predicted from the figure that the monthly AQI of Beijing will fluctuate around 50 to 80 in the next two years with a downward trend year by year.

4. Intervention Analysis

From the drawn time series chart of the AQI value, it can be observed that the AQI value of Beijing decreased significantly in 2018, which is related to the Three-Year Action Plan for Winning the War to Protect Blue Skies (hereinafter referred to as the Plan) promulgated in 2018 to a certain extent. Therefore, the intervention variable is set as x_1 . The value of the Plan is 1 when it is promulgated and zero otherwise. Based on this sequence, the intervention analysis of the policy is carried out.

4.1 View Cross-Correlation Graph

Investigate the cross-correlation diagram between sequence and intervention factors, and study the intervention mechanism of intervention variables on sequence before drawing the cross-correlation diagram between sequence and intervention factors as follows:

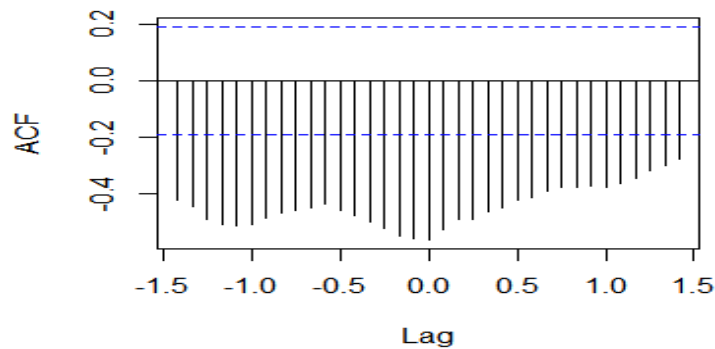


Figure 11. Cross-Correlation Coefficient Diagram between AQI of Beijing and Plan

According to the seasonal effect diagram drawn before, the promulgation of the Plan has no obvious influence on the seasonal effect of the sequence. Based on the drawn cross-correlation coefficient diagram, the cross-correlation coefficient is the largest when the intervention variable of the Plan is 0-order lag. It can be assumed that the intervention of the variable on the sequence is only horizontal without delay. According to the above analysis, we prepare to fit the intervention model as follows

$$x_t = \beta_0 + \beta_1 x_{1t} + \frac{\theta(B)}{\Phi(B)} a_t$$

Where a_t is a white noise sequence.

4.2 Difference to Smooth the Sequence

According to the previous analysis, the first-order 12-step difference of Beijing AQI can achieve differential stabilization.

4.3 Model Ordering

Combine the intervention variables with the AQI series of Beijing after difference to establish a regression model:

$$\nabla \nabla_{12} x_t = \beta_0 + \beta_1 x_{1t} + \varepsilon_t$$

The coefficient characteristics of autocorrelation and partial autocorrelation of residual sequence are investigated. The residual coefficient diagrams of autocorrelogram and partial autocorrelogram are drawn as follows:

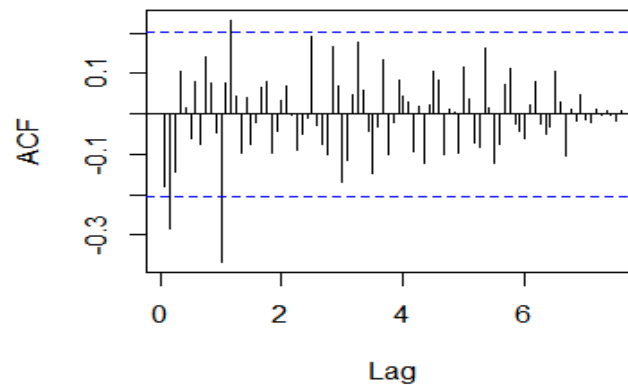


Figure 12. Autocorrelogram of Regression Residual Sequence

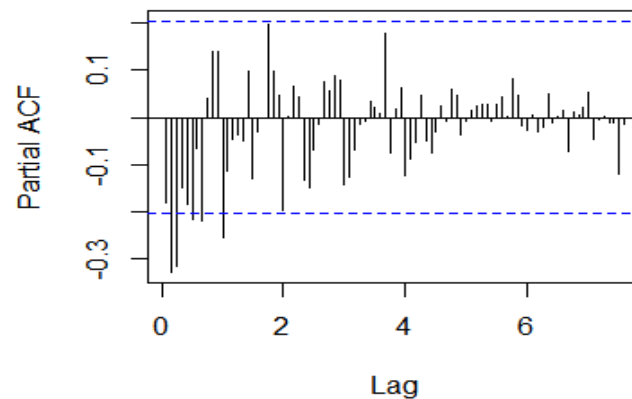


Figure 13. Partial Autocorrelogram of Regression Residual Sequence

Figure 12 and Figure 13 show that the coefficient of autocorrelation and partial autocorrelation of the residual sequence is not truncated, and seasonality still exists significantly. Examining the first 11-order autocorrelation coefficient graph and partial autocorrelation coefficient graph, it can be regarded as autocorrelation first-order truncation and partial autocorrelation trailing. Therefore, $ARIMA(0,1,1)$ can be fitted in a short time. After investigating the correlation coefficients of the delay every 12 steps, it is found that it is significantly non-zero and the correlation coefficient of delay 24 is within 2 times the standard deviation, which can be regarded as the first-order truncation of seasonal autocorrelation coefficient and the partial autocorrelation coefficient trailing. Thus, the periodic correlation $ARIMA(0,1,1)_{12}$ is fitted. To sum up, we fit the regression residual sequence $ARIMA(0,1,1) \times (0,1,1)_{12}$ as follows. The intervention model structure of Beijing AQI is fitted as:

$$\nabla V_{12}x_t = \beta_0 + \beta_1 x_{1t} + (1 - \theta_1 B)(1 - \widetilde{\theta}_{12} B)a_t$$

where a_t is the white noise sequence.

The parameter estimation results of the intervention model are obtained via the maximum likelihood estimation method. According to the regression results, we fitted the intervention model as follows:

$$x_t = x_{t-1} + x_{t-12} - x_{t-13} - 13.2297x_{1t} + (1 - B)(1 - B^{12})a_t$$

where $a_t \sim N(0, 1461)$ and the AIC value of the fitting model is 836.4

4.4 Significance Test of Fitting Model

The significance test of the fitting model is carried out and the test results are as follows:

Table 3. Significance Test of Fitting Model

X-squared	Df	P value
26.253	18	0.094
32.897	24	0.106
39.175	30	0.122

The white noise test of the residual sequence shows that the residual sequence is a white noise sequence, so the fitting model is valid.

4.5 Analyze the Intervention Coefficient of the Regression Model

Given that the coefficient of the intervention variable is -13.2297 and more than double the standard deviation, the promulgation and implementation of the plan effectively reduces the AQI value of Beijing, which shows that the plan is effective in controlling the air quality in Beijing. Moreover, the average AQI value of Beijing was 112.7429 before the promulgation of the Plan. So the average AQI value in Beijing decreased by 13.2297 after its promulgation, which was about $13.2297/112.7429=12\%$.

5. Summarize

Beijing, as the capital of China and the political and cultural center of our country, has been confronted with serious environmental problems in recent years. In particular, the air pollution in Beijing has caught the attention of people all over the country. Therefore, while promoting the development of Beijing, China has also promulgated various policies in different periods to alleviate environmental pressure, including the Beijing Clean Air Action Plan from 2013 to 2017 and the Three-Year Action Plan for Winning the War to Protect Blue Skies. On one hand, Beijing vigorously advocates green production and lifestyle, adheres to the basic strategy of harmonious coexistence between man and nature, and maintains development in environmental governance. On the other hand, it doesn't forget ecological construction under rapid development, learns from experience and lessons, and avoids making the same mistake as the developed countries that put more emphasis on development than environmental protection. As a result, people's sense of acquisition and happiness will be continuously enhanced in these ways.

Through the analysis of the time series of the AQI value of Beijing from 2014 to 2021, we can conclude that the air quality in Beijing has improved year by year from 2014 to 2021 under the vigorous national control of air quality. Especially with the implementation of Beijing's Three-Year Action Plan for Winning the War to Protect Blue Skies, its air quality has been greatly promoted and we can reasonably predict that the air quality will be further improved from 2022 to 2024 with an aim to realize the development in the breast of governance. Up against the vigorous advocacy of green development in China, we believe that protecting the environment and building a beautiful homeland will become the voluntary actions of our people, and the air quality in Beijing and even in China will be further improved. In addition, we will be able to keep the green mountains, clean waters, and blue sky with joint efforts!

References

- [1] Yu Ping. Application of Time Series Analysis in Air Quality Index (AQI) Prediction. Diss. Liaoning Normal University, 2016.
- [2] Ren Huanhuan. Research on the relationship between economic development and ambient air quality in economic zones and provincial capital cities in my country from 2014 to 2017. Diss. Lanzhou University. 2017.
- [3] Ma Pan et al. "Analysis and Fitting Research of Beijing AQI Based on Meteorological Conditions." The 31st Annual Meeting of the Chinese Meteorological Society. 2020.