

Exploration Model of Relations between the PM2.5 of the Air and Other Monitoring Indexes Based on VAR

Guo-feng Yu¹, Jia-bao Liu²

¹Department of Public Service and Management, Anhui Business Vocational College, Hefei Anhui 230041, China;

²Department of Teaching Public Courses, Anhui Xinhua University, Hefei Anhui 230088, China

Abstract—PM 2.5, as an important index about air pollution, whose form and composition is very complicated, has a negative influence on humankind. The paper, according to the meteorological data from Wu Han environmental air automatic monitoring station, conducts an quantitative analysis of independence and the 6 standards of AQI ,based on VAR and Granger causality test, draws a conclusion that inhalable particles and ozone is not the source of the increase of PM 2.5.Finally. The paper establishes VCR model between PM 2.5 and CO, SO2, NO2 ,introduces the impulse response function, explores a random perturbation of a standard deviation shock effects on different indicators for monitoring.

Keywords- PM2.5; air pollution; VAR model.

The status and variation of air quality influences human survival, life and development. It is one of the hottest issues that draw much attention of the environmental protection and the national people. Compared to the air pollution index, the air quality index can monitor the air quality status more specifically. Among these indexes, PM2.5 is a very important index. How to carry out the relevant analysis of the PM2.5 factors and discuss the PM2.5's distribution and variation rules through the collection of PM2.5 data, it is beneficial for the prediction, evaluation and empirical analysis and have theoretical and practical significance for making high efficient feasible emission reduction plan.

I. DATA SOURCE AND PREPROCESSING

A. Source of Data

To analyze the influence of PM2.5 with the monitoring data of Wuhan selected from January 1, 2013 to August 26, 2013 in the attachment.

B. Data pre-procession

In the attachment, there is some missing data about the inhalable particles on April 29, 2013 and July 7, 2013. Besides, there are some common situations of data missing that is described in the attachment. Therefore, firstly, it introduces the basic theory and techniques of the statistical treatment of data missing.

Little and Rubin[1] defined the following three different types of data-missing mechanisms;

1)Missing Completely At Random(MCAR), the data missing is irrelevant to the incomplete variable and complete variable

2)Missing At Random(MAR), the data missing completely rely on the complete variable

3)Not Missing at Random, the data missing of incomplete variable rely on the incomplete variable itself, this kind of data missing shall not be neglected. Where the complete variable refers to the variable (attribute) has no missing value. Otherwise, it will be incomplete variable. Usually, the process of missing data uses the filling method of prediction to maximize the utilization of the data obtained and construct the prediction model. In order to ensure the quality of the missing data's quality through the other data of the corresponding points, it adopts the filling method of prediction to fill the missing data and marked the inhalable particle as inhpar and imports the selected data of the period from January 1, 2013 to April 25, 2013 into Eviews 6.0 software, the results of the data such as stationary test is in the attachment 1, the fitted equation is:

$$inhpar(t) = 41.8666 + 0.5627 * inhpar(t - 1)$$

$$(0.3169) \quad (5.3167) \quad (7.4048)$$

The predicted result on April 29, 2013 is 87.4453, rounded off to 87.it is the same as the data before June 7, 2013, the fitting result of the data is

$$inhpar(t) = 36.0974 - 0.4024 * inhpar(t - 1)$$

$$(0.2006) \quad (5.7782) \quad (-6.2158)$$

Correspondingly, the supplemental result is 74.3438, rounded off to 74.

II. THE CORRELATIVE AND INDEPENDENT QUANTITATIVE ANALYSIS OF SIX MONITORING INDEXES IN AQI

For the test of the correlation and independence among the monitoring indexes, it has great importance for judging the reasonability of the monitoring index's setting and the mechanism analysis of the formation and propagation of PM 2.5 etc.

A. Quantitative Analysis of Correlation

The correlation is aim at describing the overlapped information and the dependence between different indexes, Pearson's correlation analysis is one of the most important technical methods in the multivariate data analysis. It selects 238 pairs observed data in 6 monitoring indexes of the attachment 1, leading-in Eviews6.0, respectively named as sulfur dioxide (SO2), nitrogen dioxide (NO2), particulate matter (Inhpar), carbon monoxide (CO), smell (O3), PM 2.5 (PM), the results are as follows:

TABLE I. 6 MONITORING INDEX DATA DESCRIPTIVE STATISTICS

	PM	SO2	O3	NO2	INHPAR	CO
Mean	108.737	28.1891	52.0511	63.1051	79.1610	25.8067
Median	84.5000	24.0000	46.0000	59.0000	74.0000	23.0000
Maximum	426.000	85.0000	156.000	123.000	282.000	101.000
Minimum	22.0000	5.00000	3.00000	20.0000	18.0000	10.0000
Std.Dev	74.6960	17.7239	33.7949	24.4750	33.1286	14.1098
Skewness	1.32742	0.68193	0.73441	0.50745	1.79423	1.96787
Kurtosis	4.43919	2.54436	2.96273	2.18114	9.87485	8.16541
Jarque-Bera	90.4344	20.5049	21.4088	16.8634	591.383	418.201
Probability	0.00000	0.00004	0.00002	0.00021	0.00000	0.00000
Sum	25927.0	6709.00	12388.0	15019.0	18682.0	6142.00
Sum Sq.Dev	1322340	74450.5	270677.4	141968.4	257913.9	47183.11
Observations	238	238	238	238	236	238

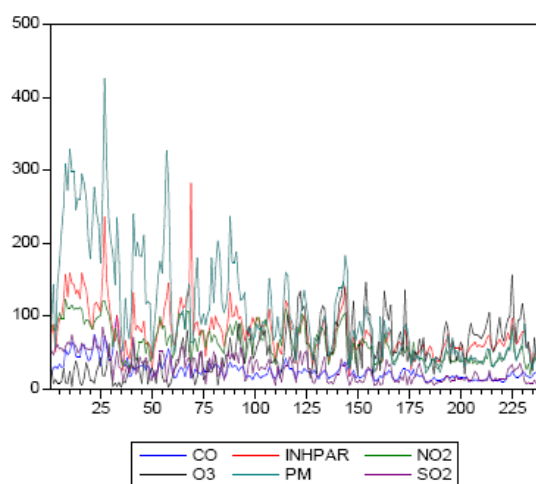


Figure 1. 6 Monitoring index data legend representation

Figure 1 and table 1 respectively are the pictorial representation and the descriptive statistics of the above six monitoring data.

TABLE II. 6 MONITORING INDEX CORRELATION ANALYSIS OF THE OBSERVED DATA

	PM	SO2	O3	NO3	INHPAR	CO
PM	1	0.724105	-0.354879	0.732391	0.778732	0.821645
SO2	0.724104	1	-0.181531	0.805142	0.6778834	0.657012
O3	-0.354879	-0.181531	1	-0.064582	-0.068503	-0.383044
NO2	0.732391	0.805142	-0.064582	1	0.727462	0.624522
INHPAR	0.778732	0.677883	-0.068503	0.727461	1	0.585986
CO	0.821646	0.657013	-0.384044	0.624522	0.585986	1

TABLE III. 6 MONITORING INDEX OBSERVATION DATA OF THE T-STATISTICS

	PM	SO2	O3	NO2	INHPAR	CO
PM	--	16.1288	-5.8313	16.5244	19.0691	22.1446
SO2	16.1288	--	-2.8358	20.8552	14.1652	13.3884
O3	-5.8313	-2.8358	--	-0.9942	-1.0548	-6.3703
NO2	16.5244	20.8552	-0.9942	--	16.2872	12.2843
INHPAR	19.0691	14.1652	-1.0548	16.2872	--	11.1093
CO	22.1446	13.3884	-6.3703	12.2843	11.1093	--

TABLE IV. RELATIONSHIP OF INDEPENDENCE BETWEEN 6 MONITORING INDICATORS

	SO2	NO2	INPPAR	CO	O3	PM2.5
SO2	--	0.0000	0.0000	0.0000	0.0000	0.0000
		co-dependent	co-dependent	co-dependent	co-dependent	co-dependent
NO2	0.0000	--	0.0000	0.0000	0.0000	0.0000
		co-dependent	co-dependent	co-dependent	co-dependent	co-dependent
INHPAR	0.0000	0.0000	--	0.0000	0.0000	0.0000
		co-dependent	co-dependent	co-dependent	co-dependent	co-dependent
CO	0.0000	0.0000	0.0000	--	0.0000	0.0000
		co-dependent	co-dependent	co-dependent	co-dependent	co-dependent
O3	0.0000	0.0000	0.0000	0.0000	--	0.0000
		co-dependent	co-dependent	co-dependent	co-dependent	co-dependent
PM2.5	0.0000	0.0000	0.0000	0.0000	0.0000	--
		co-dependent	co-dependent	co-dependent	co-dependent	co-dependent

From Table 2 we can see that PM2.5 has strong correlation with CO, SO₂ has strong correlation with NO₂ and O₃ has some weak negative correlation with other monitoring indexes. The correlation between Inhpar and PM2.5 is the strongest.

To calculate the statistical quantity value of t based on the principle of correlation test. The result is shown in Table 3.

When $\alpha = 0.05$, we can see the corresponding $t_{\alpha/2} = 1.972$, find the corresponding statistical quantity value of t in Table 3, all other monitoring indexes have linear relations with one another at different degrees except O₃, NO₂ and Inhpar do not have linear relations.

B. Quantitative Analysis of Independence

Independence test is used to determine the independence between two variables through the methods of hypothesis testing. Once confirmed the independence between two variables, we can reason out they are uncorrelated, but not vice versa. This kind of uncorrelated only can demonstrate that they have no linear relation but not mean they have no

functional relations in other forms. Through the independence test we can judge whether two monitoring indexes are independent, if yes, then we can conclude that they are not mutual influence between those two monitoring indexes.

Based on the imported monitoring variables and data, with the Eviews6.0 software, suppose confidence coefficient $\alpha = 0.05$, if the probable value of the obtained *pearson* χ^2 is < 0.05 , then we can conclude the tested two variables is not independent, otherwise they are independent. Through operating the Eviews6.0 software, we can get the specific test results demonstrated in the table below:

As is shown in the table above, the obtained *pearson* χ^2 's probable value between O₃ and NO₂, CO and inhalable particle is < 0.05 , which means they are mutual independent, It theoretically demonstrates that O₃ and these three monitoring indexes have no strong relations and the other monitoring indexes are not independent. Furthermore, it demonstrates there is some mutual influence or closer relations between them.

III. ANALYSIS OF THE RELATIONS BETWEEN PM2.5 AND OTHER MONITORING INDEXES ON THE BASIS OF VECTOR

AUTO-REGRESSION (VAR)

The core ideas of VAR model is not to take into economic theory but to the relations of different time-sequenced economic variables. The general forms are:

$$Y_t = \alpha + \sum_{i=1}^p \beta_i Y_{t-i} + \varepsilon_t, \quad \text{where}$$

$E(\varepsilon_t) = 0, E(\varepsilon_t, Y_{t-i}) = 0, i = 1, 2, \dots, p$. Y_t is the homoscedastic and stable linear stochastic process that consisted of $n \times 1$ vector quantity. β_i is $n \times n$ coefficient matrix, Y_{t-i} is the i step lagged variable of the vector quantity Y_t , ε_t is the error term, which can be regarded as random disturbance term.

Substantially, VAR is a multiple regression equations, the sequence of the required variable are all uniformly integrated variable. If there is non-stationary situation, then

we need to implement co-integrated operation to avoid the spurious regression phenomenon.

A. ADF test

In order to avoid time-sequenced “spurious regression”, it carries on a stationary test of unit root to measure whether the variable can meet the pre-condition of co-integration. It carries on unit root test on PM、CO、INHPAR、NO2、O3 and SO2 with ADP test to judge whether there is unit root or not. ASP test is used to test the stability of time sequence. If the time sequence is unstable, it will cause the spurious regression problems and can not carry on the co-integration test. So ADF test is the pre-condition of co-integration test. By using the EVIEW software we can find the following empirical data, from the above analysis, we have a modeling analysis of the quantitative relation between PM2.5 and the other five monitoring indexes. Based on the analysis above, the paper establishes the VAR model for PM2.5 with the imported monitoring index and observational data. The basic results are as follows:

TABLE V. WUHAN 6 MONITORING INDEX IN THE AQI STATIONARITY TEST RESULTS

Variable	T statistic value	probability	conclusion
PM	-3.324191	0.0149	steady
CO	-2.680730	0.0789	steady
INHPAR	-6.852862	0.0000	steady
NO2	-5.008712	0.0000	steady
O3	-6.500716	0.0000	steady
SO2	-3.696193	0.0047	steady

Table 5 indicates that, with a 10% significance level, all monitoring variable data is stable. In order to make the covering range of the index of the model big enough, we can choose the 10% significance level accordingly.

B. The Optimal Lagged Differences and Co-integration Analysis

Before the co-integration analysis, we need to confirm the model’s lag phase. Because of the limited sample space, the lagged differences shall not be too big. Currently, the most commonly used principle are AIC and SIC, the calculation method of AIC standard is:

$$AIC = \ln \frac{SSR_k}{T} + \frac{2k}{T}$$

SIC criterion of Schwarz is defined as:

$$SIC = \ln \frac{SSR_k}{T} + \frac{k(\ln T)}{T}$$

In which, k is the lag phase, T is the sample number, SSR_k is the quadratic sum of the residual value. The optimal lagged difference is determined according to the value of the AIC and SIC criterion. In general, the optimal lag phase is p , which is the value that can make the above two criterion become minimum.

Based on the selected P value of the AIC and SIC criterions, we can get the following results:

TABLE VI. AIC AND SC ALONG WITH THE CHANGE OF P

Lag number n	AIC	SC	LR statistic
1	31.33780	31.63046	-3693.529
2	31.34494	31.87332	-3662.703
3	31.21064	31.97617	-3615.250
4	31.19805	32.20216	-3582.172

As it Shown in Table 6, the minimum p value of AIC criterion is 4 while the corresponding p value is 1. According to the LR statistic, we can confirm that the optimal lagged difference is: So, the null hypothesis of lagged difference is 1 has been denied and the optimal lagged difference of VAR model is 4. The result of Jonhamson’s co-integration test shows that both the statistical quantity of Trace and the Max-Eigen can reveal at least 3 co-integration relations. That is, there are at least 3 kinds of linear combination (co-integrate vector), which makes single integrated order time sequence of the above four variable’s combination reduced. In other words, there is a long stable relation among the above variables. In which, the standardized co-integration coefficient value are shown as follows:

$$\varepsilon_t = PM - 2.3030 * SO2 + 0.5645 * NO2 - 3.6079 * CO$$

The figure below shows the result of AR root test:
Inverse Roots of AR Characteristic Polynomial

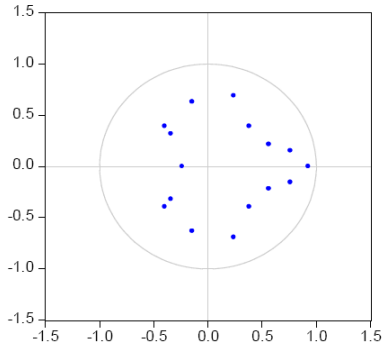


Figure 2. Inverse Roots of AR Characteristic Polynomial

AR root test is used to test the authenticity of co-integrated relations, from figure 2 we can see that, the reciprocal absolute value of AR characteristic equation's characteristic root is 1, which means it is within the unit circle and the established VAR model is stable.

C. Granger causality test

The con-integration equation reveals the long-term co-integrated relations among four variables, but the causality among them is not clear. Thus, we test the causality among those variables with Granger causality test and confirm the optimal lag phase with VAR test as the optimal lag phase of Granger causality test. The test results are shown in table 7.

TABLE VII TEST RESULT OF GRANGER CAUSALITY

Null Hypothesis:	F-Statisti	Prob.	conclusion
PM does not Granger Cause CO	5.18435	0.0005	Reject
CO does not Granger PM	4.63840	0.0013	Reject
PM does not Granger INHPAR	11.1415	2.E-05	Reject
INHPAR does not Granger PM	0.74324	0.4767	Accept
PM does not Granger NO2	0.95301	0.3871	Accept
NO2 does not Granger PM	10.8280	3.E-05	Reject
PM does not Granger O3	6.81115	0.0013	Reject
O3 does not Granger PM	6.81115	0.0013	Accept
PM does not Granger SO2	3.98615	0.0217	Reject
SO2 does not Granger PM	9.45023	0.0001	Reject

Table 7 shows the test results of Granger causality test between each pair of variables, based on the above table, we can see the monitoring index Inhpar and O3 is not the granger reason for PM2.5, that is to say, the inhalable particles and ozone is not the Granger cause of the

increasing PM2.5. Thus, the above 2 pairs of variable shall be eliminated from the subsequent VAR model to establish the vector autoregressive model between PM and COI, SO2, NO2.

D. VAR model

On the basis of the above analysis, the paper constructs a VAR model for the above four monitoring variables. The above four variables are independent variables and the corresponding lag phase is dependent variable, the data in the table determines the coefficient-related statistical quantity of VAR model. Taking the monitoring index PM 2.5 as an example, we can get the following relation equation:

$$\begin{aligned}
 PM = & 0.67 * PM(-1) + 0.012 * PM(-2) + 0.043 * PM(-3) + 0.0299 * PM(-4) \\
 & + 0.421 * SO2(-1) + 0.226 * SO2(-2) - 0.079 * SO2(-3) + 0.252 * SO2(-4) \\
 & + 0.381 * NO2(-1) + 0.099 * NO2(-2) - 0.045 * NO2(-3) - 0.328 * NO2(-4) \\
 & + 0.645 * CO(-1) - 2.061 * CO(-2) + 0.901 * CO(-3) + 0.358 * CO(-4) + 0.428
 \end{aligned}$$

IV. CORRESPONDING FUNCTIONS OF IMPULSE

Corresponding functions of impulse is used to measure the influence that a standard deviation's impact (it is named as "Impulse") of a random disturbance term of some endogenous variable on endogenous variable's current and future values. In order to describe the correlativity among different monitoring indexes, the below figure 3 indicates this influence:

The analysis of the corresponding analysis of impulse indicates that, the variation of PM2.5 basically can have positive effect on all other monitoring indexes and affect them greatly. The cyclic duration is relatively long. But it will become stable finally, that is from the long term, there is a long term and stable relations among variables. We can also obtain the similar results from other monitoring variables' impulse analysis. However, the variation among different monitoring indexes at the initial stage can have very different function director and intensity.

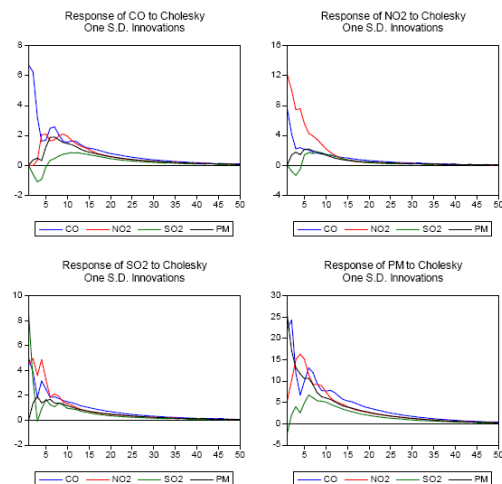


Figure 3. Impulse response function

E. Variance Decomposition

Variance decomposition is used to evaluate the contribution degree that all endogenous variables have for predicting the variance. That is to analyze the percentage of

the impact that the predicted residual vector has on different information, it also refers to the contribution percentage of the corresponding endogenous variable to the standard deviation.

Variance Decomposition of PM:Period	S.E	PM	SO2	NO2	CO
1	34.29052	100.0000	0.000000	0.000000	0.000000
2	46.53963	95.61182	2.985907	0.976478	0.425780
3	52.16391	90.22075	5.376301	2.745625	1.657326
4	56.33854	84.82639	6.270953	4.849611	4.053042
5	60.31607	81.41872	8.187593	5.797332	4.596359
6	64.02423	79.97805	10.19847	5.554042	4.269449
7	66.68339	79.20267	11.48555	5.316952	3.994824
8	68.55852	78.27862	12.50815	5.310995	3.902242
9	70.05487	77.35892	13.41211	5.331070	3.897859
10	71.29765	76.77687	14.10865	5.268967	3.845516

The variance analysis result shows, with the passing of time, PM2.5's contribution degree on the variance decomposition gradually reduced and became stable, while the other monitoring indexes' contribution rate is either in continually increasing status or relatively stable status. Based on the above table, we can see PM 2.5 has big influence on VAR model.

IV. CONCLUSION

From what I mentioned above, we discussed the correlativity between PM 2.5 and other five monitoring indexes based on VAR model. Through the unit root test, we determined the stability of all monitoring indexes' data, through the Granger causality test we confirmed the causality relations among all different monitoring indexes. Lastly, based on the previous preparation, it established VAR model and obtained the long term stable relations among other monitoring indexes except the inhalable particle and O3 are not the Granger reason for the increase of PM 2.5, and determined the specific relation equation based on the VAR model. In the end, it obtained the understanding of the function and influence of different monitoring variables on VAR model through the impulse response and variance decomposition and illustrates this kind of relations between different monitoring indexes further.

V. ACKNOWLEDGEMENTS

The research work was Funded by Natural science Foundation of Anhui under Grant No. KJ2013B015

REFERENCES

- [1] Little R.J.A., Rubin D.B, Statistical Analysis with Missing Data, (Wiley Series in Probability and Mathematical Statistics) [M], Wiley John & Sons, New York,1987.
- [2] Hu Guoding Zhang Ruichu Multivariate data analysis method of pure algebraic treatment[M] Nan Kai University Press 1990.
- [3] Gao Tiejing ,Econometric analysis and modeling; Eviews application and cases (second edition) [M], Bei Jing:Tsing Hua university Press,2009.
- [4] Li Xinling ,Correlation analysis of Xi'an climate changes and PM10 characteristics[D]. Xi'an University of Architecture and Technology, 2003.
- [5] Yang Fumo ,He Kebin ,Ma Yongliang, chen xu ,Characteristics and sources of trace elements in Beijing atmospheric[J] ,concentration about PM2.5, Environment and Science,2003,33-37.
- [6] Yang Hong, Sun Deshi, Xi'an Environmental Particulate Pollution and road traffic particulate emission mechanism [J],Journal of Chang An university natural science edition, 2002(5): 79-82.
- [7] Li Fang, Pollution characteristics of atmospheric particulate matter PM 2.5, Xi'an city and its relationship with precipitation [D]. Xi'an University of Architecture And Technology,2012.