

**RELATIONSHIP BETWEEN AIR POLLUTION STANDARD INDEX (APSI)  
IN NORTH JAKARTA AND WEST JAKARTA  
USING VECTOR AUTOREGRESSIVE (VAR) MODELING**

Sarah Sholikhatun Risma  
Badan Pusat Statistik Kota Subulussalam  
E-mail: risma.sarah@gmail.com

**ABSTRACT**

*The high level of APSI in Jakarta and surrounding regions has been categorized as an unhealthy level of air pollution, which can consequently affect the level of public health. This study aims to determine whether there is an influence of APSI pollution level from the previous period as well as the influence of APSI level on other regions from the previous period that are neighboring. VAR modeling approach used which had advantages in considering changes in data behavior over time and can also be used to explain the interaction between a number of variables. APSI level data in West Jakarta region and North Jakarta region have been stationary which are indicated by the significance of the ADF test. The analysis results found that the simultaneous testing of VAR model was significant with  $p$ -value  $< 0.05$  which can be concluded that the APSI level in West Jakarta was influenced by APSI level in the previous period and was also influenced by APSI level of North Jakarta in the previous period. Similarly, the level of APSI level of North Jakarta was influenced by APSI level in the previous period and was also influenced by APSI level of West Jakarta in the previous period.*

**Keywords:** Air Pollution, APSI, Neighboring, VAR.

**A. INTRODUCTION**

More than half population of metropolitan cities in the world have bad air quality 2.5 times higher than the standard recommendation by the UN world health agency (WHO, 2018). In 2018 Jakarta was included in the top ten cities with the worst air pollution based on data obtained from the Air Quality Index, even the air quality in Jakarta was worse than the air pollution levels of Beijing and New Delhi which are known as the cities with the highest levels of pollution in world (World Air Quality Index project, 2018). According to Greenpeace, during the period of

January to June 2017, air quality in Jabodetabek area is indicated to enter an unhealthy level for humans (Greenpeace, 2019).

Air health is a fundamental thing which is needed for undergoing the various life of human activities. Attention regarding to the environmental health is increasing along with the increasing rhythm of community activities, especially people who live in the metropolitan cities. The worse condition of air pollution level can endanger public health, especially for vulnerable groups such as children,

pregnant women and the elderly, furthermore it can increase the risk of mortality (Dandotiya, 2019). In addition, awareness and concern for the need to maintain a healthy living environment needs to be fostered from an early age to create a more harmonious life.

Air health can be indicated through the Air Pollution Standard Index (APSI) in units of  $\mu\text{g} / \text{m}^3$  (micrograms per cubic meter). APSI is an air quality report to explain how clean or polluted air quality is within a period of hours / days / months. APSI is determined based on five pollutant components, namely: Carbon Monoxide (CO), Sulfur Dioxide (SO<sub>2</sub>), Nitrogen Dioxide (NO<sub>2</sub>), Surface Ozone (O<sub>3</sub>), and Dust Particles (PM<sub>10</sub>).

The determination of APSI considers the level of air quality on the health of humans, animals, plants, buildings and aesthetic values. Local pollutant concentration in some of the cities is significantly affected by emissions from their surrounding areas (Wu et al., 2015). The flow of PM<sub>10</sub> pollutant can also flow from one region to another surrounding region (Černíkovský et al., 2016). Air pollution is one of the dynamic physical characteristics in which data can be measured over a certain time span, so

that time series analysis is needed. Analysis of the time span becomes important and relevant considering the discovery of certain phenomena that can directly affect the condition of air pollution in an area or region. Time series analysis is an analysis of a series observational data that occurs based on a time index in sequence with a fixed time interval and one of the statistical procedures applied to predict the conditions that will occur on the future in the context of decision making.

Vector Autoregressive (VAR) is one of the models in multivariate time series analysis (Sims, 2014). Granger states that if two variables have a causal relationship in which one variable affects the other variable, then the past information of a variable can help predict the other variable (Granger, 1969).

In this research, the VAR model will be applied to the APSI data which is observed every day through two Air Quality Monitoring Stations (AQMS) at two different regions which are close to each other, located on North Jakarta and West Jakarta. The geographical proximity between one region and another can affect the level of air pollution. Based on this phenomenon, the level of air pollution not only have an attachment between time, but also

have an attachment between regions, which is often known as spatial relationships.

## B. MATERIALS AND METHODS

The data which is used in this analysis are APSI levels from North Jakarta region and West Jakarta region. APSI level data are daily time series data for the period 21 April - 29 July 2017. Data obtained from the DKI Jakarta provincial government that can be downloaded via website (Provinsi DKI Jakarta, 2017). APSI data is obtained through AQMS every hour, which are averaged in the form of daily data. AQMS are located in every administrative region of DKI Jakarta Province. The AQMS of the North Jakarta region located in Kelapa Gading, while the AQMS of the West Jakarta region located in Kebon Jeruk. The analysis technique in this study uses VAR modeling with R software to assist in the calculation and tabulation process.

### VectorAutoregressive (VAR)

VAR model is a dynamic equation system in which the estimation of a variable in a certain period depends on the movement of that variable and other variables involved in the system of the previous period (Lütkepohl &

Krätzig, 2004). VAR is a model that accommodates data for dynamic structure specifications.

The VAR model is used for forecasting time series variables as well as for analyzing the dynamic impact of the disturbance factors contained in the variable system. VAR is also a very useful analytical tool, both in understanding the interrelationships between economic variables and in the formation of a structured economic model. Another advantage in using VAR modeling is the parameter estimation method using Ordinary Least Square (OLS) so it tends to be simpler than other models

For a simple system with two variables (bivariate model) with a lag of one, the simultaneous model formed is as follows:

$$\begin{aligned} Z_1(t) &= \phi_{11}Z_1(t-1) + \phi_{12}Z_2(t-1) + \varepsilon_1(t) \\ Z_2(t) &= \phi_{21}Z_1(t-1) + \phi_{22}Z_2(t-1) + \varepsilon_2(t) \end{aligned}$$

where

$Z_1(t)$  = first observed variable at time ( $t$ )

$\phi_{11}$  = coefficient of the first observed variable at time ( $t-1$ )

$Z_1(t-1)$  = first observed variable at time ( $t-1$ )

$\phi_{12}$  = coefficient of the second observed variable at time ( $t-1$ )

$Z_2(t-1)$  = second observed variable at time ( $t-1$ )

$\varepsilon_1(t)$  = error of the equation at time ( $t$ )  
 $Z_2(t)$  = second observed variable at time ( $t$ )  
 $\phi_{21}$  = coefficient of the first observed variable at time ( $t-1$ )  
 $\phi_{22}$  = coefficient of the second observed variable at time ( $t-1$ )  
 $\varepsilon_2(t)$  = error of second equation at time ( $t$ )

$$t_{hit} = \frac{\hat{\phi}_{kj}(i)}{se_{\hat{\phi}_{kj}(i)}}$$

If  $|t_{hit}| > t_{(db=T-s, \alpha/2)}$ , where  $T$  is the number of observations and  $s$  is the number of parameters expected in a equation which is  $1 + np$ , then the decision taken was to reject  $H_0$ . To test the significance of the parameters simultaneously carried out by  $F$ -test with the following hypothesis:

$H_0: A_i = 0$  or the VAR model parameter is equal to zero

$H_1$ : at least one  $A_i \neq 0$  or model parameter is not equal to zero ( $i=1,2,\dots,p$ )

By using  $\alpha = 0.05$  if the calculated  $F$ -value is greater than the value  $F_{(\alpha=0.05; k-1, n-k)}$  with  $n =$  number of observations and  $k =$  number of parameters or  $p$ -value  $< 0.05$  then  $H_0$  is rejected, meaning that the independent variable jointly affect the dependent variable which means that all parameters simultaneously significant to the model.

Before making a VAR estimation model, there are several steps that must be carried out, namely pre-estimation testing, including stationarity test and optimum lag determination. The steps in the analysis can be explained as follows:

With assumptions: (a)  $Z_1(t)$  and  $Z_2(t)$  stationary, (b)  $\varepsilon_1(t)$  and  $\varepsilon_2(t)$  error with standard deviation  $\sigma_{Z_1}^2$  dan  $\sigma_{Z_2}^2$ , (c)  $\varepsilon_1(t)$  and  $\varepsilon_2(t)$  uncorrelated.

A proper VAR model is a model with significant parameters or the parameter value is not equal to zero. The hypotheses tested in the estimated parameters of the VAR model are as follows:

$$H_0: \phi_{kj}(i) = 0$$

$$H_1: \phi_{kj}(i) \neq 0$$

Where  $\phi_{kj}(i)$  can be interpreted as the parameter coefficient of the  $j$ -th variable ( $j=1,2,\dots,n$ ) in the partial equation of the  $k$ -th variable ( $k=1,2,\dots,n$ ) for the  $i$ -th time difference ( $i=1,2,\dots,n$ ). The value of  $\phi_{kj}(i)$  estimated through the least squares method and for testing its performed using a  $t$ -test with the following statistics test:

1. Create a general description of air pollution level data of two regions with descriptive analysis
2. Carry out stationarity tests using the Augmented Dickey Fuller (ADF) Test. If the data is stationary then it can be continued for further analysis, but if the data is not stationary then a differencing process must be carried out until the data becomes stationary
3. Determine the optimum lag using ACF and PACF plots
4. Estimate parameters model and test the model, both partially and simultaneously
5. Test model assumptions to determine fitted model
6. Draw conclusions from the results of the analysis and provide suggestions for better improvement goals

## C. RESULTS AND DISCUSSION

### 1. Descriptive Analysis

Descriptive analysis is needed before continuing to the inference analysis. This aims to be able to provide an initial description of the data to be examined. Initial description

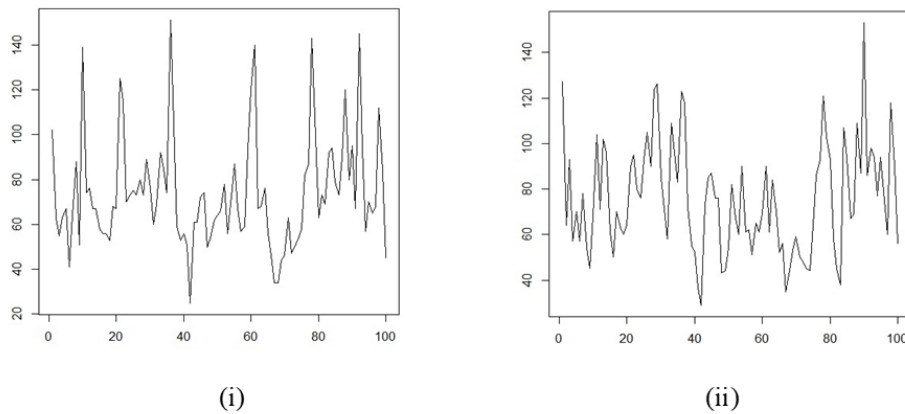
of the data from **Table 1** can be used as material for consideration to further analysis.

**Table 1.** Descriptive of Variable

Region	Min.	Max.	Average	St. Dev
North Jakarta	25	151	73.92	24.83
West Jakarta	29	153	75.53	24.22

The data used in this study amounted to 100 APSI data level of North Jakarta region and West Jakarta region. Correlation are used to understand the relationship between the APSI levels of the two regions in the study period. The Pearson correlation result for the APSI level between the regions is 0.62, which can be concluded that there is a positive relationship between the APSI level in the North Jakarta and West Jakarta regions.

The average APSI in the North Jakarta region was 73.92  $\mu\text{g} / \text{m}^3$  with a standard deviation of 24.84  $\mu\text{g} / \text{m}^3$ , while the average APSI in West Jakarta region was 75.53  $\mu\text{g} / \text{m}^3$  with a standard deviation of 24.22  $\mu\text{g} / \text{m}^3$ . The average APSI level between the two regions falls into the "medium" air pollution category.



**Figure 1.** Time Series Data Plot (i) APSI of North Jakarta; and (ii) APSI of West Jakarta

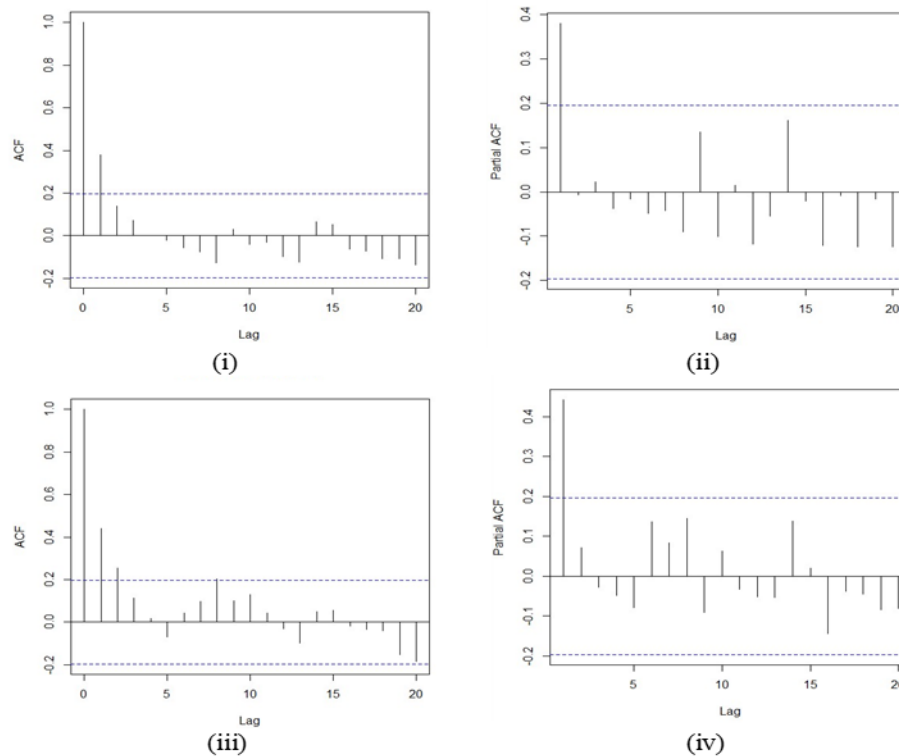
For a minimum APSI level of 25  $\mu\text{g} / \text{m}^3$  in the North Jakarta region and 29  $\mu\text{g} / \text{m}^3$  in the West Jakarta region occurred on June 1, 2017. While for the maximum APSI level in the North Jakarta region occurred on May 26, 2017 reaching 151  $\mu\text{g} / \text{m}^3$ , while the highest APSI level in the West Jakarta region occurred on July 19, 2017 reaching 153  $\mu\text{g} / \text{m}^3$  which both already classified as "unhealthy" air pollution category.

## 2. Data Stationarity

Stationary time series data is data that does not have trends, seasonal patterns, also the average and variety are constant or homogeneous over time. The time series data can be checked by looking at the time series plot.

A time series plot that fluctuates with a constant range around a constant mean shows that the time series data is stationary. Whereas the time series plot that does not fluctuate around a constant or non-fluctuating average with a constant range indicates that the time series data is not stationary. In the **Figure 1** plot, it appears that the data has shown a stationary pattern, but to be more certain the ADF test will be carried out to check the stationarity of the data.

The ADF test result for North Jakarta are significant because the p-value  $< p\text{-alpha} = 5\%$ , so it can be concluded that the APSI data for the North Jakarta region is stationary. Likewise, the ADF test results for West Jakarta showed significance because the p-value  $< p\text{-alpha} = 5\%$  so that it could be concluded that the APSI data of West Jakarta region was stationary.



**Figure 2.** (i) ACF Plot of North Jakarta; (ii) PACF Plot of North Jakarta; (iii) ACF Plot of West Jakarta; and (iv) PACF Plot of West Jakarta

### 3. Determine Optimal Lag Length

Optimal lag length of the data could be determined after checking data stationarity. The optimum lag length is important in order to determine the parameter estimator for the appropriate in time series model. One method used to determine the optimum lag length is through Auto Correlation Function (ACF) and Partial Auto Correlation Function (PACF). ACF and PACF show the association or dependence between the values of a series with the series itself at different periods, due to a time-lag.

The North Jakarta ACF plot on **Figure 2** shows that the data forms a certain pattern (sine wave), while the North Jakarta PACF plot is truncated at the 1st lag and located above the average line value. It can be concluded that data follows the VAR model with lag 1 or commonly known as VAR (1).

The same condition happened for West Jakarta ACF plot that forms sine wave patterns, while West Jakarta PACF plot is truncated on 1st lag and located above the average line value. It also can be concluded that the data follows the VAR model with lag 1 or often known as VAR (1).



#### 4. Estimation of VAR Model

##### Parameters

From the test results described earlier, it was found that the model that can be applied to APSI level data is the VAR model with a lag of 1, VAR (1). A variable is not only influenced by the

$$\text{North Jakarta}_t = 45.02 + 3.85e-01 \text{ North Jakarta}_{t-1} + 1.37e-04 \text{ West Jakarta}_{t-1}$$

$$\text{West Jakarta}_t = 34.26 + 0.28 \text{ West Jakarta}_{t-1} + 0.26 \text{ North Jakarta}_{t-1}$$

past of the variable itself but is also influenced by the past of other variables. The VAR model formed for each region can be seen in the following **Figure 2**. Output from **Table 2** can be written in equation as follows:

**Table 2.** VAR Model Estimation Results

Region	Variable	Coefficient	p-value (Partial)	F-statistic	df	p-value (Simultant)	Adjusted R <sup>2</sup>
North Jakarta <sub>t</sub>	c	45.02	4.25e-07*	8.37	96	4.44e-04*	0.13
	North Jakarta <sub>t-1</sub>	3.85e-01	1.67e-03*				
	West Jakarta <sub>t-1</sub>	1.37e-04	0.99				
West Jakarta <sub>t</sub>	c	34.26	1.28e-05*	16.28	96	8.19e-07*	0.24
	West Jakarta <sub>t-1</sub>	0.28	0.01*				
	North Jakarta <sub>t-1</sub>	0.26	0.01*				

North Jakarta region model using a test level of 5 percent, p-value of 4.44e-04\* < α = 0.05 is obtained, which means that the coefficient of the independent variable has a significant to the model simultaneously. This means that in the North Jakarta region the APSI level is influenced by the APSI level from the previous period and the APSI level from the West Jakarta region at the previous period simultaneously. In addition, it can also be seen that the North Jakarta region model has a determination coefficient of R<sup>2</sup> = 0.14, meaning that the percentage of total variation in the model that can be

explained by the response variable is only 14 percent with the standard error residuals value of 23.12.

Similarly, the West Jakarta region model using a test level of 5 percent obtained p-value 8.19e-07\* < α = 0.05, which means that the coefficient of the independent variable has a significant to the model simultaneously. This means that in the West Jakarta region the APSI level is influenced by the APSI level in the previous period and the APSI level of the North Jakarta region in the previous period simultaneously. In addition, the level of APSI in the period before and the level of APSI North



Jakarta in the previous period significantly affected the level of APSI in West Jakarta partially. For variations that can be explained by the response variable in the West Jakarta model also tends to increase by 24 percent when compared to the previous model with a standard error residual value of 20.76.

After the model is obtained, then the assumptions of the VAR model are tested include the following:

- Multivariate white noise residual test

Looking at the output from the calculation of the Portmanteu test, the p-value is  $0.27 > \alpha = 0.05$ , which means  $H_0$  accepted that the VAR (1) model has fulfilled the multivariate white noise assumption.

- Residual test with normal multivariate distribution

Looking at the output of the JB-Test, Skewness and Kurtosis calculations, the p-value of the three tests is below the significance level ( $> \alpha = 0.05$ ) so it can be concluded that  $H_0$  rejected that the VAR residual model (1) does not meet the assumption of multivariate normal distribution.

- Residual heteroscedasticity test

Looking at the output of the ARCH calculation, the p-value is  $0.204 > \alpha = 0.05$ , which means  $H_0$  accepted that the VAR (1) model has fulfilled the residual homogeneity assumption.

From the results of the VAR model residual assumptions test show that the model has fulfilled the residual homogeneity assumption and the multivariate white noise residual assumption, but has not fulfilled the normal distribution of residual assumptions.

#### **D. CONCLUSIONS**

The APSI level in the North Jakarta and West Jakarta regions are time series data that have been stationary on variance and average. The APSI level among those regions are also has high enough correlation, therefore it can be analyzed using VAR modeling. The optimum lag length of the VAR model used to describe the relationship of level of APSI pollution level between the two regions is VAR model with first order VAR (1).

From the partial test results, the APSI level for the North Jakarta region was influenced by the APSI level in the previous period, while the APSI level in the West Jakarta region in the previous

period did not statistically significantly affect the APSI level in the North Jakarta region. But on the results of simultaneous testing, North Jakarta's APSI level was proven to be significantly influenced positively by the level of APSI in the previous period and the APSI level of the West Jakarta region in the previous period. This means that if there is an increase in the APSI in the West Jakarta region, it will also be responded to by an increase in the APSI in the North Jakarta region.

However it is different from the partial test results at the APSI level for the West Jakarta region, which is significantly positively influenced by the APSI level in the previous period and the APSI level of the North Jakarta region in the previous period. For the results of simultaneous testing, similar to the North Jakarta region, the APSI level in the West Jakarta region was significantly positively influenced by both the APSI level in the previous period and the APSI level of the West Jakarta region in the previous period.

This means that if there is an increase in the APSI of the North Jakarta region, it will also be responded by an increase in the APSI of the West Jakarta region.

From the results it is also known that for each model, both the North

Jakarta and West Jakarta regions still have an adjusted  $R^2$  value which is not quite high at <50 percent. This shows that the diversity of models for each region cannot be explained by the explanatory variables of the previous period's APSI level and the APSI level between the two regions only, but there are still other variables outside the model that can explain further the APSI level in the study period.

Furthermore, in testing the model, the model must be fitted when assumptions about the residuals are met, such as multivariate white noise, normal multivariate distribution and homoscedasticity. In calculation of this AR (1) model test, it was previously known that there was one assumption that had not been fulfilled, namely the residual did not have a normal multivariate distribution, so that the AR (1) model had not been said to be fit.

This study concluded that the level of air pollution between regions that are close together will influence each other in accordance with research that has been done previously by Černíkovský et al. (2016). But the advantage of this study compared to previous studies is also considering how much influence of air pollution level from previous time of those regions.

Based on the results of this study, it is expected that the programs implemented to overcome the problem of air pollution can be managed by considering other areas that are close together and influence each other. As research conducted by Wu et al. (2016) earlier said that joint regional air pollution control could be more cost-

effective among neighboring region. The variables examined in this study are still limited to only two administrative regions of Jakarta province which are close to each other, so for further research could be conducted for the entire administrative area in Jakarta province.

## E. REFERENCES

- Černíkovský, L., Krejčí, B., Blažek, Z., & Volná, V. (2016). Transboundary air-pollution transport in the Czech-Polish border region between the cities of Ostrava and Katowice. *Central European Journal of Public Health*, 24, S45–S50. <https://doi.org/10.21101/cejph.a4532>
- Dandotiya, B. (2019). *Health Effects of Air Pollution in Urban Environment*. January, 96–115. <https://doi.org/10.4018/978-1-5225-7387-6.ch006>
- Granger, C. W. J. (1969). Investigating Causal Relations by Econometric Models and Cross-spectral Methods. *Econometrica*, Vol. 37, N, 424–438.
- Greenpeace. (2019). *Data Terkini Kualitas Udara Kota-kota di Seluruh Dunia*. <https://www.greenpeace.org/indonesia/publikasi/2217/data-terkini-kualitas-udara-kota-kota-di-seluruh-dunia/>
- Lütkepohl, H., & Krätzig, M. (2004). Applied time series econometrics. In *Applied Time Series Econometrics*. <https://doi.org/10.1017/CBO9780511606885>
- Provinsi DKI Jakarta. (2017). *Data Indeks Standar Pencemaran Udara DKI Jakarta Tahun 2017*. <https://data.jakarta.go.id/dataset/data-indeks-standar-pencemaran-udara-di-dki-jakarta>
- Sims, C. A. (2014). *Prior Restrictions The Role of Approximate Lag Estimation Distributed*. 67(337), 169–175.
- WHO. (2018). *WHO Global Ambient Air Quality Database (update 2018)*. <https://www.who.int/airpollution/data/cities/en/>
- World Air Quality Index project. (2018). *World's Air Pollution: Real-time Air Quality Index*. <https://waqi.info/>
- Wu, D., Xu, Y., & Zhang, S. (2015). Will joint regional air pollution control be more cost-effective? An empirical study of China's Beijing-Tianjin-Hebei region. *Journal of Environmental Management*, 149, 27–36. <https://doi.org/10.1016/j.jenvman.2014.09.032>