Science and Technology Indonesia
e-ISSN:2580-4391 *p*-ISSN:2580-4405
Vol. 8, No. 2, April 2023

Research Paper

Spatial Autoregressive Quantile Regression with Application on Open Unemployment Data

Ferra Yanuar[1*], Tasya Abrari[1], Izzati Rahmi HG[1], Aidinil Zetra[2]

[1]*Department of Mathematics and Data Science, Andalas University, 25163, Indonesia*
[2]*Department of Political Sciences, Andalas University, 25163, Indonesia*
***Corresponding author**: ferrayanuar@sci.unand.ac.id

Abstract
The Open Unemployment Level (OUL) is the percentage of the unemployed to the total labor force. One of the provinces with the highest OUL score in Indonesia is West Java Province. If an object of observation is affected by spatial effects, namely spatial dependence and spatial diversity, then the regression model used is the Spatial Autoregressive (SAR) model. Quantile regression minimizes absolute weighted residuals that are not symmetrical. It is perfect for use on data distribution that is not normally distributed, dense at the ends of the data distribution, or there are outliers. The Spatial Autoregressive Quantile Regression (SARQR) is a model that combines spatial autoregressive models with quantile regression. This research used the data regarding OUR in West Java in 2020 from the Central Bureau of Statistics. This study develops to modeling the Open Unemployment Level in all province in Indonesia using modified spatial autoregressive model with the quantile regression approach. This study compares the estimation results based on SAR and SARQR models to obtain an acceptable model. In this study, it was found that the SARQR model is better than SAR at dealing with the problems of dependency and diversity in spatial data modeling and is not easily affected by the presence of outlier data.

Keywords
Open Unemployment Level (OUL), Spatial Effects, SAR, SARQR, The Outlier

Received: 10 January 2023, Accepted: 11 April 2023
https://doi.org/10.26554/sti.2023.8.2.321-329

1. INTRODUCTION

Unemployment is a phenomenon that occurs in all developing countries, including Indonesia. One of the causes of unemployment is the lack of job opportunities that are not balanced by the number of job seekers in an area. This leads to an increase in the number of unemployed. An indicator to measure high unemployment in an area is the Open Unemployment Level (OUL). OUL is the percentage of unemployed in the total labor force. West Java Province is one of Indonesia's top three provinces with the highest Open Unemployment level. The OUL situation in West Java Province in August 2021 was 9.82 percent, down 0.64 percentage points compared to August 2020 (Central, 2021). However, this value is still too high above the national average. It is necessary to identify which factors are significant in reducing OUL.

Factors between regions usually tend to be involved as one factor affecting OUL in an area (Dai and Jin, 2021; Weisberg, 2014), known as spatial effects. Spatial effects are divided into two parts: spatial dependence and spatial diversity. Spatial dependence occurs due to the relationship between regions, while spatial diversity occurs due to the diversity between one region and another. If this happens, the data modeling method used is the Spatial Autoregressive (SAR) regression method (Ver Hoef et al., 2018). The existence of outlier data will also affect the modeling method used (Yanuar et al., 2023). Influential outlier data cannot be thrown away because it will eliminate important information related to the data (Yasin et al., 2020; Yasin et al., 2022). Modeling for data containing spatial effects and outliers uses the Spatial Autoregressive Quantile Regression (SARQR) method (Dai and Jin, 2021; Dai et al., 2020; Jin et al., 2016).

There has recently been an increase in the research on estimating and testing spatial autoregressive quantile models. Jin et al. (2016) explored the quantile regression approach for partially linear spatial autoregressive models with possibly varying coefficients using B-spline. (Dai et al., 2022). Dai et al. (2020) investigated fixed effects quantile regression for general spatial panel data models with both individual fixed effects and time effects based on the instrumental variable method. Dai and Jin (2021) employed the minimum distance quantile regression (MDQR) methodology for estimating the SAR panel data model with individual fixed effects. Zhang et al. (2021a) studied a penalized quantile regression for a spatial panel model with fixed effects.

To fill the research gap, this study applied a SARQR model that integelevels the spatial correlations and quantile effects to estimate the effects of regional factors on the Open Unemployment level in West Java province and how they vary with the figure of the Open Unemployment level. It aims to answer the following questions: how SAR and SARQR can produce the acceptable model of Open Unemployment level?, what are the effects of risk factors of Open Unemployment level, and how do they vary across the distribution of the Open Unemployment level?, how does the Open Unemployment level affect the above effects?. The rest of the study is organized as follows. Section 2 describes the available data used in this study. Section 3 introduces the SAR and SARQR models. Section 4 reports the modelling results and discusses the parameter estimations. Section 5 summarizes the remarkable findings and gives further directions.

2. EXPERIMENTAL SECTION

2.1 Materials

In this section, the data used in this study is secondary data obtained from the Central Bureau of Statistics for West Java Province from 1 January to 31 December in 2020 (Central, 2021). West Java Province consists of 18 regencies and 9 cities. The variable used is the percentage of the population (X_1), which is the total population in the district/city divided by the total population of the province multiplied by 100%, the percentage of poor people (X_2), the Labor Force Participation Level or LFPR (X_3), abbreviated Gross Regional Domestic Product or GRDP (X_4) and the percentage of Human Development Index or % HDI (X_5) (Candraningtyas and Santosa, 2022; Hapsari and Hasmarini, 2022). The response variable is the Open Unemployment Level or OUL .

2.2 Methods

To apply the SAR method, several stages of testing are required to ensure that the SAR method is suitable for modelling the cases raised. Multicollinearity is a condition with a correlation between the independent variables in the regression model. Multicollinearity can be detected using the Variance Inflation Factor (VIF) value as follows (Xu and Huang, 2015):

$$VIF_1 = \frac{1}{1 - R_1^2} \tag{1}$$

with R_l^2 is determinant coefficient for predictor variable l, for $l = 1, \ldots, p$. If the VIF value is less than 5, multicollinearity does not occur in the regression model (Yu et al., 2021).

A spatial weighting matrix is a matrix that describes the relationship of each region. The spatial weighting matrix is denoted by W with size $n \times n$, where n represents the number of observation areas. The spatial weighting matrix is obtained based on the results of standardized contiguity. Matrix contiguity has three types of contact, namely rook contiguity (side contact between regions), bishop contiguity (corner contact between regions) and queen contiguity (side touch and corner points between regions). Matrix contiguity has a value of 1 if between regions meet the type of contact (Dai and Jin, 2021; Yasin et al., 2020).

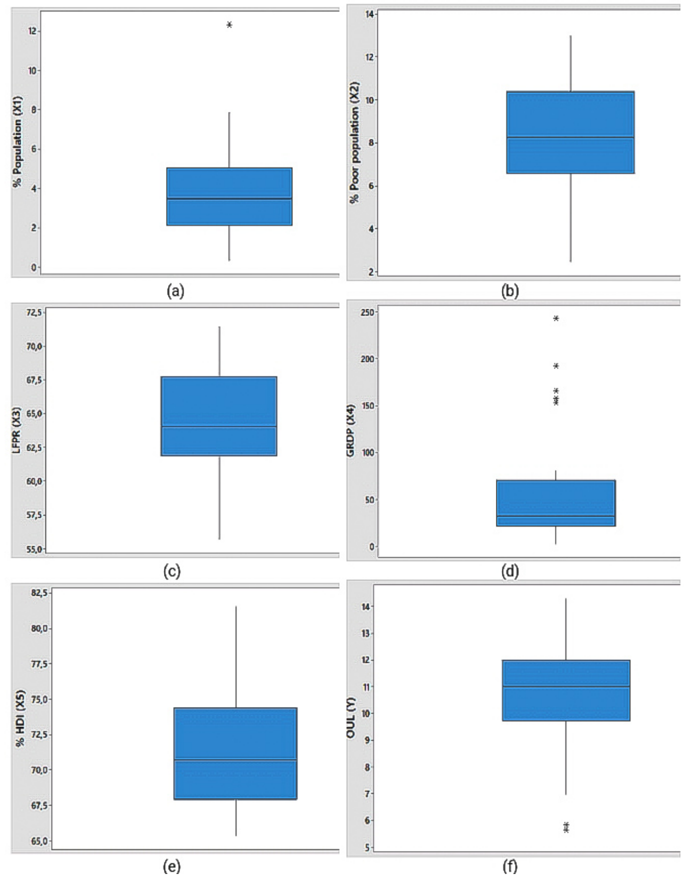

Figure 1. Box plot of variables (a) % Population (X_1), (b) % Poor population (X_2), (c) LFPR (X_3), (d) GRDP (X_4), (e) % HDI (X_5), and (f) OUL (Y)

2.3 Spatial Effects

Spatial effects consist of spatial dependence and spatial diversity. To determine the existence of spatial dependence between regions, a test was carried out with the Moran Index. Moran's index measures the relationship of observations between an area and other areas close to each other. Moran's index is defined as follows (Tribhuwaneswari et al., 2022) :

$$I = \frac{n \sum_{i=1}^{n} \sum_{j=1}^{n} w_{ij} (y_i - \bar{y})(y_j - \bar{y})}{\sum_{i=1}^{n} \sum_{j=1}^{n} w_{ij} \sum_{j=1}^{n} (y_j - \bar{y})^2} \tag{2}$$

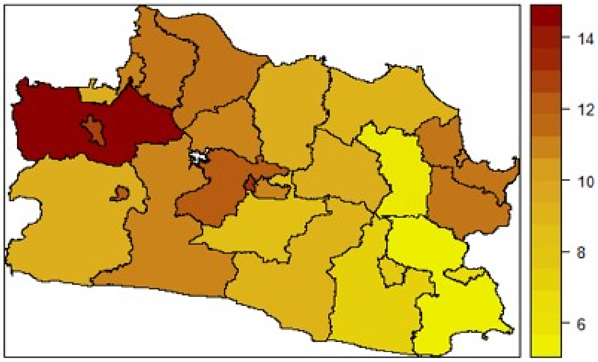

Figure 2. Spatial Distribution of Open Unemployment Level in West Sumatra (In Percentage)

$$= \frac{\sum_{i=1}^{n}\sum_{j=1}^{n} w_{ij}(y_i - \bar{y})(y_j - \bar{y})}{S^2 \sum_{i=1}^{n}\sum_{j=1}^{n} w_{ij}}$$

with i and j are the number of data (n), for $i \neq j$, y_i is the response variable for all observations i, $\bar{y}$ is the mean of y, w_{ij} is element of spatial weighting matrix W, and S^2 is sample variance. If I is positive, adjacent areas have similar values, and the data pattern tends to be clustered. If I is negative, it means that adjacent areas have different values, and the data pattern tends to spread out. If I is 0, it means no spatial dependence detected (Huang et al., 2010). The existence of spatial dependence on the dependent variable is also checked in the hypothesis model. The Lagrange Multiplier test is used to determine spatial dependence, with the general form of the Lagrange Multiplier as follows (Anselin and Anselin, 1988):

$$LM_{lag} = \frac{\left(\frac{U'WY}{S^2}\right)^2}{nP} \tag{3}$$

with

$$nP = T + \frac{(WX\beta)'M(MX\beta)}{S^2},$$
$$T = \text{trace}((W + W')W),$$
$$M = I - X(X'X)^{-1}X',$$
$$s^2 = \frac{U'U}{n}, \ U \text{ as a remainder.}$$

Where X is matrix of independent variables of size $n \times (k+1)$ and β is factor of the regression coefficient of size $(k+1)\times 1$. If the $LM\ lag > X^2{}_{\alpha,1}$, it indicates that there is a spatial dependence on the dependent variable. If a model contains spatial dependence, the modelling is carried out using a Spatial Autoregressive model (Anselin and Anselin, 1988). As for detecting spatial diversity, it can use the Breush-Pagan (BP) test with the following formula:

$$BP = \frac{1}{2}(f'X)(X'X)^{-1}(X'f) \tag{4}$$

where f is a vector where its elements are $\frac{\hat{U}_i^2}{\hat{\sigma}}$-1 while $\hat{U}_1$ is observation from the regression estimation result and $\hat{\sigma}$ is the variance of the residuals, if the $BP > X^2{}_{\alpha,k-1}$ means that there is spatial variation between observation areas at significant α. Parameter k represent the number of independent variables involves in the model.

2.4 Quantile Regression

The quantile regression method uses the approach of separating the data into certain quantile groups that may have different estimated values (Yanuar et al., 2022; Yanuar et al., 2019). The linear regression equation model for the quantile τ is as follows (Davino et al., 2013; Yanuar and Zetra, 2021):

$$Y = X\beta_\tau + U, \tag{5}$$

Y = vector of dependent variables of size $n \times 1$,
X : matrix of independent variables of size $n \times (k+1)$,
β_τ : factor of the regression coefficient of quantiles of size$(k+1) \times 1$depending on the quantile$-_\tau$,
U : vector of residuals of size $n \times 1$.

The parameter estimation using the regression quantile method is obtained by minimizing the sum of the absolute values of the errors with weights τ for positive errors and (1-τ) for negative errors, formulated as follows:

$$\arg\min_{\beta \in \theta} \sum_{i=1}^{n} \rho_\tau(y_i - x_i\beta_i)$$

with loss function ρ_τ (U)=$[\tau$-$I(U<0)]U$. Where $I(.)$ is the indicator function, its value is one when $I(.)$ is accurate and zero otherwise. While ρ_τ (U) is the loss function which is defined as follows

$$\rho_\tau(U) = \begin{cases} U\tau, & if U > 0 \\ U(\tau - 1) & otherwise \end{cases}$$

2.5 Spatial Autoregressive Model (SAR)

The SAR model is a linear model with a spatial correlation of the dependent variables. The SAR model is written as follows (Anselin and Anselin, 1988):

$$Y = \lambda WY + X\beta + U, U \sim N(0, \sigma^2 I) \tag{6}$$

Where λ represents spatial autoregressive coefficient, which shows the magnitude of spatial dependence between regions, W denotes spatial weighting matrix of size $n \times n$, and β is quantile regression coefficient vector of size $(k+1) \times 1$.

Parameter estimation of the SAR model can be estimated using the maximum likelihood estimation (MLE) method by assuming that the residual u is random variable from the normal distribution, $N(0,\sigma^2)$. Parameter estimation for β in the SAR model is obtained as follows:

$$\bar{\beta} = (X'X)^{-1}X'^{(1-\lambda W)}Y \tag{7}$$

While the estimation of parameter σ^2in the SAR model is obtained as follows:

$$\bar{\sigma}^2 = \frac{1}{n}((I-\lambda W)Y - Y\beta)'((I-\lambda W)Y - X\beta), \tag{8}$$

Parameter estimation λ can be obtained using a numerical approach (Anselin and Anselin, 1988).

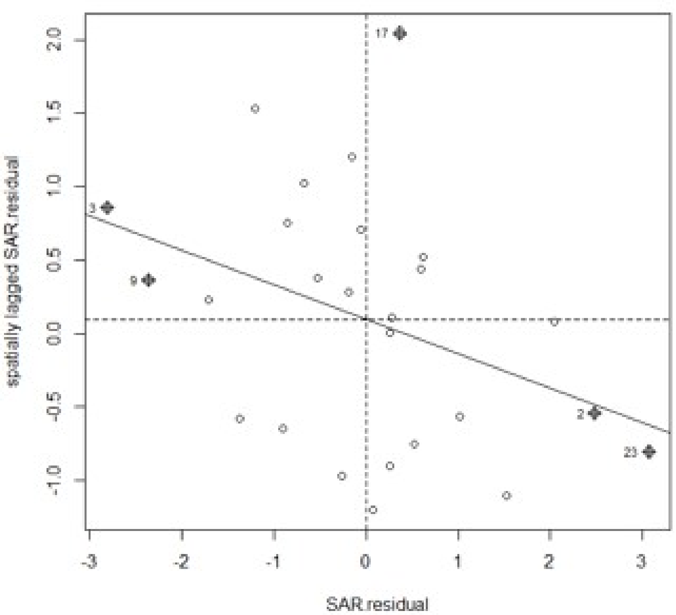

Figure 3. Spatial Outliers in the SAR Model

2.6 Spatial Autoregressive Quantile Regression Model (SARQR)

The SARQR model is a model that combines spatial autoregressive models with quantile regression. The SARQR has the spatial autoregressive coefficient (λ) and the regression vector (β), which depend on certain quantile values (τ) (McMillen, 2015; Ver Hoef et al., 2018). The development of SAR modelling on the quantile τth is specifically defined as follows (Lum and Gelfand, 2012):

$$Y = \lambda_\tau WY + X\beta_\tau + U. \tag{9}$$

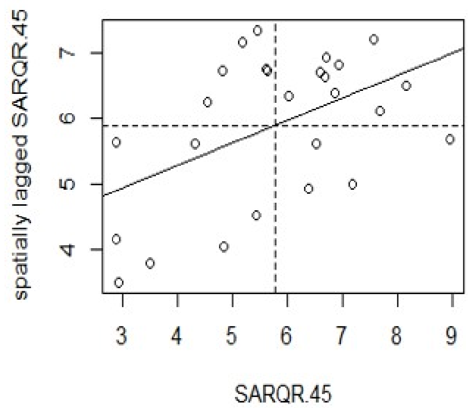

Figure 4. Spatial Outliers Test Based on SARQR Model at τ=0.45.

The value of the spatial autoregressive coefficient in the SARQR model shows the magnitude of the spatial dependence between adjacent areas. The IVQR (Instrumental Variable Quantile Regression) method is used to estimate parameters in the SARQR model (Yu et al., 2021). The assumptions used in estimating the parameters in SARQR model are as follows (Zhang et al., 2021a):

1. $P(u_i \leq 0) = \tau$, for every i=1,2,...,n
2. $\sup_{n\geq 1} \max_{1\geq i \geq n} E(u_i) \leq \mu < \infty$,
3. $U \sim N(0, \lambda^2 I)$.

Getting the parameters λ_τ and β_τ are done by minimizing Equation (10) respect to each parameter.

$$\text{argmin E} = \left[\rho_\tau\left(y_i - \lambda_\tau \sum_{i=j}^{n} \omega_{ij} y_j - X_i'\beta_\tau - g(X_i, Z_i)\right)\right] \tag{10}$$

Where $g(X_i, Z_i)$ is a linear function as instrumental variable quantile regression (IVQR). The steps for estimating the parameters based on IVQR in the SARQR model are as follows (Su and Yang, 2011; Zhang et al., 2021b):

1. Set a specific value for λ, and do the quantile regression modeling at τth quantile, which is defined as
$$(\bar{\beta}_\tau(\lambda), \bar{\gamma}\tau(\lambda)) = argmin_{\beta\gamma} Q_\tau(\beta, \lambda, \gamma)$$
2. To calculate the estimated value of the IVQR is done by minimizing the vector of the estimated variable instrument $\bar{\gamma}_\tau(\lambda)$
$$\hat{\lambda}_\tau = argmin_\lambda \hat{\gamma}_\tau(\lambda)\hat{A}(\hat{\gamma}_\tau(\lambda))'$$
where $\hat{A}$=A+$O_p(1)$, A is a positive definite matrix.
3. The estimator β is obtained in the following way.
$$\hat{\beta}_\tau = \hat{\beta}_{\tau-1}\hat{\lambda}_{\tau-1}$$
Repeat the above steps for each quantile τ. At each quantile τ, different estimating parameters are obtained.

3. RESULTS AND DISCUSSION

Table 1 and Figure 1 present respectively the descriptive and box plot of data used in this study. Figure 1 informs us that there are outliers on the population percentage (X_1), GRDP (X_4) and OUL (Y).

Furthermore, the multicollinearity test among independent variables were carried out to find out whether there was a correlation between each independent variable. The results of the multicollinearity test are presented in Table 2. Each variable has VIF value less than 5, meaning there is no multicollinearity problem between the independent variables.

The spatial weighting matrix in this study uses the queen contiguity (side contact and corner points between regions). This study's spatial units are districts/cities in West Java Province. The list of regencies/cities in the adjacent West Java is presented in Table 3.

In this section, the empirical data regarding the Open Unemployment Level in 26 regions in West Java is employed to construct the OUL model. The first step is to estimate the Moran's Index coefficient values based on Equation (2):

$$I = \frac{n \sum_{i=1}^{n} \sum_{j=1}^{n} w_{ij}(y_i - \bar{y})(y_j - \bar{y})}{\sum_{i=1}^{n} \sum_{j=1}^{n} w_{ij} \sum_{j=1}^{n} (y_j - \bar{y})^2} = 0.5528$$

It is found that I value is positive, which means that adjacent areas have similar values and data patterns tend to be clustered. The next test is to estimate the Moran's Index using Equation (2) above. It is obtained that value of Moran's Index is 4.1451 where it indicated that there is a spatial dependence between regions in West Java Province. Furthermore, the Lagrange Multiplier test was carried out to determine the spatial dependence on the dependent variable, as written in Equation (3). This study found that the Lagrange Multiplier value is 3.9572, which means that there is a spatial dependence on the dependent variable.

Additionally, the heteroscedasticity test is also determined in the hypothesis model using the Breusch-Pagan test. Based on the data used in this study, the Breusch-Pagan's value in this hypothesis model is 2.9538 (with *p-value* is 0.061). It indicates that the variances of Open Unemployment level among regions in East Java are the same.

The SAR model was carried out because there is a spatial effect in the form of spatial dependence on the dependent variable, where the results of the parameter estimation of the SAR model can be seen in Table 4.

Based on Table 4, it is found that not all independent variables have a significant influence on district/city OUL in West Java Province. The independent variables that have significant effect on OUL are the percentage of poor people (X_2) and LFPR (X_3), indicated by *p-value* for corresponding predictor are smaller than the significance level α=0.1. Meanwhile, the percentage of the population (X_1), GRDP (X_4) and % HDI (X_5) have no a significant effect on OUL.

The purpose of this regression is to build the best model that can predict the dependent variable (OUL), so the spatial autoregression model of OUL in West Java Province obtained from data analysis is:

$$\begin{aligned}\hat{y} =& 0.322WY + 13.041 + 0.143X_1 + 0.295X_2 \\ &- 0.302X_3 + 0.005X_4 + 0.142X_5\end{aligned}$$

Based on above equation, it can be explained that: $\hat{\lambda}$=0.322, means that the Open Unemployment Level of each province has an effect of 0.322 times the average Open Unemployment Level of each neighboring province. An increase of 1% in the Poor population (X_2) will increase the Open Unemployment Level by 0.295, where other variables are considered constant. The proposed model also informed us that an increase of 1% in the Labor Force Participation Level or LFPR (X_3), will decrease the Open Unemployment Level by 0.302, where other variables are considered constant. It is only the significant variables can be interpreted, the non-significant variables give no meaning.

Spatial effect testing in spatial dependence and spatial diversity was conducted again on the SAR model. This is done to check whether the SAR model can handle spatial effects properly or not. In terms of spatial diversity, *p-value* = 0.8863 > α= 0.05 means that between regions have the same variance. In spatial dependence, *p-value* = 0.0155 < α = 0.05 means that the SAR model still contains the effect of spatial dependence on the dependent variable. Furthermore, *Moran's scatterplot* is figured to detect spatial outliers in the SAR model, as provided in Figure 3. In this figure, it is found that there are five spatial outliers in the SAR model, those are observations number 2, 3, 9, 17 and 23, denoted by star. Outliers will affect the parameter estimation results and decrease model accuracy. Therefore, the OUL model obtained based on SAR method could not be accepted. SARQR model is then applied to deal with spatial effects and data containing spatial outliers to obtain more acceptable model.

SARQR modelling produces a model that may differ in each quantile. SARQR modelling uses the IVQR method as parameter estimation by minimizing the coefficients of the instrument variables for each quantile so that the optimal independent variable parameters and spatial autoregressive coefficients are obtained in the SARQR model. The results of parameter estimation for each quantile are presented in Table 5.

It can be seen in Table 5 that percentage of the population (X_1), percentage of poor population (X_2), LFPR (X_3), and GRDP (X_4) are significant in selected quantile. The spatial autoregressive coefficients on the SARQR model are varies among quantiles, some are positive, other are negative, and several are close to zero. The SARQR model which has a positive spatial autoregressive coefficient, indicates that the districts/cities in the adjacent West Java Province have a positive influence on the districts/cities in the West Java Province in that quantile.

The comparison between SAR and SARQR models based

Table 1. Descriptive Statistics on Data

Variable	Mean	Q_1	Median	Q_3
% Population (X_1)	3.85	2.13	3.49	5.03
% Poor Population (X_2)	8.40	6.53	8.27	10.39
LFPR (X_3)	64.56	61.88	64.09	67.75
GRDP (X_4)	62.10	22.25	32.25	IPM
% HDI (X_5)	71.78	67.96	70.74	74.38
OUL (Y)	10.57	9.75	11.00	12.00

Table 2. Multicollinearity Test Results Independent

Variable	VIF
% Population (X_1)	1.76
% Poor population (X_2)	2.69
LFPR (X_3)	1.08
GRDP (X_4)	1.69
% HDI (X_5)	2.72

Table 3. The Neighboring Regions in West Java

Regions	Neighboring Regions
Bogor	Suka Bumi. Cianjur. Purwakarta. Karawang. Bekasi. Sukabumi City. City of Depok. and Bekasi
Sukabumi	Bogor. Cianjur. and Sukabumi City,
Cianjur	Bogor. Sukabumi. Bandung. arrowroot Purwakarta. West
Bandung	Cianjur. Garut. Sumedang. Subang. West Bandung. Bandung. Cimahi City
Garut	Cianjur. Bandung. Tasikmalaya. Sumedang
Tasikmalaya	Garut. Ciamis. Majalengka. Sumedang. Tasikmalaya City
Ciamis	Tasikmalaya. Kuningan. Majalengka. Tasikmalaya City. Banjar
Kuningan	Ciamis. Cirebon. Majalengka
Cirebon	Kuningan. Majalengka. Indramayu. Cirebon City
Majalengka	Tasikmalaya. Ciamis. Kuningan. Cirebon. Sumedang. Indramayu
Sumedang	Bandung. Garut. Tasikmalaya. Majalengka. Indramayu. Subang
Indramayu	Cirebon. Majalengka. Sumedang. Subang
Subang	Bandung. Sumedang. Indramayu. Karawang. Purwakarta. West Bandung,
Purwakarta	Bogor. Cianjur. Subang. Karawang. West Bandung
Karawang	Bogor. Subang. Purwakarta. Bekasi
Bekasi	Bogor. Karawang. Bekasi City,
West Bandung,	Cianjur. Bandung. Subang. Purwakarta. Bandung. Cimahi
Bogor City	Bogor
Sukabumi City	Sukabumi
Bandung City	Bandung. West Bandung. Cimahi City
Cirebon City	Cirebon
Bekasi City	Bogor. Bekasi. Depok
Depok City	Bogor. Bekasi
Cimahi City	Bandung. West Bandung. City of Bandung
Tasikmalaya City	Tasikmalaya. Ciamis
Banjar City	Ciamis

Table 4. Model Estimation Results with the SAR Model

Variable	Estimated Mean	Standard Error	Z-Statistics	*p-value*
Constant	13.0411	9.9793	1.3072	0.1912
% Population (X_1)	0.1432	0.1342	1.0654	0.2873
% Poor population (X_2)	0.2951*	0.1564	1.8853	0.0593
LFPR (X_3)	-0.3022*	0.0763	-3.9491	0.0000
GRDP (X_4)	0.0052	0.0053	0.8692	0.3850
% HDI (X_5)	0.1421	0.0932	1.5332	0.1250
SAR coefficient (λ)	0.3222	0.1701	1.8882	0.0572

* Significant at the significant level α=0.1,$Z_{\alpha/2}$= 1.645

Table 5. Estimated Parameter Based on SARQR

Quantile		Parameters					
(τ)	Constants	X_1	X_2	X_3	X_4	X_5	λ_τ
0.10	24.666	-0.294	-0.058	-0.363*	0.007	0.073	0.021
0.15	14.345	-0.249	0.164	-0.304*	-0.000	0.126	0.023
0.20	14.832	-0.219	0.189	-0.305	-0.001	0.111	0.018*
0.25	-2.342	0.020	0.436	-0.157	0.009	0.139	0.015*
0.30	-2.342	0.019	0.436	-0.157	0.009	0.140	0.015*
0.35	5.601	0.125	0.410	-0.216	0.006	0.092	-0.019
0.40	11.052	0.263	0.389	-0.264	0.002	0.067	-0.001
0.45	14.659	0.197	0.426*	-0.336*	0.005	0.104	-0.002*
0.50	14.916	0.147	0.429	-0.348*	0.006	0.119	0.000
0.55	12.145	0.123	0.487	-0.339*	0.006	0.145	-0.005
0.60	9.940	0.375*	0.415	-0.330*	-0.001	0.187	0.005
0.65	13.055	0.316*	0.309	-0.288*	-0.001	0.131	0.024
0.70	17.709	0.311*	0.269	-0.289*	0.001	0.067	-0.018*
0.75	11.128	0.357*	0.278	-0.233*	-0.005	0.093	-0.018
0.80	-0.351	0.391*	0.325	-0.176	-0.009	0.187	-0.003
0.85	-2.357	0.385	0.262	-0.229	-0.016*	0.255	0.018*
0.90	-1.381	0.380	0.238	-0.236	-0.016*	0.252	0.018*
0.95	-1.381	0.380	0.238	-0.236	-0.017*	0.252	0.018*

*Significant at level α=0.1

Table 6. The Comparison between SAR and SARQR Model.

Model	Spatial Dependency Test LM Value	Spatial Dependency Test *p-value*	Spatial Diversity Test BP Value	Spatial Diversity Test *p-value*	AIC
SAR	5.8601	0.0155	3.1674	0.0751	105.2200
SARQR τ=0.10	0.0549	0.8146	1.8686	0.1716	71.6224
SARQR τ=0.15	0.0052	0.9426	1.4958	0.2213	72.3596
SARQR τ=0.20	0.0066	0.9353	1.5040	0.2146	69.7271
SARQR τ=0.25	2.7343	0.0982	3.0631	0.0801	72.0653
SARQR τ=0.30	3.6432	0.0563	3.0673	0.0799	72.3285
SARQR τ=0.35	2.2484	0.1338	0.7179	0.3968	71.1260
SARQR τ=0.40	3.1256	0.0771	0.5489	0.4588	75.6256
SARQR τ=0.45	3.2947	0.0695	0.7238	0.3949	63.2113
SARQR τ=0.50	2.7432	0.0977	0.8668	0.3518	72.1255
SARQR τ=0.55	2.4893	0.1146	0.9020	0.3423	76.9123
SARQR τ=0.60	2.9768	0.0845	0.7254	0.3944	77.1333
SARQR τ=0.65	0.0300	0.8625	1.4136	0.2345	88.5306
SARQR τ=0.70	2.4576	0.1170	1.3566	0.2411	91.0339
SARQR τ=0.75	2.6164	0.1058	1.1538	0.2827	88.1333
SARQR τ= 0.80	3.8364	0.0502	1.0257	0.3112	69.8960
SARQR τ = 0.85	1.8343	0.1756	1.0736	0.3001	77.7989
SARQR τ = 0.90	2.3478	0.1255	0.9457	0.3308	78.8910
SARQR τ = 0.95	2.3478	0.1255	0.9457	0.3308	78.8910

on the results of spatial dependency tests and spatial diversity test. The criteria for the best model is based on the smallest value of AIC (Akaike Information Criteria) (Yasin et al., 2020), provided in Table 6.

It can be informed in Tabel 6 that all *p-value* for spatial dependency test and spatial diversity test are all more than 0.05. It indicated that in these SARQR model, there is no spatial dependency on the dependent variable for each quantile. In other words, the SARQR model can overcome the problem of spatial dependence on the dependent variable. Furthermore, in the inter-regional spatial diversity test, both methods (SAR and SARQR) have resulted in the homogenous variance of inter-regional for each model, indicated by all *p-value* being higher than 0.05. It means no longer problem with inter-regional spatial diversity.

Table 6 also informs us that model SARQR at all quantiles has a smaller value of AIC than the SAR model, with a model at τ=0.45 having the smallest value of AIC. The spatial outliers will be detected in this quantile, as the example. Figure 4 shows that the SARQR quantile 0.45 model has no spatial outliers, meaning there is no longer an outlier problem in this SARQR model. The proposed model for estimating the Open Unemployment Level is based on this SARQR model at τ=0.45:

$$\hat{Y} = -0.002WY + 14.659 - 0.197X_1 + 0.426X_2 - 0.336X_3 + 0.005X_4 + 0.104X_5$$

The proposed model at this 0.45 quantile informs us that: $\hat{\lambda}$=0.002, means that the Open Unemployment Level of each province has an effect of 0.002 times the average Open Unemployment Level of each neighboring province. An increase of 1% in the Poor population (X_2) will increase the Open Unemployment Level by 0.426, where other variables are considered constant. The proposed model also informed us that an increase of 1% in the Labor Force Participation Level or LFPR (X_3), will decrease the Open Unemployment Level by 0.336, where other variables are considered constant.

4. CONCLUSION

The comparison between the SAR model and the SARQR model based on the AIC value is the SARQR model is good at predicting OUL in West Java Province. The SARQR model is also proven to deal with spatial effect problems such as spatial dependence and spatial diversity and is not easily affected by the presence of outliers. The SARQR model can provide model information for each selected quantile of the response distribution, while the SAR model can only estimate the average response model. It is recommended for further research to use other parameter estimation methods such as GML (*Quasi Maximum Likelihood*), GMM (*Generalized Method of Moments*), and 2SLS (*Two Stage Least Square*). The comparison study among those methods is also important.

5. ACKNOWLEDGMENT

This research was funded by DRPM, the Deputy for Strengthening Research and Development of the Ministry of Research and Technology/National Research and Innovation Agency of

Indonesia, following Contract Number T/23/UN.16.17/PT.01.03/PDKN-Kesehatan/2022.

REFERENCES

Anselin, L. and L. Anselin (1988). The Scope of Spatial Econometrics. *Spatial Econometrics: Methods and Models*, **3**(4); 7–15

Candraningtyas, C. P. and A. B. Santosa (2022). Open Unemployment Rate Modeling in The Province of Jawa Timur in 2019-2021 using Data Panel Regression Method: Open Unemployment Rate Modeling in The Province of Jawa Timur in 2019-2021 using Data Panel Regression Method. *Jurnal Forum Analisis Statistik (FORMASI)*, **2**(2); 92–103

Central, B. S. (2021). *Tingkat Pengangguran Terbuka (TPT) Jawa Barat sebesar 9,82 persen*, volume 5. jabar bps

Dai, X. and L. Jin (2021). Minimum Distance Quantile Regression for Spatial Autoregressive Panel Data Models with Fixed Effects. *Plos One*, **16**(12); e0261144

Dai, X., S. Li, L. Jin, and M. Tian (2022). Quantile Regression for Partially Linear Varying Coefficient Spatial Autoregressive Models. *Communications in Statistics-Simulation and Computation*, **7**(9); 1–16

Dai, X., Z. Yan, M. Tian, and M. Tang (2020). Quantile Regression for General Spatial Panel Data Models With Fixed Effects. *Journal of Applied Statistics*, **47**(1); 45–60

Davino, C., M. Furno, and D. Vistocco (2013). *Quantile Regression: Theory and Applications*, volume 988. John Wiley & Sons

Hapsari, A. C. N. and M. I. Hasmarini (2022). Analysis of The Determining Factors of Open Unemployment Rate In East Java Province 2018-2020. *Eduvest-Journal of Universal Studies*, **2**(12); 2864–2871

Huang, H., M. A. Abdel Aty, and A. L. Darwiche (2010). County Level Crash Risk Analysis in Florida: Bayesian Spatial Modeling. *Transportation Research Record*, **2148**(1); 27–37

Jin, L., X. Dai, A. Shi, and L. Shi (2016). Detection of Outliers in Mixed Regressive-spatial Autoregressive Models. *Communications in Statistics Theory and Methods*, **45**(17); 5179–5192

Lum, K. and A. E. Gelfand (2012). Spatial Quantile Multiple Regression Using the Asymmetric Laplace Process. *Bayesian Analysis*, **7**; 235–258

McMillen, D. (2015). Conditionally Parametric Quantile Regression for Spatial Data: An Analysis of Land Values in Early Nineteenth Century Chicago. *Regional Science and Urban Economics*, **55**; 28–38

Su, L. and Z. Yang (2011). Instrumental Variable Quantile Estimation of Spatial Autoregressive Models. *Research Collection School Of Economics*, **7**; 1–35

Tribhuwaneswari, A. B., A. Hapsery, and W. K. Rahayu (2022). Spatial Autoregressive Quantile Regression as A Tool for Modelling Human Development Index Factors in 2020 East Java. *AIP Publishing LLC*, **2668**(1); 070007

Ver Hoef, J. M., E. E. Peterson, M. B. Hooten, E. M. Hanks, and M. J. Fortin (2018). Spatial Autoregressive Models for Statistical Inference from Ecological Data. *Ecological Monographs*, **88**(1); 36–59

Weisberg, S. (2014). *Applied Linear Regression (Fourth Edition)*. NJ: Wiley

Xu, P. and H. Huang (2015). Modeling Crash Spatial Heterogeneity: Random Parameter Versus Geographically Weighting. *Accident Analysis & Prevention*, **75**; 16–25

Yanuar, F., A. S. Deva, and A. Zetra (2023). Length of Hospital Stay Model of COVID-19 Patients with Quantile Bayesian with Penalty LASSO. *Communications in Mathematical Biology and Neuroscience*, **2023**; Article–ID

Yanuar, F., H. Yozza, and A. Zetra (2022). Modified Quantile Regression For Modeling The Low Birth Weight. *Frontiers in Applied Mathematics and Statistics*, **2**; 58

Yanuar, F. and A. Zetra (2021). Length of Stay of Hospitalized COVID-19 Patients Using Bootstrap Quantile Regression. *IAENG International Journal of Applied Mathematics*, **51**(3); 1–12

Yanuar, F., A. Zetra, C. Muharisa, D. Devianto, A. R. Putri, and Y. Asdi (2019). Bayesian Quantile Regression Method to Construct the Low Birth Weight Model. *Journal of Physics: Conference Series*, **1245**(1); 012044

Yasin, H., A. R. Hakim, and B. Warsito (2020). Development Life Expectancy Model in Central Java using Robust Spatial Regression with M-estimators. *Communications in Mathematical Biology and Neuroscience*, **69**; 1–16

Yasin, H., B. Warsito, A. R. Hakim, and R. N. Azizah (2022). Life Expectancy Modeling Using Modified Spatial Autoregressive Model. *Media Statistika*, **15**(1); 72–82

Yu, T., F. Gao, X. Liu, and J. Tang (2021). A Spatial Autoregressive Quantile Regression to Examine Quantile Effects of Regional Factors on Crash Rates. *Sensors*, **22**(1); 5

Zhang, J., Q. Lu, L. Guan, and X. Wang (2021a). Analysis of Factors Influencing Energy Efficiency Based on Spatial Quantile Autoregression: Evidence From the Panel Data in China. *Energies*, **14**(2); 504

Zhang, Y., J. Jiang, and Y. Feng (2021b). Penalized Quantile Regression for Spatial Panel Data With Fixed Effects. *Communications in Statistics-Theory and Methods*, **6**; 1–13

