Pakistan Journal of Statistics and Operation Research

# The Construction of Unemployment Rate Model Using SAR, Quantile Regression, and SARQR Model

Ferra Yanuar[1*], Tasya Abrari[2], Izzati Rahmi HG[3]

* Corresponding Author

1. Department of Mathematics and Data Science, Andalas University, Indonesia, ferrayanuar@sci.unand.ac.id
2. Department of Mathematics and Data Science, Andalas University, Indonesia, tasya28112000@gmail.com
3. Department of Mathematics and Data Science, Andalas University, Indonesia, izzatirahmihg@sci.unand.ac.id

**Abstract**

The Unemployment Rate can be used to determine how well the labor force can be absorbed by the current field positions, making unemployment a crucial statistic in the employment sector. The Central Bureau of Statistics of Indonesia noted that in August 2022, West Java was the province with the most unemployment contributors in Indonesia. The unemployment rate tends to be affected by spatial effect and its data is a skewed distribution. The purposes of this study are to analyze the performance of Spatial Autoregressive (SAR) model, Quantile regression, and Spatial Autoregressive Quantile Regression (SARQR) model in constructing the best model of the Unemployment Rate in West Java. This study used secondary data regarding the Unemployment Rate in 2022 of 27 regencies/cities in West Java, obtained from the Central Bureau of Statistics. This study found that (1) the SARQR model outperforms the SAR and Quantile regression at dealing with the problems of dependency and diversity in spatial data modeling and is not easily affected by the presence of outlier data (2) the SARQR model could give more information regarding the effects of variables which vary with the quantiles, (3) the Unemployment Rate in West Java is affected significantly by Labor Force Participation Level.

**Key Words:** The Unemployment Rate, spatial effects, SAR, Quantile Regression, SARQR.

**Mathematical Subject Classification:** 62C10, 62F10, 62F15.

## 1. Introduction

Unemployment is a problematic issue to deal with in developing countries, including Indonesia. One of the causes of unemployment is the lack of job opportunities 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 Unemployment Rate. In Indonesia, the unemployment rate measures the number of people actively looking for a job as a percentage of the labor force. The unemployment rate in Indonesia in 2021 was the fourth highest in Southeast Asia. The official rate of the unemployment rate of Indonesia in 2021 was 3.83%. It was higher than Thailand and Philippines, which were just 0.99% and 2.63% respectively. It was only lower than Malaysia, which was 4.05 (The Global Economic.com, 2022). Among provinces in Indonesia, West Java is the province with the highest value of unemployment rate.

Factors between regions usually tend to be involved as one factor affecting the unemployment rate in an area (Dai & 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. Influential outlier data cannot be thrown away because it will eliminate important information related to the data (Yasin et al., 2020, 2022). Modeling for data containing spatial effects and outliers can be modeled using a quantile regression approach. The quantile regression method also has powerful robustness and flexibility for handling data with outliers and skewed distribution (Yanuar et al., 2019, 2022; Yanuar & Zetra, 2021;

Yu et al., 2022). A combination of SAR and quantile regression then resulting in the Spatial Autoregressive Quantile Regression (SARQR) method, can also be used to model the data which contain spatial effect and skewed distribution (Dai et al., 2020; Dai & Jin, 2021; Jin et al., 2016).

The study of estimating and testing spatial autoregressive quantile models has recently increased. The quantile regression method for partially linear spatial autoregressive models with possibly varying coefficients using B-spline was investigated by Dai et al. (Dai et al., 2016). Dai et al. (Dai et al., 2020) investigated fixed effects quantile regression for general spatial panel data models with individual fixed and time effects based on the instrumental variable method. Zhang et al. (Y. Zhang et al., 2021) studied a penalized quantile regression for a spatial panel model with fixed effects. Dai and Jin (Dai & Jin, 2021) employed the minimum distance quantile regression (MDQR) methodology for estimating the SAR panel data model with individual fixed effects.

This study implements a SARQR model, which integrates spatial correlations and quantile effects, to address a research gap by estimating the effects of regional characteristics on the unemployment rate in 27 districts/cities in West Java and how those effects vary with the unemployment rate figure. The research questions of this study are how SAR, quantile regression, and SARQR provide an acceptable unemployment rate model, what are the significant risks associated with the Unemployment Rate in West Java, and how the unemployment rate affects the above effects. Thus, the objectives of this study are to analyze the performance of Spatial Autoregressive (SAR) model, Quantile regression, and Spatial Autoregressive Quantile Regression (SARQR) model in constructing the best model of the Unemployment Rate in West Java and determine the significant factors of Unemployment Rate model.

The rest of the research is structured as follows. Section 2 introduces the SAR, Quantile regression, and SARQR models. The available data used in this investigation, results regarding the implementation for all approaches, and discussion the goodness of fit for proposed model are described in Section 3. Section 4 summarizes the remarkable findings and gives further directions.

## 2. Methodology

Several stages of testing are required to ensure that the SAR method is suitable for modeling the cases raised. The first stage is a multicollinearity test between independent variables. 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 & Huang, 2015):

$$VIF_l = \frac{1}{1 - R_l^2}, \tag{1}$$

If the VIF value is less than 5, multicollinearity does not occur in the regression model (Yu et al., 2022).

### 2.1. 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 written in Eq. (2). 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, no spatial dependence is detected (Huang et al., 2010).

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

with $w_{ij}$ is an element of spatial weighting matrix $\boldsymbol{W}$, which its form is as follows

$$\boldsymbol{W} = \begin{bmatrix} w_{11} & w_{12} & \dots & w_{1n} \\ w_{21} & w_{22} & \dots & w_{2n} \\ \dots & \dots & \dots & \dots \\ w_{n1} & w_{n2} & \dots & w_{nn} \end{bmatrix},$$

where $n$ represents the number of observation areas. The spatial weighting matrix is a matrix that describes the relationship of each region. Its element, denoted by $w_{ij}$, for $i = j = 1,2,\dots,n$ obtained based on the results of standardized contiguity (Dai & Jin, 2021; Yasin et al., 2020):

$$w_{ij} = \frac{c_{ij}}{\sum_{i=1}^{n} c_{ij}},$$

where $c_{ij}$ =1 for $i$ and $j$ have contact and $c_{ij}$ = 0 if not. 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).

The existence of spatial dependence on the dependent variable is also checked using the Lagrange Multiplier (LM) test. The general form of the Lagrange Multiplier is as follows (Anselin, 1988):

$$LM_{lag} = \frac{\left(\frac{U'\boldsymbol{W}Y}{s^2}\right)^2}{nP}, \tag{3}$$

with

$$nP = T + \frac{(\boldsymbol{WX\beta})'\boldsymbol{M}(\boldsymbol{WX\beta})}{s^2},$$
$$T = trace\big((\boldsymbol{W} + \boldsymbol{W}')\boldsymbol{W}\big),$$
$$M = \boldsymbol{I} - \boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}',$$
$$s^2 = \frac{U'U}{n}, U \text{ is residual.}$$

If the $LM_{lag} > \chi^2_{\alpha,1}$, it indicates that there is a spatial dependence on the dependent variable. If a model contains spatial dependence, the modeling is carried out using a Spatial Autoregressive model (Anselin, 1988). As for detecting spatial diversity, it can use the *Breush-Pagan* (BP) test with the following formula:

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

where $f$ is a vector where its elements are $\frac{\widehat{U}_i^2}{\widehat{\hat{\sigma}}} - 1$ while $\widehat{U}_i, i = 1,\dots,n,$ is the observation from the regression estimation result and $\hat{\sigma}$ is the variance of the residuals, if the $BP$ more than $\chi^2_{(\alpha,k-1)}$ means that there is spatial variation between observation areas. Variable $k$ is the number of predictors in the hypothesis model.

### 2.2. 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, 1988):

$$\boldsymbol{Y} = \lambda\boldsymbol{WY} + \boldsymbol{X\beta} + \boldsymbol{U}\ , \quad \boldsymbol{U} \sim N(\boldsymbol{0}, \sigma^2\boldsymbol{I}), \tag{5}$$

where $\boldsymbol{Y}$ is a vector of dependent variables of size $n \times 1$, $\boldsymbol{X}$ is matrix of independent variables of size $n \times (k+1)$, $\lambda$ represents spatial autoregressive coefficient, which shows the magnitude of spatial dependence between regions, $\boldsymbol{W}$ denotes spatial weighting matrix of size $n \times n$, $\boldsymbol{\beta}$ is regression coefficient vector of size $(k+1) \times 1$, and $\boldsymbol{U}$ represent vector of residuals of size $n \times 1$.

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

$$\widehat{\boldsymbol{\beta}} = (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'^{(\boldsymbol{I}-\lambda\boldsymbol{W})}\boldsymbol{Y}. \tag{6}$$

While the estimation of parameter $\sigma^2$ in the SAR model is obtained as follows:

$$\widehat{\boldsymbol{\sigma^2}} = \frac{1}{n}\big((\boldsymbol{I} - \lambda\boldsymbol{W})\boldsymbol{Y} - \boldsymbol{X\beta}\big)'\big((\boldsymbol{I} - \lambda\boldsymbol{W})\boldsymbol{Y} - \boldsymbol{X\beta}\big), \tag{7}$$

Parameter estimation $\lambda$ can be obtained using a numerical approach (Anselin, 1988).

### 2.3. Quantile Regression

The Quantile regression method uses the approach of separating the data into certain quantile groups that may have different estimated values. The linear regression equation model for the quantile $\tau$ is as follows (Davino et al., 2014; Yanuar, 2018; Yanuar et al., 2019; Yanuar & Zetra, 2021):

$$\boldsymbol{Y} = \boldsymbol{X\beta}_\tau + \boldsymbol{U}, \qquad (8)$$

where $\boldsymbol{\beta_\tau}$ denotes factor of the regression coefficient of quantiles of size $(k+1) \times 1$ depending on the quantile-$\tau$.

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

$$\sum_{i=1}^{n} \rho_\tau(y_i - x_i\beta_i),$$

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

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

### 2.4. 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 ($\lambda$) and the regression vector ($\boldsymbol{\beta}$), which depend on certain quantile values (τ) (Ver Hoef et al., 2018). The development of SAR modelling on the quantile $\tau$th is specifically defined as follows:

$$\boldsymbol{Y} = \lambda_\tau \boldsymbol{WY} + \boldsymbol{X\beta}_\tau + \boldsymbol{U}. \qquad (9)$$

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., 2022). The assumptions used in estimating the parameters in SARQR model are as follows (J. Zhang et al., 2021):

1. $P(u_{it} \le 0) = \tau$, for every $i = 1,2,\dots,n, t = 1,2,\dots,T$.
2. For any t, $\sup_{n\ge1} \max_{1\le i\le n} E(u_{it}) \le \bar{u} < \infty$,
3. Given a value of $u_{it}$, the conditional distribution function $F(.|\bar{u}_{it})$ has a bounded continuous conditional probability density function $f(.|\bar{u}_{it})$. Where $\bar{u}_{it} = \sum_{k\ne i}^{n}\sum_{t=1}^{T} g_{ikt}u_{kt},\ g_{ikt}$ is the element of *G*.
4. *X* is non-random. Its absolute value is uniformly bounded and contains the intercept term.
5. The instrument variable *Z* is non-random with full column rank.

Getting the parameters $\lambda_\tau$ and $\beta_\tau$ are done by minimizing equation (10) respect to each parameter,

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

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 & Yang, 2011; J. Zhang et al., 2021):

(1). Set a specific value for $\lambda$, and do the quantile regression modeling at $\tau$th quantile, which is defined as

$$\left(\hat{\beta}_\tau(\lambda), \hat{\gamma}_\tau(\lambda)\right) = argmin_{\beta,\gamma} Q_\tau(\beta, \lambda, \gamma).$$

(2). To calculate the estimated value of the IVQR is done by minimizing the vector of the estimated variable instrument $\hat{\gamma}_\tau(\lambda)$

$$\hat{\lambda}_\tau = argmin_\lambda\ \hat{\gamma}_\tau(\lambda)\hat{A}\left(\hat{\gamma}_\tau(\lambda)\right)',$$

where $\hat{A} = A + O_p(1)$, A is a positive definite matrix.

(3). The estimator $\beta$ is obtained in the following way.

$$\hat{\beta}_\tau = \hat{\beta}_{\tau-1}\,\hat{\lambda}_{\tau-1}$$

Repeat the above steps for each quantile $\tau$. At each quantile $\tau$, different estimating parameters are obtained.

## 3. Unemployment Rate Model

The data utilized in this section were secondary data from the West Java Province collected in 2022 by the Central Bureau of Statistics. There are 18 regencies and 9 cities in the West Java Province. 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 Life Expectancy or LE ($X_5$). The response variable is the Unemployment Rate. The descriptive of the corresponding data is presented in Table 1, while Figure 1 is the boxplot which informs us the existence of outlier in the data.

**Table 1**. Descriptive Statistics on Data.

| Variable | Mean | S.D | Min | Max |
|---|---|---|---|---|
| % Population ($X_1$) | 3.81 | 2.42 | 0.42 | 11.25 |
| % Poor population ($X_2$) | 8.63 | 2.88 | 2.53 | 12.77 |
| LFPR ($X_3$) | 65.98 | 2.34 | 61.80 | 69.98 |
| GRDP ($X_4$) | 61.49 | 67.11 | 3.51 | 265.13 |
| % HDI ($X_5$) | 72.75 | 1.41 | 69.95 | 75.48 |
| Unemployment ($Y$) | 8.04 | 2.01 | 3.75 | 10.78 |

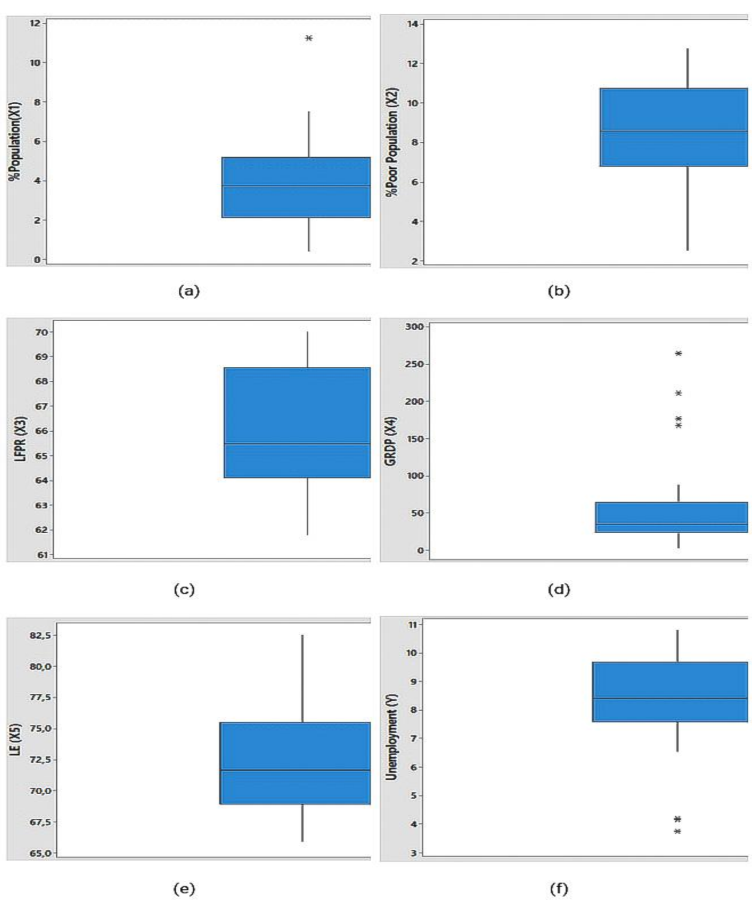

**Figure 1**. Box plot of variables (a) % Population ($X_1$), (b) % Poor population ($X_2$), (c). LFPR ($X_3$), (d). GRDP ($X_4$), (e). % LE ($X_5$), and (f). Unemployment rate ($Y$).

In order to determine whether there was a correlation between each independent variable, the multicollinearity test was performed among the independent variables. Table 2 displays the outcomes of the multicollinearity test among independent variables. If each variable's VIF value is less than 5, there will not be any issues with multicollinearity amongst the independent variables.

**Table 2**. Multicollinearity Test

| Variable | VIF |
|---|---|
| % Population ($X_1$) | 1.85 |
| % Poor population ($X_2$) | 1.75 |
| LFPR ($X_3$) | 1.16 |
| GRDP ($X_4$) | 2.19 |
| % LE ($X_5$) | 1.81 |

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. Table 3 presents the list of regencies/cities in the adjacent West Java. Meanwhile the distribution of Unemployment Rate value (in percentage) for 27 regions in West Java are presented in Figure 1.

**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 |
| Bandung | Cianjur, Garut, Sumedang, Subang, West Bandung, Bandung, Cimahi City |
| Garut | Cianjur, Bandung, Tasikmalaya, and Sumedang |
| Tasikmalaya | Garut, Ciamis, Majalengka, Sumedang, and Tasikmalaya City |
| Ciamis | Tasikmalaya, Kuningan, Majalengka, Tasikmalaya City, and Banjar |
| Kuningan | Ciamis, Cirebon, and Majalengka |
| Cirebon | Kuningan, Majalengka, Indramayu, and Cirebon City |
| Majalengka | Tasikmalaya, Ciamis, Kuningan, Cirebon, Sumedang, and Indramayu |
| Sumedang | Bandung, Garut, Tasikmalaya, Majalengka, Indramayu, and Subang |
| Indramayu | Cirebon, Majalengka, Sumedang, and Subang |
| Subang | Bandung, Sumedang, Indramayu, Karawang, Purwakarta, and West Bandung |
| Purwakarta | Bogor, Cianjur, Subang, Karawang, and West Bandung |
| Karawang | Bogor, Subang, Purwakarta, and Bekasi |
| Bekasi | Bogor, Karawang, and Bekasi City |
| West Bandung | Cianjur, Bandung, Subang, Purwakarta, Bandung, and Cimahi |
| City Bogor | Bogor |
| Sukabumi | Sukabumi |
| Bandung City | Bandung, West Bandung, and Cimahi City |
| Cirebon City | Cirebon |
| Bekasi City | Bogor, Bekasi, and Depok |
| Depok City | Bogor and Bekasi |
| Cimahi City | Bandung, West Bandung, Bandung City |
| Tasikmalaya City | Tasikmalaya and Ciamis |

| | |
|---|---|
| Banjar City | Ciamis |

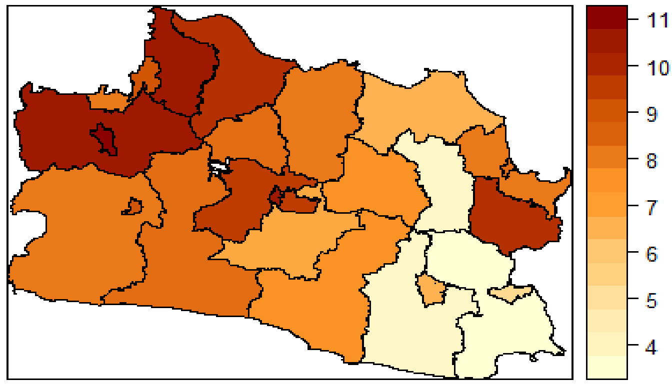


**Figure 1**. Unemployment Rate (in percentage) for 27 Regions in West Java.

In this section, the empirical data regarding the Unemployment Rate in 27 regencies/cities in West Java is employed to construct the Unemployment Rate model. The first step is to estimate the Moran's Index coefficient values based on Eq. (2). It was obtained the Moran's index, $I = 0.571$. 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 determine the spatial dependence on the dependent variable using Lagrange Multiplier test, as written in Eq. (3). This study found that the Lagrange Multiplier value is 9.177 (*p-value* is 0.002), which means that there is a spatial dependence on the dependent variable. Then, 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 1.394 (with *p-value* is 0.238). It indicates that the variances of Unemployment Rate among regions in East Java are the same, or there is no spatial diversity problem in the Unemployment Rate data.

Based on these preliminaries study, it was found that there is a spatial effect in the form of spatial dependence on the dependent variable. Thus, SAR model was used to obtain the acceptable model. Table 4 presents the results of the parameter estimation based on the SAR model.

**Table 4**. Model Estimation Results with the SAR Model.

| Variable | Estimated Mean | Standard Error | *t-value* | *p-value* |
|---|---|---|---|---|
| Constant | -22.618 | 21.374 | -1.058 | 0.290 |
| % Population ($X_1$) | -0.071 | 0.139 | -0.514 | 0.607 |
| % Poor population ($X_2$) | 0.219* | 0.112 | 1.939 | 0.050 |
| LFPR ($X_3$) | -0.195 | 0.113 | -1.711 | 0.087 |
| GRDP ($X_4$) | 0.009 | 0.005 | 1.665 | 0.096 |
| % LE ($X_5$) | 0.511* | 0.238 | 2.146 | 0.032 |
| SAR coefficient ($\lambda$) | 0.533* | 0.163 | 3.270 | 0.001 |

* Significant at the significant level $\alpha = 0.05, Z_{\alpha/2} = 1.96$

Based on Table 4, it is found that not all independent variables have a significant influence on regencies/cities' Unemployment Rate in West Java Province. The independent variables that have a significant effect on the Unemployment Rate are the percentage of poor people ($X_2$) and the percentage of LE ($X_5$). Therefore, the proposed model for the Unemployment Rate based on the SAR model is:

$$\hat{Y} = 0.533\boldsymbol{WY} - 22.618 + 0.219X_2 + 0.511X_5 \quad (11)$$

The proposed model in Eq. (11) then is checked whether the proposed SAR model can handle spatial effects properly or not. Spatial effect testing consists of spatial dependence and spatial diversity. In spatial dependence, the Lagrange Multiplier value is 0.031 (p-*value* is 0.860) meaning that the SAR model does not contain the effect of spatial dependence on the dependent variable anymore. In terms of spatial diversity, which is estimated using Eq. (4), it was obtained that the Breusch-Pagan value is 0.575 with *p-value* = 0.448. This result informs us that regions have the same variance. *Moran's scatterplot* is figured then to detect the existence of spatial outliers in the SAR model. *Moran's scatterplot* is provided in Figure 2. Looking at the figure, it is found that there are four spatial outliers in the SAR model, these are residuals from observation numbers 6, 9, 23, and 25. Therefore, the Unemployment Rate model obtained based on the SAR method in Eq. (11) could not be accepted. Thus, quantile regression and SARQR are then applied to deal with spatial effects and data containing spatial outliers to obtain a more acceptable model.

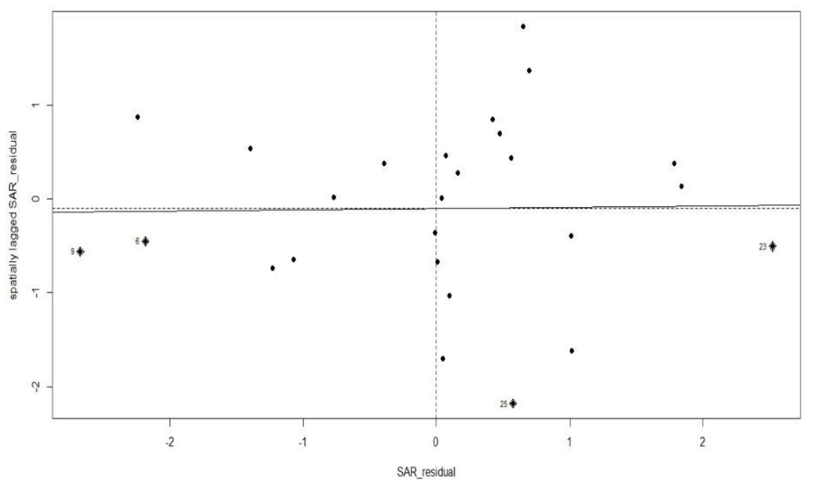


**Figure 2**. Spatial Outliers in the SAR Model.

The results for the estimated parameter based on the Quantile regression model and SARQR are presented in Table 5 and 6, respectively. In Quantile regression, the accuracy of the models crossing the entire distribution of the Unemployment Rate is predicted by setting the quantile value to $\tau = 0.1, 0.2, \ldots, 0.9$. Quantile regression and SARQR modeling produce a model that may differ in each quantile. SARQR modeling 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.

**Table 5**. Estimated Parameter Based on Quantile Regression

| Quantile | Estimated Parameters | | | | | |
|---|---|---|---|---|---|---|
| ($\tau$) | Constants | $X_1$ | $X_2$ | $X_3$ | $X_4$ | $X_5$ |
| 0.1 | -34.410 | -0.432* | 0.205 | -0.496* | 0.023* | 0.983* |
| 0.2 | -36.221 | -0.145 | 0.088 | -0.128 | 0.016 | 0.690 |
| 0.3 | -55.428 | -0.171 | 0.263 | -0.166 | 0.016 | 0.976 |
| 0.4 | -41.197 | 0.083 | 0.247 | -0.080 | 0.012 | 0.701 |
| 0.5 | 12.631 | -0.076 | 0.044 | -0.198 | 0.012 | 0.112 |
| 0.6 | 8.078 | 0.104 | 0.023 | -0.243 | 0.008 | 0.214 |
| 0.7 | 11.584 | 0.084 | -0.027 | -0.288 | 0.007 | 0.217 |
| 0.8 | -16.187 | 0.130 | -0.245 | -0.187 | 0.001 | 0.545 |
| 0.9 | -30.249 | 0.190 | -0.125 | -0.038 | -0.001 | 0.591 |

*Significant at level α= 0.05

Based on Table 5, it is informed that only at quantile $\tau = 0.1$ has significant parameters on the response. Meanwhile, the SARQR model could give more information regarding the effects of variables that vary with the quantiles. It can be seen in Table 6 that percentage of the population ($X_1$), percentage of the poor population ($X_2$), LFPR ($X_3$), GRDP ($X_4$), and the percentage of LE ($X_5$) are significant in different selected quantiles. The spatial autoregressive coefficients on the SARQR model are varies among quantiles, some are positive, others 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.

**Table 6**. Estimated Parameter Based on SARQR

| Quantile | Estimated Parameters | | | | | | |
|---|---|---|---|---|---|---|---|
| ($\tau$) | Constants | $X_1$ | $X_2$ | $X_3$ | $X_4$ | $X_5$ | $\lambda_\tau$ |
| 0.1 | -83.795* | -0.551* | 0.497* | -0.251 | 0.019* | 1.359* | 0.008* |
| 0.2 | -44.692 | -0.492 | 0.424 | -0.223 | 0.017 | 0.768 | 0.165 |
| 0.3 | -37.419 | -0.317 | 0.495 | -0.271 | 0.013 | 0.726* | -0.029 |
| 0.4 | 2.415 | -0.130 | 0.236 | -0.319 | 0.010 | 0.277 | -0.005 |
| 0.5 | -0.795 | -0.038 | 0.160 | -0.215 | 0.008 | 0.241 | 0.031 |
| 0.6 | -8.719 | 0.076 | 0.178 | -0.271* | 0.004 | 0.394 | -0.023 |
| 0.7 | -16.329 | 0.156 | 0.126 | -0.211* | 0.002 | 0.452 | -0.018* |
| 0.8 | -21.688 | 0.185 | 0.110 | -0.220* | -0.105 | 0.537 | -0.017* |
| 0.9 | -36.276 | 0.211 | 0.081 | -0.210 | -0.010 | 0.722* | 0.005* |

*Significant at level α= 0.05

**Table 7**. The Goodness of Fit test Based on SAR, Quantile Regression, and SARQR Model

| Model | Quantile | Goodness of Fit | |
|---|---|---|---|
| | | MAE | RMSE |
| SAR | - | 1.207 | 1.549 |
| Quantile Regression | $\tau = 0.1$ | 4.055 | 5.155 |
| | $\tau = 0.2$ | 3.341 | 4.186 |
| | $\tau = 0.3$ | 3.309 | 4.150 |
| | $\tau = 0.4$ | 2.227 | 2.800 |
| | $\tau = 0.5$ | 1.506 | 1.909 |
| | $\tau = 0.6$ | 1.470 | 1.814 |
| | $\tau = 0.7$ | 1.536 | 1.945 |
| | $\tau = 0.8$ | 2.578 | 3.352 |
| | $\tau = 0.9$ | 2.501 | 3.348 |
| SARQR | $\tau = 0.1$ | 1.108 | 1.547 |
| | $\tau = 0.2$ | 1.106 | 1.539 |
| | $\tau = 0.3$ | 1.104 | 1.523 |
| | $\tau = 0.4$ | 1.103 | 1.485 |
| | $\tau = 0.5$ | 1.100 | 1.482 |
| | $\tau = 0.6$ | 1.017 | 1.456 |
| | $\tau = 0.7$ | 0.990 | 1.431 |
| | $\boldsymbol{\tau = 0.8}$ | **0.903** | **1.365** |
| | $\tau = 0.9$ | 0.905 | 1.379 |

Two common measures, i.e., mean absolute error (MAE) and root mean square error (RMSE) were used to evaluate the model performance of Quantile regression and SARQR models. The lower values of MAE and RMSE are the predicted accuracy of the corresponding models (Yanuar et al., 2023). As observed in Table 7, the MAE and RMSE values at each quantile of the SARQR model were much lower than that of the SAR and Quantile regression model. This result indicates that the SARQR model outperforms the SAR and the Quantile regression model in data Unemployment Rate. This study also found that the lowest value of MAE and RMSE is based on the SARQR model at quantile $\tau = 0.8$, MAE is 0.905 and RMSE is 1.365. We could conclude the best model for the Unemployment Rate model is affected significantly by Labor Force Participation Level ($X_3$), based on quantile $\tau = 0.8$:

$$\widehat{\boldsymbol{Y}} = -0.017\boldsymbol{WY} - 21.688 - 0.220X_3 \quad (12)$$

Table 8 presents the result of the spatial effect test based on the SARQR model. This study proved that the SARQR model for all selected quantiles could overcome the dependence problem, indicated by the p-value for each quantile higher than 0.05. Furthermore, in the inter-regional spatial diversity test, SARQR has 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. The final check is the spatial outliers test using *Moran's scatterplot* will be figured based on the SARQR model. We choose the model at quantile $\tau = 0.8$, as the best model. Looking at Figure 3, this figure informs us that the SARQR quantile 0.8 model has no spatial outliers, meaning there is no longer outlier problem in this SARQR model.

**Table 8**. The Spatial Effect Test Based on SARQR Model.

| Model | Spatial Dependency Test | | Spatial Diversity Test | |
|---|---|---|---|---|
| | LM value | *p-value* | BP value | *p-value* |
| SARQR $\tau = 0.1$ | 0.149 | 0.699 | 0.076 | 0.783 |
| SARQR $\tau = 0.2$ | 0.523 | 0.469 | 0.015 | 0.901 |
| SARQR $\tau = 0.3$ | 0.758 | 0.383 | 0.039 | 0.843 |
| SARQR $\tau = 0.4$ | 1.278 | 0.258 | 0.350 | 0.554 |
| SARQR $\tau = 0.5$ | 0.683 | 0.409 | 0.716 | 0.397 |
| SARQR $\tau = 0.6$ | 0.370 | 0.543 | 0.117 | 0.732 |
| SARQR $\tau = 0.7$ | 0.560 | 0.454 | 0.415 | 0.519 |
| SARQR $\tau = 0.8$ | 0.246 | 0.620 | 0.130 | 0.719 |
| SARQR $\tau = 0.9$ | 0.232 | 0.630 | 0.215 | 0.642 |

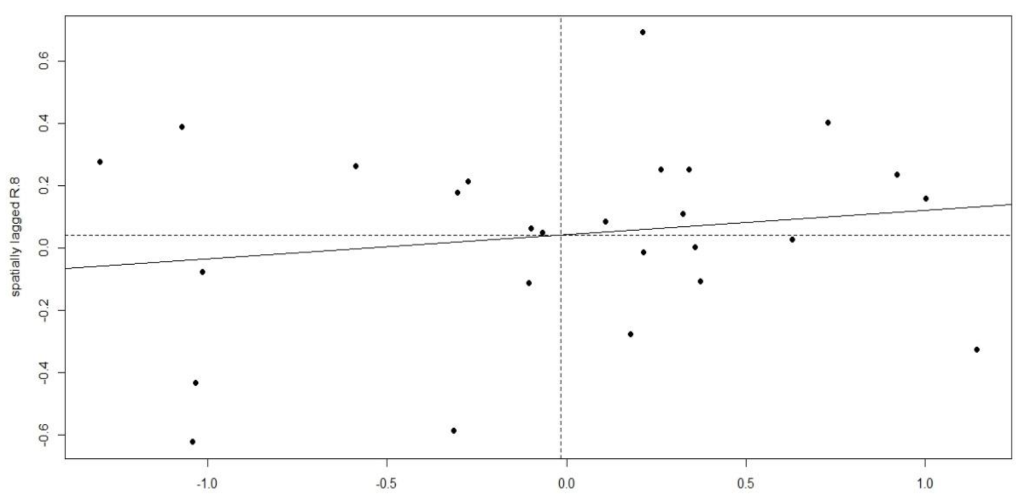


**Figure 3**. Spatial Outliers Test Based on SARQR Model at $\tau = 0.8$.

## 4. **Conclusion**

Through the empirical analysis to the Unemployment Rate of 27 regencies/cities in West Java at different quantiles, this study found that the significant risk factor of Unemployment Rate are Labor Force Participation Level and Life Expectancy. There is a negative effect of Labor Force Participation Level to the Unemployment Rate, especially at high level of Unemployment Rate. Meanwhile the Life Expectancy give positive effect at the lower quantile ($\tau = 0.1\ and\ 0.3$) of Unemployment Rate and at the quantile $\tau = 0.9$. Since, the best model is at quantile $\tau = 0.8$ where Life Expectancy does not give significant effect at this quantile, thus this factor is not involved in the proposed model.

By comparing with SAR and Quantile regression, this study shows the advantages of SARQR model in the analysis. SARQR could result the acceptable predicting model of Unemployment Rate 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.

**Acknowledgments**

This research was funded by DRTPM, Ministry of Education, Culture, Research and Technology (Kemendikbud-Ristek) of Indonesia, following Contract Number T/4/UN16.19/PT.01.03/PDKN-Kesehatan/2023.

## References

1. Anselin, L. (1988). The Scope of Spatial Econometrics. In L. Anselin, *Spatial Econometrics: Methods and Models* (Vol. 4, pp. 7–15). Springer Netherlands. https://doi.org/10.1007/978-94-015-7799-1_2
2. Dai, X., & Jin, L. (2021). Minimum distance quantile regression for spatial autoregressive panel data models with fixed effects. *PLOS ONE*, *16*(12), e0261144. https://doi.org/10.1371/journal.pone.0261144
3. Dai, X., Li, S., & Tian, M. (2016). *Quantile Regression for Partially Linear Varying Coefficient Spatial Autoregressive Models*. https://doi.org/10.48550/ARXIV.1608.01739
4. Dai, X., Yan, Z., Tian, M., & Tang, M. (2020). Quantile regression for general spatial panel data models with fixed effects. *Journal of Applied Statistics*, *47*(1), 45–60. https://doi.org/10.1080/02664763.2019.1628190
5. Davino, C., Furno, M., & Vistocco, D. (2014). *Quantile Regression Theory and Applications*. Wiley Series.
6. Huang, H., Abdel-Aty, M. A., & Darwiche, A. L. (2010). County-Level Crash Risk Analysis in Florida: Bayesian Spatial Modeling. *Transportation Research Record: Journal of the Transportation Research Board*, *2148*(1), 27–37. https://doi.org/10.3141/2148-04
7. Jin, L., Dai, X., Shi, A., & Shi, L. (2016). Detection of outliers in mixed regressive-spatial autoregressive models. *Communications in Statistics - Theory and Methods*, *45*(17), 5179–5192. https://doi.org/10.1080/03610926.2014.941493
8. Su, L., & Yang, Z. (2011). Instrumental Variable Quantile Estimation of Spatial Autoregressive Models. *Research Collection School Of Economics*, 1–35.
9. The Global Economic.com. (2022). *Unemployment rate, 2021—Country rankings*. https://www.theglobaleconomy.com/rankings/unemployment_rate/South-East-Asia/
10. Ver Hoef, J. M., Peterson, E. E., Hooten, M. B., Hanks, E. M., & Fortin, M.-J. (2018). Spatial autoregressive models for statistical inference from ecological data. *Ecological Monographs*, *88*(1), 36–59. https://doi.org/10.1002/ecm.1283
11. Weisberg, S. (2014). *Applied linear regression* (Fourth edition). Wiley.
12. Xu, P., & Huang, H. (2015). Modeling crash spatial heterogeneity: Random parameter versus geographically weighting. *Accident Analysis & Prevention*, *75*, 16–25. https://doi.org/10.1016/j.aap.2014.10.020
13. Yanuar, F. (2018). Sample size and power calculation for univariate case in quantile regression. *Journal of Physics: Conference Series*, *948*, 012072. https://doi.org/10.1088/1742-6596/948/1/012072
14. Yanuar, F., Deva, A. S., Zetra, A., & Maiyastri. (2023). Length of hospital stay model of COVID-19 patients with quantile Bayesian with penalty LASSO. *Communications in Mathematical Biology and Neuroscience*, *23*. https://doi.org/10.28919/cmbn/7881
15. Yanuar, F., Yozza, H., & Zetra, A. (2022). Modified Quantile Regression for Modeling the Low Birth Weight. *Frontiers in Applied Mathematics and Statistics*, *8*, 890028. https://doi.org/10.3389/fams.2022.890028

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

17. Yanuar, F., Zetra, A., Muharisa, C., Devianto, D., Putri, A. R., & Asdi, Y. (2019). Bayesian Quantile Regression Method to Construct the Low Birth Weight Model. *Journal of Physics: Conference Series*, *1245*, 012044. https://doi.org/10.1088/1742-6596/1245/1/012044

18. Yasin, H., Hakim, A. R., & Warsito, B. (2020). Development life expectancy model in Central Java using robust spatial regression with M-estimators. *Communications in Mathematical Biology and Neuroscience*, *69*, 1–16. https://doi.org/10.28919/cmbn/4984

19. Yasin, H., Warsito, B., Hakim, A. R., & Azizah, R. N. (2022). Life Expectancy Modeling Using Modified Spatial Autoregressive Model. *Media Statistika*, *15*(1), 72–82. https://doi.org/10.14710/medstat.15.1.72-82

20. Yu, T., Gao, F., Liu, X., & Tang, J. (2022). A Spatial Autoregressive Quantile Regression to Examine Quantile Effects of Regional Factors on Crash Rates. *Sensors*, *22*(5), 1–15. https://doi.org/10.3390/s22010005

21. Zhang, J., Lu, Q., Guan, L., & Wang, X. (2021). Analysis of Factors Influencing Energy Efficiency Based on Spatial Quantile Autoregression: Evidence from the Panel Data in China. *Energies*, *14*(2), 504. https://doi.org/10.3390/en14020504

22. Zhang, Y., Jiang, J., & Feng, Y. (2021). Penalized quantile regression for spatial panel data with fixed effects. *Communications in Statistics - Theory and Methods*, 1–13. https://doi.org/10.1080/03610926.2021.1934028