EXAMINING REGIONAL FACTORS ON MALNUTRITION RATE IN INDONESIA USING SPATIAL AUTOREGRESSIVE APPROACH

FERRA YANUAR[1,*], TASYA ABRARI[1], AIDINIL ZETRA[2], IZZATI RAHMI HG[1], DODI DEVIANTO[1],

SYARIFATUL AHDA[1]

[1]Department of Mathematics and Data Science, Andalas University, Padang, Indonesia

[2]Department of Political Sciences, Andalas University, Padang, Indonesia

Abstract: The most frequent issues in malnutrition rate modeling analysis are skewed distribution and spatial autocorrelation. Previous researches were generally focused on spatial autocorrelation between neighboring regions or auto relationships between malnutrition rates and significant factors across different quantiles of the malnutrition rate distribution, but rarely both. This study aims to estimate how contributing factors influence the malnutrition rate. The estimation is carried out by implementing the spatial autoregressive (SAR) approaches, including ordinary SAR, Robust SAR and SAR Quantile (SARQ), using 2021 data from the Health Ministry of Indonesia. The result shows that the SARQ outperforms the SAR and the Robust SAR in data fitness and prediction accuracy. The SARQ is also insensitive to outliers and skewed distribution. Estimation using SARQ provides effects of explanatory variables vary with the quantiles, while SAR and RSAR cannot do.

Keywords: spatial autoregression (SAR); robust SAR; SAR quantile; malnutrition rate; outliers.

2020 AMS Subject Classification: 62F15.

---

*Corresponding author

E-mail address: ferrayanuar@sci.unand.ac.id

1. INTRODUCTION

Nutritional issues are a public health problem that can affect all age groups. The most affected age group is the 0-5 years old (infants). Underweight children under five have a higher risk factor for stunting than adults. Growth obstacles in early age will adversely affect the quality of future generations [1], [2].

Indonesia is a prime example of the burden of malnutrition. About 1 in 3 children under five years is stunted [3]. In September 2021, UNICEF Indonesia screened 84 million children under the age of five and identified 500,000 new cases of children suffering from wasting, which carries an increased risk of death. Based on the data from the Indonesian Toddler Nutrition Status Survey in 2021, the prevalence of toddlers stunting is 24.4 percent (5.33 million). The prevalence of stunting has decreased from previous years, that is 26.92 percent in 2020. The Covid-19 pandemic, which has entered its third year, has led to an increase malnourisment leading to stunting for children under five. By 2024, Indonesia is targeting to lower the stunting rate by 14 percent. To achieve this target, a situational analysis is required to determine the malnutrition rate among number children under the age of five and to assess key determinants of the malnutrition in specific social and geographical locations [4]. This analysis will provide grounds for intervention so that the actions can be tailored to address the contextual needs in Indonesia. Studies have previously been made to appropriately analyze childhood stunting and wasting in developing countries, including Indonesia. Unfortunately, most of the analyses emphasized on modeling mean regression instead of quantile regression and did not consider the spatial effect.

It is evident that highly left-skewed distribution and outliers (e.g., minimum figure) malnutrition data due to stunting and wasting, commonly present significant spatiotemporal differences [5], [6]. There is the possibility of spatial dependence in the data, meaning that observations in one can influence observation in other areas. As Tobler states that everything is related to one another, but something adjacent is more influential [7]. Linear regression analysis ignores the existence of this spatial dependence, this method is also highly sensitive to outliers and fail to be extended to the full distribution of malnutrition, which may make these models output biased results and side

effects [8]. However, it is more attractive for researchers to understand the effects of risky factors in spatial with high malnutrition cases than in other locations. Since there is serious problem relate to stunting in these spatial, it is inevitable to implement the spatial auto regressive (SAR) method [9]. Hybrid of SAR and quantile regression (QR) is also applied here, it is assumed this method will produce better model in such of this case [10]–[14].

There has recently been an increase in the research on estimating and testing spatial auto regressive models. For instance, Dai et al. [15] explored the quantile regression approach for partially linear spatial auto regressive models with possibly varying coefficients using B-spline. Dai et al. [16] investigated fixed effects quantile regression for general spatial panel data models with both individual fixed effects and time period effects based on the instrumental variable method. Zhang et al. [17] studied a penalized quantile regression for a spatial panel model with fixed effects. Dai and Jin [9] employed the minimum distance quantile regression (MDQR) methodology for estimating the SAR panel data model with individual fixed effects.

This study applies the hybrid of spatial auto regressive and quantile approach (SARQ) to achieve the best model of malnutrition data. The proposed method will be compared with spatial auto regressive (SAR) and Robust SAR. As far as we know, the SARQ model has not been applied in any existing references on this subject. The model is important to investigate the factors impacting malnutrition incidence in Indonesia and advise decision-makers on reducing malnutrition.

The rest of the paper is organized as follows, Section 2 is about the methods and models including the variables involved in the model, SAR, Robust SAR, and SARQ methods. In Section 3, we implement the methods to construct an acceptable model regarding malnutrition factors in Indonesia.

2. PRELIMINARIES

2.1. Data

This study assumed that the percentage of low birth weight, percentage of vitamin A, percentage of health insurance, percentage to access water hygiene, and percentage of breastfeeding as factors

influencing child malnutrition, based on the study by many researchers [18]–[21]. The hypothesis model then fitted to the data set regarding malnutrition obtained from the study's results on the nutritional status (stunted) of children under five in thirty-four provinces in Indonesia in 2021 [22]. Three estimation methods are applied in this study to achieve the best model of malnutrition rate in all provinces in Indonesia. The detailed descriptive statistics about these candidate variables are recorded in Table 1.

Table 1. Descriptive Statistics of Variables

Variables	No. obs	Min	Average	Max	Std. Dev
Percentage of malnutrition (Y)	34	16.80	25.27	33.80	4.53
Percentage of health insurance (X_1)	34	11.60	38.96	73.60	14.78
Percentage access the water hygiene (X_2)	34	42.40	70.31	94.50	12.78
Percentage of breast-feeding (X_3)	34	38.70	51.97	70.40	8.55

2.2. Methods and Preliminary Analysis

A preliminary diagnostic test, i.e., multicollinearity, was applied to the hypotheses model. The variance inflation factor (VIF) was calculated to identify the statistical correlation using the following formula:

$$VIF = \frac{1}{1 - R^2},$$ (1)

with R^2 is coefficient of determination of corresponding model. The variables with VIF values greater than five would be removed from the analysis [8], [23]. It indicates that the corresponding variable has a significant correlation with others, or it is said that multicollinearity problem is due in this case. Table 2 presents the VIF values for each variable. From the table, VIF's values for each variable are less than 5, it means no multicollinearity issue in this data. Therefore, all five variables are included in the analysis.

Table 2. VIF Value

Variables	VIF Value
Percentage of health insurance (X_1)	1.219
Percentage access the water hygiene (X_2)	1.302
Percentage of breast-feeding (X_3)	1.104

The Moran's I test was applied subsequently to identify whether there is a significant spatial correlation of variables among neighbors. Moran's I is defined as follows:

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

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

where $S^2 = \frac{1}{n}\sum_{i=1}^{n}(y_i - \bar{y})^2$, with n is the number of areas (spatial units), y_i the malnutrition rate of the i-th region, w_{ij} is the element of the spatial weight matrix, which represents the adjacent relationship between the i-th region and the j-th region. It is defined as follows:

$$w_{ij} = \begin{cases} 1 & \text{If area } i \text{ is adjacent to area } j \\ 0 & \text{If area } i \text{ is not adjacent to area } j. \end{cases}$$

In this study, the queen contiguity method is applied to design the spatial weight matrix, w_{ij}. That is, if there is a common boundary between the two regions, they are considered adjacent. A spatial weight matrix can then be constructed using this method. In this study, the spatial unit is provinces in Indonesia. The adjacent provinces in Indonesia are presented in Table 3. Four provinces, i.e., Bangka Belitung Island, Bali, West Nusa Tenggara, East Nusa Tenggara, are not included in the analysis since all four provinces are the island and have no adjacent provinces. Figure 1 provides the map of malnutrition' percentage in each province in Indonesia. Each region has different characteristics and form some groups. Therefore, a spatial regression is needed in modeling the malnutrition case in Indonesia.
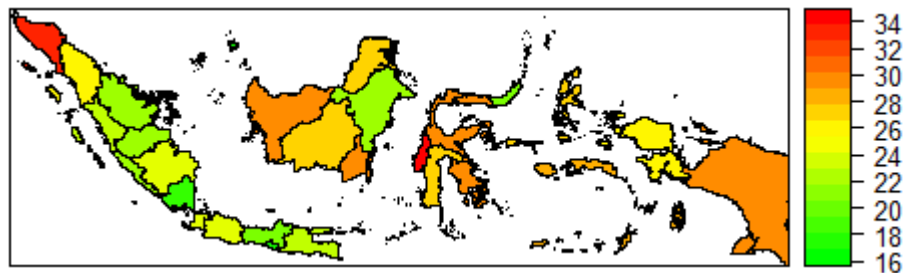
Figure 1. Spatial Distribution of Malnutrition in Indonesia (in percentage)

Table 3. The Neighboring Provinces in Indonesia

No	Province	Neighboring Province
1	Gorontalo	North Sulawesi, Central Sulawesi
2	Riau	North Sumatera, Jambi, West Sumatera
3	Jakarta	Banten, West Java
4	Riau Islands	Jambi
5	South Sulawesi	West Sulawesi, Southeast Sulawesi, Central Sulawesi
6	South Sumatera	Bengkulu, Lampung, Jambi
7	West Kalimantan	Central Kalimantan, East Kalimantan
8	Aceh	North Sumatera
9	Bengkulu	South Sumatera, Lampung, Jambi, West Sumatera
10	Lampung	Bengkulu, South Sumatera
11	Central Java	East Java, Yogyakarta, West Java
12	North Sumatera	Riau, Aceh, West Sumatera
13	Jambi	Riau, Riau Islands, South Sumatera, Bengkulu, West Sumatera
14	Banten	Jakarta, West Java
15	East Jawa	Central Java
16	Yogyakarta	Central Java
17	South Kalimantan	Central Kalimantan, East Kalimantan
18	North Sulawesi	Gorontalo
19	North Kalimantan	East Kalimantan
20	West Sulawesi	South Sulawesi, Central Sulawesi
21	North Maluku	Maluku
22	Central Kalimantan	West Kalimantan, East Kalimantan, South Kalimantan
23	West Papua	Papua
24	West Java	Jakarta, Central Java, Banten
25	Southeast Sulawesi	South Sulawesi, Central Sulawesi
26	West Sumatera	Riau, Bengkulu, North Sumatera, Jambi
27	Papua	West Papua
28	East Kalimantan	West Kalimantan, South Kalimantan, North Kalimantan, Central Kalimantan
29	Maluku	North Maluku
30	Central Sulawesi	Gorontalo, South Sulawesi, West Sulawesi, Southeast Sulawesi

The Moran's I value commonly ranges from -1 to +1. The positive and negative values of the Moran index indicate that the variable is spatial clustering or spatial dispersion [24]. If the value is close to zero or the p-value is less than 0.1, the variable does not have an easy-observed spatial pattern [8]. Thus, spatial dispersion would not be considered in this study. The Moran's I value obtained in this study is 0.3622 (p-value is 0.014), which is indicated that there is a spatial correlation of variables among neighbors.

Lagrange Multiplier $\left(LM_{lag}\right)$ test is also applied to test the existence of spatial correlation among response or indicators using the following formula:

$$LM_{lag} = \frac{\left(\frac{u'Wy}{s^2}\right)^2}{nP},$$

with

$$nP = T + \frac{(WX\beta)'M(WX\beta)}{s^2},$$

$$T = trace\big((W + W')W\big),$$

$$M = I - X(X'X)^{-1}X',$$

$$s^2 = \frac{u'u}{n},\tag{3}$$

where u is an error. The Lagrange Multiplier's value for this study is 3.388 (p-value is 0.056). This result indicates the spatial dependence of the response variable is due to the hypothesis model. It means that the malnutrition rate in this study has a spatial correlation among neighboring provinces in Indonesia. Additionally, the heteroscedasticity test is also determined in the hypothesis model using the Breusch-Pagan test:

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

The Breusch Pagan's value for this malnutrition data based on the hypothesis model is 3.499 (p-value is 0.061). It indicates that heteroscedasticity is due among provinces in Indonesia regarding the malnutrition rate.

Based on several tests above, it was concluded here that spatial correlation is due among regions in this malnutrition data. Therefore, the classical regression method using ordinary least squares

cannot be applied in this study. Several researchers suggested using the spatial autoregressive (SAR) [25] method to overcome these problems in order to obtain an acceptable model. Several researchers also applied the Robust spatial autoregressive (Robust SAR) method to construct a model with spatial autocorrelation of data and outlier problems [26], [27]. The spatial autoregressive with quantile (SARQ) method also could be implemented to solve the problem of spatial autoregressive case [8], [9], [16].

Spatial Autoregressive Model (SAR)

Spatial Autoregressive (SAR) model is one of the spatial models based on area. The spatial regression model can describe the relationship between independent and dependent variables by considering the effect of the data location. The presence of location effects on the data is represented by weights. The general model of the spatial autoregressive framework is defined as follows:

$$Y = \lambda WY + X\beta + u,$$

$$u \sim N(0, \sigma^2 I), \tag{5}$$

where Y is a vector of dependence variable, λ is a parameter of spatial lag coefficient of WY and indicates the degree of spatial autocorrelation. W is an $n \times n$ spatial weight matrix, specifying the spatial correlations between Y of different areas, with n being the number of observations. X denotes a matrix of independent variables, β presents a vector of regression coefficient of each explanatory variable, and σ^2 is a variance of u as a vector of spatial error. The parameter β which is generated to obtain a spatial regression model is estimated using the following formula:

$$\hat{\beta} = (X'X)^{-1} X'^{(I-\lambda W)} y. \tag{6}$$

The parameter, σ^2 is estimated by

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

while the parameter λ is estimated using a numerical approach.

Robust Spatial Autoregressive (Robust SAR)

The Robust spatial autoregressive approach uses the same model as the previous general model of the spatial autoregressive framework, which is defined as follows:

$$Y = \lambda WY + X\beta + u,$$

$$u \sim N(0, \sigma^2 I). \tag{8}$$

Robust SAR applies the Robust estimator M to estimate parameter β and spatial coefficient (λ). The Robust estimator S is used to minimize the following objective function:

$$min_\beta \sum_{i=1}^{n} \rho(u_i) = min_\beta \sum_{i=1}^{n} \rho\left(\frac{e_i}{s}\right) = min_\beta \sum_{i=1}^{n} \rho\left(\frac{y_i - \sum_{j=0}^{k} x_{ij}\beta_j}{s}\right) \tag{9}$$

with $s = \frac{MAD}{0.6745} = \frac{median|e_i - median(e_i)|}{0.6745}$. To minimize the equation above, we find the first derivative on β_j and then set the derivative equal to zero and solve it. After solving it, we obtain the estimate for β is $\hat{\beta} = (X^T W X)^{-1} X^T W Y$. The value for $\hat{\beta}$ is estimated iteratively, known as Iteratively Reweighted Least Square (IRLS). To estimate β using IRLS methods for $(m + 1)th$ iteration is defined as

$$\hat{\beta}^{m+1} = (X^T W^m X)^{-1} X^T W^m Y, \quad m = 0, 1, ... \tag{10}$$

Iteration will stop if the value of $\left|\hat{\beta}_j^{m+1} - \hat{\beta}_j^m\right|$ close to zero.

Spatial Autoregressive Quantile Model (SARQ)

The Quantile Regression model is developed to address the skewed distributions because of outliers. It enables us to examine the influence of explanatory variables at any quantiles of the malnutrition distribution. The effects of factors on malnutrition were not only changed with the quantiles but were also affected by potential spatial autocorrelation. The spatial autoregressive quantile regression (SARQ) model, can be used to account for spatial correlation among neighbors at both central and non-central locations [28]. The regression model of SARQ is described as:

$$Y = \lambda_\tau WY + X\beta_\tau + u, \tag{11}$$

where λ_τ and β_τ represent the spatial lag term and coefficients of the explanatory variable at the τth quantile. W is the spatial weight matrix. The parameter estimation of the SARQ model is estimated by the instrumental variable method proposed by Chernozhukov and Hansen [7], [29].

By applying quantile constraint $P(u_{i(\tau)} \leq 0|X) = \tau$, we can obtain the following equation:

$$P\left(y_i - \lambda_\tau \sum_{i=j}^{n} \omega_{ij} y_j - X_i' \beta_\tau \leq 0|X_i\right) = \tau \qquad (12)$$

Then (12) is equivalent to find $(\lambda_\tau, \beta_\tau)$ such that zero is a solution to the ordinary quantile regression:

$$0 \in argmin_{g \in G} E\left[\rho_\tau\left(y_i - \lambda_\tau \sum_{i=j}^{n} \omega_{ij} y_j - X_i' \beta_\tau - g(X_i, Z_i)\right)\right] \qquad (13)$$

Then, we assume G is defined as $G = \{g(X_i, Z_i) = \gamma' Z_i, \gamma \in \Gamma\}$, where γ is a column vector and $\Gamma \subset \mathbb{R}^m$. Let $\xi_i = (X_i, Z_i)'$, the instrumental variable $Z_i = \sum_{j=1}^{n} \omega_{ij} X_j$ is non-random with full column rank. The algorithm for solving parameter estimation of the SARQ model is summarized as follows [30]:

1. To perform a spatial autoregressive quantile regression for a given τ is defined by:

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

2. Minimize the value of $\hat{\gamma}_\tau(\lambda)$, and then get the estimator of λ_τ:

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

where $\hat{A} = A + O_p(1)$, A is a symmetric positive definite matrix, $O_p(1) = sup_\lambda |\sqrt{n} S_\tau(\lambda)|$, $S_\tau(\lambda) = I_n - \lambda_\tau W$. To find the estimator of λ_τ, the coefficient of $\hat{\gamma}_\tau(\lambda)$ of the instrument, the variable should be driven as close to zero as possible. Here, we take $\hat{A}$ to be an identity matrix, as written in Chernozhukov and Hansen [7], [29].

3. Get the estimator of β_τ:

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

Model Comparison

Three common measures are used in this study [8] to evaluate the performance of each model (SAR, Robust SAR, and SARQ). Mean absolute error (MAE); root mean square error (RMSE), and R-squared (R^2) are applied here to determine the better model. The lower the values of MAE

and RMSE are, the higher the accuracy of the corresponding models. In opposite, the higher the value of R-squared, the better the corresponding model is.

3. MAIN RESULTS

Based on a preliminary diagnostic test, it is informed that there is a positive spatial correlation between malnutrition rates in various provinces in Indonesia. In this section, we apply the above three models to analyze the spatial autocorrelation and influence factors of malnutrition in Indonesia in 2021. We consider the data from all 34 provinces of Indonesia in the analysis. The spatial autoregressive (SAR) approach is used first. Table 4 presents the estimated parameter model based on the SAR model.

Table 4. Parameter Estimate Based on Spatial Autoregressive (SAR) Method

Variable	Coefficient	S.D	Z-value	*p-value*
Intercept	18.428	7.852	2.347	0.019
Percentage of health insurance (X_1)	-0.089*	0.053	-1.675	0.093
Percentage access the water hygiene (X_2)	0.047	0.064	0.737	0.460
Percentage of breast-feeding (X_3)	-0.021	0.087	-0.244	0.807
Coefficient SAR (λ)	0.324*	0.150	2.160	0.031

*significant at $\alpha = 0.05$

From Table 4, it can be seen that the coefficient of spatial correlation (λ) of the malnutrition in this regression model is 0.324. This value is significant since the *p*-value is 0.031, less than 0.05. We then check for the existence of residual outliers using Moran's scatterplot as presented in Figure 2. The plot shows that there are three outliers in SAR's residual with observation numbers 2, 7, and 19. This results informed that the distribution of error is not homogeneous or the assumption of homogeneity of variance is violated.
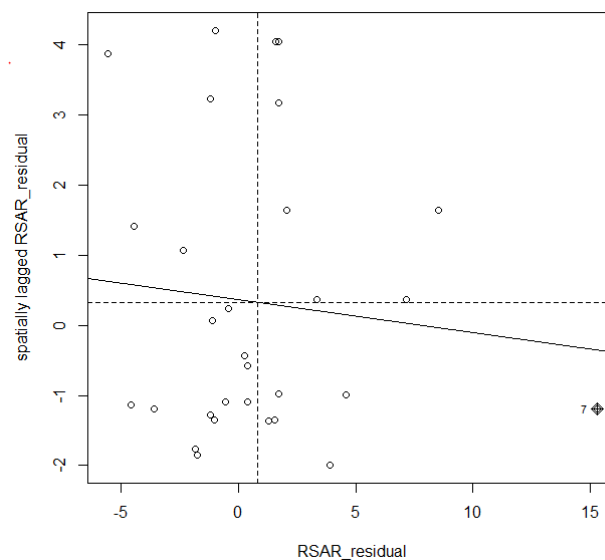
Figure 2. Moran's scatterplot for SAR Residual

We also evaluated the model performance of this SAR model based on the values of MAE, MSE, RMSE and R-squared. The values of all four criteria for this model are 2.894, 14.696, 3.834 and 26.062% respectivel. The R-squared is so small that the proposed model explains only 26.062% variance of response.

Based on several tests above, it is identified that the performance of the SAR model is not good and it could not be accepted. This study then applies the Robust SAR model to construct an acceptable model of malnutrition. The iteration processes of parameter estimation are presented in Table 5. There are seven iterations carried out, since $\left|\hat{\beta}_j^7 - \hat{\beta}_j^6\right|$ close to zero for $j = 1, 2, 3$. The parameter estimate obtained from 7th iteration of the Robust SAR model is provided in Table 6.

Table 5. Iteration Process for Parameter Estimate Using Robust SAR

Variable	Iteration Process for Parameter Estimate						
	1th	2nd	3rd	4th	5th	6th	7th
Intercept	16.371	14.261	14.165	14.507	14.764	14.952	15.065
X_1	-0.134	-0.186	-0.197	-0.200	-0.202	-0.202	-0.203
X_2	0.079	0.106	0.108	0.107	0.106	0.106	0.105
X_3	0.004	0.040	0.045	0.041	0.038	0.035	0.034
λ	0.324	0.324	0.324	0.324	0.324	0.324	0.324

Table 6. Parameter Estimate Using Robust SAR

Variables	Std. Error	Wald value	p-value
Intercept	6.169	5.963	0.014
Percentage of health insurance (X_1)	0.044	21.239	0.000
Percentage access the water hygiene (X_2)	0.049	4.595	0.032
Percentage of breast-feeding (X_3)	0.067	0.256	0.613
Coefficient SAR (λ)	0.170	3.616	0.057

*Significant at $\alpha = 0.05$

The performance parameters of the Robust SAR model are 2.101 for MAE, 2.756 for RMSE and 61.752 (R-squared). The robust SAR is better in performance than the SAR model. Moran's scatter plot based on the Robust SAR model is shown in Figure 3. By looking at Figure 3, there is an outlier inside the scatterplot, residual of observation number 7. The plot also indicates that the error distribution is not homogeneous or the assumption of homogeneity of variance is violated in this model. Thus, this model could not be accepted.

Figure 3. Moran's scatterplot for Robust SAR Residual

Therefore, Spatial Autoregressive Quantile (SARQ) approaches are applied then to obtain an acceptable model. In the SARQ work, we consider quantile $\tau = 0.10, 0.15, 0.20, 0.25, 0.30, 0.35,$ 0.40, 0.45, 0.50, 0.55, 0.60, 0.65, 0.70, 0.75, 0.80, 0.85, 0.90, 0.95. The results of parameter estimations based on the SARQ approach are shown in Table 7.

From Table 7, it can be seen that spatial autoregressive coefficient λ is significantly positive at lower quantiles ($\tau = 0.10, 0.15$) and at upper quantiles ($\tau \geq 0.70$) except at $\tau = 0.95$, with a significant level 5%. It shows that in lower cases of malnutrition, the neighboring provinces of a region positively influence it. Provinces with higher cases of malnutrition tend to have a positive influence on the surrounding provinces, which contributes to a high-high clustering distribution pattern. While in the SAR and Robust SAR models, the spatial effect coefficient λ is higher than in the SARQ model at all quantiles. The estimates of spatial effect in SAR and Robust SAR are rougher and somewhat unreasonable. Since both methods only estimate at the conditional mean of the response variable across values of the predictor variables, while SARQ could estimate at any conditional quantiles of the response variable. Table 7 also provides the AIC values for each quantile. The lower the value of AIC, the better performance of the corresponding model is. Table 8 presents the value of goodness of fit for SAR, Robust SAR and SARQ at the selected quantile. From both tables, it can be shown that the SARQ models perform best at the high tail ($\tau \geq 0.75$), at the non-central location.

Table 7. Parameter Estimate Using SARQ and AIC Value

τ^{th} quantile	Parameter Estimate					AIC value
	β_0	β_1	β_2	β_3	λ_τ	
0.10	-17.630	-0.204*	0.070	0.331*	0.006*	182.5834
0.15	6.061	-0.141*	0.013	0.059	0.018*	179.9062
0.20	14.792	-0.161*	0.002	0.006	0.000	179.2181
0.25	13.915	-0.167*	0.011	0.017	-0.011	175.8946
0.30	12.767	-0.194*	0.042	0.047	-0.010	174.6488
0.35	12.137	-0.194*	0.073	0.050	0.016	170.8287
0.40	17.894*	-0.152*	0.037	0.008	0.019	165.8496
0.45	24.305*	-0.159	0.067	-0.046	0.021	166.7232
0.50	25.923*	-0.148	0.042	-0.071	-0.012	162.1041
0.55	15.524	-0.103	0.081	-0.028	-0.011	161.4474
0.60	8.131	-0.057	0.123*	0.009	-0.018	160.5549
0.65	7.465	-0.043	0.111*	0.009	-0.017	159.0974
0.70	7.300	-0.040	0.111*	-0.009	-0.026	156.9992
0.75	4.239	0.100	-0.048	-0.008	0.072*	155.1244
0.80	1.064	0.114*	-0.037	0.052	0.053*	153.2034
0.85	0.987	0.113*	-0.035	0.052	0.009*	149.6779
0.90	8.303	0.109*	-0.005	0.109	0.306*	**146.6354**
0.95	-2.210	0.111*	0.047	0.112	-0.013	153.7761

*Significant at $\alpha = 0.05$

Table 8. The Goodness of Fit for Each Model.

Model	Criteria for Goodness of Fit			
	MAE	MSE	RMSE	R-Squared
SAR	2.894	14.696	3.834	26.062
ROBUST SAR	2.101	7.594	2.756	61.752
SARQ $\tau = 0.10$	2.648	11.694	3.420	41.100
SARQ $\tau = 0.15$	2.535	11.161	3.341	43.786
SARQ $\tau = 0.20$	2.438	9.992	3.161	49.673
SARQ $\tau = 0.25$	2.280	9.308	3.051	53.118
SARQ $\tau = 0.30$	2.126	8.789	2.965	55.732
SARQ $\tau = 0.35$	2.223	8.697	2.949	56.197
SARQ $\tau = 0.40$	2.156	8.017	2.831	59.621
SARQ $\tau = 0.45$	2.154	8.119	2.849	59.106
SARQ $\tau = 0.50$	2.149	7.863	2.804	60.398
SARQ $\tau = 0.55$	2.237	7.662	2.768	61.411
SARQ $\tau = 0.60$	2.116	7.174	2.678	63.870
SARQ $\tau = 0.65$	2.102	7.081	2.661	64.335
SARQ $\tau = 0.70$	2.062	7.043	2.654	64.528
SARQ $\tau = 0.75$	2.065	5.874	2.424	70.414
SARQ $\tau = 0.80$	1.855	5.702	2.388	71.281
SARQ $\tau = 0.85$	1.709	5.009	2.238	74.774
SARQ $\tau = 0.90$	**1.524**	**4.813**	**2.194**	**75.756**
SARQ $\tau = 0.95$	1.982	5.830	2.414	70.637

It is obtained that quantile 0.90 has the lowest value of AIC than others. MAE = 1.524 and RMSE = 2.194 at the 90th quantile are the lowest values. R-squared value at the 90th quantile is also higher compared to other quantiles $(R^2 = 75.756\%)$. As observed in Table 8 at higher quantile $(\tau \geq 0.65)$ SARQ has a much lower value of MAE and MSE than SAR and Robust SAR. At these intervals, the SARQ model also has a greater value of R-squared than the others. Moran's scatter plot at the 90th quantile, as presented in Figure 4, identified no outlier in residual anymore. This result reveals the necessity of using the quantile approach in spatial autoregressive models to handle the outliers and skewed distribution of malnutrition data. These results indicate that the SARQ model outperforms SAR and Robust SAR models in terms of fitting performance and prediction.
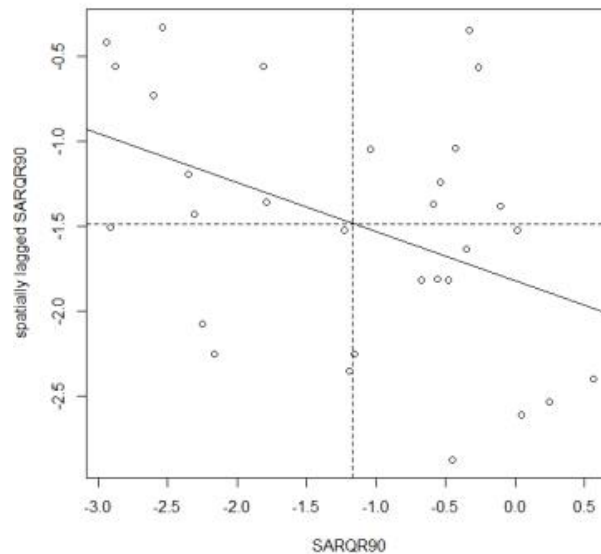
Figure 4. Moran's scatterplot based on SARQ at the 90th quantile.

Thus, the proposed model at the 90th quantile could be accepted. We also conclude here that the proposed model at 90th quantile is the best model for malnutrition data. The model of malnutrition in Indonesia is proposed as follows:

$$\widehat{Y}_i = 0.306_\tau WY_i + 0.109_\tau X_{1i} - 0.005_\tau X_{2i} + 0.306_\tau X_{3i}.$$

3. CONCLUSIONS

This study proves that by comparing the ordinary Spatial Autoregressive (SAR) model, Robust SAR, and Spatial Autoregressive Quantile (SARQ) model, SARQ is more resistant to outliers and performs better than others. SAR and Robust SAR only produce the model at an average response value, whereas both methods could not yield an acceptable model in this data analysis. Meanwhile, SARQ could estimate the model not only at the average of the response but also varies at different quantiles of the response. This study found that the proposed model at selected quantiles could be accepted based on the value of MAE, MSE, and R-squared as criteria of goodness of fit in the data analysis. Using SARQ, we could observe how the three measures vary with quantiles; specifically, the MAE and MSE are more petite at higher quantiles, while R-squared is greater. These results show that when the quantile value moves from the low tail to the high tail, the model performance improves, reaches an optimal point at quantile 0.90, and then decreases during the rest of the distribution.

Through the empirical analysis of the malnutrition rate in 30 provinces in Indonesia (4 provinces are excluded from analysis) at different quantiles, we conclude that only the Percentage of health insurance (X_1) was found as a significant indicator variable in the malnutrition data, as provided in Table 7.

CONFLICT OF INTERESTS

The author(s) declare that there is no conflict of interests.

REFERENCES

[1]   Y.F. Djoumessi, The impact of malnutrition on infant mortality and life expectancy in Africa, Nutrition. 103–104 (2022), 111760. https://doi.org/10.1016/j.nut.2022.111760.

[2]   A. Sofia, B. Susetyo, M.N. Aidi, Comparison Poisson regression and spatial autoregressive (SAR) Poisson (case study: in preventing the malnutrition factors in Java Island), Int. J. Sci. Res. Sci. Eng. Technol. 4 (2018), 234-239.

[3]   J.H. Rah, A. Melse-Boonstra, R. Agustina, et al. The triple burden of malnutrition among adolescents in Indonesia, Food Nutr. Bull. 42 (2021), S4-S8. https://doi.org/10.1177/03795721211007114.

[4]   A.H. Shibeshi, Z.G. Asfaw, Bayesian Spatial quantile interval model with application to childhood malnutrition in Ethiopia, preprint, https://doi.org/10.21203/rs.3.rs-108232/v2.

[5]   H. Yasin, A.R. Hakim, B. Warsito, Development life expectancy model in Central Java using robust spatial regression with M-estimators, Commun. Math. Biol. Neurosci. 2020 (2020), 69. https://doi.org/10.28919/cmbn/4984.

[6]   K.M. Zurnila, M. Saiful, M. Selvi, Identifying determinants of child malnutrition using spatial regression analysis, IOP Conf. Ser.: Mater. Sci. Eng. 506 (2019), 012051. https://doi.org/10.1088/1757-899x/506/1/012051.

[7]   V. Chernozhukov, C. Hansen, Instrumental variable quantile regression: A robust inference approach, J. Econometrics. 142 (2008), 379-398. https://doi.org/10.1016/j.jeconom.2007.06.005.

[8]   T. Yu, F. Gao, X. Liu, J. Tang, A spatial autoregressive quantile regression to examine quantile effects of regional factors on crash rates, Sensors. 22 (2021), 5. https://doi.org/10.3390/s22010005.

[9]   X. Dai, L. Jin, Minimum distance quantile regression for spatial autoregressive panel data models with fixed effects, PLoS ONE. 16 (2021), e0261144. https://doi.org/10.1371/journal.pone.0261144.

[10] C. Muharisa, F. Yanuar, D. Devianto, Simulation study the using of Bayesian quantile regression in nonnormal error, CAUCHY. 5 (2018), 121-126. https://doi.org/10.18860/ca.v5i3.5633.

[11] F. Yanuar, The simulation study to test the performance of quantile regression method with heteroscedastic error variance, CAUCHY. 5 (2017), 36-41. https://doi.org/10.18860/ca.v5i1.4209.

[12] F. Yanuar, H. Yozza, A. Zetra, Modified quantile regression for modeling the low birth weight, Front. Appl. Math. Stat. 8 (2022), 890028. https://doi.org/10.3389/fams.2022.890028.

[13] F. Yanuar, A. Zetra, Length-of-stay of hospitalized COVID-19 patients using bootstrap quantile regression, IAENG Int. J. Appl. Math. 51 (2021), 799-810.

[14] F. Yanuar, T. Abrari, I. Rahmi HG, et al. Spatial autoregressive quantile regression with application on open unemployment data, Sci. Technol. Indones. 8 (2023), 321-329. https://doi.org/10.26554/sti.2023.8.2.321-329.

[15] X. Dai, S. Li, M. Tian, Quantile regression for partially linear varying coefficient spatial autoregressive models, preprint, (2016). https://doi.org/10.48550/ARXIV.1608.01739.

[16] X. Dai, Z. Yan, M. Tian, et al. Quantile regression for general spatial panel data models with fixed effects, J. Appl. Stat. 47 (2019), 45-60. https://doi.org/10.1080/02664763.2019.1628190.

[17] Y. Zhang, J. Jiang, Y. Feng, Penalized quantile regression for spatial panel data with fixed effects, Commun. Stat. - Theory Methods. 52 (2021), 1287-1299. https://doi.org/10.1080/03610926.2021.1934028.

[18] O.R. Katoch, Determinants of malnutrition among children: A systematic review, Nutrition. 96 (2022), 111565. https://doi.org/10.1016/j.nut.2021.111565.

[19] N.V.N. Mbuya, S.J. Atwood, P.N. Huynh, Persistent malnutrition in ethnic minority communities of Vietnam: issues and options for policy and interventions, World Bank, Washington, DC. 2019. https://doi.org/10.1596/978-1-4648-1432-7.

[20] N. Ngoc Hien, N. Ngoc Hoa, Nutritional status and determinants of malnutrition in children under three years of age in Nghean, Vietnam, Pak. J. Nutrition. 8 (2009), 958-964. https://doi.org/10.3923/pjn.2009.958.964.

[21] C.T. Tharakan, C.M. Suchindran, Determinants of child malnutrition—An intervention model for Botswana, Nutr. Res. 19 (1999), 843–860. https://doi.org/10.1016/S0271-5317(99)00045-7.

[22] R.I. Kementrian Kesehatan, Hasil Studi Status Gizi Indonesia (SSGI)Tingkat Nasional, Propinsi, dan Kabupaten/Kota Tahun 2021, Kemenkes RI, 2021. http://www.badankebijakan.kemkes.go.id/buku-saku-hasil-studi-status-gizi-indonesia-ssgi-tahun-2021.

[23] P. Xu, H. Huang, Modeling crash spatial heterogeneity: random parameter versus geographically weighting, Accident Anal. Prevent. 75 (2015), 16-25. https://doi.org/10.1016/j.aap.2014.10.020.

[24] H. Huang, M.A. Abdel-Aty, A.L. Darwiche, County-level crash risk analysis in Florida: Bayesian spatial modeling, Transport. Res. Record. 2148 (2010), 27-37. https://doi.org/10.3141/2148-04.

[25] J.M. Ver Hoef, E.E. Peterson, et al. Spatial autoregressive models for statistical inference from ecological data,

Ecol. Monogr. 88 (2018), 36–59. https://doi.org/10.1002/ecm.1283.

[26] T. Huang, G. Saporta, H. Wang, et al. A robust spatial autoregressive scalar-on-function regression with t-distribution, Adv. Data Anal. Classif. 15 (2020), 57-81. https://doi.org/10.1007/s11634-020-00384-w.

[27] L. Jin, X. Dai, A. Shi, et al. Detection of outliers in mixed regressive-spatial autoregressive models, Commun. Stat. - Theory Methods. 45 (2016), 5179–5192. https://doi.org/10.1080/03610926.2014.941493.

[28] D. McMillen, Conditionally parametric quantile regression for spatial data: An analysis of land values in early nineteenth century Chicago, Region. Sci. Urban Econ. 55 (2015), 28-38.
https://doi.org/10.1016/j.regsciurbeco.2015.09.001.

[29] V. Chernozhukov, C. Hansen, Instrumental quantile regression inference for structural and treatment effect models, J. Econometrics. 132 (2006), 491-525. https://doi.org/10.1016/j.jeconom.2005.02.009.

[30] J. Zhang, Q. Lu, L. Guan, et al. Analysis of factors influencing energy efficiency based on spatial quantile autoregression: evidence from the panel data in China, Energies. 14 (2021), 504.
https://doi.org/10.3390/en14020504.