Available online at http://scik.org
Commun. Math. Biol. Neurosci. 2022, 2022:29
https://doi.org/10.28919/cmbn/7234
ISSN: 2052-2541

MODELLING THE PREVALENCE OF STUNTING TODDLERS USING SPATIAL AUTOREGRESSIVE WITH INSTRUMENT VARIABLE AND S-ESTIMATOR

VIEVIEN ABIGAIL DAMU DJARA[1,2,*], YUDHIE ANDRIYANA[3], LIENDA NOVIYANTI[3]

[1]Post-Graduate Program in Applied Statistics, Faculty of Mathematics and Natural Sciences, Universitas Padjadjaran, Indonesia

[2]Central Bureau of Statistics of Sumba Timur, Nusa Tenggara Timur, Indonesia

[3]Department of Statistics, Faculty of Mathematics and Natural Sciences, Universitas Padjadjaran, Indonesia

Abstract: Stunting is caused by multidimensional factors and not only caused by chronic malnutrition. Paying attention to toddler nutrition and other socio-economic variables are solutions to reduce stunting prevalence. This study uses seven explanatory variables which are indicated to have an effect on increasing or decreasing stunting prevalence. The unit analysis in this study is all districts/cities in Java island. The spatial autoregressive model (SAR) was considered. The presence of outliers can cause inaccurate parameter estimation results. Removing outliers in spatial data can change the composition of spatial effects on the data. We use the instrument variables with S-estimator to overcome the presence of these outliers. The R shiny program was developed to estimate model parameters. The results showed that the underweight variable, expenditure variable, education variable, and the variable of household waste management had a significant effect on the prevalence of stunting in Java. The results of this study also found

*Corresponding author
E-mail address: vievien20001@mail.unpad.ac.id
Received February 5, 2022

that three other variables, namely defecation behavior, the level of difficulty in accessing the public health center, and access to safe drinking water had no significant effect on the prevalence of stunting in Java. The result of the model evaluation show that instrument variable with S-estimator had a lower residual standard error and higher R-squared than the instrument variable without S-estimator.

Keywords: stunting, SAR; R shiny; instrument variable; S-estimator.

2010 AMS Subject Classification: 93A30.

1. INTRODUCTION

Stunting is a condition in which toddlers experience growth failure [1], [2]. Stunting causes the level of children's intelligence to be potentially not optimal and is more susceptible to disease [1], [2]. The causes of stunting are chronic malnutrition and recurrent infections, especially in the first 1000 days of life, from fetus to child aged 23 months [2]–[4]. Stunting will broadly inhibit economic growth, increase poverty and widen inequality [4]. According to the National Development Planning Agency (Bappenas), Indonesia will suffer annual losses of approximately 300 trillion rupiahs if the prevalence of stunted toddlers is not reduced [2]. The government has made the acceleration of stunting reduction one of the national priority programs [2]. In the national mid-term development plan (RPJMN) 2020-2024, the government has targeted to reduce the stunting rate to 14% by 2024 [1], [2].

The map of the number of stunting toddlers in Indonesia by the World Bank through the publication of the National Team for the Acceleration of Poverty Reduction (TNP2K) in 2017 shows that the highest number of stunting toddlers is in Java island, reaching 4.353.000 toddlers suffering from stunting. This increase in the prevalence of stunting in Java island will increase the burden on the government in achieving the stunting reduction target in Indonesia. Stunting reduction can be achieved by knowing the main causes of high stunting in the region. Family conditions such as finances, knowledge of toddler nutrition, handling when toddlers are sick, housing and sanitation conditions, environmental cleanliness, and easy access to health care facilities have an impact on stunting.

Stunting is caused by multi-dimensional factors and is not only caused by poor nutrition experienced by pregnant women and toddlers [4], [5]. Tiopan Sipahutar (2021) conducted a study to find stunting hotspot areas in all districts/cities in Indonesia. The results showed that there were 14 provinces that had a high distribution of stunting cases, and 3 of them were on Java island. The prevalence of stunting in districts/cities in Indonesia has a relationship with the surrounding area [3]. Due to the vast territory of Indonesia and the wide variety of regions, Indonesia needs additional information that incorporates the regional context in the analysis. Spatial analysis in the context of stunting is still not widely used in Indonesia and has not been used as a decision support system in policies or programs at the national and regional levels [3]. Because of its ability to accommodate various spatial dependency structures, econometric spatial modeling has become widely used in epidemiological studies [6].

One of the econometric spatial models is the spatial autoregressive (SAR) model. This model is often referred to as the *spatial lag* model (in Anselin and Bera, 1998) or *mixed regressive spatial autoregressive* model (in Anselin, 1988) [7]. Parameter estimation in the model is used to see how each explanatory variable influence stunting cases. In some cases, the presence of outliers causes the parameter estimates to be biased. Removing outliers in spatial analysis changes the spatial effects composition on data [8]. Therefore, a robust method against outliers can be used to estimate the model parameters. Handling outliers in the spatial model can be approached by several methods such as a least median square estimator, least trimmed squares estimator, method of moment, M-estimator, and S-estimator [9].

S-estimator has the highest breakdown point (50%) and can overcome outliers in response variables and explanatory variables. Instrument variable is used to eliminate endogeneity due to the lag of the dependent variable as a regressor in the model. The open-source software R can be used to estimate the model parameters. R shiny was developed to simplify parameter estimation in the model. Shiny is one of the R visualization packages that make it easy to create interactive web-based applications directly from R, and many researchers have worked to make R packages more user-friendly [1], [10]. This research uses R shiny to analyze the determinants of stunting toddlers using spatial autoregressive models with instrument variable and S-estimator.

2. METHOD

This study used data from Central Bureau of Statistics and basic health research reports in 2018. The units of analysis in this study are 119 districts/cities on Java island. All data used in this study are taken from the public domain.

Table 1: Response variable and explanatory variables

Variable	Notation
The prevalence of stunting toddlers	Y
The prevalence of underweight of toddlers	X_1
Percentage of population aged 3 years and over who behave correctly in defecating (BAB)	X_2
Real expenditure per capita adjusted (millions/person/year) (an indicator in the decent living dimension to measure the human development index)	X_3
The average length of schooling of the population aged 25 years and over	X_4
Difficulty level of access to public health center/ sub-health center/ mobile health center/ village midwife (%)	X_5
Percentage of households that throw waste everywhere	X_6
Percentage of households with proper drinking water	X_7

In this study, the development of R shiny was focused on the spatial autoregressive model using instrument variable without S-estimator and instrument variable with S-estimator. R Shiny is composed of two components: a UI (user interface) and a server. The UI's purpose is to display all input and output, whereas the server's purpose is to process input into output. The framework in R shiny is as follows [1], [11]:

```
library(shiny)
ui <- fluidPage( # or other layout function
  # contents of ui.R file
)
server <- function(input, output) {
  #contents of server.R file
}
shinyApp(ui = ui, server = server)
```

The stages of R Shiny developed had complied with the stages in the analysis of the econometric spatial models, namely descriptive analysis and map, detect outlier using Moran's scatterplot, classic assumption test, Lagrange multiplier test, parameter estimation using instrument variable and S-estimator, and diagnostic checking of the spatial model. Detecting outliers using Moran's scatterplot can be done by looking at quadrant II (upper left) and quadrant IV (lower right) [12], [13]. Observation in quadrant I (upper right) indicates a positive spatial autocorrelation between high values, and observation in quadrant III (lower left) indicates a positive spatial autocorrelation among low values (i.e. a spatial clustering with a similar level) [14]. The Breusch-pagan test, Durbin-Watson (DW) test, and the VIF multicollinearity test are used to test the homoscedasticity, independence, and multicollinearity, respectively, with a 5% alpha. The Euclidean distance method was used to determine spatial weighted. Two locations with coordinates (u_i, v_i) and (u_j, v_j), whose distance will be measured using the Euclidean distance with the following formula [1], [6]:

$$d_{ij} = \sqrt{\left(u_i - u_j\right)^2 (v_i - v_j)^2} \tag{1}$$

Inverse distance:

$$w_{ij} = \begin{cases} \dfrac{1}{d_{ij}^{\propto}} & ,if\ i \neq j \\ 0 & ,if\ i = j \end{cases} \tag{2}$$

generally, α = {1,2}

Spatial autocorrelation or spatial dependence can be tested using Moran's I with the following formula [1], [6], [15]:

$$I = \frac{n}{\sum_{i=1}^{n}\sum_{j=1}^{n} w_{ij}} \frac{\sum_{i=1}^{n}\sum_{j=1}^{n} w_{ij}\,(y_i - \bar{y})(y_j - \bar{y})}{\sum_{i=1}^{n}(y_i - \bar{y})^2} \tag{3}$$

Where n is the number of spatial units, y_i is the observation variable at location i, w_{ij} is elements of the spatial weight matrix W, and I is Moran's global coefficient. Moran's I values range $[-1,1]$. Negative values of the I statistic means a negative autocorrelation (i.e. a high stunting area is surrounded by low areas or vice versa). Positive autocorrelation means the existence of clusters of similar values (high or low) (Goodchild, 1986) [15]. Local Moran's I can be used to identify spatial dependencies on each unit with the following formula [1], [6]:

$$I_i = \frac{y_i - \bar{y}}{\sum_{i=1}^{n}(y_i - \bar{y})^2 / n} \sum_{j=1}^{n} w_{ij}(y_j - \bar{y}) \tag{4}$$

The null hypothesis for autocorrelation is $I = E(I)$ no spatial dependence. The formula of test statistics can be written as follows:

$$Z(I) = \frac{I - E(I)}{\sqrt{Var(I)}} \tag{5}$$

with,

$$E(I) = -\frac{1}{n-1}$$

$$Var(I) = \frac{n^2.S_1 - n.S_2 + 3.S_0^2}{(n^2-1)S_0^2} - [E(I)]^2$$

$$S_0 = \sum_{i=1}^{n}\sum_{j=1}^{n} w_{ij} \quad ; \; S_1 = \frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n}(w_{ij} + w_{ji})^2 \quad ; \; S_2 = \sum_{i=1}^{n}(\sum_{j=1}^{n} w_{ij} + \sum_{j=1}^{n} w_{ji})^2$$

Reject the null hypothesis at significance level $\propto$ if $Z(I) > Z_{1-\propto}$. There are several spatial econometric models, so we can use Lagrange Multiplier (LM) test as the initial criteria for model selection. Formally, SAR model is expressed in equation (6) where y is a $n \times 1$ vector of observation on the dependent variable, X is a $n \times k$ matrix of observation on explanatory variables, W is a $n \times n$ spatial weight matrix, ε is a $n \times 1$ vector of i.i.d. error terms, ρ is the spatial autoregressive coefficient, and β a $k \times 1$ vector of regression coefficient [16].

$$\mathbf{y} = \rho\mathbf{Wy} + \mathbf{X\beta} + \boldsymbol{\varepsilon} \;\; ; \; \boldsymbol{\varepsilon} \sim \text{iid}(0, \sigma^2\mathbf{I}) \tag{6}$$

LM test for Spatial Autoregressive model can be written as follows:

$$\text{LM}_{\text{lag}} = \frac{(\hat{\boldsymbol{\varepsilon}}^{\mathbf{t}}\mathbf{Wy})^2}{\text{s}^2((\mathbf{WX\beta})^{\text{t}}\mathbf{M}(\mathbf{WX\beta}) + \text{Ts}^2)} \tag{7}$$

with,

$$\mathbf{M} = \mathbf{I} - \mathbf{X}(\mathbf{X}^{\text{t}}\mathbf{X})^{-1}\mathbf{X}^{\text{t}} \quad ; \quad \text{T} = \text{tr}\,((\mathbf{W}^{\text{t}} + \mathbf{W})\,\mathbf{W}) \; ; \quad s^2 = \frac{\hat{\boldsymbol{\varepsilon}}^{\text{t}}\hat{\boldsymbol{\varepsilon}}}{\text{n}}$$

Reject the null hypothesis if $\mathrm{LM}_{\mathrm{lag}} > \mathrm{X}^2_{(\alpha,1)}$ or $p - value < \propto$, so the SAR model can be used. Parameter in the model can be estimate by Ordinary Least Square, Maximum Likelihood estimator and Instrument Variables. For some purpose, eq. (6) can be rewritten as [17]:

$$\mathbf{y} = \mathbf{X\beta} + \rho\mathbf{Wy} + \boldsymbol{\varepsilon}$$

$$\mathbf{y} = [\mathbf{X} \quad \mathbf{Wy}] \times \begin{bmatrix} \boldsymbol{\beta} \\ \rho \end{bmatrix} + \boldsymbol{\varepsilon} = \mathbf{Z\theta} + \boldsymbol{\varepsilon} \tag{8}$$

where $\mathbf{Z}$ is $n \times (k+1)$ matrix and $\boldsymbol{\theta}$ is $(k+1) \times 1$ vector. Based on instrument variables (IV) approach, $\mathbf{Q}$, that are strongly correlated with the original variables $\mathbf{Z}$, but asymptotically uncorrelated with the error term, parameter $\boldsymbol{\theta}$ in the eq. (8) can be estimate by:

$$\boldsymbol{\theta}_{\mathrm{IV}} = [\mathbf{Q}^{\mathrm{t}}\mathbf{Z}]^{-1}\mathbf{Q}^{\mathrm{t}}\mathbf{y} \tag{9}$$

The number of instruments (q) will be larger than the number of parameters in the model, so that eq. (9) is not well defined [17]. Therefore, similar to least squares (LS) approach to estimation, the problem can be formulated as a minimization of the problem bellow $\left(\min \phi(\theta) = \frac{\partial\phi(\theta)}{\partial\boldsymbol{\theta}^{\mathrm{t}}}\right)$ [17]:

$$\phi(\theta) = (\mathbf{y} - \mathbf{Z\theta})^{\mathrm{t}}\mathbf{Q}(\mathbf{Q}^{\mathrm{t}}\mathbf{Q})^{-1}\mathbf{Q}^{\mathrm{t}}(\mathbf{y} - \mathbf{Z\theta}) \tag{10}$$

$$\boldsymbol{\theta}_{\mathrm{IV}} = \left[\mathbf{Z}^{\mathrm{t}}\mathbf{P_Q}\mathbf{Z}\right]^{-1}\mathbf{Z}^{\mathrm{t}}\mathbf{P_Q}\mathbf{y} \tag{11}$$

with $\mathbf{P_Q} = \mathbf{Q}(\mathbf{Q}^{\mathrm{t}}\mathbf{Q})^{-1}\mathbf{Q}^{\mathrm{t}}$ and $\mathbf{P_Q Z} = \mathbf{Q}.\{(\mathbf{Q}^{\mathrm{t}}\mathbf{Q})^{-1}\mathbf{Q}^{\mathrm{t}}\mathbf{Z}\}$ with the term in brackets as the familiar OLS estimate for a regression of $\mathbf{Z}$ on $\mathbf{Q}$. Therefore, with $\mathbf{Z_p} = \mathbf{P_Q Z}$ as the predicted values of the $\mathbf{Z}$, the IV estimator in eq. (11) can also be expressed as [17]:

$$\boldsymbol{\theta}_{\mathrm{IV}} = \left[\mathbf{Z_p}^{\mathrm{t}}\mathbf{Z_p}\right]^{-1}\mathbf{Z_p}^{\mathrm{t}}\mathbf{y} \tag{12}$$

$$\boldsymbol{\theta}_{\mathrm{IV}} = \left[\mathbf{Z_p}^{\mathrm{t}}\mathbf{Z}\right]^{-1}\mathbf{Z_p}^{\mathrm{t}}\mathbf{y} \tag{13}$$

Due to the idempotency of the projection matrix, eq. (12) can be written as eq. (13). Eq. (13) is the familiar 2SLS estimator. It is equivalent to OLS on the predicted values for the explanatory variables obtained from an auxiliary regression on a fixed set of exogenous instruments [17]. Eq. (12) is the familiar OLS estimate for a regression of $\mathbf{Z_p}$ on y.

The inaccuracy of parameter estimation of spatial regression model can be caused by outlier observation [8]. Rousseeuw & Yohai (1984) proposed a parameter estimator for robust regression with S-estimator [9], [18]. The S-estimator method uses the residual standard deviation to overcome the weaknesses of the median method [9], [19]. In this study, S-estimator is applied in the second stage when $\mathbf{Z_p}$ is used to estimate parameter $\boldsymbol{\theta}$ with the instrument variable $\mathbf{Q} = [\mathbf{X} \quad \mathbf{WX} \quad \mathbf{W^2X}]$.

Consider the linier model [19] :

$$\boldsymbol{y}_i = \beta_0^t \boldsymbol{x}_i + \boldsymbol{r}_i \ , \qquad i = 1, \dots, n \tag{14}$$

Based on eq. (12) and eq. (14), the linier model is rewritten as follows:

$$\boldsymbol{y}_i = \theta_0^t \boldsymbol{Z}_{\boldsymbol{p}_i} + \boldsymbol{r}_i \ , \qquad i = 1, \dots, n \tag{15}$$

where $\boldsymbol{Z}_{\boldsymbol{p}_i} = \left(\boldsymbol{Z}_{\boldsymbol{p}_{i1}}, \dots, \boldsymbol{Z}_{\boldsymbol{p}_{ip}}\right)^t$ is matrix of explanatory variables and θ_0 is vector of regression coefficient, and the r_i's vector of the residuals $\boldsymbol{r} = (r_1, \dots, r_n)^t$. S-estimator proposed by Rousseeuw & Yohai (1984) is defined as [9], [18]–[20]:

$$\hat{\theta}_n = arg\,min \sum_{i=1}^{n} \rho\left(\frac{r_i(\theta)}{\hat{s}}\right) \tag{16}$$

with $\hat{s} = s_M\left(r(\hat{\theta}_n)\right)$ and ρ is objective function Tukey's bi-square. Huber (1964) define M-scale as $s_M(\boldsymbol{r})$:

$$\frac{1}{n}\sum_{i=1}^{n} \rho\left(\frac{r_i(\theta)}{s_M(\boldsymbol{r})}\right) = b \tag{17}$$

Iterative algorithm used in the computation of $s\left(\boldsymbol{r}(\theta_j)\right)$, starting from $\hat{s} = MAD(\boldsymbol{r}(\theta_j))$ [19].

$$\hat{s} = \frac{median|e_i - median(e_i)|}{0.6745} \tag{18}$$

The S-estimate based on a ρ function in the bi-square family with functions [19], [21]:

$$\rho(u_i) = \begin{cases} \dfrac{u_i^2}{2} - \dfrac{u_i^4}{2c^2} + \dfrac{u_i^6}{6c^4} & , |u_i| \leq c \\ \dfrac{c^2}{6} & , |u_i| > c \end{cases} \tag{19}$$

where $u_i = {r_i(\theta)}/{\hat{s}}$. To obtain an S-estimates with breakdown point 0.5, the $c = 1.547$ and $b =$

0.5 [19], [21]. Differentiating eq. (16), estimating equations for S-estimator [19]:

$$\sum_{i=1}^{n} \psi\left(\frac{r_i(\hat{\theta}_n)}{\hat{s}}\right) \boldsymbol{Z}_{\boldsymbol{p}_i} = 0 \tag{20}$$

where,

$$\psi(u_i) = \rho'(u_i) = \begin{cases} u_i\left[1-\left(\frac{u_i}{c}\right)^2\right]^2 & , |u_i| \le c \\ 0 & , |u_i| > c \end{cases} \tag{21}$$

The S-estimator function is obtained based on $w(u) = {\psi(u)}/{u}$, so that $w(u)$ is the iteratively re-weighted least-square (IRWLS) with the following function:

$$w(u_i) = \begin{cases} \left[1-\left(\frac{u_i}{c}\right)^2\right]^2 & , |u_i| \le c \\ 0 & , |u_i| > c \end{cases} \tag{22}$$

Based on eq. (20), S-estimator of spatial autoregressive model is as follow [22], [23]:

$$\boldsymbol{\theta}_{\mathbf{s}} = \left[\mathbf{Z_p}^{\mathrm{t}}\mathbf{W}\mathbf{Z_p}\right]^{-\mathbf{1}}\mathbf{Z_p}^{\mathrm{t}}\mathbf{W}\mathbf{y} \tag{23}$$

The weighting matrix $\mathbf{W}$ is a symmetric matrix of size n based on the weighting function in eq. (22). The elements of the matrix are $w_{ij} = w_i(u_i)$ for $i = j$ and 0 for other elements. This weighting matrix is different from the spatial weighting matrix $\mathbf{W}$. The value in the matrix $\mathbf{W}$ indicates that the $i-$th observation is an outlier or not, so a large $w_i(u_i)$ value indicates that the observation is an outlier observation [9]. To distinguish them in writing, the weighting matrix of the S-estimator will be written as $\mathbf{W}_s$.

$$\boldsymbol{\theta}_{s} = \left[\mathbf{Z_p}^{\mathrm{t}}\mathbf{W}_s\mathbf{Z_p}\right]^{-\mathbf{1}}\mathbf{Z_p}^{\mathrm{t}}\mathbf{W}_s\mathbf{y} \tag{24}$$

Barrera and Yohai (2006) have attention to increases the speed of the resampling algorithm for computing S-estimator. Given and estimate $\hat{\theta}^{(0)}$ of the regression coefficients, the local improvement step (I-step) for S-estimator is defined as follows [19] :

1. Compute initial scale $s_0 = s_M\left(\boldsymbol{r}(\hat{\theta}^{(0)})\right)$ and the weight $w_{\mathrm{i}}(\mathrm{u_i}) = w\left({r_i(\hat{\theta}^{(0)})}/{s_0}\right)$

2. Define $\hat{\theta}^{(1)}$ as the weighted LS-estimator.

I-step is one step of the step of the IRWLS algorithm to solve eq. (20) starting from $\hat{\theta}^{(0)}$. Simple version of the fast-S algorithm is as follows [19] :

1. Draw N random subsamples of size p and let $\hat{\theta}_j, j = 1, \dots, N$ be the coefficient of each subsamples.
2. improve each of these candidates applying k I-steps and let $\hat{\theta}_j^C$ be the resulting improved candidates.
3. For each $\hat{\theta}_j^C$ compute the M-scale $s_j = s_M\left(\boldsymbol{r}\left(\hat{\theta}_j^C\right)\right)$ and keep the improved candidates with the best t scales $(1 \leq t \leq N)$. Call $\left(\hat{\theta}_j^B, s_j^B\right),\ j = 1, \dots, t$.
4. Apply the I-step to each $\hat{\theta}_j^B$ until convergence, obtaining $\left(\hat{\theta}_j^F, s_j^F\right)$ where $s_j^F = s\left(\boldsymbol{r}\left(\hat{\theta}_j^F\right)\right)$
5. The final estimate is the $\hat{\theta}_j^F$ associated with the smallest s_j^F.

Barrera and Yohai (2006) modify step 2 and step 3. In step 2, modify of the I-step as follows: given a candidate $\hat{\theta}_j$ the improved vector $\hat{\theta}_j^C$ and replace $s\left(\boldsymbol{r}\left(\hat{\theta}_j\right)\right)$ by an approximated value obtained at the rth step of any iterative algorithm, starting from eq. (18). The modify in step 3 is as follows [19]:

(a) Compute $s_m = s\left(\boldsymbol{r}\left(\hat{\theta}_m^C\right)\right),\ m = 1, \dots, t$ and let $A_t = max_{1 \leq m \leq t} s_m$ and $I_t = \{1, \dots, t\}$.

(b) Suppose that we have already examined r candidates, where $r > t$. I_r is the set of the t best scale estimates found after examining these r candidates and A_r is the maximum of these scales. $\hat{\theta}_{r+1}^C$ will be include in I_{r+1} if $s\left(\boldsymbol{r}\left(\hat{\theta}_{r+1}^C\right)\right) < A_r$ and this equivalent with eq. (17):

$$\sum_{i=1}^{n} \rho\left(\frac{r_i\left(\hat{\theta}_{r+1}^C\right)}{A_r}\right) < b \qquad (25)$$

Then compute $s\left(\boldsymbol{r}\left(\hat{\theta}_{r+1}^C\right)\right)$ only if eq. (25) holds. This change allows the computation of the best candidate t with a smaller number of scale calculations [19].

Parameter estimation results can be used to calculate the direct and indirect effects of the spatial autoregressive model. The formula of direct and indirect effects of the spatial autoregressive model is as follows [15]:

Table 2: Direct effect and spillover effect of SAR model

Direct effect	spillover effect
Diagonal elements from the matrix $(I-\rho W)^{-1}\beta_k$	Non-diagonal elements from the matrix $(I-\rho W)^{-1}\beta_k$

3. Results and Discussion

The average prevalence of stunting in Java island reached 30.14% with the lowest prevalence of stunting in Sleman district (14.7%), and the highest prevalence of stunting in Sampang district (47.92%). According to WHO, the prevalence of stunting becomes a problem when the prevalence reaches 20% or more [24]–[26]. Figure 1 shows that there are only 9 districts/cities that are in the low category (dark blue zone) with stunting prevalence below 20%, while 110 other districts/cities experience stunting problems spread across all provinces in Java island.

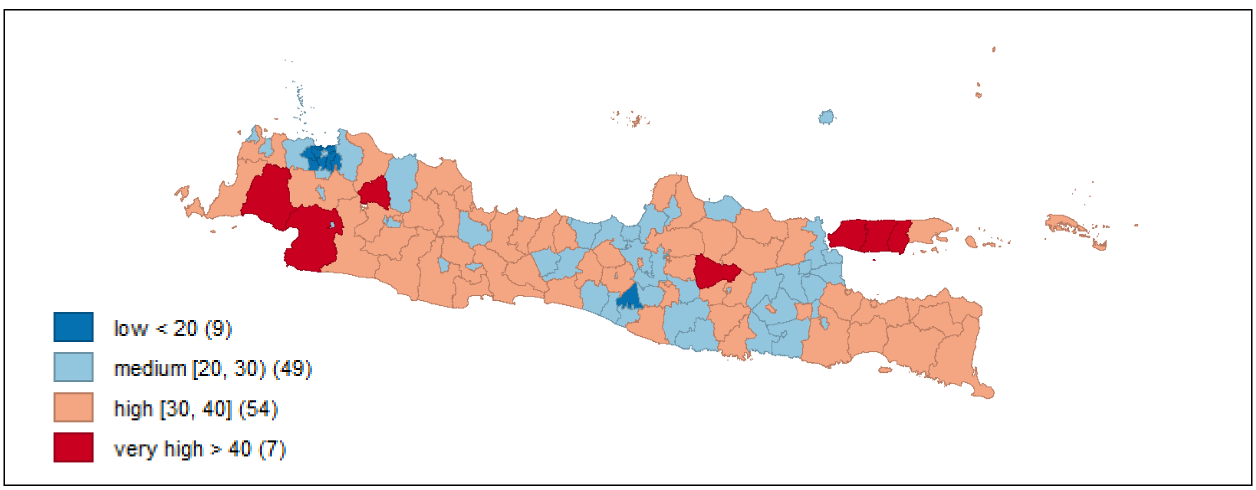

Figure 1. Map of the prevalence of stunting toddlers in Java island

Table 3 shows that the probability of Moran's I is 0.000, so the null hypothesis is rejected, which means that there is spatial autocorrelation on prevalence of stunting in Java island. Moran's index is 0.141 indicating that there is a positive autocorrelation. High stunting area will be surrounded by high areas as well, and low stunting area will be surrounded by low areas as well. The results of the robust Lagrange Multiplier and Breusch-Pagan tests show that the spatial effect is significant for lag $(\rho \neq 0)$ and not significant for error $(\lambda = 0)$, and there is no spatial heterogeneity, so the spatial lag or spatial autoregressive (SAR) model can be used.

Table 3. The results of the spatial autocorrelation test and the spatial effect test

Statistic test	**p-value**
Global Moran's I	0,000 (I = 0,141)
$\mathrm{LM}_{\mathrm{rlag}}$	0,028
$\mathrm{LM}_{\mathrm{rerr}}$	0,309
Breusch-Pagan	0,461

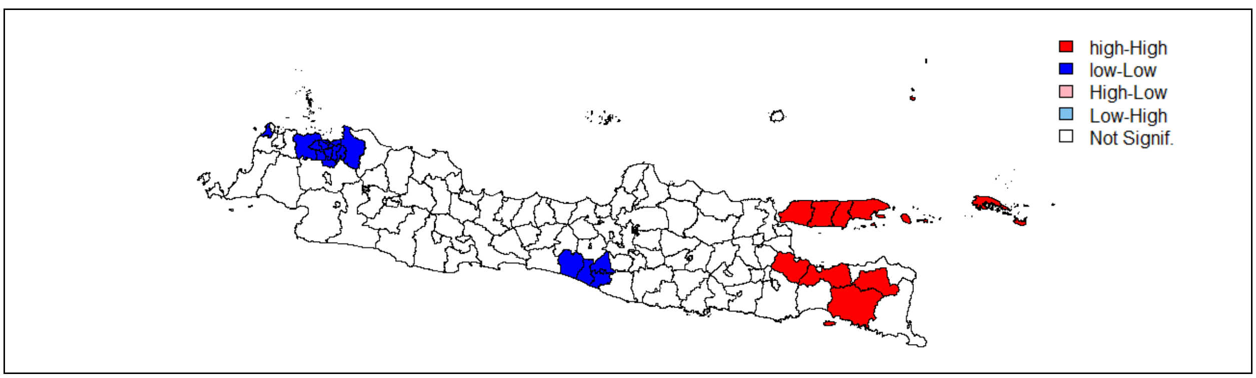

Figure 2: Local Moran's I

Based on the results of local Moran's I test which is visualized in Figure 2, it can be seen that there is significant spatial autocorrelation in high stunting areas and low stunting areas. Table 4 shows that there are lower outliers and upper outliers so that parameter estimation is carried out using a robust estimator, namely the S-estimator. Table 5 and Table 6 illustrates the result of parameter estimation.

Table 4. Outlier detection results using Moran scatter plot

variable	Lower outlier	Upper outlier
y: stunting	-	-
X_1: underweight	-	Tangerang city, Jakarta Timur city
X_2: defecation behavior	-	Probolinggo city
X_3: expenditure	Bekasi	Surabaya city, Bandung city
X_4: education	Bogor	Cirebon city, Madiun city
X_5: Difficulty level of access to public health center	Serang city, Tangerang Selatan city, Cilegon city	-
X_6: waste management	Magelang city, Cilegon city	Magelang
X_7: proper drinking water	-	Bekasi, Depok city, Bekasi city, Tangerang city

Table 5. Model parameter estimation results using instrument variable without S-estimator

	Estimate	Std. Error	t-value	p-value
(Intercept)	-9,123	14,794	-0,617	0,539
X_1: underweight	0,248	0,109	2,269	0,025
X_2: defecation behavior	0,092	0,063	1,471	0,144
X_3: expenditure	-0,273	0,285	-0,959	0,339
X_4: education	-1,307	0,521	-2,508	0,014
X_5: Difficulty level of access to public health center	0,129	0,045	2,833	0,005
X_6: waste management	0,139	0,097	1,432	0,155
X_7: proper drinking water	0,008	0,032	0,256	0,799

	Estimate	Std. Error	t-value	p-value
Rho	1,204	0,396	3,038	0,003
p-value of F- statistic: 0,000				
R-squared: 0,426				
Residual standard error: 4,362				

Table 6. Model parameter estimation results using instrument variable with S-estimator

	Estimate	Std. Error	t-value	p-value
(Intercept)	18,343	7,259	2,527	0,013
X_1: underweight	0,211	0,055	3,863	0,000
X_2: defecation behavior	-0,026	0,031	-0,844	0,401
X_3: expenditure	-0,664	0,139	-4,748	0,000
X_4: education	-0,944	0,264	-3,579	0,000
X_5: Difficulty level of access to public health center	0,035	0,023	1,540	0,126
X_6: waste management	0,161	0,045	3,564	0,000
X_7: proper drinking water	-0,009	0,015	-0,663	0,509
Rho	0,841	0,189	4,455	0,000
p-value of F- statistic: 0,000				
R-squared: 0,995				
Residual standard error: 1,794				

$$\hat{y}_i = 0.841 \sum_{j=1, i\neq j}^{n} w_{ij}\, y_j + 18{,}343 + 0{,}211X_{1i} - 0{,}026X_{2i} + 0{,}664X_{3i} - 0{,}944X_{4i} + 0{,}035X_{5i} + 0{,}161X_{6i} - 0{,}009X_{7i}$$

According to the results in Tables 5 and 6, instrument variable with S-estimator has a lower residual standard error and a higher R-squared value than the instrument variable without S-estimator. From the F test statistics in table 6, it can be seen that simultaneously all explanatory variables have a significant effect on the prevalence of stunting in Java. From the t-test statistics, it can be seen that partially the prevalence of underweight and household waste management by dumping it in any place has a significant and positive effect on increasing the prevalence of stunting. Correct behavior in defecating and having access to safe drinking water has a negative effect on decreasing stunting prevalence, but these two variables do not have a significant effect. The variable of expenditure and education has a negative and significant effect on decreasing stunting prevalence. Furthermore, the difficulty level of access to the public health center/ sub-health center/ mobile health center/ village midwife has a positive effect on increasing the prevalence of stunting, but this variable does not have a significant effect.

The impact of each explanatory variable shown in table 6 is global average and there is a coefficient $\hat{\rho}$, so that the interpretation of the impact of each explanatory variable cannot be done easily. The impact of each explanatory variable will be interpreted using the impact measure in Table 7.

Tabel 7. Direct effect, spillover effect, and total effect of SAR Model

	direct effect	spillover effect	total effect
X_1: underweight	0,222	0,009	0,231
X_2: defecation behavior	-0,028	-0,001	-0,029
X_3: expenditure	-0,698	-0,029	-0,727
X_4: education	-0,994	-0,042	-1,036
X_5: Difficulty level of access to public health center	0,037	0,002	0,039
X_6: waste management	0,169	0,007	0,176
X_7: proper drinking water	-0,0104	-0,0004	-0,0108

The direct effect in table 7 above can be interpreted as a change in the average prevalence of stunting in an area due to a change in the causal factor of one unit in the area, assuming other causal factors are in a constant state (ceteris paribus). The spillover effect is interpreted as a change in the average prevalence of stunting in an area due to a change in the causal factor of one unit in the surrounding area, assuming the other causal factors are in a constant state (ceteris paribus). the total effect is interpreted as a change in the average prevalence of stunting in an area due to a change in the causal factor of one unit in the area and a change in the causal factor in the surrounding area, assuming the other causal factors are in a constant state (ceteris paribus).

Each percent increase in underweight prevalence in an area increases the average stunting prevalence in that area by 0.222%. Each percent increase in underweight prevalence in the surrounding areas increases the average stunting prevalence in that area by 0.009%. Each percent increase in underweight prevalence in Java increases the average stunting prevalence in Java by 0.231%. Poor nutrition in an area can be caused by infectious diseases related to the high incidence of infectious diseases and poor environmental health [27], which in turn these factors not only affect stunting in the area but also stunting in the surrounding area. The prevalence of underweight has a positive and significant effect in increasing the prevalence of stunting, so these districts/cities need to pay attention to the condition of underweight, both in their area and surrounding areas.

Each percent increase in the population who behaves correctly in defecating in an area reduces the average stunting prevalence in that area by 0.028%. Each percent increase in the population who behaves correctly in defecating in the surrounding areas reduces the average stunting prevalence in that area by 0.001%. Each percent increase in the population who behaves correctly in defecating in Java reduces the average stunting prevalence in Java by 0.029%. Seeing the importance of defecation behavior on health, the government has targeted zero percent open defecation in the 2020-2024 RPJMN. According to research conducted by UNICEF and WHO were more than 370 Indonesian toddlers died due to bad behavior open defecation. Open defecation can increase the risk of stunted physical growth in children [28].

Real expenditure per capita adjusted is the only indicator in the decent living dimension to measure the human development index. The greater the per capita expenditure or the greater the purchasing power of the people, the more capable the community is to achieve a decent life. One of the causes of stunting is the inability of households to meet the needs of toddlers such as nutritious food and health services for children under five. According to Tuft (2001) and WHO (2007), stunting is caused by the inability of households to meet food needs both in terms of quality and quantity, and health care [29]. Each unit increase in real expenditure per capita adjusted in an area reduces the average stunting prevalence in that area by 0,698%. Each unit increase in real expenditure per capita adjusted in the surrounding areas reduces the average stunting prevalence in that area by 0.029%. Each unit increase in real expenditure per capita adjusted in Java reduces the average stunting prevalence in Java by 0.727%. Population expenditure not only has an impact on the economy in the region, but also has an impact on the economy of the surrounding area. This happens because the population's expenditure is not only to consume goods and services available in their area, but also goods and services available in other areas. The existence of this relationship can lead to a better economy in an area which can reflect the better household economy in that region. The better the economy of a household, the more capable the household is to meet the nutritional needs of children under five, so that stunting in the area can be reduced.

Toddlers' needs for food and health are very dependent on their families. The theory that reveals that higher education will make a person know more about health problems, and have insightful knowledge about the types of food and good sources of nutrition for family consumption [30]. Based on this, it is hoped that the higher a person's education, the more knowledgeable they will be on the needs of good nutrition for toddlers. Each unit increase in the average length of schooling in an area reduces the average stunting prevalence in that area by 0,994%. Each unit increase in the average length of schooling in the surrounding areas reduces the average stunting prevalence in that area by 0.042%. Each unit increase in the average length of schooling in Java reduces the average stunting prevalence in Java by 1.036%.

Knowledge can be obtained from anywhere and is not limited by a certain space. This is what causes a person's higher education in other areas to reduce stunting in a certain area. This knowledge sharing can occur both directly when meeting friends/family from other regions, and indirectly through online media, and various other social relationships that may occur. Education is also related to awareness to utilize health facilities, to be able to interact more effectively with health care providers and to be easier to comply with the input/suggestions given [30]. So, this illustrates that the more educated a person is, the better they know what is good for toddlers, such as the use of health facilities when giving birth, complying with the advice given for the good of the child's growth and development starting from the gestation period.

One of the causes of unresolved public health problems is difficult access to health services [31]. Improving the quality of health services must pay attention not only from the perspective of service providers but also from the perspective of the community as users [31]. The high level of difficulty in accessing available health facilities in an area can cause public health problems, one of which is the problem of stunting. The health facilities available in an area can not only be used by people living in that area, but also people from other areas. Utilization of health facilities in other areas can occur during family visits, when access to health facilities outside the area is easier to reach than in the area itself, and other social relationships that may occur. Therefore, the ease of access to health facilities offered in an area can not only reduce the incidence of stunting in the region, but also the incidence of stunting in other areas. Each percent increase in the difficulty level of access to public health center/ sub-health center/ mobile health center/ village midwife in an area increases the average stunting prevalence in that area by 0.037%. Each percent increase in difficulty level of access to public health center/ sub-health center/ mobile health center/ village midwife in the surrounding areas increases the average stunting prevalence in that area by 0.002%. Each percent increase in difficulty level of access to public health center/ sub-health center/ mobile health center/ village midwife in Java increases the average stunting prevalence in Java by 0.039%.

Waste management by dumping it in any place not only has an impact on the health of the people in an area, but also has an impact on the health of the people in the surrounding area. The waste problem requires serious attention from various parties. Waste that is not handled properly will have an impact on the environment, the potential for greater disasters such as floods, and a decline in the quality of public health [32]. There is a significant relationship between environmental conditions and the incidence of stunting and the average stunting toddler comes from families with unfavorable environmental conditions [33]. Each percent increase in households that manage waste by dumping it in any place in an area increases the average stunting prevalence in that area by 0.169%. Each percent increase in households that manage waste by dumping it in any place in the surrounding areas increases the average stunting prevalence in that area by 0.007%. Each percent increase in households that manage waste by dumping it in any place in Java increases the average stunting prevalence in Java by 0.176%.

The existence of social relations between people from various regions such as family relationships, work relations, travel, and so on, can make people from other regions enjoy drinking water available in an area. Adequate drinking water has a close relationship with health, especially stunting [34]. The intervention of providing safe drinking water has a contribution to stunting prevention, and the government has targeted 100% access to safe drinking water by 2024 [34]. Each percent increase in households that have access to proper drinking water in an area reduces the average stunting prevalence in that area by 0.0104%. Each percent increase in households that have access to proper drinking water in the surrounding areas reduces the average stunting prevalence in that area by 0.0004%. Each percent increase in households that have access to proper drinking water in Java reduces the average stunting prevalence in Java by 0.0108%.

Due to the fact that R is a free piece of software with an open source code, it can be shared and improved by users who wish to do so [1], [10]. R shiny assists users in data processing, provides efficient and accurate results, and does so by minimizing errors in formulating R code for spatial autoregressive model using instrument variable and S-estimator. Figure 3 illustrates the menus contained within this R shiny of SAR. https://statistikterapan.shinyapps.io/Robust_SAR/ is the URL for the developed of R shiny.

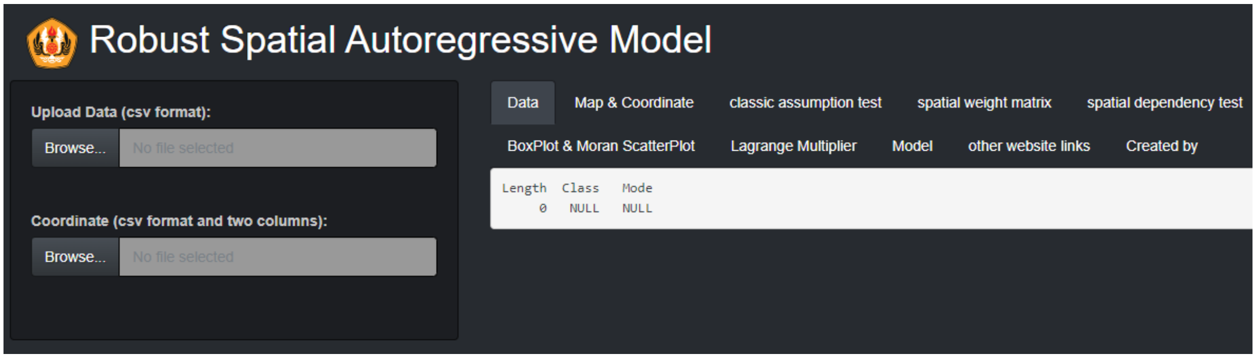

Figure 3. R shiny development result

4. CONCLUSION

The instrument variable with S-estimator is better to use when there are outliers because it can reduce the residual standard error and increase the coefficient of determination R-squared. The seven explanatory variables used in the model simultaneously have a significant effect on the prevalence of stunting in Java so that the prevalence of stunting in Java can be reduced by taking these seven factors into account. Four of the seven explanatory variables partially have a significant effect on the prevalence of stunting in Java, so the acceleration of stunting reduction can be carried out by taking into account these four factors. The four factors are underweight prevalence, education, expenditure (decent living), and waste management. The existence of a spillover effect illustrates that the incidence of stunting in every district/city in Java is not only influenced by changes in the causal factors in the area but is also influenced by changes in the causal factors in the surrounding area. Therefore, it is necessary to have good cooperation between districts/cities in Java island to achieve the stunting reduction target.

CONFLICT OF INTERESTS

The authors declare that there is no conflict of interests.

REFERENCES

[1] V.A.D. Djara, I.G.N.M. Jaya, The spatial econometrics of stunting toddlers in Nusa Tenggara Timur Province 2019, Commun. Math. Biol. Neurosci. 2021 (2021), 82. https://doi.org/10.28919/cmbn/6584.

[2] BPS, Laporan indeks khusus penanganan stunting 2018-2019, (2020). https://www.bps.go.id/publication/2020/12/01/fa48ee93a717baed2370d84a/laporan-indeks-khusus-penanganan-stunting-2018-2019.html.

[3] T. Sipahutar, T. Eryando, M.P. Budhiharsana, et al. Finding stunting hotspot areas in seven major islands using spatial analysis: For the acceleration of stunting prevention in Indonesia, medRxiv 2021.03.31.21254736. (2021). https://doi.org/10.1101/2021.03.31.21254736.

[4] TNP2K, 100 Kabupaten/Kota Prioritas untuk Intervensi Anak Kerdil (stunting), vol. 1. Jakarta Pusat: TNP2K, 2017.

[5] Sutarto, D. Mayasari, R. Indriyani, Stunting, Faktor Resiko dan Pencegahannya, J Agromed. 5 (2018), 540-545. http://repository.lppm.unila.ac.id/id/eprint/9767.

[6] I.G.N.M. Jaya, Y. Andriyana, Analisis Data Spasial Perspektif Bayesian. Alqaprint Jatinangor, Bandung, 2020.

[7] G. Arbia. Spatial econometrics: statistical foundations and applications to regional convergence. Springer, Berlin. 2006

[8] A.R. Hakim, B. Warsito, H. Yasin, Live expectancy modelling using spatial durbin robust model, J. Phys.: Conf. Ser. 1655 (2020), 012098. https://doi.org/10.1088/1742-6596/1655/1/012098.

[9] W.C. Mastuti, A. Djuraidah, E. Erfiani, Robust spatial regression model on original local government revenue in Java 2017, Indones. J. Stat. Appl. 4 (2020), 68–79. https://doi.org/10.29244/ijsa.v4i1.573.

[10] L. de F. Ramalho, W.R. de C. Segundo, R-shiny as an interface for data visualization and data analysis on the brazilian digital library of theses and dissertations (BDTD), Publications. 8 (2020), 24. https://doi.org/10.3390/publications8020024.

[11] C. Beeley, S. R. Sukhdeve, Web application development with R using shiny, Packt Publishing Ltd. Birmingham, 2018.

[12] S. Shekhar, C.T. Lu, P. Zhang, A unified approach to detecting spatial outliers. GeoInformatica, 7 (2003), 139–166. https://doi.org/10.1023/A:1023455925009.

[13] T. Wuryandari, A. Hoyyi, D.S. Kusumawardani, D. Rahmawati, Identifikasi autokorelasi spasial pada jumlahpengangguran di jawa tengah menggunakan indeks moran, Media Stat. 7 (2014), 1–10. https://doi.org/10.14710/medstat.7.1.1-10.

[14] L. Anselin, R.J.G.M. Florax, S.J. Rey, eds., Advances in spatial econometrics: methodology, tools and applications, 1. ed., softcover version of original hardcover ed. 2004, Springer, Berlin, 2010.

[15] K. Kopczewska, Applied spatial statistics and econometrics, Routledge, New York, 2021.

[16] G. Arbia, B.H. Baltagi, Spatial econometrics: methods and applications, Physica-Verlag, Heidelberg, 2009.

[17] L. Anselin, Spatial econometrics: methods and models, Kluwer Academic Publishers, Amsterdam,1988.

[18] Y. Susanti, H. Pratiwi, S. Sulistijowati H., T. Liana, M estimation, S estimation, and MM estimation in robust regression, Int. J. Pure Appl. Math. 91 (2014), 349-360. https://doi.org/10.12732/ijpam.v91i3.7.

[19] M. Salibian-Barrera, V.J. Yohai, A fast algorithm for S-regression estimates, J. Comput. Graph. Stat. 15 (2006), 414–427. https://doi.org/10.1198/106186006X113629.

[20] H. Musyarofah, H. Yasin, T. Tarno, Robust spatial autoregressive untuk pemodelan angka harapan hidup provinsi jawa timur, J.Gauss. 9 (2020), 26–40. https://doi.org/10.14710/j.gauss.v9i1.27521.

[21] P.J. Rousseeuw, A.M. Leroy, Robust regression and outlier detection, Wiley, New York, 1987.

[22] W. Setiawan, N.N. Debataraja, E. Sulistianingsih, Metode estimasi-S pada analisis regresi robust dengan pembobotan Tukey bisquare, Bul. Ilmiah Mat, Stat, Terapannya (Bimaster). 8 (2019), 289-296. https://doi.org/10.26418/bbimst.v8i2.32354.

[23] [1]N. Atikah, D.L. Afifah, N. Kholifia, Robust spatial regression model in City Minimum Wages (CMW) in East Java 2018, Adv. Soc. Sci. Educ. Human. Res. 528 (2021), 315-322. https://doi.org/10.2991/assehr.k.210305.045.

[24] G. Apriluana, S. Fikawati, Analisis Faktor-Faktor Risiko terhadap Kejadian Stunting pada Balita (0-59 Bulan) di Negara Berkembang dan Asia Tenggara, Media Penelitian dan Pengembangan Kesehatan (Media of Health Research and Development), 28 (2018), 247–256. https://doi.org/10.22435/mpk.v28i4.472.

[25] D. Izwardi, Kebijakan dan Strategi Penanggulangan Masalah Gizi, Widyakarya Nas. Pangan dan Gizi XI tahun 2018 (2018), pp. 1–34.

[26] R.I. Kemenkes, Buku Saku Pemantauan Status Gizi Tahun 2017, Kementerian Kesehatan Republik Indonesia, Jakarta (2018).

[27] A. S. Hartono, N. A. Zulfianto, M. Rachmat, Surveilans Gizi. 2017. http://perpus.poltekkesjkt2.ac.id/setiadi/index.php?p=show_detail&id=3332.

[28] Dinkes provinsi Sumatera Utara, Buang Air Besar Sembarangan (BABS), 2017. http://dinkes.sumutprov.go.id/artikel/buang-air-besar-sembarangan-babs (accessed Jan. 11, 2022).

[29] I. Batubara, S. Juwarni, Faktor-Faktor Yang Berhubungan Dengan Kejadian Stunting di Kecamatan Sayurmatinggi Kabupaten Tapanuli Selatan, J. Reprod. Heal. 3 (2018), 90-98.

[30] D. Wahyuni, R. Fitrayuna, Pengaruh sosial ekonomi dengan kejadian stunting di desa kulau tambang kampar," Preportif J. Kesehat. Masy. 4 (2020), 20–26. https://doi.org/10.31004/prepotif.v4i1.539.

[31] H. Megatsari, A. Dwi Laksono, I. Akhsanu Ridlo, M. Yoto, and A. Nur Azizah, Perspektif Masyarakat tentang Akses Pelayanan Kesehatan, Bul. Penelit. Sist. Kesehat. 21 (2018), 247–253. https://doi.org/10.22435/hsr.v21i4.231

[32] Disperkimta, Dampak Lingkungan Kotor dan Polusi Sampah, 2019. https://disperkimta.bulelengkab.go.id/informasi/detail/artikel/dampak-lingkungan-kotor-dan-polusi-sampah-32 (accessed Jan. 21, 2022).

[33] N. Mukaramah, M. Wahyuni, Hubungan Kondisi Lingkungan dengan Kejadian Stunting pada Balita di Rt 08, 13 dan 14 Kelurahan Mesjid Kecamatan Samarinda Seberang 2019, Borneo Student Res. 1 (2020), 750–754.

[34] Kemenkopmk, Menko PMK: Ketersediaan Air Bersih Mampu Cegah Stunting, 2021. https://www.kemenkopmk.go.id/menko-pmk-ketersediaan-air-bersih-mampu-cegah-stunting (accessed Jan. 21, 2022).