SIMULTANEOUS SPATIAL OF POVERTY AND HDI USING GS2SLS

Nur Jihan Salsabila[1], Nur Azizah[2], Dwi Endah Kusrini[3], Destri Susilaningrum[4]
[1,2,3,4]Institut Teknologi Sepuluh Nopember; Keputih, Sukolilo District, Surabaya, East Java, (031)5947274
Email: [2]azizah@its.ac.id

Abstract

The SDGs program encourages change towards sustainable development which makes poverty alleviation the main goal. Poverty is a person's inability to meet the minimum standard of living that hinders the welfare of an individual. The benchmark for welfare is the Human Development Index (HDI). It is suspected that there are spatial influences between regions because, in terms of territoriality, the province of East Java has similarities in the value of the percentage of poor people and HDI in nearby areas. Poverty and HDI and vice versa have a relationship that affects each other, so modeling is done with a system of simultaneous similarities. This work used a queen contiguity weight matrix and the Generalized Spatial Two Stage Least Squares (GS2SLS) approach to analyze spatial simultaneous equations. This method can cope with autocorrelation and heteroskedasticity. The data used are the percentage of poor people and HDI as well as variables from previous studies that are thought to significantly affect poverty and HDI in 38 Regencies/Cities of East Java in 2019. The results showed that there was a negative reciprocal relationship between the percentage of poor people and HDI. The spatial effect is positive and significant on the HDI variables with GS2SLS Spatial Autoregressive (SAR) modeling, while the percentage of poor people without spatial effects is so modeled with Two Stage Least Square (2SLS). HDI and GRDP growth rates significantly affect the percentage of poor people, while HDI is significantly influenced by the percentage of poor people and population density.

Keywords: 2SLS, GS2SLS, HDI, Poverty, SDGs

INTRODUCTION

East Java Province, has the second highest GRDP in Indonesia after DKI Jakarta province, but the handling of poverty in East Java Province which is rudimentary and not fully successful can be seen from the relatively high poverty rate. It was noted at the Badan Pusat Statistik (BPS) in 2019 that the number of poor people in Indonesia reached 24,785 million people with 16.36 percent of them domiciled in East Java Province or equivalent to 4,056 million people, which is 10.2 percent of the total population in East Java Province. Although from 2015 to 2019 it decreased, this figure is still above the percentage of poor people nationally, which is 9.22 percent [1].

Every nation deals with poverty, and emerging nations like Indonesia are particularly at risk. The SDGs program is a new development agreement that promotes transformation to sustainable development based on equality and human rights to promote social, economic, and environmental development, making the reduction of poverty an important objective [2].

According to Kuncoro, poverty is a person's inability to meet minimum living standards such as clothing, food, and shelter [3]. This will have an impact on the welfare of the individual. The population belonging to the poor group has limitations in production factors, so low productivity and income generated are far from sufficient, which results in basic needs such as food, clothing, and housing being difficult to meet. Other needs are also limited, which results in it will be difficult to enter human development in areas with a relatively large number of poor people. By BPS data from 2015 to 2019, the HDI of East Java has grown every year, while the percentage of poor people has decreased. This condition shows that an increase in the HDI value is followed by a decrease in the percentage of the poor population.

Several studies have been conducted on the analysis of factors affecting poverty and HDI, including [4-7] in their research on the analysis of factors affecting poverty, it is known that HDI has a significant effect on poverty. Meanwhile, Rohmah and Trunajaya in their research on the analysis of factors affecting HDI can be concluded that poverty has a significant effect on HDI [8-9]. This shows that poverty and HDI affect each other. The results of research by Sofilda, Hamzah, and Sholeh reinforce the statement that poverty and HDI have a negative reciprocal relationship and significantly affect each other two[10].

Based on the background that has been presented, modeling is not enough with a single equation because there are reciprocal relationships between variables. Relationships between variables in simultaneous equations can provide more comprehensive information about interrelated economic problems[11].

When looking at the geography of East Java province's poverty and HDI, districts on the island of Madura (Sampang, Bangkalan, Pamekasan, and Sumenep regencies) have a disproportionately high percentage of the poor and a nearly same percentage. When viewed on the map, these regions are side by side, so it is suspected that there is a functional relationship of dependence that is observed in the socio-economic criteria, human resources, economic development planning policies, and so on. In accordance with the first law on geography, it reads "Everything is interconnected with one another, but something more adjacent will have more effect than something far away" [12]. This suggests that economic activity in one region affects the resources of other neighboring regions, resulting in spatial dependence. Research by Rosa, Maiyastri, and Yozza (2021) and Laswinia (2016) on the spatial effects on poverty and HDI data and the conclusion was reached that poverty and HDI have spatial lag effects[13-14].

This study is different from the research of Sofilda, et al.[10], because this study pays attention to spatial effects. This study is also different from the research of Rosa, Maiyastri, and Yozza and Laswinia [13-14] which is only a single regression by pays attention to spatial effects. By focusing on spatial factors, this study seeks to understand how the percentage of the impoverished population and the HDI are related to one another. One of the characteristics of spatial regression models is the interdependence between regions, which makes estimating the model more difficult. When the observation unit and response variables are interconnected between regions, it is said that there is spatial lag and there may be a correlation of errors. This condition causes Ordinary Least Square (OLS) to produce inconsistent estimation. The method used in the study was GS2SLS with a queen contiguity weight matrix.

Estimation using the GS2SLS method will produce consistent and normal asymptotic estimates, as well as being computationally simpler [15]. The GS2SLS method has three stages in which there are generalized moment and Cochran-Orcutt methods, which can overcome autocorrelation and heteroskedasticity.

THEORETICAL BASIS

This study uses the *Generalized Spatial Two Stage Least Squares* (GS2SLS) method which is a limited *information estimation* method if there is a spatial effect. The GS2SLS estimation method is used to estimate the parameters in the *i*th equation. The GS2SLS estimation process consists of three stages, [16] namely:

1) Two Stage Least Squares Estimation

Autoregressive spatial elements, namely vectors $\overline{\mathbf{Y}}_{j,n}$, intersect with an error ($\mathbf{u}_{j,n}$) which results in δ_j cannot be estimated consistently with OLS because $E\left[\overline{\mathbf{Y}}_{j,n}\mathbf{u}'_{j,n}\right] \neq 0$, the estimation of parameter δ_j with the Two Stage Least Square method uses the instrument matrix $\mathbf{H}_n$.

i. Regressing OLS between $\mathbf{Z}_{j,n}$ and $\mathbf{H}_n$, so that it is obtained

$$\mathbf{Z}_{j,n} = \mathbf{P}_H \mathbf{Z}_{j,n}; \text{with } \mathbf{P}_H = \mathbf{H}_n(\mathbf{H}'_n\mathbf{H}_n)^{-1}\mathbf{H}'_n,\ j = 1,\dots M \tag{1}$$

and it can be known $\mathbf{Z}_{j,n} = (\mathbf{Y}_{j,n}, \mathbf{X}_{j,n}, \overline{\mathbf{Y}}_{j,n})$ where $\mathbf{Y}_{j,n} = \mathbf{P}_H\mathbf{Y}_{j,n}$ and $\overline{\mathbf{Y}}_{j,n} = \mathbf{P}_H\overline{\mathbf{Y}}_{j,n}$.

ii. Regressing OLS between $\mathbf{y}_n$ with $\mathbf{Z}_{j,n}$, so that it is obtained

$$\delta_{j,n} = (\mathbf{Z}'_{j,n}\mathbf{Z}_{j,n})^{-1}\mathbf{Z}'_{j,n}\mathbf{y}_{j,n},\ j = 1,\dots,M \tag{2}$$

It is found that $\delta_{j,n}$ is a consistent estimator $\delta_{j,n}$, but does not provide information about the spatial correlation of the components of the $\mathbf{u}_j$. disturbance. The estimated 2SLS of the first stage is obtained from residual

$$\mathbf{u}_{j,n} = \mathbf{y}_{j,n} - \mathbf{Z}_{j,n}\delta_{j,n},\ j = 1,\dots,M \tag{3}$$

2) Estimation of *Autoregressive Spatial* Parameters from the *Disturbance* Process(λ)

Spatial Autoregressive parameters are estimated using the *generalized moments* method in the *disturbance* process of each *j*-th equation for *j*=1,…,*M* [17]. Equation $\mathbf{Y}_n = \mathbf{Y}_n\mathbf{B} + \mathbf{X}_n\mathbf{C} + \overline{\mathbf{Y}}_n\mathbf{\Lambda} + \mathbf{U}_n$ used to support this method so that it is obtained $\hat{\lambda}_i$ and $\hat{\sigma}^2_j$ which is an estimator for λ_i and σ^2_j. Equation (4) used is $y_{j,n} = \mathbf{Z}_{j,n}\delta_j + \mathbf{u}_{j,n}$,

$$\mathbf{u}_{j,n} = \lambda_j\mathbf{W}_n\mathbf{u}_{j,n} + \varepsilon_{j,n}, \tag{4}$$

where

$\mathbf{Z}_{j,n} = (\mathbf{Y}_{j,n}, \mathbf{X}_{j,n}, \mathbf{W}_n\mathbf{Y}_{j,n})$ and $\boldsymbol{\delta}_j = (\boldsymbol{\beta}'_j, \boldsymbol{\gamma}'_j, \boldsymbol{\rho}'_j)',\ j = 1, 2, \dots, M$

. Persamaan on the second line and can be described as follows.

$$\mathbf{u}_{j,n} - \boldsymbol{\lambda}_j\mathbf{W}_n\mathbf{u}_{j,n} = \varepsilon_{j,n} \text{ or } \mathbf{u}_{j,n} - \boldsymbol{\lambda}_j\overline{\mathbf{u}}_{j,n} = \varepsilon_{j,n},\ j = 1,\dots,M \tag{5}$$

If equation (5) is multiplied by weighting $\mathbf{W}_n$ will generate equations (6).

$$\mathbf{W}_n\mathbf{u}_{j,n} - \boldsymbol{\lambda}_j\mathbf{W}^2_n\mathbf{u}_{j,n} = \mathbf{W}_n\varepsilon_{j,n} \text{ or } \overline{\mathbf{u}}_{j,n} - \boldsymbol{\lambda}_j\overline{\overline{\mathbf{u}}}_{j,n} = \overline{\varepsilon}_{j,n},\ j = 1,\dots,M \tag{6}$$

Contrasting equations (5) and (6) and multiplying equations (5) and (6) so that three new equations are formed which are then divided by *n* observations, shown in equation (7).

$$\begin{gathered}
\frac{\varepsilon'_{j,n}\varepsilon_{j,n}}{n} = \frac{\mathbf{u}'_{j,n}\mathbf{u}_{j,n}}{n} + \boldsymbol{\lambda}^2_j\frac{\overline{\mathbf{u}}'_{j,n}\overline{\mathbf{u}}_{j,n}}{n} - 2\boldsymbol{\lambda}_j\frac{\mathbf{u}'_{j,n}\overline{\mathbf{u}}_{j,n}}{n} \\
\frac{\overline{\varepsilon}'_{j,n}\overline{\varepsilon}_{j,n}}{n} = \frac{\overline{\mathbf{u}}'_{j,n}\overline{\mathbf{u}}_{j,n}}{n} + \boldsymbol{\lambda}^2_j\frac{\overline{\overline{\mathbf{u}}}'_{j,n}\overline{\overline{\mathbf{u}}}_{j,n}}{n} - 2\boldsymbol{\lambda}_j\frac{\overline{\overline{\mathbf{u}}}'_{j,n}\overline{\mathbf{u}}_{j,n}}{n} \\
\frac{\varepsilon'_{j,n}\overline{\varepsilon}_{j,n}}{n} = \frac{\overline{\mathbf{u}}'_{j,n}\overline{\mathbf{u}}_{j,n}}{n} + \boldsymbol{\lambda}^2_j\frac{\overline{\mathbf{u}}'_{j,n}\overline{\overline{\mathbf{u}}}_{j,n}}{n} - \boldsymbol{\lambda}_j\frac{[\mathbf{u}'_{j,n}\overline{\overline{\mathbf{u}}}_{j,n} + \overline{\mathbf{u}}'_{j,n}\overline{\mathbf{u}}_{j,n}]}{n}
\end{gathered} \tag{7}$$

$E(n^{-1}\varepsilon'_{j,n}\varepsilon_{j,n}) = \sigma^2_j$ with σ^2_j is the *j-th* diagonal element of $\mathbf{\Sigma}$. $E(n^{-1}\overline{\varepsilon}'_{j,n}\overline{\varepsilon}_{j,n}) = \sigma^2_j n^{-1} tr(\mathbf{W}'_n\mathbf{W}_n)$ and $E(n^{-1}\varepsilon'_{j,n}\overline{\varepsilon}_{j,n}) = \sigma^2_j n^{-1} tr(\mathbf{W}_n) = 0$. If $\alpha_j = (\lambda_j, \lambda^2_j, \sigma^2_j)';$, $\gamma_{j,n} = n^{-1}[E(\mathbf{u}'_{j,n}\mathbf{u}_{j,n}), E(\overline{\mathbf{u}}'_{j,n}\overline{\mathbf{u}}_{j,n}), E(\mathbf{u}'_{j,n}\overline{\mathbf{u}}_{j,n})]'$, j = 1, …, *M*, so the equation of the moment form can be generated as follows.

$$\gamma_{j,n} = \mathbf{\Gamma}_{j,n}\alpha_j \tag{8}$$

where $\mathbf{\Gamma}_{j,n}$ is the shape of the moment according to

$$\mathbf{\Gamma}_{j,n} = n^{-1}\mathrm{E}\left\{\begin{matrix}
2\mathbf{u}'_{j,n}\overline{\mathbf{u}}_{j,n} & -\overline{\mathbf{u}}'_{j,n}\overline{\mathbf{u}}_{j,n} & n \\
2\overline{\overline{\mathbf{u}}}'_{j,n}\overline{\mathbf{u}}_{j,n} & -\overline{\overline{\mathbf{u}}}'_{j,n}\overline{\overline{\mathbf{u}}}_{j,n} & tr(\mathbf{W}'_n\mathbf{W}_n) \\
(\mathbf{u}'_{j,n}\overline{\overline{\mathbf{u}}}_{j,n} + \overline{\mathbf{u}}'_{j,n}\overline{\mathbf{u}}_{j,n}) & -\overline{\mathbf{u}}'_{j,n}\overline{\overline{\mathbf{u}}}_{j,n} & 0
\end{matrix}\right\}$$

, to j = 1,…,*M*

and $\gamma_{j,n}$ is a vector *of disturbance*, so the value of λ_j and σ^2_j in α_j can be known through equations $\alpha_j = \mathbf{\Gamma}^{-1}_{j,n}\gamma_{j,n}$. *General approach* to obtaining estimators for $\mathbf{\Gamma}_{j,n}$ and $\gamma_{j,n}$ which is provided $\mathbf{G}_{j,n}$ and $\mathbf{g}_{j,n}$ is [17],

$$G_{j,n} = \frac{1}{n}\begin{bmatrix} 2\tilde{\mathbf{u}}'_{j,n}\tilde{\bar{\mathbf{u}}}_{j,n} & \tilde{\bar{\mathbf{u}}}'_{j,n}\tilde{\bar{\mathbf{u}}}_{j,n} & n \\ 2\tilde{\bar{\bar{\mathbf{u}}}}'_{j,n}\tilde{\bar{\mathbf{u}}}_{j,n} & -\tilde{\bar{\bar{\mathbf{u}}}}'_{j,n}\tilde{\bar{\bar{\mathbf{u}}}}_{j,n} & tr(\mathbf{W}_n'\mathbf{W}_n) \\ (\mathbf{u}'_{j,n}\tilde{\bar{\bar{\mathbf{u}}}}_{j,n} + \tilde{\bar{\mathbf{u}}}'_{j,n}\tilde{\bar{\mathbf{u}}}_{j,n}) & -\tilde{\bar{\mathbf{u}}}'_{j,n}\tilde{\bar{\bar{\mathbf{u}}}}_{j,n} & 0 \end{bmatrix}$$

and

$$\mathbf{g}_{j,n} = n^{-1}[\tilde{\mathbf{u}}'_{j,n}\tilde{\mathbf{u}}_{j,n}, \tilde{\bar{\mathbf{u}}}'_{j,n}\tilde{\bar{\mathbf{u}}}_{j,n}, \tilde{\mathbf{u}}'_{j,n}\tilde{\bar{\mathbf{u}}}'_{j,n}]' \tag{9}$$

Estimation of the first stage with 2SLS based on equation (3) obtained residual $\mathbf{u}_{j,n} \cdot \tilde{\bar{\mathbf{u}}}_{j,n} = \mathbf{W}_n\tilde{\mathbf{u}}_{j,n}$, dan $\tilde{\bar{\bar{\mathbf{u}}}}_{j,n} = \mathbf{W}_n\tilde{\bar{\mathbf{u}}}_{j,n} = \mathbf{W}_n^2\tilde{\mathbf{u}}_{j,n}$.

The empirical form of equation (8) is obtained by the following equation.

$$\mathbf{g}_{j,n} = \mathbf{G}_{j,n}\alpha_j + \zeta_{j,n} \tag{10}$$

$\zeta_{j,n}$ is a regression residual vector. The generalized moment estimators for λ_j and σ_j^2 are $\tilde{\lambda}_j$ dan $\tilde{\sigma}_j^2$ which are the least squares nonlinear estimators by minimizing the residual squares of the equation (8) with the first derivative against α_j, so that it is obtained:

$$\mathbf{G}'_{j,n}\mathbf{g}_{j,n} = \mathbf{G}'_{j,n}\mathbf{G}_{j,n}\alpha_j$$

Then it can be estimated for α_j, namely

$$\hat{\alpha}_j = (\tilde{\lambda}_j, \tilde{\sigma}_j) = \left[\mathbf{G}'_{j,n}\mathbf{G}_{j,n}\right]^{-1}\mathbf{G}'_{j,n}\mathbf{g}_{j,n} \tag{11}$$

3) Generalized Spatial Two Stage Least Square Estimation

The first stage still does not pay attention to the correlation *of error disturbance* so that in the third stage in the estimation of the parameters of the spatial simultaneous equation is to obtain an estimation according to the equation $\mathbf{Y}_n = \mathbf{Y}_n\mathbf{B} + \mathbf{X}_n\mathbf{C} + \overline{\mathbf{Y}}_n\mathbf{\Lambda} + \mathbf{U}_n$. The result of the second stage, namely $\tilde{\lambda}_j$ as an estimate of λ_j, is substituted into the model that Cochran-Orcutt previously transformed. If μ is a scalar and it is defined that $\mathbf{y}^*_{j,n}(\mu) = \mathbf{y}_{j,n} - \mu\mathbf{W}_n\mathbf{y}_{j,n}$ and $\mathbf{Z}^*_{j,n}(\mu) = \mathbf{Z}_{j,n} - \mu\mathbf{W}_n\mathbf{Z}_{j,n}$, j = 1, …, M. Application of the *Cochran-Orcutt* transformation to the equation (4) obtained equations (12).

$$\mathbf{y}^*_{j,n}(\lambda_j) = \mathbf{Z}^*_{j,n}(\lambda_j)\boldsymbol{\delta}_j + \varepsilon_{j,n} \tag{12}$$

It is assumed that the moment λ_j is known based on the results of the estimation of the second stage, the GS2SLS estimation for δ_j and $\hat{\delta}_{j,n}$ is based on the 2SLS process as in the first stage, obtained

$$\hat{\delta}_{j,n} = \left[\hat{\mathbf{Z}}^*_{j,n}(\lambda_j)'\mathbf{Z}^*_{j,n}(\lambda_j)\right]^{-1}\hat{\mathbf{Z}}^*_{j,n}(\lambda_j)'\mathbf{y}^*_{j,n}(\lambda_j),\ j = 1,\dots,M \tag{13}$$

$\hat{\mathbf{Z}}^*_{j,n}(\lambda_j) = \mathbf{P}_H\mathbf{Z}^*_{j,n}(\lambda_j)$ with $\mathbf{P}_H = \mathbf{H}_n(\mathbf{H}_n'\mathbf{H}_n)^{-1}\mathbf{H}_n'$.

However, if the results of the analysis show that the data has no spatial effect, the method used is 2SLS (*Two Stage Least Square*). *Two Stage Least Square* (2SLS) is a method used to obtain an estimate of the structural coefficient of the *reduced-form* coefficient estimated in the structural equation identified over *identified*. In addition 2SLS can also be used to assess the exact structural equations identified. In 2SLS, independent variables (which correlate with errors) are replaced with their own estimated values. As the name implies, this method includes two consecutive applications of OLS[18].

Here are steps 2SLS, for the example given the equation contained in the equation (14) and (15).

$$Y_{1n} = \beta_{11}Y_{2n} + \gamma_{11}X_{1n} + \gamma_{12}X_{2n} + \gamma_{13}X_{3n} + \varepsilon_{1n} \tag{14}$$

$$Y_{2n} = \beta_{21}Y_{1n} + \gamma_{21}X_{4n} + \gamma_{22}X_{5n} + \gamma_{23}X_{6n} + \varepsilon_{2n} \tag{15}$$

Stage 1. Regresses between Y_{1n} with all predetermined variables in the system using OLS, to avoid a correlation between Y_{1n} with ε_{2n}, so that it is obtained:

$$Y_{1n} = \hat{\Pi}_0 + \hat{\Pi}_1X_{1n} + \hat{\Pi}_2X_{2n} + \hat{\Pi}_3X_{3n} + \hat{\Pi}_4X_{1n} + \hat{\varepsilon}_{1n} \tag{16}$$

or

$$Y_{1n} = \hat{Y}_{1n} + \hat{\varepsilon}_{1n} \quad (17)$$

Consider the equation (16) is a reduced form regression equation because it is only an exogenous or predetermined variable.

Stage 2. Substitution of Y_{1n} on the equation (17) into the equation (15), so that it can be written

$$\begin{aligned} Y_{2n} &= \beta_{20} + \beta_{21}(\hat{Y}_{1n} + \hat{\varepsilon}_{1n}) + \gamma_{21}X_{4n} + \gamma_{22}X_{5n} + \gamma_{23}X_{6n} + \varepsilon_{2n} \\ &= \beta_{20} + \beta_{21}\hat{Y}_{1n} + \gamma_{21}X_{4n} + \gamma_{22}X_{5n} + \gamma_{23}X_{6n} + (\beta_{21}\hat{\varepsilon}_{1n} + \varepsilon_{2n}) \\ &= \beta_{20} + \beta_{21}\hat{Y}_{1n} + \gamma_{21}X_{4n} + \gamma_{22}X_{5n} + \gamma_{23}X_{6n} + \varepsilon^{*}_{2n} \end{aligned} \quad (18)$$

METODE PENELITIAN

The data used in this study was obtained from the website and annual publications of the Badan Pusat Statistik (BPS) of East Java, as well as the decree of the Governor of East Java No.665 concerning the District/City Minimum Wage for East Java in 2019. The type of data used is cross-section data. The object of this study was 38 regencies/cities observed in East Java in 2019. The selection of exogenous variables against the percentage of poor people and HDI is based on previous research references, these variables in the study it is divided into endogenous variables and exogenous variables shown in Table 1.

Table 1. Research Variables

Variable	Notation	Meaning	Denomination
Endogenous	P_0	Percentage of Poor Population	Percent
	IPM	Human Development Index	Percent
Exogenous	LPDRB	Gross Regional Domestic Product Rate	Percent
	TPT	Open Unemployment Rate	Percent
	UMK	District/City Minimum Wage	Million Rupiah
	AML	Percentage of Households That Have Access to Adequate Drinking Water Sources	Percent
	RK	Dependency Ratio	-
	KP	Population Density	Soul/Km^2

The relationship between the variables used in this study is shown in Figure 1. According to the findings of earlier studies, the variables lpdrb, UMK, and TPT have a significant impact on the proportion of the population who live in poverty, and the variables RK, KP, and AML have a significant impact on HDI.

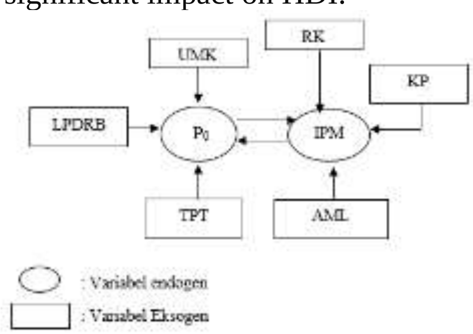

Figure 1. Scheme of Relationships Between Variables

An important aspect that needs to be considered before doing spatial modeling is the spatial weighting used. This study uses a contiguity approach as a weighing matrix.

Figure 2. Map of Regencies/Cities in East Java

Table 2. Research Areas

No	Regencies/Cities
1	Pacitan Regency
2	Ponorogo Regency
3	Trenggalek Regency
4	Tulungagung Regency
5	Blitar Regency
6	Kediri Regency
7	Malang Regency
8	Lumajang Regency
9	Jember Regency
10	Banyuwangi Regency
11	Bondowoso Regency
12	Situbondo Regency
13	Probolinggo Regency
14	Pasuruan Regency
15	Sidoarjo Regency
16	Mojokerto Regency
17	Jombang Regency
18	Nganjuk Regency
19	Madiun Regency
20	Magetan Regency
21	Ngawi Regency
22	Bojonegoro Regency
23	Tuban Regency
24	Lamongan Regency
25	Gresik Regency
26	Bangkalan Regency
27	Sampang Regency
28	Pamekasan Regency
29	Sumenep Regency
30	Kediri City
31	Blitar City
32	Malang City
33	Probolinggo City
34	Pasuruan City
35	Mojokerto City
36	Madiun City
37	Surabaya City
38	Batu City

Modeling the level of poverty and HDI in East Java with spatial simultaneous equations is carried out through several stages, namely:

1) Creating a model formula of simultaneous structural equations.

   The formulation of endogenous variables used in this study refers to research conducted by Sofilda et al.,[10]. Based on empirical evidence from the research of Laswinia and Rosa et al., which states the existence of spatial *lag* in the percentage of poor people and HDI [13-14], the hope of model specifications in this study uses two endogenous variables and six exogenous variables with equations (19) and (20).

$$P0_i = \gamma_{10} + \beta_1 IPM_i + \gamma_{11} LPDRB_i + \gamma_{12} TPT_i + \gamma_{13} \log UMK_i + \rho_1 \sum_{i=1}^{n} \sum_{j=1}^{n} w_{ij} P0_i + \varepsilon_1 \tag{19}$$

   with the expected coefficient $\beta_1, \gamma_{11}, \gamma_{13}$ <0 and γ_{12}>0

$$IPM_i = \gamma_{20} + \beta_2 P0_i + \gamma_{21} AML_i + \gamma_{22} \log RK_i + \gamma_{23} \log KP_i + \rho_2 \sum_{i=1}^{n} \sum_{j=1}^{n} w_{ij} IPM_i + \varepsilon_2 \tag{20}$$

   with the expected coefficient $\beta_2, \gamma_{22}, \gamma_{23}$ <0 and γ_{21}>0

2) Model identification to find out if structural equations are identified or not.

   A system of simultaneous equations is considered identifiable if identified *exactly identified* and *over identified*. A system of simultaneous equations if the values of the parameters cannot be precisely predicted from the *reduced form* equation of the system of simultaneous equations n then, the system of simultaneous equations is considered unidentified (*unidentified* or *underidentified*) [20].

Exactly identified occurs when the parameter estimation value of the reduced form equation of the system of simultaneous equations can be estimated by only one equation, while *over identified* occurs when the parameter estimation value of the *reduced form* equation can be estimated with more than one equation [18].

An equation can be considered identified, if the two conditions the order condition of identifications and *the rank condition of identifications* are met [18]. An equation can be considered identified, if the two conditions the order condition of identifications and *the rank condition of identifications* are met .

M = endogenous variables in the model as a whole

m = a particular equation's endogenous variable count.

K = total amount of fixed variables in the model

k = amount of variables in an equation that are known in advance

The order condition of identifications provides information about exactly *identified* or *overidentified*. *The order condition* is said to be *exactly identified* if *K-k = m-1* and *overidentified* if *K-k > m-1*.

The rank condition of identifications provides information about whether or not an equation is identified. *The rank condition* is said to be *identified* if the rank of the matrix $(\mathbf{A})$ equals M-1 and *unidentified* if matrix rank $(\mathbf{A})$ less than M-1, with matrix $(\mathbf{A})$ is a matrix formed from the coefficients of the entire variable that is excluded from the equation under consideration but is included in other equations of the model.

This study only uses two structural equations so that for the matrix rank (**A**) produced is a row matrix where it has no inverse and the determinant of the matrix is not zero, so the equation can be said to be identified.

3) Determine the spatial weight matrix according to information about the map of regencies/cities in East Java.

The weight matrix used is *queen contiguity* because the information obtained is more complete by requiring a grouping of regions that have side or angle intersections of the region. Generally, a transformation is performed to get the number of rows equal to one. The formation of the weight matrix in this study refers to the Figure and the results of the matrix attached to Appendix 1.

4) Conducting a *Hausman* simultaneous test on the model to find out if it has a simultaneous relationship between structural equations.

According to [21] the simultaneous test aims to prove that a system of equation models actually has simultaneous relationships between its structural equations, using hypotheses,

$H_0 : E(\varepsilon|\mathbf{Z}) = \mathbf{0}$ (endogenous variables do not correlate with *error*)

$H_1 : E(\varepsilon|\mathbf{Z}) \neq \mathbf{0}$ (endogenous variables correlated with *error*)

The OLS method cannot be used if the endogenous variable correlates with *error*, because the estimation results are consistent but inefficient.

Hausman's simultaneous testing has several steps, namely:

i. Regressing each endogenous variable with the *reduced form* equation and save the value as an endogenous variable and also store the residual value u $\hat{Y}_n$ n obtained from the

ISSN No. 1978-3787
Open Journal Systems

difference between the estimated value of the endogenous variable and the observation value.

ii. Substitution of endogenous variables on structural equations with the results of estimates and residual values obtained.

iii. Regressing along with other free variables on structural equations.

The Significance Test uses an F-test or t-test for one regression coefficient of u_n. If the test produces a significant coefficient then the null hypothesis is rejected. If the test shows that the residual variable is insignificant then it can be concluded that there are enough reasons not to reject the initial hypothesis, meaning that there is an element of simultaneity with other equations.

1) Perform spatial dependency tests with Moran's Index and *Lagrange Multiplie* test.

a. Moran's Index

The Moran's index is annotated with *I*, first put forward by Patrick Alfred Pierce Moran (1948, 1950) which is only used on dependent changers, then expanded by Cliff and Ord (1981) so that the Moran index can be used on the error of the linear regression model [22]. Moran's index formula (*I*) with weighting has been standardized amd shown in the equation (21)

$$I = \frac{\boldsymbol{e}'\mathbf{W}\boldsymbol{e}}{\boldsymbol{e}'\boldsymbol{e}} \tag{21}$$

where $\boldsymbol{e}$ is the residual vector 2SLS. The hypothesis used in this test is

H_0 : $I = 0$ (No spatial dependencies)

H_1 : $I > 0$ (There are spatial dependencies)

The test statistics used are shown on the equation (22)

$$Z(I) = \frac{I - E(I)}{\sqrt{V(I)}} \tag{22}$$

with

$$E(I) = tr(\mathbf{MW})/(n-k)$$

$$V(I) = [tr(\mathbf{MWMW}') + tr(\mathbf{MW})^2 + (tr(\mathbf{MW}))^2]/d - E(I)^2$$

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

$$d = (n-k)(n-k+2)$$

$Z(I)$ normally distributed, to a predetermined degree of *significant* α, the null hypothesis is rejected if $|Z(I)| > Z_{\alpha/2}$ which means there are spatial dependencies.

b. Lagrange Multiplier Test (LM Test)

Testing using the *Lagrange Multiplier* can be called a one-way test in the sense that test statistics are designed to test the specification that one hypothesis is true at a time when the other is considered zero [12]. The LM test also allows selecting spatial error models or spatially lagged models .

The existence of spatial lag dependency, so spatial model testing with hypotheses is carried out [23],

$H_0 : \rho = 0$ (there is no spatial lag dependency)

$H_1 : \rho \neq 0$ (there is a spatial lag dependency)

LM test statistics for the spatial lag dependency test formulated by [23] are shown in the equation (23).

$$LM_{lag} = \frac{\left(\frac{\boldsymbol{e}'\mathbf{W}_n\boldsymbol{y}}{s^2}\right)^2}{\left((\mathbf{W}_n\mathbf{X}\gamma)'\left(\mathbf{I} - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}^{-1}\right)(\mathbf{W}_n\mathbf{X}\gamma) + Ts^2\right)}$$

$$= \frac{\left(\frac{\boldsymbol{e}'\mathbf{W}_n\boldsymbol{y}}{s^2}\right)^2}{Q}, LM \square \chi^2_{(1)} \tag{23}$$

The null hypothesis is rejected if $LM_{lag} > \chi^2_{(\alpha,1)}$, meaning there is a spatial lag dependency.

Once the conclusion is reached on whether there is a spatial lag dependency, it is important to test for the presence of spatial error dependency with a hypothesis

$H_0 : \lambda = 0$ (There is no spatial error dependency)

$H_1 : \lambda \neq 0$ (There is a spatial error dependency)

LM test statistics for error spatial dependency tests formulated by Burridge [24] are shown on the equation (24).

$$LM_{error} = \frac{\left(\frac{\boldsymbol{e}'\mathbf{W}_n\boldsymbol{e}}{s^2}\right)^2}{T}, \; \boldsymbol{LM} \sim \chi^2_{(1)} \tag{24}$$

The null hypothesis is rejected if $LM_{error} > \chi^2_{(\alpha,1)}$, meaning there are spatial dependencies of errors.

Both LM tests can yield significant conclusions, so it is not yet clear which model was used. The robust test against the misspesification is a *Robust* LM test that can be more precise to identify which spatial regression model is used. LM test statistics are modified at the time of conducting tests on $\boldsymbol{\rho} = \mathbf{0}$ and $\boldsymbol{\lambda} \neq \mathbf{0}$ and vice versa [25].

Modification to LM test statistics for Robust LM_{lag} as follows

$$RLM_{lag} = \frac{\left(\frac{\boldsymbol{e}'\mathbf{W}_n\boldsymbol{y}}{s^2} - \frac{\boldsymbol{e}'\mathbf{W}_n\boldsymbol{e}}{s^2}\right)^2}{s^{-2}Q - T} \tag{25}$$

Testing with hypothesis $\boldsymbol{\rho} = \mathbf{0}$ and $\boldsymbol{\lambda} \neq \mathbf{0}$ (*Robust* LM_{error}) then the modification will be as follows

$$RLM_{error} = \frac{\left(\frac{\boldsymbol{e}'\mathbf{W}_n\boldsymbol{e}}{s^2} - TQ^{-1}\boldsymbol{e}'\mathbf{W}_n\boldsymbol{y}\right)^2}{T - T^2 s^2 Q^{-1}} \tag{26}$$

Which one

$T = tr((\mathbf{W}_n' + \mathbf{W}_n)\mathbf{W}_n)$.

$s^2 = n^{-1}\boldsymbol{e}'\boldsymbol{e}$

with

$\mathbf{W}_n$: Spatial weight matrix of size $n \times n$

s^2 : estimated value of *the Variance of the Two Stage Least Squares* (2SLS) model

$\mathbf{H}_n$: Variable instrument matrix

$\boldsymbol{e}$: residual vector 2SLS

2) Estimation of simultaneous parameters of the equation of poverty rate and HDI with spatial effects using the GS2SLS method and if there is no spatial effect, it is estimated using the 2SLS method.
3) Parameter significance test.

The significific test is used to determine whether there is a meaningful influence between the predictor variables on the variable $x_1, x_2, ..., x_k$ respon *y*. There are two stages in the significance test, the first is the simultaneous test with the hypothesis,

$H_0 : \theta_1 = \theta_2 = ... = \theta_k = 0$ (Exogenous variables have no significant effect on endogenous variables)

$H_1 :$ there is at least one $\theta_j \neq 0$ (Exogenous variables have a significant effect on endogenous variables)

θ_j is a substitute for the parameter $\beta_i, \gamma_{ij}, \rho_i$ with the *F*-test statistics shown in equation (27) and a significant degree of α predetermined, the null hypothesis is rejected if $F > F_{\alpha,(k,n-k-1)}$ which means that simultaneously the predictor variable has a significant effect on the response variable.

$$F = \frac{MS_{regresi}}{MS_{error}} \tag{27}$$

Secondly, if there is at least one $\theta_j \neq 0$ then it can proceed to a partial test with a hypothesis,

$H_0 : \theta_j = 0$ (The *j-th* exogenous variable has no significant effect on the endogenous variable)

$H_1 : \theta_j \neq 0$ (The *j-th* exogenous variable has a significant effect on the endogenous variable)

ISSN No. 1978-3787
Open Journal Systems

The test statistics used in the partial test are shown in the equation (28)

$$t_0 = \frac{\hat{\theta}_j}{se(\hat{\theta}_j)} \tag{28}$$

The null hypothesis is rejected if $|t_0| > t_{\alpha/2, n-k-1}$, so it is concluded that the *j-th* exogenous variable has a significant effect on the endogenous variable [26].

4) Calculating the coefficient of determination using the residual final stage of parameter estimation.

The coefficient of determination denoted by R^2. Theaddition of the model on simultaneous equations can be evaluated using the residual value of the estimation of the final stage parameters by applying the equation (13). So that R^2 for spatial simultaneous equations can be shown in the equation (29)[27]. .

$$R^2_{sim,j} = \left[\tilde{\delta}_j' \mathbf{Z}^{*}_{j,n}{}' \boldsymbol{y}^*_j - n[E(\boldsymbol{y}^*_j)]^2 \right] / \left[\boldsymbol{y}^{*\prime}_j \boldsymbol{y}^*_j - n[E(\boldsymbol{y}^*_j)]^2 \right] \tag{29}$$

5) Checking the residual assumptions of the equation model obtained.

a. Heteroskedasticity Examination

Heteroskedasticity occurs when the variants of the population of dependent variables are not constant. According to [18], the problem of heteroskedasticity is common in *cross- sectional* data where a person usually deals with members of the population at a certain point in time. One way to detect the presence of heteroskedasticity with the Glejser test, the way it is to regress the absolute value of the residual with all the predictor variables, with the hypothesis,

$H_0 : \sigma_1^2 = \sigma_2^2 = \cdots = \sigma_k^2 = \sigma^2$ (No heteroskedasticity)

$H_1 :$ minimal ada satu $\sigma_j^2 \neq \sigma^2$ (There is heteroskedasticity)

with F-test statistics and a predetermined significant level of α, the null hypothesis is rejected if $F > F_{\alpha,(k,n-k-1)}$ which means the variance is not homogeneous or heteroskedasticity.

b. Normal Distribution Check

The normal distribution examination aims to find out whether the assumption of normality in residuals has been met, in this study *the Jarque-Fallow* (*JB*) test was used. The *JB* normality test in samples 20 to 1000 is more accurate when compared to the Kolmogorov Smirnov test [28], in addition to that according to [18], the *JB* test it is more accurately used when the sample is large. The *JB* test is an asymptotic test (large sample) based on a residual with a hypothesis,

H_0 = Residual normally distributed

H_1 = Residual non-normal distribution

First calculated the size of the skewness and kurtosis of the residual and using the test statistics shown in the equation (30).

$$\text{JB} = n\left[\frac{S^2}{6} + \frac{(L-3)^2}{24}\right] \tag{30}$$

where n is the sample size, S is the skewness coefficient which is the third moment of residual to its mean i.e. $S = E(\hat{\varepsilon}_j - \bar{\varepsilon})^3$, and L is the kurtosis coefficient which is the fourth moment of residual to the mean i.e. $L = E(\hat{\varepsilon}_j - \bar{\varepsilon})^4$. Normally distributed variables have S = 0 and L = 3, in this case the *JB* value is expected to be zero.

JB statistics follow the distribution of chi-squares with the degree of free is two. If the p-*value* calculated from the *JB* statistics is more than a significant level (α) then the null hypothesis fails to be rejected so it is concluded that the distributed residual is normal, but if the p-value is less than the significant level (α) then the null hypothesis is rejected and the residual is not normally distributed [18].

6) Interpreting based on the formed model.

Draw conclusions based on the results obtained.

RESULTS AND ANALYSIS

3.1. Model Identification

Model identification is performed to determine the estimation method to be used. Identification can indicate the presence or absence of the possibility of obtaining structural parameters i.e. a system of simultaneous equations derived from the reduced form parameters. The equation that can be processed is if the model is exactly identified or over identified. The order condition of the equation under study can be seen in Table 3.

Table 3. Simultaneous Equation Model Identification

Equation	K-k	m-1	Decision
(3.13)	6-3	2-1	*Over identified*
(3.14)	6-3	2-1	*Over identified*

The results of the order condition examination showed that the equations in the model system of simultaneous equations are categorized as over-identified equations because K-k > m-1. This study only uses two structural equations so that the resulting matrix rank (A) is a row matrix where it has no inverse and determinants of non-zero matrices, so that the equation can be said to be identified or the equation has met the requirements of sufficient rank condition.

3.2. Hasuman Simultaneity Test

The simultaneity test is carried out to empirically prove the existence of simultaneous relationships between its structural equations. T-statistics is used because there is only one endogenous variable as an exogenous variable in its structural equations. Hypotheses used

$H_0 : E(\varepsilon|\boldsymbol{Z}) = \boldsymbol{0}$ (endogenous variables are not correlated with errors)

$H_1 : E(\varepsilon|\boldsymbol{Z}) \neq 0$ (endogenous variables correlate with errors)

The level is significantly 10% so that H_0 is rejected if $t > t_{(0.1;df)}$ or p-value < 0.1, and the results of the simultaneousity test between the percentage of poor people and HDI are presented in Table 4.

Table 4. Hausman Simultaneousity Test Results

Equation	**Variable**	***t*-Statistics**	***p*-*value***	**Conclusion**
(3.13)	P_0	-2,29	0,029	There Are Simultaneous
(3.14)	IPM	-2,33	0,026	There Are Simultaneous

The results of the simultaneous tests demonstrate that both equations have simultaneous effects, thus the study of the simultaneous equations can move further.

3.3. Spatial Dependency Test

There are two spatial dependency tests, namely the Moran's Index and the Lagrange Multiplier. The Moran's Index is used to test the presence of spatial autocorrelation with hypothesis,

$H_0 : I = 0$ (there is no autocorrelation between regions)

$H_1 : I \neq 0$ (there is autocorrelation between regions)

however, both equations will be tested by Lagrange Multiplier to detect the spatial effect of lag or spatial error with hypothesis,

$H_0 : \rho = 0$ (there is no spatial lag dependency)

$H_1 : \rho \neq 0$ (there is a spatia lag dependency)

$H_0 : \lambda = 0$ (there is no spatial error dependency)

$H_1 : \lambda \neq 0$ (there is a spatial error dependency)

The results of the spatial dependency test with queen contiguity weights are presented in Table 5.

Table 5. Result of Spatial Dependency Test

Test	**P_0 Equation**		**IPM Equation**	
	Test Statistics	***p*-*value***	**Test Statistics**	***p*-*value***

Moran's I	1,318	0,187	2,426	0,015
LM *lag*	2,595	0,107	8,998	0,002
LM *error*	1,164	0,280	3,910	0,048

Table 3 shows that the results of Moran's test have autocorrelation between regions in the HDI equation at a significant level of 10% while in equation P0 there is no autocorrelation between regions so that estimates are carried out using the 2SLS method. The results of the LM test show that the HDI equation is detected to contain spatial lag and spatial error so that the best simultaneous spatial modeling of HDI perception based on the goodness of the model is GS2SLS Spatial Autoregressive (SAR).

3.4. Estimation Of Spatial Simultaneous Equation Parameters

Modeling on the percentage equation of the poor population (P_0) with the 2SLS procedure and the Human Development Index (HDI) with the spatial simultaneous equation system procedure, namely the GS2SLS Spatial Autoregresive (SAR). The spatial weighting used is the queen contiguity weight. Obtained parameter estimates presented in Table 6.

Table 6. Estimasi Parameter

Variabel Dependen	Variabel	Parameter	Koefisien
P_0	*Intercept*	γ_{10}	57,436
	IPM	β_1	-0,584
	LPDRB	γ_{11}	-1,257
	TPT	γ_{12}	0,080
	UMK	γ_{13}	1,042
	R-sq	76,88%	
IPM	*Intercept*	γ_{20}	28,108
	P_0	β_2	-0,282
	AML	γ_{21}	0,051
	RK	γ_{22}	-3,128
	KP	γ_{23}	2,635
	Rho	ρ_2	0,494
	R-sq	87,07%	

A parameter significance test is carried out θ_j that is a substitute for the parameter $\beta_i, \gamma_{ij}, \rho_i$, starting with testing simultaneously with the following hypothesis.

Hypothesis of simultaneous test of equation P_0:

$H_0 : \theta_{11} = \theta_{12} = \theta_{13} = \theta_{14} = 0$ (there is no influence of HDI, LPDRB, UMK, and TPT on the percentage of poor people)

$H_1 : \theta_{1j} \neq 0$ (there is at least one influence of HDI, LPDRB, UMK, and TPT on the percentage of poor people)

Hypothesis of simultaneous test of HDI equations:

$H_0 : \theta_{21} = \theta_{22} = \theta_{23} = \theta_{24} = \theta_{25} = 0$ (no influence of P0, AML, RK, KP, and spatial effects on HDI)

$H_1 : \theta_{2j} \neq 0$ (there is at least one influence of P0, AML, RK, KP, and spatial effects on HDI)

The null hypothesis is rejected if $F > F_{0.1(df1, df2)}$ or p-value < α with a significant degree of 10%. The statistical results of the test are shown in Table 7.

Table 7. Concurrent Test of Parameters

Dependence Variable	*F-Statistik*	*F*Tabel	p-*Value*	Decision
P_0	27,43	2,123	0,000	H_0 is rejected
IPM	76,32	2,036	0,000	H_0 is rejected

Table 5 shows that there is at least one influence of HDI, GRDP growth (LPDRB), Unemployment (TPT), and District/City Minimum Wage (UMK) on the percentage of poor people (P_0), as well as HDI is influenced by at least one of P_0, Access to Decent Drinking

Water Sources (AML), Dependency Ratio (RK) and Population Density (KP) so that it proceeds to partial tests with the following hypothesis.

Partial test hypothesis of equation P_0:

$H_0:\theta_{1j}=0$, j=1,2,3,4 (There is no influence of the jth exogenous variable on the percentage of the poor population)

$H_1:\theta_{1j}\neq 0$, j=1,2,3,4 (There is an influence of the jth exogenous variable on the percentage of the poor population)

Partial test hypothesis of the HDI equation:

$H_0:\theta_{2j}=0$, j=1,2,3,4,5 (No influence of jth exogenous variable on HDI)

$H_1:\theta_{2j}\neq 0$, j=1,2,3,4,5 (There is an influence of the jth exogenous variable on HDI)

A significant level of 10% is set so that H_0 is rejected if $t > t_{(0.1,df)}$ or p-value < 0.1. The results of the partial test of the parameters are presented in Table 8.

Table 8. Partial Test of Model Parameters

Dependence Variable	Variable	Parameter	*t*-Statistics	*P-Value*
P_0	*Intercept*	γ_{10}	7,976	0,000*
	IPM	β_1	-4,600	0,000*
	LPDRB	γ_{11}	-3,355	0,002*
	TPT	γ_{12}	0,186	0,854
	UMK	γ_{13}	0,607	0,547
IPM	*Intercept*	γ_{20}	1,109	0,267
	P_0	β_2	-1,896	0,058*
	AML	γ_{21}	0,686	0,492
	RK	γ_{22}	-0,652	0,514
	KP	γ_{23}	5,050	0,000*
	Rho	ρ_2	3,975	0,000*

*) significant at the level of 10%

Table 6 shows that the HDI and GRDP growth rate (LPDRB) have a value of more than $t_{(0.1;33)}$ of 1,692 so that the reject H_0 which means it has a significant effect on the percentage of the poor. The percentage of poor population (P_0), population density (KP) and spatial effects has a value of more than $t_{(0.1;32)}$ of 1,694 so that the rejection of H_0 which means a significant effect on HDI in East Java in 2019 at a significant level of 10%, so that re-modeling is carried out by issuing insignificant variables into the model. The results of the re-modeling estimates presented in Table 9 are obtained.

Table 9. Estimation of Parameters with Significant Variables

Variabel Dependen	Variabel	Parameter	Koefisien
P_0	*Intercept*	γ_{10}	55,4991
	IPM	β_1	-0,540
	LPDRB	γ_{11}	-1,270
	R-sq	76,46%	
IPM	*Intercept*	γ_{20}	31,75
	P_0	β_2	-0,448
	KP	γ_{23}	2,609
	Rho	ρ_2	0,369
	R-sq	86,43%	

Partial test hypothesis of equation P_0:

$H_0:\theta_{1j}=0$, j=1.2 (No influence of jth exogenous variable on the percentage of poor population)

$H_1:\theta_{1j}\neq 0$, j=1,2 (There is an influence of the jth exogenous variable on the percentage of the poor population)

Partial test hypothesis of the HDI equation:

$H_0:\theta_{2j}=0$, j=1,2,3 (No influence of jth exogenous variable on HDI)

$H_1:\theta_{2j}\neq 0$, j=1,2,3 (There is an influence of the jth exogenous variable on HDI)

Table 10. Partial Test of Model Parameters

Dependence Variable	Variable	Parameter	*t*-Statistics	P-*Value*
P_0	*Intercept*	γ_{10}	8,730	0.000*
	IPM	β_1	-5,272	0.000*
	LPDRB	γ_{11}	-3,646	0.001*
IPM	*Intercept*	γ_{20}	2,603	0.009*
	P_0	β_2	-2,960	0.003*
	KP	γ_{23}	5,174	0.000*
	Rho	ρ_2	2,628	0.008*

*) significant at the level of 10%

Table 10 shows that all variables have significantly entered into the model, so the model shown in equations 31 dan 32.

$$P_{0_i} = 55,499 - 0,540 IPM_i - 1,270 LPDRB_i + \varepsilon_1 \quad (31)$$

$$IPM_i = 31,75 - 0,448 P_{0_i} + 2,609 \log KP_i + 0,369 \sum_{i=1}^{n}\sum_{j=1}^{m} w_{ij} IPM_i + \varepsilon_2 \quad (32)$$

The results of parameter estimation using the GS2SLS method of the SAR model show that the Human Development Index (HDI) has a negative and significant effect on the percentage of the poor (P_0), and vice versa, namely the percentage of poor people (P_0) has a negative and significant effect on the Human Development Index (HDI). This shows that there is simultaneousity and negative interplay between P_0 and HDI. These results are in accordance with previous studies Sofilda, et al. which stated that HDI and Poverty affect each other significantly between the two and have a negative reciprocal relationship[10].

When compared to the strength of the relationship direction, the estimation results show that HDI stronger the affects percentage of poor people than the percentage of poor people affects HDI. This can be seen from the coefficient value for the percentage of poor people of -0.448 while the value of the coefficient for HDI is -0.540, meaning that in cateris paribus conditions the percentage of those living in poverty will decrease by 0.54 percent for every 1 percent increase in the HDI value. In the same way, a 1 percent drop in the percentage of the poor will result in a 0.448 percent gain in the HDI number, cateris paribus.

At a significant threshold of 0.1, the expansion of PDRB significantly affected the decline in the percentage of the poor population (P_0). The LPDRB coefficient is equal to -1.27, which implies that, cateris paribus, an increase in LPDRB of 1% will result in a reduction in the percentage of the population living in poverty of 1.27 percentage points. In order to comprehend the economic dynamics of a region by examining the acceleration of its economy, changes in the GRDP value of a region are evaluated using constant prices, which can demonstrate the rate of economic growth. This means that as economic progress accelerates, demand for goods and services rises along with it. As a result, as people's requirements for goods and services increase, poverty—which is always characterized by people's inability to meet basic necessities—is indirectly reduced.

According to cateris paribus assumptions, the percentage of the poor (P_0) is positively correlated with the open unemployment rate (TPT), hence the higher the TPT value, the greater the percentage of the poor (P_0). TPT does not significantly change the percentage of the population that is considered to be poor, indicating that either some jobless individuals do not have low incomes or that not all unemployed individuals do.

The Regency/City Minimum Wage (UMK) has a positive effect on the percentage of poor people (P_0) and is not significant, because 60.64% of the population of East Java in 2019 was the informal sector. It is common knowledge that the minimum wage only applies to the formal sector and is not imposed on the unofficial sector.

The proportion of homes with access to adequate drinking water sources (AML) has a positive relationship with HDI, which means that as the number of families with access to adequate drinking water sources rises, so does the HDI value. The proportion of households having access to adequate drinking water sources (AML) is not significant to HDI because, even with sufficient access to clean water, HDI cannot be significantly impacted by access to clean water in the absence of public knowledge of a healthy lifestyle.

The Dependency Ratio (RK) negatively affects the HDI, meaning that if the dependency ratio increases, it will reduce the HDI value. The dependency ratio is not significant to HDI because the population of East Java in 2019 at the productive age (15 years-64 years) can still afford to be able to bear the costs for the unproductive age (0 years-14 years or 65+ years).

At a substantial level of 0.1, population density (KP) has a favorable and significant impact on HDI. The value of the population density coefficient is 2,609, meaning that in cateris paribus, if the population density increases by 1%, it will increase HDI value by 2,609 per 100 percent. This shows that a dense population can be the basic capital of development supported by adequate facilities, it will have the potential to be in the development process.

The *rho* coefficient in the HDI equation is positively marked and significant at a significant level of 0.1 which means that there is an influence of spatial lag on adjacent areas that will increase the human development index (HDI) in a region by 0.369, thus the *rho* coefficient indicates that if an area is surrounded by *m* other regions, then the influence of each region that surrounds it can be measured by 0.369 divided by the number of regions that surround it.

The percentage of the poor (P_0) equation's coefficient of determination is 76.46 percent. This indicates that variations in the HDI and LPDRB variables can account for 76.46 percent of the variation in the percentage variable of the poor (P_0), while other variables that are not part of the poor population percentage model can influence or account for the remaining 23.54 percent of the variation (P_0). As a result, the coefficient of determination in the endogenous variable equation of HDI is 86.43 percent, indicating an 86.43 percent contribution from the variable influence of population density and the percentage of the impoverished population (P0) on HDI. The HDI model does not account for the remaining 13.57 percent, which is affected or explained by other factors.

3.5. Pemeriksaan Asumsi Residual

Examination of regression assumptions on the spatial simultaneous equation of data on the percentage of poor people and HDI in East Java based on the results of the examination of heterogeneity assumptions and normal distribution will be explained in the following points.

a. Examination of Heterogeneity Assumptions

The Glejser test is used to examine the assumption of heterogeneity of the residual equation of the percentage of the poor population and HDI with the hypothesis,

$H_0 : \sigma_1^2 = \sigma_2^2 = ... = \sigma_k^2 = \sigma^2$ (there is no heteroskedasticity)

H_1 : there is at least one $\sigma_j^2 \neq \sigma^2$ (there is a heteroskedasticity)

with a significant degree of 10% so that H_0 is rejected if $t > t_{(0,1,df)}$ or p-value < 0.1. The results of the Glejser test analysis are presented in Table 11.

Table 11. Examination of Heteroskedasticity Assumptions

Equation	*F*-Statistics	p-*Value*	Decision
(31)	1,18	0,319	H_0 is accepted
(32)	1,48	0,237	H_0 is accepted

Table 11 Shows that in general the GS2SLS SAR model with queen contiguity

ISSN No. 1978-3787
Open Journal Systems

weights shows that the p-value is more than 0.1 so that the residual of the equation of the percentage of the poor and HDI populations has met the assumption of homoskedasticity.

b. Normal Distribution Assumption Check
The *Jarque-Bera test is* used for the examination of the assumption of the normal distribution of residual equations of the percentage of the poor and HDI population with hypotheses,
H_0 = Normally distributed residual
H_1 = Residual non-normal distribution
with a significant degree of 10% so that H_0 is rejected if $JB > x^2_{(0.1.2)}$ or p-value < 0.1. The results of the Jarque-Bera test analysis are presented in Table 12.

Table 12. Normal Distribution Assumption Check

Equation	*JB*-Statistics	p-*Value*	Decision
(31)	0,140	0,932	H_0 is accepted
(32)	2,664	0,264	H_0 is accepted

Table 12 shows that in general the GS2SLS SAR model with queen contiguity weights shows that the p-value is more than 0.1 so that the residual of the equation of the percentage of poor people and HDI has been normally distributed.

PENUTUP

Conclusion

The conclusion obtained from the results of the analysis and discussion on the data on the percentage of poor people (P_0) and the Human Development Index (HDI) in East Java, namely in the variable percentage of poor people, the spatial effect is insignificant so that it is modeled with Two Stage Least Square (2SLS) and produces a coefficient of determination of 76,46%. Spatial dependence of lag is positive and significant only on HDI variables with GS2SLS Spatial Autoregressive (SAR) modeling which has a coefficient of determination of 86,43%. The modeling results using 2SLS in the equation of the percentage of poor people found that HDI has a significant effect on the percentage of poor people while GS2SLS SAR for HDI found that the percentage of poor people has a significant effect on HDI. This proves that there is a negative reciprocal relationship between the percentage of poor people and HDI with the spatial lag effect, with HDI stronger affecting the percentage of poor people than the percentage of poor people affecting HDI. While the significant LPDRB variable affects the percentage of poor people with a negative relationship direction, while the significant variable that affects HDI is population density.

Saran

Suggestions for future research can be developed by combining cross-sectional data and time series data whose parameter estimation uses a simultaneous information system model full of spatial effects, and should include other variables that can affect poverty and HDI.

DAFTAR PUSTAKA

[1]Badan Pusat Statistik of East Java. 2020. East Java Province in Numbers, BPS of East Java.
[2]Sustainable Development Goals 2021. What is SDGs, SDGs. https://www.sdg2030indonesia.org/page/8-apa-itu, accessed: December 28th, 2021.
[3]Kuncoro, M. 2000. Ekonomi Pembangunan: Teori, Masalah dan Kebijakan, UPP AMP YKPN.
[4]Saleh, S. 2002. Ekonomi Pembangunan Faktor-Faktor Penentu Tingkat Kemiskinan Regional Di Indonesia. Jurnal Ekonomi Pembangunan. Vol. 7, Issue 2.
[5]Pratama, Y. C. 2014. Analisis Faktor-Faktor Yang Mempengaruhi Kemiskinan Di Indonesia. Bisnis Dan Manajemen, 4(2).
[6]Widodo, A., Waridin, and Maria, J. 2011. Analisis Pengaruh Pengeluaran Pemerintah Di Sektor Pendidikan Dan Kesehatan Terhadap Pengentasan Kemiskinan Melalui

Peningkatan Pembangunan Manusia Di Provinsi Jawa Tengah, Dinamika Ekonomi Pembangunan, 1(1).

[7]Zuhdiyati, N., and Kaluge, D. 2017. "Analisis Faktor-Faktor Yang Mempengaruhi Kemiskinan Di Indonesia Selama Lima Tahun Terakhir (Case Studies in 33 Provinces).

[8]Rohmah, R. N. 2017. Analisis Faktor-Faktor Yang Mempengaruhi Indeks Pembangunan Manusia Di Provinsi Jawa Tengah Tahun 2012-2014.

[9] Trunajaya, I. P. 2015. Faktor-Faktor Yang Mempengaruhi Indeks Pembangunan Manusia Di Seluruh Provinsi Indonesia. Institut Pertanian Bogor.

[10]Sofilda, E., Hamzah, M. Z., and Sholeh, S. 2013. Human Development And Poverty In Papua Province (An Analysis Of Simultaneous Approach On Panel Data Regression), OIDA International Journal of Suistainable Development, 6(6).

[11]Gujarati, D. 2015. Econometrics by example, Palgrave. 2nd Ed.

[12]Anselin, L. 1988b. Spatial Econometrics: Methods and Models, Kluwer Academic, 1st Ed., Vol 4.

[13]Rosa, M., Maiyastri, M., and Yozza, H. 2021. Pemodelan Indeks Pembangunan Manusia Menggunakan Analisis Regresi Spasial Di Provinsi Jawa Timur, Jurnal Matematika UNAND, 9(4), 347. https://doi.org/10.25077/jmu.9.4.347-356.2020.

[14]Laswinia, V. D. 2016. Pattern Analysis of Relationship Analysis Between Low-Lived Population Percentage and Environmental , Economic , and Social Factors in Indonesia Using Spatial Regression, Institut Teknologi Sepuluh Nopember.

[15]Kelejian, H. H., and Prucha, I. R. 1998. A Generalized Spatial Two-Stage Least Squares Procedure for Estimating a Spatial Autoregressive Model with Autoregressive Disturbances, Journal of Real Estate Finance and Economics, Vol. 17.

[16]Kelejian, H. H., and Prucha, I. R. 2004. Estimation of simultaneous systems of spatially interrelated cross sectional equations, Journal of Econometrics, 118(1–2), 27–50. https://doi.org/10.1016/S0304-4076(03)00133-7.

[17]Kelejian, H. H., and Prucha, I. R. 1999. A Generalized Moments Estimator For The Autoregressive Parameter In A Spatial Model, International Economic Review, Vol. 40, Issue 2.

[18]Gujarati, D. N. 2004. Basic econometrics, McGraw Hill.

[19]Minister of Manpower Regulation No: Per-01/Men/1999

[20]Setiawan and Kusrini, D. E. 2010. Econometrika, ANDI.

[21]Hausman, J. A. 1978. Specification Tests in Econometrics, The Econometrics Society, 46(6), 1251–1271.

[22]Cliff, A., and Ord, J. 1981. Spatial Processes, Models and Applications, Pion.

[23]Anselin, L. 1988a. A Test For Spatial Autocorrelation In Seemingly Unrelated Regressions, In Economics Letters, Vol 28.

[24]Burridge, P. 1980. On the Cliff-Ord test for spatial correlation, Journal of the Royal Statistical Society, 42, 107–108.

[25]Anselin, L., Bera, A. K., Florax, R., and Yoon, M. J. 1996. Simple diagnostic tests for spatial dependence, Regional Science and Urban Economics, 26(1), 77–104. https://doi.org/10.1016/0166-0462(95)02111-6.

[26]Montgomery, D. C., Peck, E. A., and Vining, G. G. 2012. Introduction To Linear Regression Analysis, Wiley. 5th Ed.

[27]Greene, W. H. 2018. Econometric Analysis, Pearson.. 8th Ed.

[28]Tomsik, R. 2019. Power Comparisons of Shapiro-Wilk, Kolmogorov-Smirnov and Jarque-Bera Test, Scholars Scitech Research Organization, 3(3).

HALAMAN INI SENGAJA DIKOSONGKAN