Simultaneous Spatial Regression Modeling of Panel Data Using the Two Stage Estimation Method for Sustainable Development Indicators in East Java, 2019-2022

Tenti Amelya [1], Rahma Fitriani [1], Henny Pramoedyo [1]

[1]Department of Statistics, Faculty of Mathematics and Natural Sciences, Universitas Brawijaya, Indonesia 65145

Corresponding Author Email ID: tentiamelya@gmail.com

Article Info
Page Number: 2128 – 2140
Publication Issue:
Vol 72 No.1 (2023)

Abstract

The relationship of mutual dependence (simultaneity) between endogenous variables, namely the Human Development Index (HDI) and the Environmental Quality Index (EQI), cannot be modeled in a single equation, but there are two equations in a system of simultaneous equations. Each of these equations cannot be estimated separately without including information from other equations. The purpose of this study is to model panel data spatial regression simultaneously. The panel data used consists of 38 regencies/cities in East Java in 2019-2022. Parameter estimation was performed using the Two Stage Estimation method in simultaneous panel data spatial regression. Based on the results of the Chow test, Hausman test, and Lagrange multiplier test, it was found that the fixed effects model is more suitable for modeling HDI, while the random effects model is better for modeling EQI. Both models have a significant F statistic, indicating that they are able to explain the simultaneity between HDI and EQI. Variables that have a significant effect on HDI are GRDP, the percentage of health complaints, and the ratio of the number of student teachers. As for EQI, the only variable that has a significant effect is population density. Estimation of the structural model results that there is no spatial dependence between districts/cities on the HDI and EQI structural models. The F statistic value for the first equation is 37.53 with a significance level of 0.000, and the R-square value is 63.26%. For the second equation, the value of the F statistic is 6.08 with a significance level of 0.000, and the R-square value is 17.23%.

Article History
Article Received: *15 November 2022*
Revised: *24 December 2022*
Accepted: *18 January 2023*

Keywords: panel regression, spatial autoregressive, simultaneous model, sustainable development

Introduction

Regression analysis is a statistical method used to determine a causal relationship between one or several variables on other variables. In statistics, the commonly used regression model is the single equation model. However, sometimes there is interdependence or interdependence between variables in some models, which results in a two-way relationship. A model like this is called a simultaneous equation model [1]. In the simultaneous equation model, there is more than one equation that forms a system of equations. The distinctive feature of this model is that the dependent variable in one equation can become the independent variable in another equation. Therefore, the terms independent and dependent variables are not properly used in simultaneous equations. Instead, the terms endogenous variables (dependent variables) and predetermined variables (independent variables that are determined beforehand) are used.

An important assumption in classical linear regression is that there is no correlation between error and the independent variable. If there is such a correlation, the model parameter estimators will be inconsistent. This often occurs in modeling systems of simultaneous equations because the endogenous variables in one equation become independent variables in another equation, so that the possibility of correlation with errors is quite large (endogeneity). In modeling a system of simultaneous equations using Ordinary Least Squares (OLS), there are three problems that arise due to endogeneity, namely estimators become biased and inconsistent, hypothesis testing becomes invalid, and forecasts become biased and inconsistent.

Several previous studies have applied the simultaneous equation model in various contexts. For example, [1] applied a simultaneous equation parameter estimator using the Two Stage Least Square (2SLS) method to the equation of national income and money circulation in Indonesia in 1981-2008. [3] used the 2SLS method to analyze the relationship between poverty and GRDP in each province in Indonesia in 2010. [4] used panel data to model simultaneous regression with the 2SLS estimation technique to analyze the relationship between the price index received and paid by farmers in Indonesia in 2013-2015.

In regression analysis, cases are often encountered where the observed value at one location depends on another location, which is called spatial dependency. Therefore, a model that pays attention to the effect of this spatial dependency is needed, which is called a dependent spatial model. Several previous studies have modeled spatial dependencies using panel data. For example, [5] modeled the economic growth of districts/cities in North Sumatra Province in 2012-2014 using the autoregression spatial model (SAR) and the spatial error model (SEM) with the Fixed Effect Model (FEM) and Random Effect Model (REM) estimation techniques. Windi (2019) modeled the district/city human development index in Central Java Province in 2003-2017 using the autoregression spatial model (SAR) and the FEM and REM estimation techniques. However, there is no research that examines the parameter estimation of the spatial regression model of simultaneous panel data using the 2SLS technique.

In this research, modeling the reciprocal relationship between the Human Development Index (HDI) and the Regency/City Environmental Quality Index (EQI) in East Java Province in 2019-2022. Sustainable development has three dimensions, namely economic, social and environmental. HDI is used to measure economic and social development achievements, while EQI is used to measure environmental development achievements. Several previous studies have found a relationship between HDI and EQI. [6] found that HDI has a negative and significant effect on EQI in Indonesia. [7] found that population density and land transportation have a significant effect on EQI. [8] found that GRDP has a negative effect on the environmental quality of provinces in Indonesia. [9] found that the gini ratio, the percentage of poor people, and the ratio of the number of teacher students had a significant effect on HDI in East Java. [10] found that health complaints had a negative and significant effect on HDI in East Java. [11] found that GRDP had a positive and significant effect on HDI in the Districts/Cities of Aceh Province. However, there is no research that examines the influence of EQI on HDI.

economic, social and environmental patterns need to be considered from a spatial perspective between regions. Spatial dependencies arise because activities in one region affect resources in other adjacent areas. In this study, panel data will be used to model the relationship between HDI and EQI in the districts/cities of East Java Province. Panel data combines cross section and time series data, thus enabling heterogeneity control and comparison Unbalanced of individual conditions in different time periods. Parameter estimation using the two-stage estimation method will be used in this simultaneous panel data spatial regression model. This research is expected to provide further understanding of the simultaneous relationship between economic, social and environmental development in the context of sustainable development.

Research Method

Data Sources and Research Variables

The data source used in this research is secondary data from the publications of the Central Bureau of Statistics and the East Java Province Environmental Service. The type of data used is panel data. The object of research is 38 districts/cities in East Java Province which were observed in 2019-2022. The variables used in this study are divided into endogenous and exogenous variables. The endogenous variables used in this study are the human development index and the environmental quality index. While the exogenous variables in this study are gross regional domestic product, health complaints, teacher-pupil ratio, population density, and access to proper drinking water.

Model Specifications

The formulation of the variables used in this study refers to the concept of sustainable development which consists of three dimensions namely, economic, social and environmental. Furthermore, the selection of exogenous variables for each endogenous variable is based on previous research literature in economic, social and environmental studies. The model formulation in this study with 2 endogenous variables and 7 exogenous variables with the model specifications as follows.

$$ln\,HDI_{i,t} = \alpha_0 + \lambda_1 \sum_{j=1}^{N} w_{ij}\, ln\,EQI_{j,t} + \alpha_1\, ln\,EQI_{i,t} + \alpha_2\, ln\,GRDP_{i,t} + \alpha_3\, ln\,HC_{i,t} + \alpha_4\, ln\,STR_{i,t} + e1_{i,t} \quad (1)$$

$$ln\,EQI_{i,t} = \beta_0 + \lambda_2 \sum_{j=1}^{N} w_{ij}\, ln\,HDI_{j,t} + \beta_1\, ln\,HDI_{i,t} + \beta_2\, ln\,GRDP_{i,t} + \beta_3\, ln\,PD_{i,t} + \beta_4\, ln\,ACDW_{i,t} + e2_{i,t} \quad (2)$$

The form of each endogenous variable as a function of all exogenous variables is as follows.

$$\begin{aligned} ln\,HDI_{i,t} =& \\ &(\pi_{10} + \beta_0\pi_{12}) + (\pi_{11} + \beta_1\pi_{12}) \sum_{j=1}^{N} w_{ij}\, ln\,HDI_{j,t} + (\beta_2\pi_{12}) \sum_{j=1}^{N} w_{ij}\, ln\,GRDP_{j,t} + \\ &(\beta_3\pi_{12}) \sum_{j=1}^{N} w_{ij}\ \ (\beta_4\pi_{12}) \sum_{j=1}^{N} w_{ij}\, ln\,ACDW_{j,t} + (\pi_{13})\, ln\,GRDP_{i,t} + (\pi_{14})\, ln\,HC_{i,t} + (\pi_{15})\, ln\,STR_{i,t} + \\ &(\pi_{16})\, ln\,PD_{i,t} + (\pi_{17})\, ln\,ACDW_{i,t} + \left(\pi_{12} \sum_{j=1}^{N} w_{ij}\, e2_{j,t} + v1_{i,t}\right) \qquad (3) \end{aligned}$$

$ln\, EQI_{i,t} =$
$(\pi_{20} + \alpha_0\pi_{21}) + (\alpha_1\pi_{21} + \pi_{22})\sum_{j=1}^{N} w_{ij}\, ln\, EQI_{j,t} + (\alpha_2\pi_{21})\sum_{j=1}^{N} w_{ij}\, ln\, GRDP_{j,t} +$
$(\alpha_3\pi_{21})\sum_{j=1}^{N} w_{ij}\, ln\, HC_{j,t} + (\alpha_4\pi_{21})\sum_{j=1}^{N} w_{ij}\, ln\, STR_{j,t} + (\pi_{23})\, ln\, GRDP_{i,t} + (\pi_{24})\, ln\, HC_{i,t} + (\pi_{25})\, ln\, STR_{i,t} +$
$(\pi_{26})\, ln\, PD_{i,t} + (\pi_{27})\, ln\, ACDW_{i,t} + \left(\pi_{21}\sum_{j=1}^{N} w_{ij}\, e1_{j,t} + v2_{i,t}\right)$ (4)

From the equation above, it can be identified that the endogenous variable as a function of all exogenous variables is the SARAR model with Spatial lag of X (SLX).

Spatial Weighting Matrix Specifications

Spatial regression modeling to model the human development index and environmental quality index in East Java uses a spatial queen contiguity weighting matrix. The selection of the queen contiguity matrix is based on the consideration of the asymmetrical shape of the districts/cities in East Java. The calculation of the queen contiguity matrix is carried out based on the administrative boundaries that connect neighboring regencies/cities. In the queen contiguity matrix, each element has a value of 1 or 0 indicating the presence or absence of boundaries between two spatial units, such as districts or cities. However, in the spatial queen contiguity weighting matrix, the weights for each element are calculated by taking into account the number of regencies or cities bordering certain regencies or cities, divided by the total number of regencies or cities in the region. The use of the queen contiguity spatial weighting matrix makes it possible to identify the spatial relationship between districts/cities in East Java in the spatial simultaneous equation between HDI and EQI.

Data Analysis Method

To achieve the research objectives using a simultaneous panel data spatial regression model, this research was conducted through structured stages. The first stage is to identify the research problem and determine the appropriate model specifications. The second stage involves the identification of a model which consists of three possibilities, namely the underidentified or exactly identified model which requires the addition of exogenous variables, as well as the overidentified model which can be directly continued to the next step. Furthermore, the third stage involves the formation of a contiguity matrix (C) using the Queen Contiguity criteria and a standardized spatial weighting matrix (W) also using the Queen Contiguity criteria. The fourth stage is to estimate the first stage of two-stage parameters in equations (3) and (4) with the SARAR concept with spatial lag of X. The fifth stage involves estimating the second stage of two-stage parameters in equations (1) and (2) with the panel regression concept with spatial lag of X, using the endogenous variable estimators obtained in the previous step. Finally, the sixth stage is to compare the CEM, FEM, and REM models to determine which model is most appropriate in explaining the simultaneity of panel data.

Panel Data SARAR Models

The panel spatial regression model is a panel regression model that involves spatial effects. According to [12] the linear regression model on panel data where there is interaction between the spatial units, it will have a spatial lag variable in the response variable or a process spatial variable in the error. The Spatial Autoregressive Fixed Effect and Random Effect models can each be expressed as equations (5) and (6) below.

$\boldsymbol{Y} = \lambda(\boldsymbol{I_T} \otimes \boldsymbol{W}_N)\boldsymbol{Y} + \boldsymbol{X\beta} + (\boldsymbol{i}_T \otimes \boldsymbol{I}_N)\boldsymbol{\alpha_i} + \boldsymbol{u}$ (5)

$\boldsymbol{Y} = \lambda(\boldsymbol{I_T} \otimes \boldsymbol{W}_N)\boldsymbol{Y} + \boldsymbol{X\beta} + (\boldsymbol{i}_T \otimes \boldsymbol{I}_N)\boldsymbol{\alpha} + \boldsymbol{u}$ (6)

$\boldsymbol{u} = \rho(\boldsymbol{I_T} \otimes \boldsymbol{W}_N)\boldsymbol{u} + \boldsymbol{\varepsilon}$

with

λ : Spatial coefficient autoregressive

ρ : Coefficient of spatial error

$\boldsymbol{W}_N$: Standardized spatial weighting matrix with size $N \times N$

$\boldsymbol{Y}$: The dependent variable matrix with size $NT \times 1$

$\boldsymbol{X}$: The independent variable matrix with size $NT \times K$

$\boldsymbol{\beta}$: Slope coefficient vector with size $K \times 1$

$\boldsymbol{i}_T$: A vector of size $T \times 1$ where each element contains 1

$\boldsymbol{I}_N$: The identity matrix with size $N \times N$

$\boldsymbol{\alpha_i}$: Individual effect vectors are constant in time and differ per unit

$\boldsymbol{\alpha}$: The general intercept vector of the model for cross-sectional units and time

$\boldsymbol{\varepsilon}$: Vector error with size $NT \times 1$

$\boldsymbol{u}$: The disturbance vector with size $NT \times 1$

Simultaneous Equation Models

According to [13], simultaneous equations are a system of equations that are interconnected in one set. Between variables have a two-way relationship so that there is more than one equation of the endogenous variable or dependent variable. Each simultaneous equation is composed by three variables, namely endogenous variables, predetermine variables, and error variables. In [14] general form a system of simultaneous equations with as many Gendogenous variables as $Y_1, Y_2, \dots, Y_G$ and as many as Kexogenous variables that is $X_1, X_2, \dots, X_K$ can be written as follows.

$$\begin{bmatrix} \alpha_{11} & \alpha_{12} & \cdots & \alpha_{1G} \\ \alpha_{21} & \alpha_{22} & \dots & \alpha_{2G} \\ \vdots & \vdots & \ddots & \vdots \\ \alpha_{G1} & \alpha_{G2} & \dots & \alpha_{GG} \end{bmatrix} \begin{bmatrix} Y_{1it} \\ Y_{2it} \\ \vdots \\ Y_{Git} \end{bmatrix} + \begin{bmatrix} \beta_{11} & \beta_{12} & \cdots & \beta_{1K} \\ \beta_{21} & \beta_{22} & \dots & \alpha_{2K} \\ \vdots & \vdots & \ddots & \vdots \\ \beta_{G1} & \beta_{G2} & \dots & \beta_{GK} \end{bmatrix} \begin{bmatrix} X_{1it} \\ X_{1it} \\ \vdots \\ X_{Kit} \end{bmatrix} = \begin{bmatrix} \varepsilon_{1it} \\ \varepsilon_{2it} \\ \vdots \\ \varepsilon_{Git} \end{bmatrix} \quad (7)$$

or

$\boldsymbol{\Gamma Y}_{it} + \boldsymbol{BX}_{it} = \boldsymbol{E}_{it}$(8)

with

$\boldsymbol{\Gamma}$: Coefficient matrix of endogenous variable parameter with size $G \times G$

$\boldsymbol{B}$: Coefficient matrix of exogenous parameters with size $G \times K$

$\boldsymbol{Y}_{it}$: Vector endogenous variable with size $G \times 1$

$\boldsymbol{X}_{it}$: exogenous vector with size $K \times 1$

$\boldsymbol{E}_{it}$: The measuring vector $G \times 1$ of the structural disturbance

Y_{jit} : The measuring vector $NT \times 1$ of the dependent variable with the $j, j = 1, \dots, G$

X_{kit} : The measuring vector of observations $NT \times 1$ on exogenous variables, fork $= 1,2, \dots, K$

ε_{jit} : The size of the disturbance vector $NT \times 1$ in the th equation j

Simultaneous Panel Data Spatial Regression Models

The single spatial equation introduced by Cliff and Ord in 1970 can be expanded into a simultaneous equation for more than one equation of the dependent variable that is correlated with each other. Based on [15] the spatial lag or spatial autoregressive simultaneous equation model for a number of Gequations, it can be described as follows.

$$\boldsymbol{Y_{it}} = \boldsymbol{\Gamma Y_{it}} + \boldsymbol{BX_{it}} + \boldsymbol{\Lambda}(\boldsymbol{I_T} \otimes \boldsymbol{W}_N)\boldsymbol{Y_{it}} + \boldsymbol{E_{it}} \quad (9)$$

where

$\boldsymbol{Y_{it}} = (Y_{1it}, \dots, Y_{Git})$

$\boldsymbol{X_{it}} = (X_{1it}, \dots, X_{Kit})$

$\boldsymbol{E_{it}} = (\varepsilon_{1it}, \dots, \varepsilon_{Git})$

$\boldsymbol{\Lambda} = diag_{j=1}^{G}(\rho_j)$

with

$\boldsymbol{\Lambda}$: Lag-sized spatial parameter matrix $G \times G$

Result and Discussion

Identification of Simultaneous Equation Models

Model identification is needed to determine the estimation method to be used. Identification of the simultaneous equation model can be done by identifying the order condition and rank condition.

Table 1Identification of the Order Condition of the Equation System

Equality	K-k	g-1	Status
(1)	$7-4$	$2-1$	Overidentified
(2)	$7-4$	$2-1$	Overidentified

Information:

g : The number of endogenous variables in a particular equation

K : The total number of predetermined variables in the model

k : The number of predetermined variables in a particular equation

Table 1 shows the results of identifying the order condition of the simultaneous equation system. All equations in the model are categorized as overidentified equations so that parameter estimation can be performed using the 2SLS method.

Table 2Identification of Equation System Rank Condition

Equality	Y_1	Y_2	X_1	X_2	X_3	X_4	X_5	X_6	X_7
(1)	1	β_1	ρ_1	β_2	β_3	β_4	0	0	0
(2)	γ_1	1	0	γ_2	0	0	ρ_2	γ_3	γ_4

Description:

Y_1 : $ln\, HDI_{i,t}$

Y_2 : $ln\, EQI_{i,t}$

X_1 : $\sum_{j=1}^{n} w_{ij}\, ln\, EQI_{j,t}$

X_2 : $ln\, GRDP_{i,t}$

X_3 : $ln\, HC_{i,t}$

X_4 : $ln\, STR_{i,t}$

X_5 : $\sum_{j=1}^{n} w_{ij}\, ln\, HDI_{j,t}$

X_6 : $ln\, PD_{i,t}$

X_7 : $ln\, ACDW_{i,t}$

Based on Table 2, it is concluded that the simultaneous equation model fulfills the rank condition because in each structural equation in the simultaneous model one non-zero determinant can be formed from the variable coefficient which is not contained in the equation, but is contained in other equations in the simultaneous equation model.

The 2SLS method includes two consecutive OLS applications. The 2SLS method applied in this study used the SARAR Panel SLX Model in the estimation of stage 1 and the Panel SLX Model in the estimation of stage 2. The estimation of stage 1 was carried out with the maximum likelihood estimate (MLE) and the estimation of stage 2 was carried out with the OLS for the Common Effect model, the estimation in for Fixed Effect models, and GLS for Random Effect Models. Therefore, the 2SLS naming method was not appropriate, so it was changed to a two-stage estimation method.

Estimation of Reduced Model **Parameters**

Estimation of reduced model parameters using spatial regression panel data with Spatial Lag of X (SLX), there are three approaches, namely Common Effect Spatial Autoregressive with Autoregressive Disturbances (SARAR-CE), Fixed Effect Spatial Autoregressive with Autoregressive Disturbances (SARAR-FE), and Random Effect Spatial Autoregressive with

Autoregressive Disturbances (SARAR-RE). The reduced model using the SARAR panel with Spatial Lag of X using the Queen Contiguity Matrix is as follows.

SARAR Common Effects:

$$\ln HDI_{i,t} = -1{,}596 + 0{,}107\ \sum_{j=1}^{N} w_{ij} \ln HDI_{j,t} + 0{,}006\ \sum_{j=1}^{N} w_{ij} \ln GRDP_{j,t} + 0{,}013\ \sum_{j=1}^{N} w_{ij} \ln PD_{j,t} + 0{,}710\ \sum_{j=1}^{N} w_{ij} \ln ACDW_{j,t} + 0{,}005\ \ln GRDP_{i,t} - 0{,}018 \ln HC_{it} - 0{,}001\ \ln STR_{i,t} + 0{,}025\ \ln PD_{i,t} + 0{,}417 \ln ACDW_{i,t} - 0{,}589\ \sum_{j=1}^{N} w_{ij} \quad (10)$$

$$\ln EQI_{i,t} = 1{,}043 + 0{,}436\ \sum_{j=1}^{N} w_{ij} \ln EQI_{j,t} + 0{,}002\ \sum_{j=1}^{N} w_{ij} \ln GRDP_{j,t} + 0{,}089\ \sum_{j=1}^{N} w_{ij} \ln HC_{j,t} + 0{,}436\ \sum_{j=1}^{N} w_{ij} \ln STR_{j,t} + 0{,}009 \ln GRDP_{i,t} + 0{,}052\ \ln HC_{i,t} + 0{,}015\ \ln STR_{i,t} - 0{,}053\ \ln PD_{i,t} - 0{,}037\ \ln ACDW_{i,t} - 0{,}346\ \sum_{j=1}^{N} w_{ij}\, u2_{j,t} \quad (11)$$

SARAR Fixed Effect:

$$\ln HDI_{i,t} = \hat{\pi}_{1i} + 0{,}907\ \sum_{j=1}^{N} w_{ij} \ln HDI_{j,t} + 0{,}0008\ \sum_{j=1}^{N} w_{ij} \ln GRDP_{j,t} + 0{,}0004\ \sum_{j=1}^{N} w_{ij} \ln PD_{j,t} + 0{,}032\ \sum_{j=1}^{N} w_{ij} \ln ACDW_{j,t} + 0{,}016\ \ln GRDP_{i,t} - 0{,}0003\ \ln HC_{i,t} - 0{,}007\ \ln STR_{i,t} + 0{,}00004\ \ln PD_{i,t} + 0{,}005 \ln ACDW_{i,t} - 0{,}661\ \sum_{j=1}^{N} w_{ij}\, u1_{j,t} \quad (12)$$

$$\ln EQI_{i,t} = \hat{\pi}_{2i} + 0{,}557\ \sum_{j=1}^{N} w_{ij} \ln EQI_{j,t} + 0{,}022\ \sum_{j=1}^{N} w_{ij} \ln GRDP_{j,t} + 0{,}136\ \sum_{j=1}^{N} w_{ij} \ln HC_{j,t} + 0{,}443\ \sum_{j=1}^{N} w_{ij} \ln STR_{j,t} - 0{,}179 \ln GRDP_{i,t} + 0{,}081\ \ln HC_{i,t} + 0{,}144\ \ln STR_{i,t} - 0{,}018\ \ln PD_{i,t} + 0{,}195 \ln ACDW_{i,t} - 0{,}568\ \sum_{j=1}^{N} w_{ij}\, u2_{j,t} \quad (13)$$

With $\hat{\pi}_{1i}$ and $\hat{\pi}_{2i}$ which are different for each district/city.

SARAR Random Effect:

$$\ln HDI_{i,t} = 6{,}665 - 0{,}611\ \sum_{j=1}^{N} w_{ij} \ln HDI_{j,t} + 0{,}001\ \sum_{j=1}^{N} w_{ij} \ln GRDP_{j,t} - 0{,}0002\ \sum_{j=1}^{N} w_{ij} \ln PD_{j,t} + 0{,}027\ \sum_{j=1}^{N} w_{ij} \ln ACDW_{j,t} + 0{,}005\ \ln GRDP_{i,t} - 0{,}00006\ \ln HC_{i,t} - 0{,}007 \ln STR_{i,t} + 0{,}0002\ \ln PD_{i,t} + 0{,}012 \ln ACDW_{i,t} + 0{,}967\ \sum_{j=1}^{N} w_{ij}\, u1_{j,t} \quad (14)$$

$$\ln EQI_{i,t} = 0{,}876 + 0{,}468\ \sum_{j=1}^{N} w_{ij} \ln EQI_{j,t} + 0{,}008\ \sum_{j=1}^{N} w_{ij} \ln GRDP_{j,t} + 0{,}103\ \sum_{j=1}^{N} w_{ij} \ln HC_{j,t} + 0{,}406\ \sum_{j=1}^{N} w_{ij} \ln STR_{j,t} + 0{,}006 \ln GRDP_{i,t} + 0{,}068\ \ln HC_{i,t} + 0{,}034 \ln STR_{i,t} - 0{,}044\ \ln PD_{i,t} - 0{,}067\ \ln ACDW_{i,t} - 0{,}420\ \sum_{j=1}^{N} w_{ij}\, u2_{j,t} \quad (15)$$

Structural Model Parameter Estimation

Estimation of structural model parameters using panel data regression with Spatial Lag of X (SLX) there are three approaches, namely Common Effect Model (CEM), Fixed Effect Model (FEM), and Random Effect Model (REM). Estimation of the parameters of the HDI and EQI structural models is carried out by substituting the appropriate HDI and EQI variable estimates based on equation (9) to equation (15) to the original HDI and EQI variables which are on the right side of each structural equation. The spatial regression model for simultaneous panel data on the HDI and EQI equations is as follows.

Common Effects Models:

$$\ln HDI_{i,t} = 5{,}043 - 0{,}357\ \sum_{j=1}^{N} w_{ij} \ln EQI_{j,t} - 0{,}135\ \ln EQI_{i,t} + 0{,}006 \ln GRDP_{i,t} - 0{,}012\ \ln HC_{i,t} + 0{,}134\ \ln STR_{i,t} \quad (16)$$

$$\ln EQI_{i,t} = 8{,}427 - 0{,}838 \sum_{j=1}^{N} w_{ij} \ln HDI_{j,t} - 0{,}279 \ln HDI_{i,t} + 0{,}007 \ln GRDP_{i,t} - 0{,}059 \ln PD_{i,t} + 0{,}076 \ln ACDW_{i,t} \quad (17)$$

Fixed Effects Models:

$$\ln HDI_{i,t} = \hat{\alpha}_{0i} - 0{,}014 \sum_{j=1}^{N} w_{ij} \ln EQI_{j,t} - 0{,}0002 \ln EQI_{i,t} + 0{,}141 \ln GRDP_{i,t} - 0{,}004 \ln HC_{i,t} - 0{,}056 \ln STR_{i,t} \quad (18)$$

$$\ln EQI_{i,t} = \hat{\beta}_{0i} - 0{,}436 \sum_{j=1}^{N} w_{ij} \ln HDI_{j,t} - 0{,}041 \ln HDI_{i,t} - 0{,}342 \ln GRDP_{i,t} - 0{,}024 \ln PD_{i,t} + 0{,}309 \ln ACDW_{i,t} \quad (19)$$

Random Effect Models:

$$\ln HDI_{i,t} = 3{,}946 - 0{,}063 \sum_{j=1}^{N} w_{ij} \ln EQI_{j,t} - 0{,}022 \ln EQI_{i,t} + 0{,}070 \ln GRDP_{i,t} - 0{,}005 \ln HC_{i,t} + 0{,}064 \ln STR_{i,t} + \hat{\mu}_{1i} \quad (20)$$

$$\ln EQI_{i,t} = -9{,}636 - 1{,}414 \sum_{j=1}^{N} w_{ij} \ln HDI_{j,t} + 3{,}458 \ln HDI_{i,t} + 0{,}007 \ln GRDP_{i,t} - 0{,}056 \ln PD_{i,t} - 0{,}016 \ln ACDW_{i,t} + \hat{\mu}_{2i} \quad (21)$$

Model Selection

After conducting simultaneous panel data spatial regression analysis for the East Java region on the endogenous HDI and EQI variables with the Common Effect, Fixed Effect, and Random Effect approaches, the best model that is feasible to use to model HDI and EQI in East Java in 2019-2022 is selected.

Table 3 Results of the Chow Test, Hausman Test, and Lagrange Multiplier Test

Model	**Chow test**		**Hausman test**		**Lagrange Multiplier Test**	
	Stats. Test	**p-values**	**Stats. Test**	**p-values**	**Stats. Test**	**p-values**
HDI	695.11	0.000	72.82	0.000	-	-
EQI	1.409	0.088	-	-	0.859	0.354

Based on Table 3, with such a small p-value on the Chow Test for the HDI model provides evidence that in this problem the Fixed Effect model is better than the Common Effect model for modeling HDI in East Java. Furthermore, the Hausman test was carried out to find out whether the Fixed Effect model was better than the Random Effect model. The Hausman test also produces a very small p-value so that it can be concluded that the Fixed Effect model is better than the Random Effect model in modeling the proximity and characteristics of each region for HDI in East Java. While in the EQI model, the Chow test produces a p-value greater than 0.05 so that it can be concluded that the Common Effect model is better than the Fixed Effect model for modeling EQI in East Java. Then, a Lagrange Multiplier Test was performed to compare the Common Effect and

Random Effect models. This test results that the Common effect model is better than the Random Effect model for modeling EQI in East Java.

The selected HDI and EQI models are represented by equations (18) and (21), respectively. The complete results of estimating the selected model parameters can be seen in the following table.

Table 4 Parameter Estimation Results of Fixed Effect Model HDI

Variable	Coefficient	p-values	F	R-Square
Intercepts	β_{0i}		37,534	63.26%
W ln EQI	-0.014	0.356	(0.000)	
ln EQI	-0.0002	0.981		
ln GRDP	0.141	0.000		
ln HC	-0.004	0.017		
ln STR	-0.056	0.000		

Based on Table 4, the effect of the spatial *lag variables* EQI and EQI on HDI in East Java is not statistically significant with respective coefficients of -0.014 (p -*value* = 0.356) and -0.0002 (*p-value* = 0.981). The insignificant effect of spatial *lag* EQI indicates that the HDI in a district/city is not affected by the EQI of surrounding districts/cities. However, there is a significant effect of the GRDP variable with a positive coefficient of 0.141 (*p-value* = 0.000), indicating that for every 1 percent increase in GRDP, the HDI will increase by 0.141 percent. The health complaints variable also has a significant negative effect with a coefficient of -0.004 (*p-value* = 0.017), indicating that for every 1 percent increase in health complaints, the HDI will decrease by 0.004 percent. In addition, the student teacher ratio variable has a significant negative effect with a coefficient of -0.056 (*p-value* = 0.000), indicating that for every 1 percent increase in the teacher student ratio, the HDI will decrease by 0.056 percent. The R- *square* of this model reaches 63.26% , which indicates that the independent variables in the model are able to explain 63.26% of the variation in HDI East Java.

In the selected HDI model, the intercept has a different value for each district/city. The differences are displayed by the HDI fixed effect components presented in Table 5.

Table 5 Intercept Fixed Effect Model HDI

Region	intercept (β_{0i})	Region	intercept (β_{0i})
Bangkalan Regency	2.992	Pasuruan Regency	2.806
Banyuwangi Regency	2.93	Ponorogo Regency	3.111

Blitar Regency	3.04	Probolinggo Regency	2.976
Bojonegoro Regency	2.879	Sampang Regency	2.984
Bondowoso Regency	3.051	Sidoarjo regency	2.93
Gresik Regency	2.915	Situbondo Regency	3.082
Jember Regency	2.884	Sumenep Regency	2.968
Jombang Regency	3.054	Trenggalek Regency	3.118
Kediri Regency	3.046	Tuban Regency	2.921
Lamongan Regency	3.041	Tulungagung Regency	3.059
Lumajang Regency	2.981	Batu City	3.219
Madiun Regency	3.14	Blitar City	3.376
Magetan Regency	3.167	Kediri City	2.971
Malang Regency	2.896	Madiun City	3.289
Mojokerto Regency	2.961	Malang city	3.074
Nganjuk Regency	3.097	Mojokerto City	3.363
Ngawi Regency	3.123	Pasuruan City	3.304
Pacitan Regency	3.106	Probolinggo City	3.234
Pamekasan Regency	3.065	Surabaya City	2.797

Table 6 Parameters Estimation Results of Common Effect Model EQI

Variable	**Coefficient**	**p-values**	**F**	**R-Square**
Intercepts	8,427	0.001	6,077	17.23%
W ln HDI	-0.838	0.148	(0.000)	
ln HDI	-0.279	0.266		
ln GRDP	0.007	0.518		
ln KP	-0.059	0.000		
ln AAML	0.077	0.673		

Based on Table 6, the spatial *lag* variables HDI, HDI, GRDP, and access to adequate drinking water do not have a significant effect on EQI in East Java, with their respective coefficients -0.838 (p-value = 0.1 48), -0.279 (p -value = 0 .266), 0.007(p-value = 0.518), and 0.077 (p-value = 0.673). The effect of the spatial *lag* of the HDI which is not significant indicates that the EQI in a district/city is not affected by the HDI of the surrounding districts/cities. However, the density variable shows a statistically significant effect with a coefficient of -0.05 9 (*p-value* = 0.000), indicating that for every 1 percent increase in population density, the EQI will decrease by 0.059 percent. The R- *square* of this model reaches 17.23 % , which indicates that the independent variables in the model are able to explain 17.23 % of the variation in East Java's EQI.

Conclusion

Based on the analysis and discussion conducted, several conclusions can be drawn. First, the estimation of the simultaneous panel spatial regression model to model the relationship between the Human Development Index (HDI) and the Environmental Quality Index (EQI) in East Java shows that the fixed effects model is more suitable for modeling HDI, while the common effect model is better for modeling EQI. Both models have significant F statistical values, indicating that the two models are suitable for expressing the simultaneity between HDI and EQI. Second, the variables that have a significant effect on HDI are GRDP, the percentage of health complaints, and the number of student-teacher ratios, while only the population density variable has a significant effect on EQI. Third, each district/city has specific characteristics in the HDI equation. Finally, there is no spatial dependence between districts/cities in the HDI and EQI structural models, which means that the HDI in a district/city is not affected by EQI in the surrounding districts/cities, and vice versa.

Reference

[1] Gujarati, J. (2012). A comprehensive induction system: A key to the retention of highly qualified teachers. *The Educational Forum*, *76*(2), 218–223.

[2] Wulandari, S. (2010). Penerapan metode 2SLS (two stage least square) pada model persamaan simultan untuk persamaan pendapatan nasional dan peredaran uang. (Thesis, Universitas Sebelas Maret).

[3] Bekti, R. D., David, D., Gita, N., Priscillia, P., & Serlyana, S. (2014). Model Persamaan Simultan pada Analisis Hubungan Kemiskian dan PDRB. *ComTech: Computer, Mathematics and Engineering Applications*, *5*(2), 810–817.

[4] Falah, B. Z., Mustafid, M., & Sudarno, S. (2016). Model Regresi Data Panel Simultan dengan Variabel Indeks Harga yang Diterima dan yang Dibayar Petani. *Jurnal Gaussian*, *5*(4), 611–621.

[5] Purba, O. N. (2016). Modeling *of Economic Growth North Sumatera with Spatial Econometrics Panel Data*. (Doctoral dissertation, Institut Teknologi Sepuluh Nopember).

[6] Ramadhantie, S. S., Ramadhan, M. J., & Hasibuan, M. A. (2021). Pengaruh Indeks Pembangunan Manusia Terhadap Indeks Kualitas Lingkungan Hidup di Indonesia Menggunakan Regresi Data Panel. *Ekologia*, *21*(1), 35–43.

[7] Hidayati, A. Z. (2022). Analisis Faktor-Faktor Yang Mempengaruhi Indeks Kualitas Lingkungan Hidup (IKLH) Di Indonesia Tahun 2017-2019. *Jurnal Medika Hutama*, *3*(02 Januari), 2327–2340.

[8] Pranoto, A. W. (2016). Pengaruh Desentralisasi Fiskal, PDRB, Kepadatan Penduduk, dan Penanaman Modal Asing terhadap Kualitas Lingkungan Hidup (Studi pada Provinsi di Indonesia Periode 2009-2013. http://etd.repository.ugm.ac.id/

[9] Utami, E. R. (2017). Penerapan Generalized Spatial Three Stage Least Square (GS3SLS) Pada Persamaan Simultan Spasial Untuk Pemodelan Pertumbuhan Ekonomi Jawa Timur (Doctoral dissertation, Institut Teknologi Sepuluh Nopember).

[10] Novitasari, N. I., Suharno, S., & Arintoko, A. (2021). Pengaruh Keluhan Kesehatan terhadap Indeks Pembangunan Manusia Di Provinsi Jawa Timur. *Jurnal Ilmiah Universitas Batanghari Jambi*, *21*(1), 239–244.

[11] Muliza, M., Zulham, T., & Seftarita, C. (2017). Analisis pengaruh belanja pendidikan, belanja kesehatan, tingkat kemiskinan dan PDRB terhadap IPM di provinsi Aceh. *JPED (Jurnal Perspektif Ekonomi Darussalam)(Darussalam Journal of Economic Perspectives)*, *3*(1), 51–69.

[12] Elhorst, J. P. (2003). Specification and estimation of spatial panel data models. *International Regional Science Review*, *26*(3), 244–268.

[13] Gujarati, D. N. (2004). Basic *Econometrics*. Perseus Books.

[14] Seddighi, H., Lawler, K. A., Lawler, K., & Katos, A. v. (2000). *Econometrics: A practical approach*. Psychology Press.

[15] Kelejian, H. H., & Prucha, I. R. (2004). Estimation of simultaneous systems of spatially interrelated cross sectional equations. Journal *of Econometrics*, 118(1–2), 27–50.