International Journal of Mathematical, Engineering and Management Sciences
Vol. 9, No. 3, 537-556, 2024
https://doi.org/10.33889/IJMEMS.2024.9.3.028

Ram Arti Publishers

# A Spatial Error Model in Structural Equation for the Human Development Index Modeling

**Anik Anekawati**
Faculty of Teacher Training and Education,
Universitas Wiraraja, Sumenep, Indonesia.
*Corresponding author*: anik@wiraraja.ac.id

**Purhadi**
Department of Statistics,
Institut Teknologi Sepuluh Nopember, Surabaya, Indonesia.
E-mail: purhadi@statistika.its.ac.id

**Mohammad Rofik**
Faculty of Economics and Business,
Universitas Wiraraja, Sumenep, Indonesia.
E-mail: rofik@wiraraja.ac.id

**Syaifurrahman Hidayat**
Faculty of Health Science,
Universitas Wiraraja, Sumenep, Indonesia.
E-mail: dayat.fik@wiraraja.ac.id

(Received on September 23, 2023; Revised on December 4, 2024 & January 24, 2024: Accepted on March 9, 2024)

**Abstract**
Spatial regression, particularly the Spatial Error Model (SERM), was utilized in prior studies to analyze Human Development Index (HDI) modeling. However, the studies were unable to determine which dimension among the three defined by the UN and BPS had the significant impact on HDI, as they constructed models based on the indicators used for the interpretation of the dimensions. Therefore, a comprehensive analysis combining spatial regression and Structural Equation Modeling (SEM), known as spatial SEM, was deemed necessary. This is the reason the current study aimed to develop SERM-SEM modeling holistically. The model parameters were estimated using the Generalized Method of Moments (GMM). To assess spatial dependency, the Lagrange Multiplier (LM) method was employed, with a distinct model error distribution compared to the error distribution of the traditional spatial model. The result of the LM test development showed that, under the null hypothesis, the LM test statistics followed a distribution. The results of the SERM-SEM model development were applied to HDI modeling using data in 2022 with three latent variables, namely a Long and Healthy Life (LHL), Knowledge (Know_L), and a Decent Standard of Living (DLS) (based on UN standards). The assessment of the outer model in SEM was based on the loading factor values that exceed 0.5 and their significance. This evaluation aimed to identify indicators that effectively explained or measured latent variables, so it got the revised model in SEM. These indicators are LHL2 and LHL 4 to form LHL. DLS1 and DLS3 are indicators to make up DLS, and for Know_L, they are K2 and K3. The revised SEM model was analyzed using spatial. The results of the spatial dependency test showed that the HDI model significantly led to the SERM-SEM model. Knowledge and a decent standard of living variables significantly influence HDI.

**Keywords-** Structural equation modeling, Spatial, Spatial error model, Human development index, Spatial SEM.

## 1. Introduction

The Human Development Index (HDI) was determined by the United Nations (UNDP, 2022) based on three main components, namely a long and healthy life, knowledge (measured by two indicators), and a decent standard of living. It is crucial to acknowledge that these same components are also used to measure HDI in Indonesia (BPS, 2022a). Studies on HDI modeling and its relationship with various indicators are based on these three dimensions (BPS, 2022a; UNDP, 2022). Some of these studies have utilized spatial analysis because it is believed that HDI exhibits spatial interdependence among regions.

Wati & Khikmah (2020) developed the HDI modeling using the Spatial Error Model (SERM), which considered five indicators, including poverty, student-teacher ratio, school participation rate, inflation, and open unemployment. In modeling HDI, Darsyah et al. (2018) employed spatial analysis with four independent variables such as life expectancy, expected length of school, average school duration, and real per capita consumption. The study revealed the superiority of SERM over the Ordinary Least Square (OLS) and Spatial Autoregressive Model (SAR), as evidenced by the lowest Akaike Information Criterion (AIC) score. In addition, Pramesti & Indrasetianingsih (2018) investigated the factors influencing HDI at the district level in East Java, Indonesia, utilizing four indicators, which are per capita expenditures, morbidity, the percentage of impoverished people, and mean years of schooling. The results of the spatial dependency test led to both SAR and SERM, with the SERM yielding the lowest AIC score. Rahma (2020) carried out spatial modeling with five indicators believed to impact HDI. The indicators were life expectancy, expected years of schooling, the number of impoverished people, per capita expenditure, and the open unemployment rate. The spatial dependency test, employing the Lagrange Multiplier (LM), supported the suitability of the SERM model. Niranjan (2020) investigated spatial disparities in human development using eleven indicators, including the number of doctors per 1,000 people, the proportion of anemic pregnant women, the proportion of underweight children born, the proportion of malnourished children, male labor force participation rate, urbanization, the rate of school dropouts, per-capita spending on health care, and average rainfall. Consequently, it was discovered that the LM spatial dependency test led to the SAR model.

None of the studies mentioned above provide answers regarding the particular dimension that truly affects HDI among the three defined by UN and BPS, as they constructed models based on the indicators used. Jeong & Yoon (2018) recognized a limitation, which relied on spatial regression with an indicator-based approach to identify vulnerability factors influenced by natural disasters. This viewpoint seemed to be at odds with the goal of the study, which was to determine the effects of the economic losses brought on by natural disasters on the aspects of the economic, social, and environmental vulnerability in 230 South Korean local communities. Each aspect represented a latent variable measured through several indicators. However, an indicator-based spatial regression hindered the ability to pinpoint which variables or aspects influenced this vulnerability. Ahmed et al. (2022) proposed a solution, suggesting the regression of components resulting from Principal Component Analysis (PCA). However, this method had its weaknesses, including the absence of an evaluation of the contribution of each indicator suspected of forming the components or latent variables.

Dimensions or components represent manifestations of several indicators that cannot be directly measured. These dimensions, which are also known as latent variables, can be modeled optimally by using Structural Equation Modeling (SEM). SEM will be used to address at least two significant questions. First, SEM will empirically confirm the construct's suitability based on indicators conceptualized as measurable variables of the construct (Joreskog & Sorbom, 1993). Second, SEM estimates the causal relationship among latent variables, which are then expressed in specific structural equations and shown in a path diagram (Joreskog & Sorbom, 1996). However, it is important to note that SEM comprising spatial data remains a relatively unexplored area.

Anselin (1988) developed a spatial model using cross-section data. At least three models have been developed, namely the Spatial Autoregressive Model (SAR), the Spatial Error Model (SERM), and the Spatial Autoregressive Moving Average (SARMA). Numerous methods for estimating spatial model parameters have been widely used. The estimation method of Generalized Methods of Moment (GMM) is used to overcome the biases associated with maximum likelihood estimation (MLE) (Anwar et al., 2020; Januardi & Utomo, 2017). The generalized spatial two-stage least square (GS2SLS) estimation method is used to obtain a consistent estimator (Saputro et al., 2019). Nevertheless, there is a paucity of studies utilizing location sample units that are hypothesized to have both spatial influence and involve causal relationships among latent variables.

This study offers a unique perspective, considering that HDI was influenced by three dimensions or latent variables, measured through appropriate indicators, with the sample unit being location. The analysis combines spatial regression and SEM, creating what is known as spatial SEM.

There are two options for incorporating spatial weights into SEM, namely at the structural or measurement model levels. Incorporating spatial weights at the measurement model level generally consists of the analysis of multivariate spatial data (Wang & Wall, 2003), as variables often exhibit correlations when measured within the same spatial area. Each variable tends to correlate with its location due to similar geographic characteristics. However, this approach differs from earlier explorations, which primarily focus on multivariate spatial data within the measurement model (Christensen & Amemiya, 2002; Hogan & Tchernis, 2004; Wang & Wall, 2003). Some studies have modeled regression relationships among latent variables, including spatial weights in a structural model (Congdon, 2008; Liu et al., 2005; Oud & Folmer, 2008).

Oud & Folmer (2008) proposed an enhanced approach to the spatial dependence model, advocating the placement of spill-over effects within the structural model. This modeling offered greater flexibility and information richness than the conventional practice of assigning spatial weights to the measurement model. Parameter estimation was carried out through the Full Information Maximum Likelihood (FIML) method. It presented computational complexity. The methodology did not include a spatial dependency test and the estimation of latent variables. In a separate study, Anekawati et al. (2017) explored education quality modeling by estimating latent variables using the Partial Least Square (PLS) method, performed parameter estimation via the Maximum Likelihood Estimation (MLE), and conducted spatial dependency tests using the Lagrange Multiplier (LM). During the LM test, the model error distribution was assumed to match the spatial model of Anselin (1988). As a result, this modeling result ultimately yielded the SAR-SEM.

The benefit of this research is that it contributes significantly to the existing literature by comprehensively developing the SERM-SEM. Given the predominant focus of prior HDI modeling studies on the SERM model, this study endeavors to comprehensively develop the SERM-SEM, namely a spatial model in the form of the Spatial Error Model (SERM), which involves analyzing the relationship among latent variables (SEM). In this context, parameter estimation employs the generalized method of moments (GMM), and the spatial dependency test is executed through the Lagrange Multiplier (LM) method. It is noteworthy that the LM test necessitates knowledge of the model error distribution. Consequently, this study presents a new perspective in SERM-SEM modeling by using the model error distribution, distinct from the error distribution of the traditional spatial model. In SERM-SEM, the variables are substituted by the factor scores that the latent variable estimation results from SEM. These latent variables were estimated using the weighted least squares (WLS) method and subsequently integrated into the spatial framework. The error distribution of SERM-SEM was derived from the result of latent variable estimation in SEM, which is distinct from Anselin (1988). This SERM-SEM model will be instrumental in addressing issues pertaining to the aforementioned HDI modeling challenges, namely HDI modeling based on dimensions, no longer based on indicators.

## 2. Materials and Methods

This section covers associated literature studies Structural Equation Modeling (SEM), spatial model, spatial SEM, latent variable estimation and research methodology.

### 2.1 Structural Equation Modeling (SEM)

Bollen (1989) introduced Structural Equation Modeling (SEM) as comprising two fundamental models, including the measurement and structural. In the measurement model, the relationship between manifest variables (indicators) and exogenous variables is shown in Equation (1) and the relationship with endogenous variables is written in in Equation (2). On the other hand, the structural model describes the interactions among latent variables as in Equation (3).

$$\mathbf{x}_{A\times 1} = \mathbf{\Lambda_x \xi} + \boldsymbol{\delta}^* \tag{1}$$

$$\mathbf{y}_{B\times 1} = \mathbf{\Lambda_y \eta} + \boldsymbol{\varepsilon}^* \tag{2}$$

$$\boldsymbol{\eta}_{q\times 1} = \mathbf{B\eta} + \mathbf{\Gamma}^*\boldsymbol{\xi} + \boldsymbol{\zeta} \tag{3}$$

where, $\boldsymbol{\delta}^* \sim N(\mathbf{0}, \mathbf{\Theta}_{\delta})$ and $\boldsymbol{\varepsilon}^* \sim N(\mathbf{0}, \mathbf{\Theta}_{\varepsilon^*})$. $\boldsymbol{\eta}$ notation is a vector of endogenous random variables, $\boldsymbol{\xi}$ represents a vector of exogenous random variables, $\mathbf{B}$ denotes a coefficient matrix that shows the effect of the relationship between endogenous latent variables on other endogenous, and $\mathbf{\Gamma}^*$ indicates a coefficient matrix showing the effect of the relationship between exogenous latent variables ($\boldsymbol{\xi}$) on endogenous ($\boldsymbol{\eta}$). Meanwhile $\boldsymbol{\zeta}$ refers to a random error vector, with an expectation score of zero. The $q$ index depicts the number of endogenous latent variables and $p$ index is its exogenous numbers.

$\mathbf{y}$ and $\mathbf{x}$ notations are vectors of the observed variables, while $\mathbf{\Lambda_y}$ and $\mathbf{\Lambda_x}$ are coefficient matrix indicating the relationship between $\mathbf{y}$ and $\boldsymbol{\eta}$ as well as $\mathbf{x}$ and $\boldsymbol{\xi}$. $\boldsymbol{\varepsilon}^*$ and $\boldsymbol{\delta}^*$ notations are error vectors determined from the measurement of $\mathbf{y}$ and $\mathbf{x}$. This error is the addition of a component representing the imperfection of the indicators in measuring the related latent variables.

The indicators number of the *i*-th exogenous latent variable is $a_i$ where $i = 1, 2, 3, \ldots, p$. The total number of indicators on the exogenous latent variables is calculated as follows $\sum_{i=1}^{p} a_i = A$, where $A$ denotes the outcome. Generally, the covariance matrix of the measurement error from the observed variables $\mathbf{x}$ with $p$ exogenous latent variables and the total number of indicators $A$ is determined as $\mathbf{\Theta}_{\delta A\times A} = diag\left(\sigma^2_{\delta_{(1)1}}, \sigma^2_{\delta_{(2)1}}, \cdots, \sigma^2_{\delta_{(a1)1}}, \sigma^2_{\delta_{(1)2}}, \sigma^2_{\delta_{(2)2}}, \cdots, \sigma^2_{\delta_{(a2)2}}, \cdots, \sigma^2_{\delta_{(1)p}}, \sigma^2_{\delta_{(2)p}}, \cdots, \sigma^2_{\delta_{(ap)p}}\right)$.

The indicators number of the j-th endogenous latent variable is denoted by $b_j$, where $j = 1, 2, 3, \ldots, q$ $j = 1, 2, 3, \ldots, q$ and the total number is calculated using $\sum_{j=1}^{q} b_j = B$. In general, the covariance matrix of the measurement error from the observed variables $\mathbf{y}$ with the number of endogenous latent variables is one, while the total number of indicators $B$ is $\mathbf{\Theta}_{\varepsilon^* B\times B} = diag\left(\sigma^2_{\varepsilon_1^*}, \sigma^2_{\varepsilon_2^*}, \cdots, \sigma^2_{\varepsilon_B^*}\right)$.

The measurement model describes the relationship between a group of indicators and its latent variable. There are three ways to build relationships between the latent variable and its indicators (Trujillo, 2009):

the formative model, reflective, and multiple effect indicators for multiple causes (MIMIC). The formative model forms latent variable relationships caused or formed by each indicator. The reflective model creates a relationship where the latent variable is considered the cause of the indicators, or the indicators reflect or manifest the latent variable. Meanwhile, the MIMIC model is a combination of formative and reflective models. This study uses a reflective measurement model. Trujillo (2009) wrote the reflective measurement model equation as Equation (1) and Equation (2).

Assumptions that must be met in the structural Equation (1), Equation (2), and Equation (3) include $E(\boldsymbol{\eta})=\mathbf{0}$, $E(\boldsymbol{\xi})=\mathbf{0}$, $E(\boldsymbol{\zeta})=\mathbf{0}$, $E(\boldsymbol{\varepsilon}^*)=\mathbf{0}$ $E(\boldsymbol{\varepsilon}^*)=\mathbf{0}$, $E(\boldsymbol{\delta}^*)=\mathbf{0}$, $E(\boldsymbol{\xi}\boldsymbol{\zeta}^{'})=\mathbf{0}$, where $\boldsymbol{\xi}$ is not correlated with $\boldsymbol{\xi}$ and the matrix $(\mathbf{I}-\mathbf{B})$ is nonsingular.

## 2.2 Spatial Model

Anselin (1988) developed a spatial model using cross-section spatial data, expressed as follows:

$$\mathbf{y}^*_{T\times 1}=\mathbf{X}\boldsymbol{\beta}+\lambda\mathbf{W}\mathbf{y}^*+\mathbf{u} \tag{4}$$

with $\mathbf{u}_{T\times 1}=\rho\mathbf{M}\mathbf{u}+\boldsymbol{\varepsilon}$ and $\boldsymbol{\varepsilon} \sim N(\mathbf{0},\sigma^2_{\varepsilon}\mathbf{I})$. From the above equation, $\mathbf{y}^*$ represents a vector of endogenous variable with partial dependency, $\mathbf{x}$ is the exogenous variable matrix, $\boldsymbol{\beta}$ denotes the parameter vector of regression model, and $\lambda$ indicates the autoregressive spatial coefficient with value $|\lambda|<1$. Furthermore, $\rho$ is the spatial coefficient in error $\boldsymbol{\varepsilon}$ with its value $|\rho|<1$, $\mathbf{W}$ and $\mathbf{M}$ are the spatial weighting matrix of the diagonal elements zero, $\mathbf{u}$ is the regression error vector assuming it has random spatial effects and autocorrelated errors, and $\boldsymbol{\varepsilon}$ signifies the error vector. $T$ index is the number of observation or locations (t=1, 2, 3, …, T), while $p$ is the number of variables $\mathbf{x}$.

In instances where $\rho=0$ and $\lambda=0$, an OLS linear regression model is obtained, which is devoid of spatial effects. Subsequently, when $\rho=0$ and $\lambda\neq 0$, SAR is obtained with $\mathbf{y}^*_{T\times 1}=\mathbf{X}\boldsymbol{\beta}+\lambda\mathbf{W}\mathbf{y}^*+\mathbf{u}$. If $\rho\neq 0$ and $\lambda=0$, then it is used to gain SERM, which is $\mathbf{y}^*_{T\times 1}=\mathbf{X}\boldsymbol{\beta}+\mathbf{u}$ or $\mathbf{y}^*=\rho\mathbf{M}\mathbf{y}^*+\mathbf{X}\boldsymbol{\beta}-\rho\mathbf{M}\mathbf{X}\boldsymbol{\beta}+\boldsymbol{\varepsilon}$. When $\rho\neq 0$ and $\lambda\neq 0$, the obtained SARMA is determined using Equation (4).

## 2.3 Spatial Structural Equation Modeling (Spatial SEM)

The formulation of a latent spatial lag (SAR) by Oud & Folmer (2008) consisted the regression of latent variables measured through multiple indicators within the Multiple Indicators Multiple Causes (MIMIC) framework. In this transformation, the variable $\mathbf{y}$, signifying the vector with observations on the dependent variable, was replaced with $\boldsymbol{\eta}$ and latent variable $\boldsymbol{\xi}=\mathbf{x}$, thereby introducing spatial dependency into the structural model. The equations used to establish SAR and SERM were written as $\boldsymbol{\eta}=\rho\mathbf{W}\boldsymbol{\eta}+\boldsymbol{\Gamma}^*\mathbf{x}+\boldsymbol{\zeta}$ and $\boldsymbol{\eta}=\boldsymbol{\Gamma}^*\mathbf{x}+\boldsymbol{\varepsilon}$, respectively, with $\boldsymbol{\varepsilon}=(\mathbf{I}-\lambda\mathbf{M})^{-1}\boldsymbol{\zeta}$.

The SEM incorporated latent variables that cannot be directly measured. In the spatial SEM, these latent variables are substituted with factor scores. Additionally, both endogenous and exogenous factor scores are derived from the estimation results of the respective latent variables. This article references the spatial SEM model outlined in the study conducted by Oud & Folmer (2008). Still, it presents different things for several purposes, including (i) estimating the factor scores of latent variables through WLS, (ii) excluding the use of MIMIC, (iii). introducing spatial dependence to latent variables (structural model), and (iv). employing the $\mathbf{W}$ matrix-sized $T \times T$, indicating spatial dependence between observations or locations. The spatial SEM is generally expressed in accordance with the notation adapted to the model in section 2.2, as follows: $\boldsymbol{l}_{T\times1} = \boldsymbol{K}\boldsymbol{\beta} + \lambda\mathbf{W}\boldsymbol{l} + \mathbf{u}$, where $\mathbf{u}_{T\times1} = \rho\mathbf{M}\mathbf{u} + \boldsymbol{\varepsilon}$ or $\mathbf{u}_{T\times1} = \left(\mathbf{I} - \rho\mathbf{M}\right)^{-1}\boldsymbol{\varepsilon}$.

$\boldsymbol{l}$ is a vector of the endogenous factor score that has spatial dependence, $\boldsymbol{K}$ signifies the matrix of the exogenous factor score, $\boldsymbol{\beta}$ represents a vector of the parameter regression, and $\lambda$ denotes an autoregressive spatial coefficient with the value $|\lambda| < 1$. Furthermore, $\rho$ is a spatial coefficient in error with the value $|\rho| < 1$, $\mathbf{W}$ and $\mathbf{M}$ are a spatial weighting matrix with diagonal elements zero, $\mathbf{u}$ denotes a regression error vector, assumed to have both random spatial effect and spatially autocorrelated errors, $\boldsymbol{\varepsilon}$ indicates an error vector. The $T$ index is the number of observations or locations (t=1, 2, 3, …, T), and $p$ is the number of $\boldsymbol{K}$ variables.

When $\rho = 0$ and $\lambda \neq 0$, the spatial autoregressive model on the SEM (SAR-SEM) be expressed as $\boldsymbol{l}_{T\times1} = \boldsymbol{K}\boldsymbol{\beta} + \lambda\mathbf{W}\boldsymbol{l} + \boldsymbol{\varepsilon}$. This model assumes that the autoregressive model is only on the endogenous factor score ($\boldsymbol{l}$). When $\rho \neq 0$ and $\lambda = 0$, the SERM-SEM is obtained as follows:

$$\boldsymbol{l}_{T\times1} = \boldsymbol{K}\boldsymbol{\beta} + \mathbf{u} \tag{5}$$

SEM is usually a two-step procedure, namely the Exploratory Factor Analysis (EFA) and Confirmatory Factor Analysis (CFA). In order to find latent variables based on indicators, EFA classifies data items into matching latent variables accurately and without the need for a predefined hypothesis (Li et al., 2021). The spatial SEM focuses more on the measurement model, namely, to determine which indicators to use for construct measurement, so it involves EFA. Hair et al. (2021) explained that the measurement theory described how to use indicators to measure constructs. After the measurement model is confirmed, it doesn't move on to testing the structural theory. Latent variables are estimated and modeled in a spatial model. As a result, in spatial SEM modeling, it is essential to build a measuring model based on conceptual theories. In addition, the relationships among latent variables must also be built based on strong theoretical concepts. The relationship among latent variables must be derived from theory and logic (Hair et al., 2021)

The assumptions in the spatial SEM model are: 1) the spatial weight matrix $\mathbf{W}_T = \mathbf{M}_T$; 2) ddiagonal elements of the spatial weight matrix $\mathbf{W}_T$ are zero; 3) the matrix $\left(\mathbf{I} - \lambda\mathbf{W}_T\right)$ and $\left(\mathbf{I} - \rho\mathbf{W}_T\right)$ are nonsingular with $|\lambda| < 1$ and $|\rho| < 1$; 4) $\varepsilon_i$ has identical distribution (has property of independent); and 5) model has property of linear in variables.

## 2.4 The Estimation of Latent Variables

Estimation of latent variables is based on a measurement model that has been evaluated so that the indicators that form the latent variable are valid. The measurement model uses a reflective model as in Equation (1) and Equation (2). Estimation of latent variables, both endogenous and exogenous latent, uses weighted least squares (WLS). The estimation results of latent variables using the WLS method indicated that the equation of measurement model for the exogenous latent variable is in line with Equation (1), where $\boldsymbol{\delta}^* \sim N(\mathbf{0}, \boldsymbol{\Theta}_\delta)$ assuming $\boldsymbol{\Lambda}_x$ and $\boldsymbol{\Theta}_\delta$ were constants. Therefore, the vector distribution of the observed variable $\mathbf{x}$ was $\mathbf{x} \sim N(\boldsymbol{\Lambda}_{\mathbf{x}}\boldsymbol{\xi}, \boldsymbol{\Theta}_\delta)$. The estimation results of $\boldsymbol{\xi}$ by using the WLS method was expressed as follows:

$$\sum_{t=1}^{T}\hat{\boldsymbol{\xi}}_t=\left(\boldsymbol{\Lambda}_x^{'}\boldsymbol{\Theta}_\delta^{-1}\boldsymbol{\Lambda}_x\right)^{-1}\left(\boldsymbol{\Lambda}_x^{'}\boldsymbol{\Theta}_\delta^{-1}\right)\sum_{t=1}^{T}\mathbf{x}_t \tag{6}$$

From the equation above, matrix $\left(\boldsymbol{\Lambda}_x^{'}\boldsymbol{\Theta}_\delta^{-1}\boldsymbol{\Lambda}_x\right)$ is a diagonal matrix with an element value of zero. Meanwhile, the distribution of the observations random matrix from the observed variable vector $\mathbf{x}$ is $\mathbf{X}\sim N_{A,T}\left(\boldsymbol{\Lambda}_\mathbf{x}\boldsymbol{\xi}_\mathbf{t}\mathbf{e}^{'},\boldsymbol{\Theta}_\delta\otimes\mathbf{I}_T\right)$, where $\mathbf{e}_{T\times1}=(1,\ldots,1)^{'}$. When Equation (6) is turned into a matrix-formed equation with each contain $\mathbf{x}$ and $\hat{\xi}$, as well as assumed matrix $\begin{pmatrix}\hat{\xi}_{11} & \hat{\xi}_{12} & \cdots & \hat{\xi}_{12}\\ \hat{\xi}_{21} & \hat{\xi}_{22} & \cdots & \hat{\xi}_{2T}\\ \vdots & \vdots & \ddots & \vdots\\ \hat{\xi}_{p1} & \hat{\xi}_{p2} & \cdots & \hat{\xi}_{pT}\end{pmatrix}=\boldsymbol{K}^{'}_{p\times T}$, then Equation (6) can be simplified into $\boldsymbol{K}^{'}_{p\times T}=\left(\boldsymbol{\Lambda}_\mathbf{x}^{'}\boldsymbol{\Theta}_\delta^{-1}\boldsymbol{\Lambda}_\mathbf{x}\right)^{-1}\left(\boldsymbol{\Lambda}_\mathbf{x}^{'}\boldsymbol{\Theta}_\delta^{-1}\right)\mathbf{X}$, thereby producing the distribution of $\boldsymbol{K}$ as

$$\mathrm{K}_{T\times p}\sim N_{T,p}\left(\mathbf{e}\boldsymbol{\xi}_t^{'},\mathbf{I}_T\otimes\left(\boldsymbol{\Lambda}_\mathbf{x}^{'}\boldsymbol{\Theta}_\delta^{-1}\boldsymbol{\Lambda}_\mathbf{x}\right)^{-1}\right) \tag{7}$$

The equation of measurement model with endogenous latent variables was determined using Equation (2) with the assumptions $\boldsymbol{\Lambda}_y$ and $\boldsymbol{\Theta}_{\varepsilon^*}$ constant. Consequently, the distribution of observed variable vector $\mathbf{y}$ was $\mathbf{y}\sim N\left(\boldsymbol{\Lambda}_\mathbf{y}\eta,\boldsymbol{\Theta}_{\varepsilon^*}\right)$. The estimation results of $\eta$ by using WLS method calculated in Equation (8) as follows.

$$\sum_{t=1}^{T}\hat{\eta}_t=\left(\boldsymbol{\Lambda}_\mathbf{y}^{'}\boldsymbol{\Theta}_{\varepsilon^*}^{-1}\boldsymbol{\Lambda}_\mathbf{y}\right)^{-1}\boldsymbol{\Lambda}_\mathbf{y}^{'}\boldsymbol{\Theta}_{\varepsilon^*}^{-1}\sum_{t=1}^{T}\mathbf{y}_t \tag{8}$$

The distribution of the observations random matrix from the observed variable vector $\mathbf{y}$ is $\mathbf{Y}\sim N_{B,T}\left(\boldsymbol{\Lambda}_y\eta_t\mathbf{e}^{'},\boldsymbol{\Theta}_{\varepsilon^*}\otimes\mathbf{I}_T\right)$, where $\mathbf{e}_{T\times1}=(1,\ldots,1)^{'}$. When Equation (7) is re-written into a matrix form assuming $\boldsymbol{l}^{'}_{1\times T}=\left(\hat{\eta}_1\hat{\eta}_2\ldots\hat{\eta}_T\right)$, where $\boldsymbol{l}^{'}_{1\times T}=\left(\boldsymbol{\Lambda}_\mathbf{y}^{'}\boldsymbol{\Theta}_{\varepsilon^*}^{-1}\boldsymbol{\Lambda}_\mathbf{y}\right)^{-1}\boldsymbol{\Lambda}_y^{'}\boldsymbol{\Theta}_{\varepsilon^*}^{-1}\mathbf{Y}$, the distribution of $\boldsymbol{l}$ was obtained as shown in Equation (9).

$$\mathrm{l}_{T\times1}\sim N_T\left(\mathbf{e}\eta_t,\left(\boldsymbol{\Lambda}_\mathbf{y}^{'}\boldsymbol{\Theta}_{\varepsilon^*}^{-1}\boldsymbol{\Lambda}_\mathbf{y}\right)^{-1}\mathbf{I}\right) \tag{9}$$

## 2.5 Methodology

The development process for SERM-SEM model consisted of the following steps (Figure 1).

The first step is the process of obtaining the error distribution for the SERM-SEM model. The error distribution is obtained based on the expectation value and variance of the model in Equation (5). The second step is the process of estimating SERM-SEM model parameters. Parameter estimation uses the GS2SLS procedure in the second stage, namely the generalized method of moments (GMM). At this stage, it will obtain the first, second, and third-moment equations and then minimize the sum of squared residuals. The third step is the process of developing the spatial dependency test using the LM test. This step needs the ln likelihood function with the error distribution of the resulting model in step 1. Breusch & Pagan (1980) defined the LM test statistic as $\mathrm{LM}=\hat{\mathbf{D}}^{'}\hat{\boldsymbol{\psi}}^{-1}\hat{\mathbf{D}}$, where $\hat{\boldsymbol{\psi}}^{-1}$ is the element of the information matrix inverse. The information matrix is obtained from the first and second derivatives of the ln likelihood function for each parameter. The value of $\hat{\mathbf{D}}_\rho$ is obtained from the first derivative of the ln likelihood function to $\rho$.

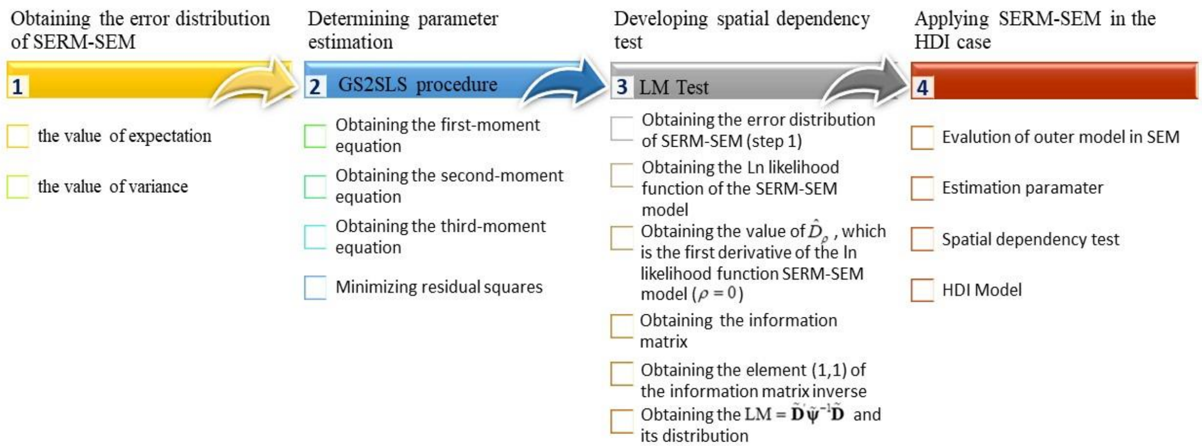


**Figure 1.** The steps of SERM-SEM model development.

The last step is the application of the SERM-SEM model for HDI modeling. The latent variables that influence HDI are determined based on UNDP (2022) and BPS (2022a), namely Long and Healthy Life (LHL), Knowledge (Know_L), and Decent Living Standard (DLS). The indicators used to build each latent variable are based on previous research. Prior to estimating latent variables, estimating the parameters of the SERM-SEM model, and testing the spatial dependency, the outer model in SEM is evaluated first. It's because latent variables that are constructed using reliable and valid indicators are given priority by the model.

## 3. Results

In this section, the study's results will be presented to obtain mathematical formulas related to model error distribution, parameter estimation, and spatial dependency testing. The error distribution of the SERM-SEM model is obtained based on the value of expectation and variance. Parameter estimation of the SERM-SEM model uses the GS2SLS procedure in the second stage, namely the generalized method of moments (GMM). Spatial dependency testing using the Lange Multiplier test.

### 3.1 The Error Distribution of SERM-SEM

SERM-SEM, as represented by Equation (5) with a corresponding distribution of $\boldsymbol{K}$ and $\boldsymbol{l}$ indicated in Equation (7) and Equation (9) can be rewritten as:

$$\boldsymbol{l}_{T\times1} = \boldsymbol{K}\boldsymbol{\beta} + \left(\mathbf{I} - \rho\mathbf{W}\right)^{-1}\boldsymbol{\varepsilon} \quad (10)$$

This model is spatial regression with $\boldsymbol{l}$ representing the response variable and $\boldsymbol{K}$ indicating the random independent variable, resulting in $\boldsymbol{l}$ is $f\left(\boldsymbol{l} \mid \boldsymbol{K}\right)$. This suggests that the variable $\boldsymbol{K}$ is now fixed rather than random. The error matrix from Equation (10) is $\boldsymbol{\varepsilon}_{T\times1} = \left(\mathbf{I} - \rho\mathbf{W}\right)\left(\boldsymbol{l} - \boldsymbol{K}\boldsymbol{\beta}\right)$, and the distribution is formulated based on the value of expectation and variance, namely $E\left(\varepsilon\right) = \left(\mathbf{I} - \rho\mathbf{W}\right)\left(\mathbf{e}\boldsymbol{\eta} - \boldsymbol{K}\boldsymbol{\beta}\right)$ and $\operatorname{var}\left(\boldsymbol{\varepsilon}\right) = \left(\mathbf{I} - \rho\mathbf{W}\right)\left(\boldsymbol{\Lambda}_{\mathbf{y}}'\boldsymbol{\Theta}_{\boldsymbol{\varepsilon}}^{-1}\boldsymbol{\Lambda}_{\mathbf{y}}\right)^{-1}\mathbf{I}\left(\mathbf{I} - \rho\mathbf{W}\right)'$. As a result, the error distribution of SERM-SEM is expressed as follows:

$$\boldsymbol{\varepsilon} \sim N_{T,1}\left((\mathbf{I}-\rho\mathbf{W})(\mathbf{e}\boldsymbol{\eta}-\mathrm{K}\boldsymbol{\beta}),\boldsymbol{\Theta}_{\rho}\right) \tag{11}$$

where, notation of $\mathbf{I}$ is an identity matrix, $\rho$ is a spatial coefficient in error, $\mathbf{W}$ is a spatial weighting matrix, $\mathbf{e}$ is a vector whose members are all 1, namely $\mathbf{e}_{T\times1}=(1,\ldots,1)'$, $\boldsymbol{\eta}$ is a vector of endogenous random variables, $\boldsymbol{K}$ signifies the matrix of the exogenous factor score, $\boldsymbol{\beta}$ represents a vector of the parameter regression, and $\boldsymbol{\Theta}_{\rho}$ is a matrix of variance-covariance, namely $\boldsymbol{\Theta}_{\rho}=(\mathbf{I}-\rho\mathbf{W})\left(\boldsymbol{\Lambda}_{\mathbf{y}}'\boldsymbol{\Theta}_{\boldsymbol{\varepsilon}^*}^{-1}\boldsymbol{\Lambda}_{\mathbf{y}}\right)^{-1}\mathbf{I}(\mathbf{I}-\rho\mathbf{W})'$. The notation of $\boldsymbol{\Lambda}_{\mathbf{y}}$ is coefficient matrix indicating the relationship between $\mathbf{y}$ and $\boldsymbol{\eta}$. $\boldsymbol{\Theta}_{\boldsymbol{\varepsilon}^*}$ is the covariance matrix of the measurement error from the observed variables $\mathbf{y}$. $\boldsymbol{\Theta}_{\rho}$ is assumed that the variable $\boldsymbol{K}$ is not correlated with $\boldsymbol{\varepsilon}$ (error) and $\mathrm{cov}(\boldsymbol{l},\boldsymbol{K})\neq\mathbf{0}$.

### 3.2 The Parameter Estimation of SERM-SEM

The estimation of parameters in spatial SEM employed a procedure called the GS2SLS developed by Kelejian & Prucha (1998; 1999). The parameter estimation for SERM-SEM was conducted using GMM method, building upon the result of SAR-SEM. The result of SAR-SEM model was represented by $\boldsymbol{l}_i$ score, namely $\hat{\boldsymbol{l}}_{T\times1}=\mathbf{Z}\hat{\boldsymbol{\delta}}+\boldsymbol{\varepsilon}$ and

$$\hat{\boldsymbol{\delta}}_{(p+2)\times1}=\left(\hat{\boldsymbol{\beta}}' \mid \hat{\lambda}\right)'=\left(\hat{\mathbf{Z}}'\hat{\mathbf{Z}}\right)^{-1}\hat{\mathbf{Z}}'\boldsymbol{l} \tag{12}$$

where, $\hat{\mathbf{Z}}_{T\times(p+2)}=\mathbf{P}_{\mathbf{H}}\mathbf{Z}$, $\mathbf{Z}_{T\times(p+2)}=(\boldsymbol{K}\mid\mathbf{W}\boldsymbol{l})$, $\mathbf{P}_{\mathbf{H}_{T\times T}}=\mathbf{H}\left(\mathbf{H}'\mathbf{H}\right)^{-1}\mathbf{H}'$, and $\mathbf{H}_{T\times2(p+1)}=(\boldsymbol{K}\mid\mathbf{W}\boldsymbol{K})$.

The residual value, designated as $\hat{\mathbf{u}}$, which serves as an observation vector for the random variable $\mathbf{u}$ on the SERM-SEM, where $\hat{\mathbf{u}}=\boldsymbol{l}_i-\boldsymbol{l}_i$. The error equation of SERM-SEM model in Equation (11) is used to obtain the value of $\boldsymbol{\varepsilon}=\mathbf{u}-\rho\mathbf{W}\mathbf{u}$. When $\bar{\mathbf{u}}=\mathbf{W}\mathbf{u}$, the error is written $\boldsymbol{\varepsilon}=\mathbf{u}-\rho\bar{\mathbf{u}}$. The equation of first moment is obtained by squaring $\boldsymbol{\varepsilon}=\mathbf{u}-\rho\bar{\mathbf{u}}$ and then divided by $T$ for all elements. The second-moment equation is obtained by multiplying each element in $\boldsymbol{\varepsilon}=\mathbf{u}-\rho\bar{\mathbf{u}}$ with $\mathbf{W}$. By taking $\bar{\boldsymbol{\varepsilon}}=\mathbf{W}\boldsymbol{\varepsilon}$ and $\bar{\bar{\mathbf{u}}}=\mathbf{W}\bar{\mathbf{u}}$, the result obtained is $\bar{\boldsymbol{\varepsilon}}=\bar{\mathbf{u}}-\rho\bar{\bar{\mathbf{u}}}$, which is then squared, and each element is divided by $T$. The third-moment equation is obtained by multiplying $\boldsymbol{\varepsilon}=\mathbf{u}-\rho\bar{\mathbf{u}}$, followed by dividing each element by $T$. The three moment equations are written into the matrix as follows

$$\begin{bmatrix} 2T^{-1}E\left(\mathbf{u}'\bar{\mathbf{u}}\right) & -T^{-1}E\left(\bar{\mathbf{u}}'\bar{\mathbf{u}}\right) & 1 \\ 2T^{-1}E\left(\bar{\mathbf{u}}'\bar{\bar{\mathbf{u}}}\right) & -T^{-1}E\left(\bar{\bar{\mathbf{u}}}'\bar{\bar{\mathbf{u}}}\right) & Tr\left(\mathbf{W}'\mathbf{W}\right) \\ T^{-1}E\left(\mathbf{u}'\bar{\bar{\mathbf{u}}}+\bar{\mathbf{u}}'\bar{\mathbf{u}}\right) & -T^{-1}E\left(\bar{\mathbf{u}}'\bar{\bar{\mathbf{u}}}\right) & 0 \end{bmatrix}\begin{bmatrix} \rho \\ \rho^2 \\ tr\left(\boldsymbol{\Theta}_{\rho}\right) \end{bmatrix}-\begin{bmatrix} T^{-1}E\left(\mathbf{u}'\mathbf{u}\right) \\ T^{-1}E\left(\bar{\mathbf{u}}'\bar{\mathbf{u}}\right) \\ T^{-1}E\left(\mathbf{u}'\bar{\mathbf{u}}\right) \end{bmatrix}=\begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}$$

. This expression can be summarized as $\boldsymbol{\Gamma}\boldsymbol{\alpha}-\boldsymbol{\gamma}=\mathbf{0}$.

One estimator that required determination is $\rho$, while $\boldsymbol{\Theta}_{\rho}$ contains $\rho$ obtained from $\boldsymbol{\Gamma}$ and $\boldsymbol{\gamma}$. From this previous equation, $\mathbf{G}$ is computed as the $\boldsymbol{\Gamma}$ estimator, which is expressed in Equation (13) and Equation (14).

$$\hat{\boldsymbol{\Gamma}} = \mathbf{G} = \begin{bmatrix} 2T^{-1}\hat{\mathbf{u}}'\mathbf{W}\hat{\mathbf{u}} & -T\hat{\mathbf{u}}'\mathbf{W}'\mathbf{W}\hat{\mathbf{u}} & 1 \\ 2T^{-1}\hat{\mathbf{u}}'\mathbf{W}'\mathbf{W}^2\hat{\mathbf{u}} & -T\hat{\mathbf{u}}'(\mathbf{W}^2)'\mathbf{W}^2\hat{\mathbf{u}} & Tr\left(\mathbf{W}'\mathbf{W}\right) \\ T^{-1}\left(\hat{\mathbf{u}}'\mathbf{W}^2\hat{\mathbf{u}} + \hat{\mathbf{u}}'\mathbf{W}'\mathbf{W}\hat{\mathbf{u}}\right) & -T\hat{\mathbf{u}}'\mathbf{W}'\mathbf{W}^2\hat{\mathbf{u}} & 0 \end{bmatrix} \tag{13}$$

$$\hat{\boldsymbol{\gamma}} = \mathbf{g} = \frac{1}{T}\begin{bmatrix} \hat{\mathbf{u}}'\hat{\mathbf{u}} \\ \hat{\mathbf{u}}'\mathbf{W}'\mathbf{W}\hat{\mathbf{u}} \\ \hat{\mathbf{u}}'\mathbf{W}\hat{\mathbf{u}} \end{bmatrix} \tag{14}$$

The estimator of $\boldsymbol{\gamma}$ is denoted as $\mathbf{g}$. The residual value can be re-written as $\hat{\mathbf{u}}_{T\times 1} = \boldsymbol{l} - \hat{\boldsymbol{l}}$ or $\hat{\mathbf{u}}_{T\times 1} = \boldsymbol{l} - \mathbf{Z}\hat{\boldsymbol{\delta}}$. The value of $\hat{\boldsymbol{\delta}}$ is consistent with Equation (12) and $\mathbf{Z}_{T\times(p+2)} = \left(\boldsymbol{K} \mid \mathbf{W}\boldsymbol{l}\right)$. The empirical equation of moment condition can be re-written as follows:

$$\mathbf{g}_{3\times 1} = \mathbf{G}\boldsymbol{\alpha} - \mathbf{v} \tag{15}$$

The definition of GMM in this case is the outcome of minimizing the sum of squared residuals, abbreviated as $\mathbf{v}'\mathbf{v}$. The value of $\mathbf{v}_{3\times 1} = \mathbf{G}\boldsymbol{\alpha} - \mathbf{g}$ is obtained from Equation (15) leading to the determination of residual squares, expressed as $\mathbf{v}'\mathbf{v}_{1\times 1} = \boldsymbol{\alpha}'\mathbf{G}'\mathbf{G}\boldsymbol{\alpha} - \boldsymbol{\alpha}'\mathbf{G}'\mathbf{g} - \mathbf{g}'\mathbf{G}\boldsymbol{\alpha} + \mathbf{g}'\mathbf{g}$, where $\boldsymbol{\alpha}'\mathbf{G}'\mathbf{g}$ denoted symmetrical scalar, resulting in $\mathbf{v}'\mathbf{v}_{1\times 1} = \boldsymbol{\alpha}'\mathbf{G}'\mathbf{G}\boldsymbol{\alpha} - 2\left(\mathbf{G}'\mathbf{g}\right)'\boldsymbol{\alpha} + \mathbf{g}'\mathbf{g}$. The estimated value of $\boldsymbol{\alpha}$ is determined by minimizing residual squares, namely $\frac{\partial \mathbf{v}'\mathbf{v}}{\partial \boldsymbol{\alpha}} = 0$, where the estimated value of $\boldsymbol{\alpha}$ is

$$\hat{\boldsymbol{\alpha}}_{3\times 1} = \left(\mathbf{G}'\mathbf{G}\right)^{-1}\mathbf{G}'\mathbf{g} \text{ or } \hat{\boldsymbol{\alpha}}_{3\times 1} = \left[\hat{\rho} \quad \hat{\rho}^2 \quad tr\left(\hat{\boldsymbol{\Theta}}_{\rho}\right)\right]'_{1\times 3} = \left(\mathbf{G}'\mathbf{G}\right)^{-1}\mathbf{G}'\mathbf{g} \tag{16}$$

The notation of $\boldsymbol{\alpha}$ is a vector that contains parameters, namely $\left(\rho \quad \rho^2 \quad tr\left(\boldsymbol{\Theta}_{\rho}\right)\right)'$, $\rho$ is a spatial coefficient in error, $\boldsymbol{\Theta}_{\rho}$ is a matrix of variance-covariance, namely $\boldsymbol{\Theta}_{\rho} = \left(\mathbf{I} - \rho\mathbf{W}\right)\left(\boldsymbol{\Lambda}_{\mathbf{y}}'\boldsymbol{\Theta}_{\boldsymbol{\varepsilon}^*}^{-1}\boldsymbol{\Lambda}_{\mathbf{y}}\right)^{-1}\mathbf{I}\left(\mathbf{I} - \rho\mathbf{W}\right)'$. $\mathbf{G}$ matrix value is determined by following Equation (13) and $\mathbf{g}$ matrix value through Equation (14).

### 3.3 The Spatial Dependency Test of SERM-SEM

**Theorem:** If SERM-SEM model and the error distribution, as represented by Equation (5) or Equation (10), and Equation (11), then Langrange Multiplier test is $\mathrm{LM}_{\rho} = \frac{p\left(\tilde{\boldsymbol{\varepsilon}}'\mathbf{W}\tilde{\boldsymbol{\varepsilon}}\right)^2}{D}$ and under the null hypothesis, the LM test statistics followed a distribution of $LM_{\rho} \sim \chi^2_{(1)}$ where $p_{1\times 1} = \left(\boldsymbol{\Lambda}_{\mathbf{y}}'\boldsymbol{\Theta}_{\varepsilon^*}^{-1}\boldsymbol{\Lambda}_{\mathbf{y}}\right)$, $D_{1\times 1} = \left(\mathbf{e}\hat{\eta}_t - \boldsymbol{K}\hat{\boldsymbol{\beta}}\right)'\mathbf{W}\mathbf{W}\left(\mathbf{e}\hat{\eta}_t - \boldsymbol{K}\hat{\boldsymbol{\beta}}\right)$, and $\tilde{\boldsymbol{\varepsilon}}_{T\times 1} = \left(\boldsymbol{l} - \boldsymbol{K}\hat{\boldsymbol{\beta}}\right)$.

**Proof:** In Equation (10) SERM-SEM is accompanied by the error equation $\boldsymbol{\varepsilon}_{T\times 1} = \left(\mathbf{I} - \rho\mathbf{W}\right)\left(\boldsymbol{l} - \boldsymbol{K}\boldsymbol{\beta}\right)$. When $\mathbf{A}_{T\times T} = \left(\mathbf{I} - \rho\mathbf{W}\right)$, the Jacobian for the equation is $J = \left|\frac{\partial \boldsymbol{\varepsilon}}{\partial \boldsymbol{l}}\right| = \left|\mathbf{I} - \rho\mathbf{W}\right| = \left|\mathbf{A}\right|$. By employing the Gaussian function, the likelihood $\boldsymbol{l}$ in SERM-SEM can be obtained by substituting $\boldsymbol{\varepsilon}$ and multiplying by the

Jacobian, as follows: $L(\rho,\boldsymbol{\beta},\boldsymbol{\Theta};\boldsymbol{l})=|\boldsymbol{\Theta}|^{-1/2}|\mathbf{A}|\exp\left[-\frac{1}{2}p(\mathbf{A}\boldsymbol{l}-\mathbf{A}\boldsymbol{K}\boldsymbol{\beta})'(\mathbf{A}\mathbf{A}')^{-1}(\mathbf{A}\boldsymbol{l}-\mathbf{A}\boldsymbol{K}\boldsymbol{\beta})\right]$. On the other hand, the function of ln likelihood for SERM-SEM is:

$$\mathrm{L}(\rho,\boldsymbol{\beta},\boldsymbol{\Theta};\boldsymbol{l})=-\frac{1}{2}\ln|\boldsymbol{\Theta}|+\ln|\mathbf{A}|-\frac{1}{2}p(\mathbf{A}\boldsymbol{l}-\mathbf{A}\boldsymbol{K}\boldsymbol{\beta})'(\mathbf{A}\mathbf{A}')^{-1}(\mathbf{A}\boldsymbol{l}-\mathbf{A}\boldsymbol{K}\boldsymbol{\beta}) \tag{17}$$

with $\boldsymbol{\Theta}_{T\times T}=p^{-1}\mathbf{A}\mathbf{A}'$, $p_{1\times 1}=\left(\boldsymbol{\Lambda}_y'\boldsymbol{\Theta}_{\varepsilon^*}^{-1}\boldsymbol{\Lambda}_y\right)$, $\boldsymbol{\Lambda}_{\mathbf{y}}$ is coefficient matrix indicating the relationship between $\mathbf{y}$ and $\boldsymbol{\eta}$, $\boldsymbol{\Theta}_{\boldsymbol{\varepsilon}^*}$ is the covariance matrix of the measurement error from the observed variables $\mathbf{y}$, and $\mathbf{A}_{T\times T}=(\mathbf{I}-\rho\mathbf{W})$.

The first derivative of the ln likelihood function, as in Equation (17) to $\rho$ is

$$\left.\frac{\partial\mathrm{L}(\rho,\boldsymbol{\beta},\boldsymbol{\Theta};\boldsymbol{l})}{\partial\rho}\right|_{1\times1}=p(\boldsymbol{l}-\boldsymbol{K}\boldsymbol{\beta})'\mathbf{A}^{-1}\mathbf{W}(\boldsymbol{l}-\boldsymbol{K}\boldsymbol{\beta}) \tag{18}$$

and the first derivative of the ln likelihood function to $\boldsymbol{\beta}$ is $\left.\frac{\partial\mathrm{L}(\rho,\boldsymbol{\beta},\boldsymbol{\Theta};\boldsymbol{l})}{\partial\boldsymbol{\beta}}\right|_{(p+1)\times1}=p\boldsymbol{K}'(\boldsymbol{l}-\boldsymbol{K}\boldsymbol{\beta})$. The second derivative of the ln likelihood function to $\rho$ and $\boldsymbol{\beta}$, respectively, is

$$\left.\frac{\partial^2\mathrm{L}(\rho,\boldsymbol{\beta},\boldsymbol{\Theta};\boldsymbol{l})}{\partial\rho^2}\right|_{1\times1}=-p(\boldsymbol{l}-\boldsymbol{K}\boldsymbol{\beta})'\mathbf{A}^{-1}\mathbf{W}\mathbf{A}^{-1}\mathbf{W}(\boldsymbol{l}-\boldsymbol{K}\boldsymbol{\beta}) \tag{19}$$

$$\left.\frac{\partial^2\mathrm{L}(\rho,\boldsymbol{\beta},\boldsymbol{\Theta};\boldsymbol{l})}{\partial\rho\partial\boldsymbol{\beta}'}\right|_{1\times(p+1)}=-2p(\boldsymbol{l}-\boldsymbol{K}\boldsymbol{\beta})'\mathbf{A}^{-1}\mathbf{W}\boldsymbol{K} \tag{20}$$

$$\left.\frac{\partial^2\mathrm{L}(\rho,\boldsymbol{\beta},\boldsymbol{\Theta};\boldsymbol{l})}{\partial\boldsymbol{\beta}\partial\boldsymbol{\beta}'}\right|_{(p+1)\times(p+1)}=-p\boldsymbol{K}'\boldsymbol{K} \tag{21}$$

$$\left.\frac{\partial^2\mathrm{L}(\rho,\boldsymbol{\beta},\boldsymbol{\Theta};\boldsymbol{l})}{\partial\boldsymbol{\beta}\partial\rho}\right|_{(p\times1)\times1}=\mathbf{0} \tag{22}$$

Breusch & Pagan (1980) defined the LM test statistic as $\mathrm{LM}=\hat{\mathbf{D}}'\hat{\boldsymbol{\psi}}^{-1}\hat{\mathbf{D}}$, where $\hat{\boldsymbol{\psi}}^{-1}$ is the element of the information matrix inverse, $\tilde{\boldsymbol{\psi}}_{\theta}$, with size $p\times p$. The information matrix contains elements which are the expectation value of the second derivative for each parameter estimated as follow: $\tilde{\boldsymbol{\psi}}_{\theta}=E\left[\frac{\partial^2L(\theta)}{\partial\theta\partial\theta'}\right]$. The information matrix for SERM-SEM at $\rho=0$ is $\tilde{\boldsymbol{\psi}}_{\theta}=\begin{pmatrix}\tilde{\boldsymbol{\psi}}_{\rho\rho} & \tilde{\boldsymbol{\psi}}_{\rho\boldsymbol{\beta}}\\ \tilde{\boldsymbol{\psi}}_{\boldsymbol{\beta}\rho} & \tilde{\boldsymbol{\psi}}_{\boldsymbol{\beta}\boldsymbol{\beta}'}\end{pmatrix}$. The elements of the information matrix determined by the expectation value of the second derivative in Equation (19), Equation (20), Equation (21), and Equation (22), as well as when $\rho=0$ are as follows: $\tilde{\boldsymbol{\psi}}_{\rho\rho\ 1\times1}=p(\mathbf{e}\eta_t-\boldsymbol{K}\boldsymbol{\beta})'\mathbf{W}\mathbf{W}(\mathbf{e}\eta_t-\boldsymbol{K}\boldsymbol{\beta})$, $\tilde{\boldsymbol{\psi}}_{\boldsymbol{\beta}\boldsymbol{\beta}\ (p+1)\times(p+1)}=p\boldsymbol{K}'\boldsymbol{K}$, $\tilde{\boldsymbol{\psi}}_{\rho\boldsymbol{\beta}\ 1\times(p+1)}=2p(\mathbf{e}\eta_t-\boldsymbol{K}\boldsymbol{\beta})'\mathbf{W}\boldsymbol{K}$, and $\tilde{\boldsymbol{\psi}}_{\boldsymbol{\beta}\rho\ (p+1)\times1}=\mathbf{0}$. The inverse element of the information matrix in the first row and column is $\tilde{\psi}^{-1}_{\rho\rho_{1\times1}}=\left(\tilde{\psi}_{\rho\rho}-\tilde{\boldsymbol{\psi}}_{\rho\beta}\tilde{\boldsymbol{\psi}}^{-1}_{\beta\beta}\tilde{\boldsymbol{\psi}}_{\beta\rho}\right)^{-1}$ and when simplified, it becomes $\tilde{\boldsymbol{\psi}}^{-1}_{\rho\rho_{1\times1}}=p^{-1}\left[(\mathbf{e}\eta_t-\boldsymbol{K}\boldsymbol{\beta})'\mathbf{W}\mathbf{W}(\mathbf{e}\eta_t-\boldsymbol{K}\boldsymbol{\beta})\right]^{-1}$.

The statistic value of the LM test for SERM-SEM is $\mathrm{LM}_{\rho} = \hat{\mathbf{D}}_{\rho}^{'}\hat{\boldsymbol{\psi}}_{\rho\rho}^{-1}\hat{\mathbf{D}}_{\rho}$. The test is carried out based on $H_0$, leading to the value of $\hat{\mathbf{D}}_{\rho}$ representing the first derivative of the ln likelihood function to $\rho$ as in Equation (18) with $\rho = 0$, as follows: $\hat{D}_{\rho_{1\times1}} = p\left(\boldsymbol{l} - \boldsymbol{K}\hat{\boldsymbol{\beta}}\right)^{'}\mathbf{W}\left(\boldsymbol{l} - \boldsymbol{K}\hat{\boldsymbol{\beta}}\right)$. Since $\left(\boldsymbol{l} - \boldsymbol{K}\boldsymbol{\beta}\right)$ represents an error from the OLS regression, then $\left(\boldsymbol{l} - \boldsymbol{K}\hat{\boldsymbol{\beta}}\right) = \tilde{\boldsymbol{\varepsilon}}_{T\times1}$ and $\hat{D}_{\rho_{1\times1}} = p\tilde{\boldsymbol{\varepsilon}}^{'}\mathbf{W}\tilde{\boldsymbol{\varepsilon}}$. The statistic value of the LM test is expressed as $\mathrm{LM}_{\rho} = \frac{p\left(\tilde{\boldsymbol{\varepsilon}}^{'}\mathbf{W}\tilde{\boldsymbol{\varepsilon}}\right)^2}{\left(\mathbf{e}\hat{\eta}_t - \boldsymbol{K}\hat{\boldsymbol{\beta}}\right)^{'}\mathbf{W}\mathbf{W}\left(\mathbf{e}\hat{\eta}_t - \boldsymbol{K}\hat{\boldsymbol{\beta}}\right)}$. By assuming $\left(\mathbf{e}\hat{\eta}_t - \boldsymbol{K}\hat{\boldsymbol{\beta}}\right)^{'}\mathbf{W}\mathbf{W}\left(\mathbf{e}\hat{\eta}_t - \boldsymbol{K}\hat{\boldsymbol{\beta}}\right) = D_{1\times1}$, the statistic value of the LM test can be written as:

$$\mathrm{LM}_{\rho} = \frac{p\left(\tilde{\boldsymbol{\varepsilon}}^{'}\mathbf{W}\tilde{\boldsymbol{\varepsilon}}\right)^2}{D} \tag{23}$$

where, $\tilde{\boldsymbol{\varepsilon}} = \left(\boldsymbol{l} - \boldsymbol{K}\hat{\boldsymbol{\beta}}\right)$, $D = \left(\mathbf{e}\hat{\eta}_t - \boldsymbol{K}\hat{\boldsymbol{\beta}}\right)^{'}\mathbf{W}\mathbf{W}\left(\mathbf{e}\hat{\eta}_t - \boldsymbol{K}\hat{\boldsymbol{\beta}}\right)$, $\boldsymbol{l}$ is a vector of the endogenous factor score that has spatial dependence, $\boldsymbol{K}$ signifies the matrix of the exogenous factor score, $\boldsymbol{\beta}$ represents a vector of the parameter regression, $\mathbf{e}$ is a vector whose members are all 1, namely $\mathbf{e}_{T\times1} = \left(1,\ldots,1\right)^{'}$, $\boldsymbol{\eta}$ is a vector of endogenous random variables, and $\mathbf{W}$ is a spatial weighting matrix.

The statistic of LM test was defined by Breusch & Pagan (1980), as follows: $\mathrm{LM} = \tilde{\mathbf{D}}^{'}\tilde{\boldsymbol{\psi}}^{-1}\tilde{\mathbf{D}}$. When $H_0$: $\left[\lambda, \rho, \boldsymbol{\alpha}^{'}\right] = 0$, it imply that the p parameters from $\boldsymbol{\alpha}$ are related with non-constraints. Anselin (1988) established that $\mathrm{LM} = \tilde{\mathbf{D}}^{'}\tilde{\boldsymbol{\psi}}^{-1}\tilde{\mathbf{D}}$ followed a distribution of $\chi^2_{(2+p)}$, hence, the SERM-SEM model with $H_0$: $\rho = 0$ obtained $LM_{\rho} \sim \chi^2_{(1)}$ or $\frac{p\left(\tilde{\boldsymbol{\varepsilon}}^{'}\mathbf{W}\tilde{\boldsymbol{\varepsilon}}\right)^2}{D} \sim \chi^2_{(1)}$.

## 4. Discussion and Applying SERM-SEM on HDI Modeling

Figure 2 is the SEM model for HDI, which connects latent variables with their indicators and among latent variables. The relationship among latent variables must be derived from theory and logic (Hair et al., 2021). Hence, this study developed an HDI model based on three latent variables as generally measured by UNDP (2022) and BPS (2022a), namely Long and Healthy Life (LHL), Knowledge (Know_L), and Decent Living Standard (DLS). The latent variables must be built by indicators based on strong theoretical concepts. In consequence, in this study, each latent variable comprised multiple indicators drawn from previous studies, namely Darsyah et al. (2018), Niranjan (2020), Pramesti & Indrasetianingsih (2018), Rahma (2020), and Wati & Khikmah (2020). The Long and Healthy Life (LHL) is constructed by four indicators. The Knowledge (Know_L) is constructed by six indicators. The Decent Living Standard (DLS) is constructed by three indicators. The full HDI model can be seen in Figure 2.

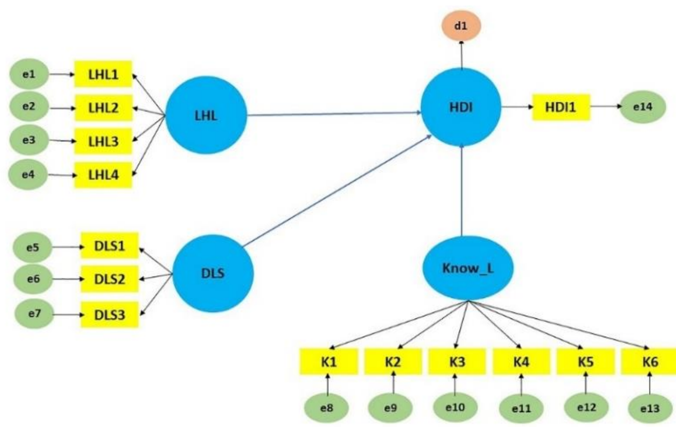


**Figure 2.** Full model of HDI.

The indicators, notations, sources for each dimension or variable are shown in Table 1.

**Table 1.** Variable, notation, and source for HDI model.

| Dimension/Latent Variable | Notation | Indicator | Notation | Source |
|---|---|---|---|---|
| Namely Long and Healthy Life | LHL | Life Expectancy | LHL1 | BPS (2023a) |
| | | The Percentage of Individuals with Health Insurance | LHL2 | BPS (2022b) |
| | | The Percentage of Access to Clean Drinking Water in Households | LHL3 | BPS (2023a) |
| | | The Percentage of Feasible Sanitation Access in Households | LHL4 | BPS (2023a) |
| Knowledge | Know_L | Mean Years of Schooling | K1 | BPS (2023a) |
| | | Expected Years of Schooling | K2 | BPS (2023a) |
| | | The Pure Participation Rate for High School | K3 | Kemendikbud-Ristek (2023) |
| | | The Percentage of Literacy Rate Among the Population Aged 15 And over | K4 | BPS (2022b, 2023a, 2023b), Kemendikbud-Ristek (2023) |
| | | The Percentage of The Population Aged 7-24 Years with Schooling Status | K5 | BPS (2022b) |
| | | The student-Teacher Ratio in High School | K6 | BPS (2023a) |
| Decent Living Standards | DLS | Average Expenditure Per Capita Per Month | DLS1 | BPS (2022b) |
| | | Percentage of The Non-Poor Population | DLS2 | BPS (2023a) |
| | | Average Net Salary of Employees Per Month | DLS3 | BPS (2023b) |
| Human Development Index | HDI | Human Development Index | HDI | BPS (2022b, 2023a, 2023b) Kemendikbud-Ristek (2023) |

**Table 2.** Code and regency/city of east Java province.

| Code | Regency/City | Code | Regency/City | Code | Regency/City | Code | Regency/City |
|---|---|---|---|---|---|---|---|
| 3501 | Pacitan | 3511 | Bondowoso | 3521 | Ngawi | 3572 | Blitar |
| 3502 | Ponorogo | 3512 | Situbondo | 3522 | Bojonegoro | 3573 | Malang |
| 3503 | Trenggalek | 3513 | Probolinggo | 3523 | Tuban | 3574 | Probolinggo |
| 3504 | Tulungagung | 3514 | Pasuruan | 3524 | Lamongan | 3575 | Pasuruan |
| 3505 | Blitar | 3515 | Sidoarjo | 3525 | Gresik | 3576 | Mojokerto |
| 3506 | Kediri | 3516 | Mojokerto | 3526 | Bangkalan | 3577 | Madiun |
| 3507 | Malang | 3517 | Jombang | 3527 | Sampang | 3578 | Surabaya |
| 3508 | Lumajang | 3518 | Nganjuk | 3528 | Pamekasan | 3579 | Batu |
| 3509 | Jember | 3519 | Madiun | 3529 | Sumenep | | |
| 3510 | Banyuwangi | 3520 | Magetan | 3571 | Kediri | | |

The data for this study were collected from the Central Bureau of Statistics of the Republic of Indonesia and the Ministry of Education, Culture, Research, and Technology. The sample is East Java province, which includes 38 regencies/cities, as shown in Table 2.

The assessment of the outer model in SEM, as shown in Figure 2, is based on the outer loading values and their significance. The evaluation of the outer model used Smart-PLS Software. This evaluation aimed to identify indicators that effectively explained or measured latent variables, with results summarized in Table 3.

**Table 3.** Loading factor and P-value for outer model.

| Indicator <- Latent Variable | Loading Factor | P -values | Updated Model |
|---|---|---|---|
| DLS1 <- DLS | 0.955 | 0.000 | Maintained |
| DLS2 <- DLS | 0.395 | 0.013 | Removed |
| DLS3 <- DLS | 0.932 | 0.000 | Maintained |
| K1 <- Know_L | 0.492 | 0.033 | Removed |
| K2 <- Know_L | 0.798 | 0.008 | Maintained |
| K3 <- Know_L | 0.868 | 0.000 | Maintained |
| K4 <- Know_L | 0.392 | 0.073 | Removed |
| K5 <- Know_L | 0.15 | 0.513 | Removed |
| K6 <- Know_L | 0.488 | 0.001 | Removed |
| LHL1 <- LHL | 0.161 | 0.566 | Removed |
| LHL2 <- LHL | 0.925 | 0.000 | Maintained |
| LHL3 <- LHL | 0.016 | 0.956 | Removed |
| LHL4 <- LHL | 0.74 | 0.001 | Maintained |

An indicator was considered valid if it exhibited a loading factor exceeding 0.50 (Hair et al., 2010). Strong convergent validity requires that every indicator under a latent variable have high loadings of greater than 0.6. (Dash & Paul, 2021; Hair et al., 2017; Shi & Maydeu-Olivares, 2020). Indicators with loading factor values below 0.60 are excluded from the model, resulting in the revised model shown in Figure 3. Figure 3 contains three latent variables, namely Long and Healthy Life (LHL), Knowledge (Know_L), and Decent Living Standard (DLS). The Percentage of Individuals with Health Insurance (LHL2) and The Percentage of Feasible Sanitation Access in Households (LHL4) are the two indicators that make up The Long and Healthy Life (LHL). Two indicators make up the Knowledge (Know_L): Expected Years of Schooling (K2) and The Pure Participation Rate for High School (K3). The two indicators that make up The Decent Living Standard (DLS) are Average Expenditure Per Capita Per Month (DLS1) and Average Net Salary of Employees Per Month (DLS3). The model employed next is the revised version shown in Figure 3.

Figure 4 shows the HDI levels in East Java Province, Indonesia. Sampang Regency has the lowest HDI index (63.39), while Surabaya City has the highest (82.74). HDI levels are categorized into four clusters: low, medium, high, and extremely high levels with progressively darker color gradations. Nine cities/regencies are part of the low HDI level cluster, eleven cities/regencies comprise the medium cluster, eleven cities/regencies are part of the high cluster, and seven cities/regencies are part of the very high cluster.

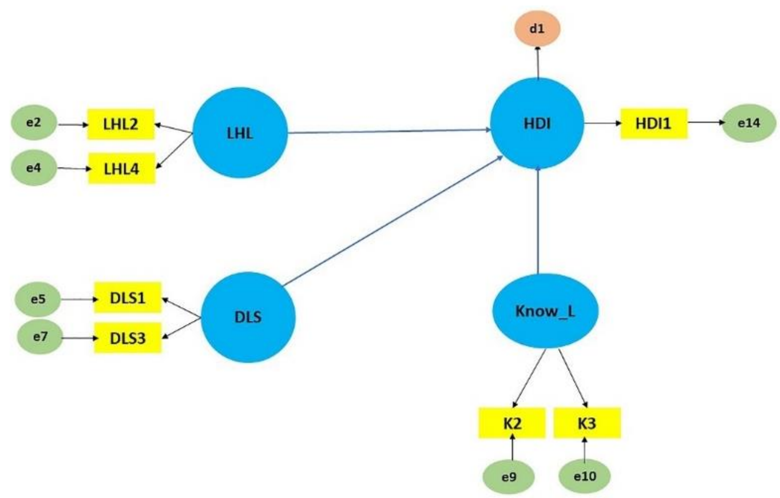


**Figure 3.** Updated model of HDI.

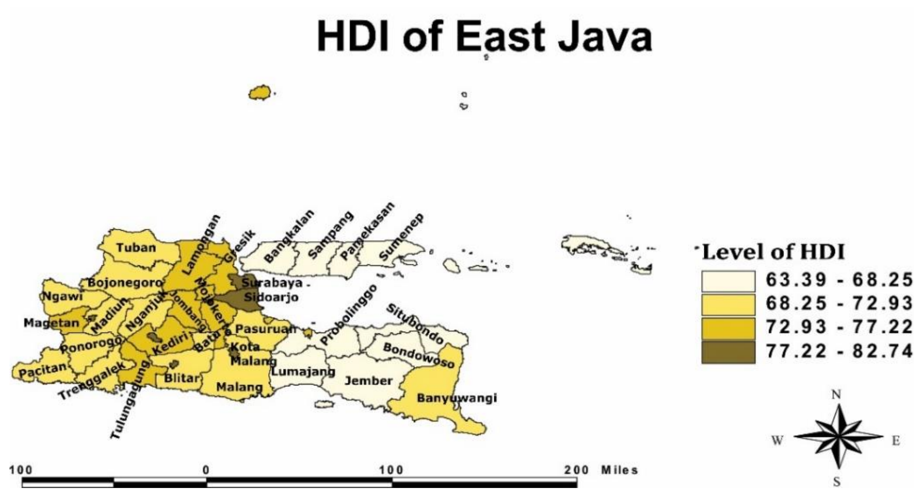


**Figure 4.** HDI Level in East Java province.

The next step is the estimation of exogenous and endogenous latent variables as in Equation (6) and Equation (8) using MATLAB Software. The SERM-SEM model employing queen contiguity weights is formulated using Equation (10). The spatial dependency test for SERM-SEM is conducted using the following Equation (23), and the calculation result is shown in Table 4. This value is compared to the chi-square distribution with one degree of freedom and an alpha value of 5%, leading to SERM-SEM. The testing process used MATLAB Software.

**Table 4.** Lagrange multiplier test result.

| Test | $LM_\rho$ | Result |
|---|---|---|
| LM (SERM-SEM) | 4,61E+24 | reject the null hypothesis |

The estimation parameters of SERM-SEM, as in Equation (12) and Equation (16) using MATLAB Software, can be seen in Table 5. The values are compared to the t-distribution with thirty-seven degrees of freedom and an alpha value of 5%, resulting in significant results, except for The Long and Healthy Life (LHL) variable. Decent Living Standard (DLS) has a significant effect on HDI. The findings of this study corroborate those of Arriani and Chotib (2021), which found that the variables of poverty, GDRP per capita, and unemployment had a significant impact on HDI and spatial effects. Combining variables of poverty, GDRP per capita, and unemployment describes a Decent Living Standard (DLS). Knowledge (Know_L) has a significant effect on HDI. Knowledge has a significant effect on HDI. This result corroborates the findings of Wafitrah et al. (2023), which found that an educated workforce influenced HDI but that government spending on education did not. A Long and Healthy Life (LHL), as represented by The Percentage of Individuals with Health Insurance (LHL2) and The Percentage of Feasible Sanitation Access in Households (LHL4), has a significant no effect on HDI. For a number of cases in developing countries, like Indonesia, LHL does not support HDI. Obstacles to health insurance (LHL2) exist in developing countries, including administrative issues in Indonesia (Banerjee et al., 2021) and tax issues in a number of developing nations (Jamal et al., 2022). Watson et al. (2021) conclude that health insurance is not a magic bullet for enhancing equity in the health sector, based on publications spanning 20 developing countries. Even though the availability of sanitation facilities might improve a decent life, there were challenges to Feasible Sanitation Access (LHL4) in developing countries, particularly in Indonesia (Budiono & Purba, 2022).

**Table 5.** The estimation results of parameter and spatial error-SEM coefficient.

| Variable | Coefficient | Result |
|---|---|---|
| Spatial Error-SEM - $\rho$ | -0.387 | Valid |
| $\beta_0$ | 8.002 | Valid |
| Decent Living Standards (DLS) - $\beta_1$ | 0.226 | Valid |
| Knowledge (Know_L) - $\beta_2$ | 0.053 | Valid |
| Long and Healthy Life (LHL) - $\beta_3$ | 0.001 | Invalid/No Effect |

In general, HDI modeling in SERM-SEM can be expressed as follows:

$$HDI_i = 8.002 + 0.226DLS_i + 0.053Know\_L_i + 0.001LHL_i - 0.387\sum_{j=1, i\neq j}^{38} W_{ij}u_j \ .$$

HDI model for several regencies/cities are as follow:

(i) Pemekasan Regency has two neighbors, namely Sampang and Sumenep Regency

$$HDI_{[3528]} = 8.002 + 0.226DLS_{[3528]} + 0.053Know\_L_{[3528]} + 0.001LHL_{[3528]} - 0.387\left(u_{[3527]} + u_{[3529]}\right)$$

(ii)Bangkalan Regency has one neighbor, namely Sampang Regency

$$HDI_{[3526]} = 8.002 + 0.226DLS_{[3528]} + 0.053Know\_L_{[3526]} + 0.001LHL_{[3526]} - 0.387u_{[3527]}$$

(iii) Kediri Regency has five neighbours, namely, Kediri city, Jombang, Tulungagung, Blitar, and Nganjuk

$$HDI_{[3506]} = 8.002 + 0.226DLS_{[3528]} + 0.053Know\_L_{[3506]} + 0.001LHL_{[3506]} + \\ - 0.387\left(u_{[3504]} + u_{[3505]} + u_{[3517]} + u_{[3518]} + u_{[3571]}\right)$$

In Figure 4, the HDI level is classified into four levels: very high, high, medium, and low. Visually, a regency/city with a very high HDI category has neighboring regencies/cities with very high or high HDI categories. On the other hand, a regency/city with a low HDI category has neighboring regencies/cities with low or medium HDI categories. It aligns with the SERM-SEM model, in which the model error value of each adjacent neighbor influences each model in a regency /city. Therefore, each region/city has a close relationship that mutually influences neighboring regencies/cities intersecting at a point or line.

The spatial error coefficient is negative. It suggested the reverse of the typical spillover effect (Anekawati et al., 2020). A regency/city gets spillover spatial effects (not giving) from neighboring regencies/cities. For instance, as shown in Figure 4, the City of Mojokerto with a high HDI level gets an overflow of HDI effects from the regencies /cities that intersect with it, namely Sidoarjo (very high-HDI level), Gresik (high-HDI level), Lamongan (high-HDI level), Jombang (high-HDI level), Batu (high-HDI level), and Pasuruan (medium-HDI level). Therefore, this finding encourages collaborative action among regencies /cities to improve the HDI level. It supports Arriani and Chotib (2021) conclusion that cooperative action was needed to improve the HDI of Central Java's regions and cities and strengthen the SDGs' implementation.

## 5. Conclusion

In conclusion, this study comprehensively developed the SERM-SEM, namely a spatial model in the form of the Spatial Error Model (SERM), which involved analyzing the relationship among latent variables (SEM). This model was developed to address the limitations of earlier studies by developing HDI modeling based on indicators rather than dimensions or latent variables. In the SERM-SEM, the variables in the SERM were substituted by the factor scores that the latent variable estimation resulted from SEM. These latent variables were estimated using the weighted least squares (WLS) method and subsequently integrated into the spatial framework. The SERM-SEM gave a new perspective by using the model error distribution, distinct from the error distribution of the traditional spatial model. The error distribution was formulated based on the value of expectation and variance from the SERM-SEM model. As a result, the error distribution of SERM-SEM was $\boldsymbol{\varepsilon} \sim N_{T,1}\left((\mathbf{I}-\rho\mathbf{W})(\mathbf{e}\boldsymbol{\eta}-\mathrm{K}\boldsymbol{\beta}), \boldsymbol{\Theta}_{\rho}\right)$. Parameter estimation in this context employed the generalized method of moments (GMM), which yielded parameter estimations of $\hat{\boldsymbol{\delta}}_{(p+2)\times 1}=\left(\hat{\boldsymbol{\beta}}' \mid \hat{\lambda}\right)'=\left(\hat{\mathbf{Z}}'\hat{\mathbf{Z}}\right)^{-1}\hat{\mathbf{Z}}'\boldsymbol{l}$ and $\left[\rho \quad \rho^2 \quad tr\left(\boldsymbol{\Theta}_{\rho}\right)\right]'_{1\times 3}=\left(\mathbf{G}'\mathbf{G}\right)^{-1}\mathbf{G}'\mathbf{g}$. As a consequence, the SERM model was formed from the result of estimating latent variables in SEM, and a new error distribution was obtained. It is noteworthy that the LM test necessitates knowledge of the model error distribution. In this study, the spatial dependency test for the SERM-SEM used the new error distribution. The development of the spatial dependency test using LM for SERM-SEM with $H_0$: $\rho=0$ led to $\frac{p\left(\tilde{\boldsymbol{\varepsilon}}'\mathbf{W}\tilde{\boldsymbol{\varepsilon}}\right)^2}{D} \sim \chi^2_{(1)}$.

The results of developing mathematical formulas for the SERM-SEM model were applied to HDI modeling. This study developed an HDI model based on three latent dimensions, namely long and healthy life (LHL), knowledge (Know_L), and decent living standards (DLS). Each latent variable comprised multiple indicators drawn from previous studies. The assessment of the outer model in SEM was based on the outer loading values and their significance. This evaluation aimed to identify indicators that effectively explained or measured latent variables, so it got the revised model in SEM. The revised SEM model was a model that had latent variables LHL, DLS, and Know_L, each of which had two valid and significant indicators. The revised SEM model was modeled into the SERM model using steps of the results of developing the SERM-SEM model, namely estimating latent variables, constructing the SERM-SEM model, estimating model parameters, and testing spatial dependency. The HDI modeling produced the SERM-SEM model denoted as $HDI_i = 8.002 + 0.226DLS_i + 0.053Know\_L_i + 0.001LHL_i - 0.387\sum_{j=1, i\neq j}^{38} W_{ij}u_j$ .

## 6. Limitation and Recommendation

The limitation of this study is that a significant test has yet to be developed simultaneously. Recommendations for future research are to develop a significance test simultaneously using the maximum likelihood ratio test (MLRT). This study estimated latent variables using GMM (parametric statistics). Another recommendation for future research is that the model parameters estimation employs the maximum likelihood test (MLE) to be in sync with the latent variable estimation method, both of which use parametric statistics.

### Conflict of Interest

The authors declare no conflict of interest.

### Acknowledgments

The authors would like to thank the Ministry of Education, Culture, Research, and Technology of Indonesia (Kemendikbud-Ristek) for providing a research grant to carry out the scheme PDKN research with grant number 077/E5/PG.02.00.PL/2023; 025/SP2H/PT-L/LL7/2023; 001/LPPM/PP-04/UNIJA/V/2023.

## References

Ahmed, N., Hussain, B., & Bhatti, A.A. (2022). Impacts of fiscal decentralization on spatial disparities across districts of Punjab, Pakistan. *Journal of Business & Economics*, *14*(1), 31-43.

Anekawati, A., & Otok, B.W. (2017, September). Modelling of the education quality of a high schools in Sumenep regency using spatial structural equation modelling. In *Journal of Physics: Conference Series* (Vol. 890, No. 1, p. 012094). IOP Publishing. Kuantan, Pahang, Malaysia. https://doi.org/10.1088/1742-6596/890/1/012094.

Anekawati, A., Otok, B.W., Purhadi, & Sutikno. (2020). Exploring the related factors in education quality through spatial autoregressive modeling with latent variables: A rural case study. *Education Research International*, *2020*, 1-10. https://doi.org/10.1155/2020/8823186.

Anselin, L. (1988). *Spatial econometrics: methods and models* (Vol. 4). Kluwer Academic Publisher. http://link.springer.com/10.1007/978-94-015-7799-1.

Anwar, R., Djuraidah, A., & Wigena, A.H. (2020). Comparison of maximum likelihood and generalized method of moments in spatial autoregressive model with heteroskedasticity. *Proceedings of the 1st International Conference on Statistics and Analytics*. *Bogor, Indonesia*. https://doi.org/10.4108/eai.2-8-2019.2290489.

Arriani, R.R. & Chotib. (2021). The correlation of SDG 1 and 8 and spatial effect of human development index in central java. *IOP Conference Series: Earth and Environmental Science*, *940*(1), 012063. https://doi.org/10.1088/1755-1315/940/1/012063.

Banerjee, A., Finkelstein, A., Hanna, R., Olken, B.A., Ornaghi, A., & Sumarto, S. (2021). The challenges of universal health insurance in developing countries: experimental evidence from Indonesia's national health insurance. *American Economic Review*, *111*(9), 3035-3063. https://doi.org/10.1257/aer.20200523.

Bollen, K.A. (1989). *Structural equations with latent variables*. John Wiley & Son. New York.

BPS. (2022a). *Indeks Pembangunan manusia provinsi Jawa Timur 2022*. BPS Provinsi Jawa Timur.

BPS. (2022b). *Welfare statistics of Jawa Timur province*, *2022*. BPS-Statistics of Jawa Timur Province.

BPS. (2023a). *Jawa Timur province in figures 2023*. BPS-Statistics of Jawa Timur Province.

BPS. (2023b). *labor force situation in Jawa Timur province august 2022*. BPS-Statistics of Jawa Timur Province.

Breusch, T.S., & Pagan, A.R. (1980). The Lagrange multiplier test and its applications to model specification in econometrics. *The Review of Economic Studies*, *47*(1), 239-253. https://doi.org/10.2307/2297111.

Budiono, S., & Purba, J.T. (2022). Reducing poverty strategy through educational participation, clean water, and sanitation in Indonesia. *Jurnal Ekonomi Dan Bisnis*, *25*(1), 177-198. https://doi.org/10.24914/jeb.v25i1.4158.

Christensen, W.F., & Amemiya, Y. (2002). Latent variable analysis of multivariate spatial data. *Journal of the American Statistical Association*, *97*(457), 302-317. https://doi.org/10.1198/016214502753479437.

Congdon, P. (2008). A spatial structural equation model for health outcomes. *Journal of Statistical Planning and Inference*, *138*(7), 2090-2105. https://doi.org/10.1016/j.jspi.2007.09.001.

Darsyah, M.Y., Suprayitno, I.J., Otok, B.W., & Ulama, B.S. (2018). Spatial modeling for human development index in central Java. *South East Asia Journal of Contemporary Business, Economics and Law*, *16*(5), 36-41.

Dash, G., & Paul, J. (2021). CB-SEM vs PLS-SEM methods for research in social sciences and technology forecasting. *Technological Forecasting and Social Change*, *173*, 121092. https://doi.org/10.1016/j.techfore.2021.121092.

Hair, J.F., Babin, B.J., & Krey, N. (2017). Covariance-based structural equation modeling in the *journal of advertising*: Review and recommendations. *Journal of Advertising*, *46*(1), 163-177. https://doi.org/10.1080/00913367.2017.1281777.

Hair, J.F., Black, W.C., Babin, B.J., & Anderson, R.E. (2010). *Multivariate data analysis*. Pearson Prentice Hall. New York.

Hair, J.F., Hult, G.T.M., Ringle, C.M., Sarstedt, M., Danks, N.P., & Ray, S. (2021). *Partial least squares structural equation modeling (pls-sem) using r: A workbook*. Springer International Publishing. Switzerland. https://doi.org/10.1007/978-3-030-80519-7.

Hogan, J.W., & Tchernis, R. (2004). Bayesian factor analysis for spatially correlated data, with application to summarizing area-level material deprivation from census data. *Journal of the American Statistical Association*, *99*(466), 314-324. https://doi.org/10.1198/016214504000000296.

Jamal, M.H., Abdul, A.F.A., Aizuddin, A.N., & Aljunid, S.M. (2022). Successes and obstacles in implementing social health insurance in developing and middle-income countries: A scoping review of 5-year recent literatures. *Frontiers in Public Health*, *10*, 918188. https://doi.org/10.3389/fpubh.2022.918188.

Januardi, R.W., & Utomo, A.P. (2017). Application of spatial error model using gmm estimation in impact of education on poverty alleviation in Java, Indonesia. *Communications in Science and Technology*, *2*(2), 53-63.

Jeong, S., & Yoon, D. (2018). Examining Vulnerability factors to natural disasters with a spatial autoregressive model: The case of South Korea. *Sustainability*, *10*(5), 1651. https://doi.org/10.3390/su10051651.

Joreskog, K.G., & Sorbom, D. (1993). *LISREL 8: Structural equation modelling with the SIMPLIS command language*. Scientific Software International Inc.

Joreskog, K.G., & Sorbom, D. (1996). *Lisrel 8: User's reference guide*. Scientific Software International Inc.

Kelejian, H.H., & 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*, *17*(1), 99-121.

Kelejian, H.H., & Prucha, I.R. (1999). A generalized moments estimator for the autoregressive parameter in a spatial model. *International Economic Review*, *40*(2), 509-533. https://doi.org/10.1111/1468-2354.00027.

Kemendikbud-Ristek. (2023). *APK/APM (Angka Partisipasi Kasar / Angka Partisipasi Murni) Tahun 2022/2023*. Pusat Data dan Teknologi Informasi, Kemendikbud, Ristek.

Li, J., Sawaragi, T., & Horiguchi, Y. (2021). Introduce structural equation modelling to machine learning problems for building an explainable and persuasive model. *SICE Journal of Control, Measurement, and System Integration*, *14*(2), 67-79. https://doi.org/10.1080/18824889.2021.1894040.

Liu, X., Wall, M.M., & Hodges, J.S. (2005). Generalized spatial structural equation models. *Biostatistics*, *6*(4), 539-557. https://doi.org/10.1093/biostatistics/kxi026.

Niranjan, R. (2020). Spatial inequality in human development in India- A case study of Karnataka. *Sustainable Futures*, *2*, 100024. https://doi.org/10.1016/j.sftr.2020.100024.

Oud, J.H.L., & Folmer, H. (2008). A structural equation approach to models with spatial dependence. *Geographical Analysis*, *40*(2), 152-166. https://doi.org/10.1111/j.1538-4632.2008.00717.x.

Pramesti, W., & Indrasetianingsih, A. (2018). East Java human development index modeling with spatial regression approach. In *1st International Conference on Social Sciences* (pp. 1494-1498). Atlantis Press. Bali, Indonesia. https://doi.org/10.2991/icss-18.2018.312.

Rahma, A. (2020). Human development index modelling in Indonesia using spatial error model approach. In *ICBAE 2020: Proceedings of the 2nd International Conference of Business, Accounting and Economics, Purwokerto, Indonesia* (p. 472). European Alliance for Innovation. https://doi.org/10.4108/eai.5-8-2020.2301175.

Saputro, D.R.S., Muhsinin, R.Y., Widyaningsih, P., & Sulistyaningsih. (2019). Spatial Autoregressive with a spatial autoregressive error term model and its parameter estimation with two-stage generalized spatial least square procedure. *Journal of Physics: Conference Series*, *1217*(1), 012104. https://doi.org/10.1088/1742-6596/1217/1/012104.

Shi, D., & Maydeu-Olivares, A. (2020). The effect of estimation methods on SEM fit indices. *Educational and Psychological Measurement*, *80*(3), 421-445. https://doi.org/10.1177/0013164419885164.

Trujillo, G.S. (2009). *Pathmox approach: Segmentation trees in partial least squares path modeling*. Universitat Politecnica de Catalunya.

UNDP. (2022). *Human development report 2021/2022: Uncertain times, unsettled lives shaping our future in a transforming world*. United Nations Development Programme.

Wafitrah, S.R., Rahmatia, & Saudi, N.D.S. (2023). Analysis of the human development index and factors affecting it in Eastern Indonesia. *Jurnal Ekonomi*, *12*(1), 1486-1493.

Wang, F., & Wall, M.M. (2003). Generalized common spatial factor model. *Biostatistics*, *4*(4), 569-582. https://doi.org/10.1093/biostatistics/4.4.569.

Wati, A.D.A., & Khikmah, L. (2020). Modeling Spatial error model (SEM) on human development index (IPM) In central Java 2018. *Journal of Intelligent Computing and Health Informatics*, *1*(2), 48. https://doi.org/10.26714/jichi.v1i2.6341.

Watson, J., Yazbeck, A.S., & Hartel, L. (2021). Making health insurance pro-poor: Lessons from 20 developing countries. *Health Systems & Reform*, *7*(2), e1917092. https://doi.org/10.1080/23288604.2021.1917092.