Media Statistika

p-ISSN 1979 – 3693 e-ISSN 2477 – 0647
MEDIA STATISTIKA 17(2) 2024: 209-220
http://ejournal.undip.ac.id/index.php/media_statistika

ANALYZING SOCIO-ECONOMIC RECOVERY ON SUMATRA ISLAND POST-COVID-19: A SPATIAL DURBIN MODEL APPROACH

Ibrah Hasanah Lubis, Saiful Mahdi, Munawar Munawar
Department of Statistics, Syiah Kuala University, Banda Aceh, Indonesia

e-mail: munawar@usk.ac.id

DOI: 10.14710/medstat.17.2.209-220

Article Info:
Received: 8 September 2024
Accepted: 21 August 2025
Available Online: 10 October 2025

Keywords:
Socio-Economic; Spatial Durbin Model; Google Mobility, COVID-19; Panel Regression
.

Abstract: The COVID-19 outbreak was designated as a public health emergency that disturbed the world from January 2020 to May 2023 by the World Health Organization. This outbreak has drastically changed the order of socio-economic life. According to data Gross Domestic Product, it was recorded to grow by 5.03% in 2023 according to data from the Central Statistics Agency, which is still slightly below the pre-pandemic level of 5.17% in 2018. At the regional level, only 6 provinces experienced a higher Gross Regional Domestic Product growth rate in 2023 compared to 2018. These figures highlight the need for recovery efforts to be made to restore the condition of the community and the environment so that the socio-economic activities of the community can run well again. This study uses Google mobility report data and panel data spatial regression analysis to determine the factors that influence socio-economic recovery on the island of Sumatra and how the influence between regions in the recovery effort. The data used is panel data for 273 observation days in eight provinces. By integrating spatial panel data methods with mobility-based proxies, this approach offers a new framework that is rarely applied in studies of post-COVID-19 socio-economic recovery in Sumatra.

1. INTRODUCTION

Coronavirus Disease 2019 (COVID-19) is an outbreak that was initially detected in Wuhan, Hubei, China, in December 2019 and was designated as a pandemic by the World Health Organization (WHO) on March 11, 2020 (Bland et al., 2024). The COVID-19 pandemic is a global pandemic that has spread throughout the world, including Indonesia (Setiadi et al., 2022). Since the declaration of COVID-19 as a pandemic, the entire socio-economic structure in Indonesia has undergone drastic changes (Torres-Favela & Luna, 2024). All social activities have been restricted to suppress the rate of spread of COVID-19 in Indonesia, and one of the restricted activities is population movement or mobility (Maloney et al., 2023).

The government's efforts had significant negative impacts on people's socio-economic activities (Utomo & Nugroho, 2022). The consequences were profound—people were confined to their homes, forced to work and study remotely, and many businesses had to close or lay off employees (Mahendra et al., 2021). Moreover, data from the Central Statistics Agency (BPS) shows that Indonesia's Gross Domestic Product (GDP) growth rate dropped sharply to -2.07% in 2020. Although it recovered to 5.05% by the end of the pandemic in 2023, this figure remains slightly below the pre-pandemic level of 5.17% in 2018. In addition, only three provinces—North Maluku, Central Sulawesi, and Papua—avoided negative GRDP growth in 2020 when the pandemic began. While all provinces experienced positive GRDP growth by 2023 compared to 2020, the majority had not yet reached pre-pandemic levels. Specifically on the island of Sumatra, only two provinces—Riau and the Riau Islands—recorded higher GRDP growth in 2023 compared to 2018.

The limited understanding of spatial factors influencing socio-economic recovery in Sumatra hinders the development of targeted post-pandemic recovery policies. Based on this background, researchers are trying to identify and explore factors that can restore the community's socio-economy by including spatial effects in their analysis. The use of spatial indicators helps measure recovery progress in disaster-affected areas, and common patterns of social recovery can be identified through cross-comparative analysis using GPS datasets (Romanillos et al., 2021). The findings aim to provide valuable insights for policymakers in designing targeted and data-driven strategies to support future socio-economic recovery after the crisis.

The analysis employs spatial panel data regression, which includes the Spatial Autoregressive Model (SAR), the Spatial Error Model (SEM), and the Spatial Durbin Model (SDM). SAR models assume that the dependent variable in one region is influenced by the dependent variables in the adjacent regions (Saputro et al., 2018). SEM models capture spatial dependence in error terms, suggesting that unobserved factors affecting one region may also affect neighbouring regions (Nurjanah et al., 2023). SDM extends SAR by incorporating both direct and indirect effects, allowing for a more comprehensive analysis of spatial spillovers in both the dependent and independent variables (Hasiru et al., 2022). The criteria for choosing the best model using the highest adjusted R^2 and the lowest Akaike Information Criterion (AIC) values.

2. LITERATURE REVIEW

2.1. Spatial Panel Regression

Spatial regression is an advancement of classical linear regression models that incorporates the influence of location or space on the data being analyzed. The spatial regression method accounts for the dependence or influence of each unit of observation that is close to one another. Thus, the nearest neighbors of each unit are determined to assess their spatial effects. In spatial regression, a spatial dependence coefficient is used to measure the strength of the dependence between one spatial unit and its neighbouring areas (Astuti et al., 2020).

The panel data spatial regression model includes autoregressive components that occur at both lags and spatial errors. When there is a spatial autoregressive process at the lag, the model is referred to as the SAR. In this model, the dependent variable in one spatial unit depends on the dependent variable in neighboring spatial units. The equation for the spatial autoregressive model is found in Equation (1).

$$\boldsymbol{y}_{it} = \boldsymbol{\rho} \sum_{\boldsymbol{j=1}}^{\boldsymbol{N}} \boldsymbol{W}_{ij}\boldsymbol{y}_{jt} + \boldsymbol{\beta x'}_{it} + \boldsymbol{\mu}_i + \boldsymbol{e}_{it} \quad (1)$$

The indices i and j are used to distinguish between different locations or spatial units.

The SEM is a model in which individual spatial unit errors are correlated with errors in neighbouring spatial units. The SEM is found in Equation (2) and (3) (Elhorst, 2014):

$$\boldsymbol{y}_{it} = \boldsymbol{x}_{it}\boldsymbol{\beta} + \boldsymbol{\mu}_i + \boldsymbol{u}_{it} \quad (2)$$

$$\boldsymbol{u}_{it} = \boldsymbol{\lambda} \sum_{j=1}^{N} \boldsymbol{W}_{ij} \boldsymbol{u}_{jt} + \boldsymbol{e}_{it} \tag{3}$$

The variable t represents the temporal dimension in panel data, used to capture spatial dependence and autocorrelation over time. Next, the SAR model is extended to the SDM, which considers spatial lag dependence in both the dependent and independent variables. The equation for the SDM model is found in Equation (4):

$$\boldsymbol{y}_{it} = \boldsymbol{\rho} \sum_{j=1}^{N} \boldsymbol{W}_{ij} \boldsymbol{y}_{jt} + \boldsymbol{\beta} \boldsymbol{x}'_{it} + \boldsymbol{W} \boldsymbol{X}_t \boldsymbol{\theta} + \boldsymbol{e}_{it} \tag{4}$$

2.2. Spatial Weighting Matrix

Spatial regression is an advancement of classical linear regression models that incorporates location or spatial effects on the analyzed data. It requires an autocorrelation approach. To measure spatial autocorrelation, a spatial weighting matrix is needed to establish a spatial relationship between individual observation units, denoted by $\boldsymbol{W}$ (Gao, 2021). The structure of the spatial weight matrix can be seen in Equation (5) (O'Sullivan & Unwin, 2010):

$$\boldsymbol{W} = \begin{bmatrix} w_{11} & w_{12} & \dots & w_{1j} \\ w_{21} & w_{22} & \dots & w_{2j} \\ \vdots & \vdots & \ddots & \vdots \\ w_{i1} & w_{i2} & \dots & w_{ij} \end{bmatrix} \tag{5}$$

where $\boldsymbol{W}$ and w_{ij} are the spatial weighted matrix on the i observation with other j observations. The spatial weight matrix in Equation (5) contains values of 0 and 1, with 0 representing the absence of nearest neighbors and 1 representing their presence for the observed cross-sectional unit.

2.3. Assumption Test

2.3.1. Spatial Autocorrelation Test

Spatial autocorrelation can be tested using Moran's Index. Moran's Index assesses spatial autocorrelation by considering the interactions between observations at adjacent spatial locations, using a spatial weighting matrix. The equation for Moran's Index is as follows (Gao, 2021):

$$I = \frac{\sum_{i=1}^{n} \sum_{j=1}^{n} w_{ij} (x_i - \bar{x})(x_j - \bar{x})}{s^2 \sum_{i=1}^{n} \sum_{j=1}^{n} w_{ij}} \tag{6}$$

with $s^2 = \frac{\sum_{i=1}^{n} (x_i - \bar{x})^2}{n}$ and $\bar{x} = \frac{\sum_{i=1}^{n} x_i}{n}$. The I value ranges from [-1, 1], with a higher Moran's Index indicating a stronger spatial correlation. Spatial patterns can be determined using the following criteria based on Moran's Index: if $I > 0$, there is a positive spatial correlation; if $I < 0$, there is a negative spatial correlation; and if $I = 0$, there is no spatial correlation.

2.3.2. Spatial Dependence Test

The appropriate spatial panel model can be selected by testing for spatial dependencies using the Lagrange Multiplier (LM) test, which is applied to both spatial lags and errors. There is also an advanced version called the Robust Lagrange Multiplier. The test statistics used for the LM test on spatial lags are presented in Equation (7):

$$LM_{lag} = \frac{[\boldsymbol{e}^{\boldsymbol{T}} (\boldsymbol{I}_{\boldsymbol{T}} \otimes \boldsymbol{W}) \boldsymbol{Y} / \hat{\sigma}^2]^2}{\boldsymbol{J}} \tag{7}$$

with $\boldsymbol{J}$ in Equation (7) obtained by:

$$\boldsymbol{J} = \left[\left((\boldsymbol{I}_{\boldsymbol{T}} \otimes \boldsymbol{W}) \boldsymbol{X} \widehat{\boldsymbol{\beta}} \right)^{\boldsymbol{T}} (\boldsymbol{I}_{\boldsymbol{NT}} - \boldsymbol{X} (\boldsymbol{X}^{\boldsymbol{T}} \boldsymbol{X})^{-1} \boldsymbol{X}^{\boldsymbol{T}}) (\boldsymbol{I}_{\boldsymbol{T}} \otimes \boldsymbol{W}) \boldsymbol{X} \widehat{\boldsymbol{\beta}} + \boldsymbol{T} \boldsymbol{T}_{\boldsymbol{W}} \hat{\sigma}^2 \right] \tag{8}$$

and $\quad T_W = tr(\boldsymbol{WW} + \boldsymbol{W^T W})$ (9)

$\boldsymbol{e}$ is the residual vector from a pooled regression model without spatial or time specification effects or from a panel data model with spatial or time period fixed effects, $\boldsymbol{I}_{NT}$ is the identity matrix of dimension $N \times T$ where T is the number of time periods and N is the number of cross-sectional units, $\boldsymbol{X}$ is a matrix of explanatory variables and $\hat{\boldsymbol{\sigma}}$ estimated variance of residuals.

The test statistics used for the LM test on spatial errors are in Equation (10):

$$LM_{error} = \frac{[\boldsymbol{e^T}(\boldsymbol{I_T} \otimes \boldsymbol{W})\boldsymbol{e}/\hat{\sigma}^2)]^2}{\boldsymbol{T} \times \boldsymbol{T_W}} \tag{10}$$

If the value of $LM_{error} > \chi^2$ or the *p-value* $< \alpha$, it indicates a spatial dependency effect in the errors and the appropriate model to use includes spatial errors. The same applies to the LM lag. Additionally, each Spatial Autoregressive Model (SAR), Spatial Error Model (SEM), and Spatial Durbin Model (SDM) model has a corresponding panel model, resulting in common effect, fixed effect, and random effect SAR, SEM, and SDM.

2.4. Significance Test and Goodness of Fit

The significance test is one of the most important stages in a study. The significance test is used to determine whether the hypothesis made at the beginning of the study will be accepted or rejected. The test statistic used is the *Z*-test (Hidayati et al., 2024).

The Goodness of Fit test is a statistical criterion used in spatial panel data models to evaluate the adequacy of the model in explaining the observed data. Two commonly used measures in this context are the Adjusted R^2 and the Akaike Information Criterion (AIC). The Adjusted R^2 indicates how well the model explains the variation in the dependent variable, with values ranging from 0 to 1. The closer the value is to 1, the better the model fits the data.

Meanwhile, the AIC aims to achieve a balance between the model's explanatory power and its complexity. A lower AIC value suggests a better model, as it indicates a good fit with fewer estimated parameters. The AIC can be calculated using Equation (11) (Cavanaugh & Neath, 2019):

$$AIC = -N\ log\left(\frac{RSS}{N}\right) + 2K \tag{11}$$

Here N is the number of observations, RSS is the residual sum of squares where is the sum of the squared differences between observed and predicted values, and K is several of estimated parameters.

2.5. Related Works

Spatial regression may be used in the case of socio-economic recovery. Yang et al. (2021) using the Dynamic Spatial Lag Model (DSLM) indicate that the number of populations, government intervention, the scale of primary, secondary and tertiary industries, road density, number of community health services, number of medical beds, number of new cases have a significant effect on socio-economic recovery in China. Wang et al. (2021) using the Spatial Durbin Model (SDM) indicate that Information and communication technology (ICT) has a significant effect on socio-economic development in focal provinces but harms development in adjacent provinces in China.

Khotiwan et al. (2023), using Exploratory Spatial Data Analysis (ESDA) and the Spatial Durbin Model (SDM) indicate initial per capita income, physical capital investment, road infrastructure, population growth, and education have significant direct effects on economic growth and furthermore initial per capita income and education also have significant spillover effects on the inter-regional economic growth. Astuti et al. (2024) develop a spatial panel model to address the impact of inflation on national economic stability and explore the inter-provincial relationships that influence inflation dynamics across expenditure groups. Using two models: The Spatial Autoregressive Model with Random Effects (SAR-RE) before the crisis and the Spatial Error Model with Random Effects (SEM-RE), the results reveal significant spatial interdependence in provincial inflation data, underscoring the role of spatial factors in economic analysis.

Although research utilizing spatial panel data analysis has been widely conducted, studies focusing on Sumatra Island at the provincial level remain limited. Moreover, while the combination of spatial panel data and socio-economic analysis has been commonly explored, research specifically addressing socio-economic recovery after disasters—particularly COVID-19—in the context of Sumatra is still scarce. This study aims to fill that gap and serve as a reference for policymakers in formulating more effective strategies for socio-economic recovery.

3. MATERIAL AND METHOD

3.1. Data

The research used secondary data from several sources, including the 2020 Provincial Publication in Figures by the Indonesian Central Bureau of Statistics, the Ministry of Health website, the Indonesian Internet Service Providers Association (APJII) internet survey reports, and the Google Mobility Reports website. This study uses 14 explanatory variables and 1 dependent variable, namely the social recovery index (Table 1). Population mobility recovery is used as a proxy for social recovery, and the Social Recovery Index (SRI) is constructed by dividing the mobility intensity in 2022 by the mobility intensity in 2021, as shown in Equation (12):

$$SRI = \Delta Y_t = \frac{Intensity\ of\ Mobility_{\ t}}{Intensity\ of\ Mobility_{\ t-1}} \tag{12}$$

where ΔY_t is defined as the ratio between the mobility intensity at time t and that time $t-1$, reflecting the relative change in mobility.

The research will focus on recovery efforts and use comparative data on Sumatra Island for the explanatory variables to demonstrate the impact of recovery. The explanatory variables will use 2019 data, while the dependent variable will use 2022 data compared to 2021. The data used is panel data, with the cross-sectional unit consisting of eight observation units (excluding the island provinces of Bangka Belitung and Riau Islands due to limitations in the construction of the spatial weighting matrix) and the time series unit covering 273 days, resulting in a total of 2,184 observations. The Social Recovery Index (SRI) data used in this study are daily observations from January 1, 2022—when Indonesia had reached a state of herd immunity (with 70% of the population having received the first dose of the vaccine)—until September 30, 2022, which marks the latest available data. These data are compared to corresponding daily data from 2021.

Missing values were imputed using the median. After identifying numerous missing values, an outlier detection test is conducted to determine the most appropriate simple imputation method—choosing among the mean, median, or mode. Mean imputation is favoured for its simplicity and computational efficiency, while median and mode imputations offer greater robustness against outliers and skewed data distributions. Therefore, if outliers are present, the median is considered the most suitable simple imputation method (Alam et al., 2023). There is a stage for detecting multicollinearity and outliers. Multicollinearity is identified using VIF (Variance Inflation Factor) values, where independent variables with VIF values greater than 10 are considered to exhibit multicollinearity and are therefore excluded from the modelling process. Outlier detection is performed using standardized residual values. An observation is considered an influential outlier if its $|DFFITS|$ value exceeds the designated cutoff threshold; in which case it will be excluded from the modelling process.

Table 1. Variables

Variables		Period
ΔY_t	Social Recovery Index	Comparison of 2021 and 2022 from January 01st - September 30th
X_1	Number of populations	During 2019 (time in variance variable)
X_2	Number of areas	During 2019 (time in variance variable)
X_3	Population density	During 2019 (time in variance variable)
X_4	Expenditure budget	During 2019 (time in variance variable)
X_5	Primary GRDP	Quarterly during 2019
X_6	Secondary GRDP	Quarterly during 2019
X_7	Tertiary GRDP	Quarterly during 2019
X_8	Number of internet users	During 2019 (time in variance variable)
X_9	Road density	During 2019 (time in variance variable)
X_{10}	Number of health workers per capita	During 2019 (time in variance variable)
X_{11}	Number of health facilities	During 2019 (time in variance variable)
X_{12}	Number of COVID-19 positive cases	01st January – 30th September 2022
X_{13}	Average years of schooling	During 2019 (time in variance variable)
X_{14}	Number of people receiving vaccines	01st January – 30th September 2022

3.2. Methods

The stages of analysis are as follows: (a) data collection from multiple sources and data imputation using the median; (b) data preprocessing includes median imputation, as the dataset contains outliers, and outlier detection is carried out using standardized residual values; (c) testing for multicollinearity and outlier effects. Variables showing signs of multicollinearity or influential outliers are excluded from the analysis; (d) displaying the Social Recovery Index (SRI) distribution map for one day across nine months, resulting in nine SRI distribution maps; (e) selection of the panel data model through three tests: the Chow test, comparing the Common Effect Model (CEM) and the Fixed Effect Model (FEM); the Hausman test, comparing the FEM and the Random Effect Model (REM); and the Lagrange Multiplier test, comparing the FEM and CEM; (f) form a spatial weight matrix and perform a spatial autocorrelation test using Moran's test; (g) testing spatial dependency with LM to see which spatial model can be used; (h) spatial panel model estimation and model selection by comparing R^2 and Akaike information criterion (AIC) values; (i) Partial testing to see which explanatory variables have a significant effect on social recovery.

4. RESULTS AND DISCUSSION

4.1. Data Pre-processing

Outlier testing identified 183 observations as outliers, of which 133 were classified as influential outliers. Since the data is panel data, there is consistency in the day and date across each cross-sectional unit, and 38 observations share the same day, resulting in 95 total observations being deleted. As the data is balanced panel data, when one row representing a specific day in one province is removed, the same day in other provinces is also removed. In total, 760 data points, or 34.80%, were deleted, leaving 1,424 data points from the original dataset. After the data removal stage, multicollinearity detection was performed in Table 2.

Table 2. Multicollinearity Test Results

Variables	VIF	Variables	VIF
X_1: Population	2,142.94	X_8: Internet user	1,557.05
X_2: Number of areas	3,860.50	X_9: Road density	1,068.65
X_3: Population density	2,168.08	X_{10}: Health workers	354.38
X_4: Budget expenditure	1,110.86	X_{11}: Health Facility	omitted
X_5: Primary GRDP	5,219.94	X_{12}: Covid-19 New Cases	11.48
X_6: Secondary GRDP	2,195.94	X_{13}: Average year of schooling	omitted
X_7: Tertiary GRDP	2,372.99	X_{14}: Receiving Vaccine	1.22

of the 14 explanatory variables used, seven exhibited multicollinearity and were excluded from the analysis. As a result, the study will proceed with 1,424 data points and seven explanatory variables: number of areas, budget expenditure, primary GRDP, tertiary GRDP, road density, number of health workers per capita, and number of people receiving vaccines.

4.2. Panel Model Selection

This stage is carried out to select the best panel model among the CEM, FEM and REM. This stage is carried out using 1,424 observations, seven explanatory variables and one dependent variable. The best model selection and model estimation are in Table 3.

Table 3. Estimation of Panel Data and Selection of the Best Panel Data Model

Variables	CEM		FEM		REM	
Intercept	-0.002	(0.017)	-0.028	(0.011)	-0.002	(0.016)
X_2	0.316	$(0.041)^{***}$	0.316	$(0.041)^{***}$	0.316	$(0.041)^{***}$
X_4	0.439	$(0.033)^{***}$	-0.450	$(0.033)^{***}$	-0.439	$(0.033)^{***}$
X_5	-0094	$(0.037)^{***}$	-0.095	$(0.037)^{***}$	-0.094	$(0.037)^{***}$
X_7	$-0,053$	$(0.020)^{***}$	-0.079	$(0.021)^{***}$	-0.053	(0.020)
X_9	0.397	$(0.041)^{***}$	0.397	$(0.041)^{***}$	0.397	$(0.041)^{***}$
X_{10}	-0.224	$(0.028)^{***}$	-0.221	$(0.028)^{***}$	-0.224	$(0.028)^{***}$
X_{14}	-0.072	$(0.014)^{***}$	0.023	(0.022)	-0.076	(0.015)
R^2	0.23		0.24		0.23	
Lagrang Multiplier (REM-CEM)			$\chi^2 = 401.820$		${\chi^2}_{0.1;1} = 2.705$	
The Chow Test (CEM – FEM)			$\mathrm{F} = 42.527$		$\mathrm{F}_{0.1;\,3,1424} = 2.088$	
The Hausman Test (FEM – REM)			$\chi^2 = 32.290$		${\chi^2}_{0.1;7} = 12.017$	

Description: standard error in parentheses; Significant at: *** α=0.01; ** α=0.05; * α=0.10

Based on the results in Table 3, the Lagrange multiplier test shows that the best model is REM because $\chi^2 = 401.820 > {\chi^2}_{0.1;1} = 2.705$, implies the rejection of H_0, suggesting that REM is more appropriate for use. Based on the Chow test $\mathrm{F} = 42.527 > \mathrm{F}_{0.1;\,3,1424} = 2.088$, this also implies rejection of H_0, suggesting FEM is more appropriate for use. Based on the Hausman test also a rejection of H_0 and suggesting the best model is FEM, so that of the three tests the best model chosen is FEM.

4.3. Autocorrelation and Dependence Spatial Tests

Autocorrelation testing was conducted using Moran's index on 178 days that were free from outlier issues. The results show that not all days exhibit significant spatial autocorrelation in Table 4.

Table 4. Result of Autocorrelation Test

Day	*Moran's I*	*p-value*	Spatial Pattern
1	-0.1346	0.4803	Negative
2	-0.2279	0.7604	Negative
3	0.1149	0.0600 *	Positive
⋮	⋮	⋮	⋮
270	0.0273	0.2404	Positive
272	0.0119	0.2460	Positive
273	-0.1476	0.5099	Negative

The number of days with significant spatial autocorrelation varies depending on the significance level: with α = 0.01, there are five significant days; with α = 0.05, there are 41 significant days; and with α = 0.1, there are 72 significant days. Although not all days exhibit significant spatial autocorrelation, the study can still proceed, as the assumption of spatial autocorrelation is sufficiently met.

Lagrange Multiplier (LM) tests and robust LM forms were applied to detect the presence of spatial lag and spatial error dependence. The significant results of both the LM-lag and LM-error tests indicated the need to consider models that account for spatial dependence in both forms. Consequently, the Spatial Autoregressive Model (SAR) and the Spatial Error Model (SEM) were deemed appropriate. Furthermore, the Spatial Durbin Model (SDM) was included in the analysis because it extends the SAR model by incorporating spatial lags of both the dependent and independent variables. The result of the dependency test is shown in Table 5.

Table 5. Spatial Dependency Test Results

Test	*p-value*	Decision
LM lag	0.0333	There are lag dependence
LM error	0.0005	There is error dependence
Robust LM lag	0.1203×10^{7}	There are lag dependence (robust model)
Robust LM error	0.5532×10^{7}	There is error dependence (robust model)

4.4. Estimation and Selection of the Best Spatial Panel Model

The next stage will be spatial panel analysis to select the best model by comparing three models, namely SAR fixed effect, SEM fixed effect and SDM fixed effect. The results obtained are in Table 6.

Table 6. Estimation of Panel Spatial Model and Selection of The Best Model

Variables	SAR Fixed Effect		SEM Fixed Effect		SDM Fixed Effect	
X_2	0.329	$(0.038)^{***}$	0.357	$(0.041)^{***}$	-12.360	$(1.492)^{***}$
X_4	-0.466	$(0.031)^{***}$	-0.471	$(0.030)^{***}$	3.441	$(1.454)^{***}$
X_5	-0.089	$(0.034)^{***}$	-0.105	$(0.037)^{***}$	4.340	$(0.773)^{***}$
X_7	-0.085	$(0.019)^{***}$	-0.104	$(0.019)^{***}$	3.989	$(1.066)^{***}$
X_9	0.419	$(0.038)^{***}$	0.441	$(0.037)^{***}$	-10.585	$(1.810)^{***}$
X_{10}	-0.210	$(0.026)^{***}$	-0.231	$(0.029)^{***}$	5.204	$(1.244)^{***}$
X_{14}	0.028	(0.021)	0.062	$(0.022)^{***}$	0.079	$(0.020)^{***}$
$W \times X_2$					-4.903	$(2.950)^{*}$
$W \times X_4$					-42.476	$(9.401)^{***}$
$W \times X_5$					0.600	(1.691)
$W \times X_7$					12.143	$(1.703)^{***}$
$W \times X_9$					41.495	$(5.840)^{***}$
$W \times X_{10}$					-0.070	$(.)$

Table 6. Estimation of Panel Spatial Model and Selection of The Best Model

$W \times X_{14}$					0.153	$(0.024)^{***}$
ρ or λ	−0.240	$(0.039)^{***}$	−0.200	$(0.041)^{***}$	−0.210	$(0.036)^{***}$
Adjusted R^2 (%)	0.19		0.18		0.27	
AIC	2492.561		2472.691		2262.149	

Description: standard error in parentheses; Significant at: *** $\alpha = 0.01$; ** $\alpha = 0.05$; * $\alpha = 0.10$; W is the impact of neighbors.

Based on the results in Table 6, the value of the spatial correlation coefficient containing lag and error is significant and negative for all models with a value range of -0.20 to -0.24, this indicates that the 8 provinces have a negative spatial correlation. Even though the correlation coefficient value is relatively low, the use of a spatial model remains relevant. As Elhorst (2014) emphasized, when the spatial dependence coefficient is small or weak (ρ or λ are relatively small in magnitude), applying a spatial model is still appropriate. This is because, even if the spatial weights matrix is not perfectly specified, the resulting parameter estimates will not deviate substantially. The SDM fixed effect has a spatial correlation value that is slightly lower than the SAR fixed effect but has the highest R^2 value and lowest AIC value compared to the three models, so the SDM fixed effect is the best model for estimating factors affecting social recovery after herd immunity on the island of Sumatra.

The number of areas has a negative and significant influence on social recovery in both the province itself and its neighboring provinces. This means that the larger the area of a province, the lower the social recovery rate in that province and in its neighboring provinces. Budget expenditure, primary GRDP, and tertiary GRDP, which fall under the category of government intervention, have a positive and significant influence on their respective provinces. Regarding neighboring provinces, budget expenditure has a negative and significant effect on social recovery, primary GRDP has a positive but insignificant effect, and tertiary GRDP has a positive and significant effect on the social recovery of neighboring provinces. These results align with the research conducted by T. Yang et al. (2023). Road density has a negative and significant effect on the social recovery rate in the province itself but a positive and significant effect on the social recovery rate of neighboring provinces. The results of this study do not align with the research conducted by Zhang et al. (2022).

The number of health workers per capita has a positive and significant effect on the social recovery rate in neighboring provinces but a negative and insignificant effect on its neighbors. This means that the more the number of health workers per capita, the social recovery in a province is also expected to increase, and these results are in line with research conducted by Serrao-Neumann et al. (2018). The number of people receiving vaccine dose 1 has a positive and significant effect on social recovery in the province itself and also its neighboring provinces. This shows that the more people in a province who receive vaccinations, the more the social recovery rate in a province can increase, and these results are in line with research (Pribadi et al., 2021).

The coefficient of determination or adjusted R^2 obtained is 27.15%, meaning that the seven independent variables can explain 27.15% of the variation in the social recovery index, while the remaining variation is explained by other variables not included in the model. Ozili & Peterson (2022) examined the phenomenon of low R^2 values in social research. Their study suggests that a low R^2 value is not necessarily a negative outcome. This is because, in many cases, the primary goal of social science modelling is not to predict human behaviour but rather to assess whether certain independent variables have a significant influence on the dependent variable. Therefore, a low R^2 value—at least 0.1 or just 10 %—can be considered acceptable, provided that some or most of the independent variables are statistically significant.

Previous studies have examined the socio-economic impacts of COVID-19 in Indonesia. For instance, Kibtiah and Medeleine (2023) highlighted changes in consumption behaviour during the post-COVID-19 period, emphasizing the importance of government policies to support both online and offline merchants. However, their study did not employ statistical analysis to quantitatively assess these changes. In contrast, Brata et al. (2021) used regression analysis to investigate whether COVID-19 exacerbated socioeconomic inequality across provinces in Indonesia. Their findings revealed disparities between urban and rural areas, as well as spatial inequality based on the Gini index, and recommended a more equitable distribution of subsidies. Despite their contributions, both studies present certain limitations. Kibtiah and Medeleine (2023) primarily focused on behavioural aspects without empirical modelling, while Brata et al. (2021) addressed regional inequality but did not consider spatial dependency in their analysis. This study seeks to fill those gaps by employing statistical models to identify the determinants of social recovery in post-COVID-19 Sumatra. By incorporating spatial dependency, panel data, and a constructed social recovery index, this study provides a more comprehensive understanding of the socio-economic recovery process following the pandemic.

5. CONCLUSION

Factors affecting social recovery after herd immunity on the island of Sumatra can be explained using the Spatial Durbin Model (SDM), which has an R^2 value of 0.27. Seven variables significantly affect the province itself, while five variables impact neighboring provinces. The seven variables that significantly affect the SRI in the region are the number of areas (X_2), the amount of budget expenditure (X_4), the amount of primary GRDP (X_5), the amount of tertiary GRDP (X_7), road density (X_9), the number of health workers per capita (X_{10}), and the number of people receiving vaccines (X_{14}). The five variables that significantly affect neighboring areas are the number of areas (X_2), the amount of budget expenditure (X_4), the amount of tertiary GRDP (X_7), road density (X_9), and the number of people receiving vaccines (X_{14}).

This study uses the SRI as the dependent variable. Since SRI data is not available from official Indonesian government institutions, it is obtained using the Google Mobility report data, which only calculates movement trends in public locations. Future research could develop studies using inter-provincial mobility data to make the spatial dependence between regions more evident.

REFERENCES

Alam, S., Ayub, M. S., Arora, S., & Khan, M. A. (2023). An investigation of the Imputation Techniques for Missing Values in Ordinal Data Enhancing Clustering and Classification Analysis Validity. *Decision Analytics Journal*, *9*, 100341. https://doi.org/10.1016/j.dajour.2023.100341

Astuti, A. B., Pramoedyo, H., Astutik, S., & Setiarini, A. N. D. (2024). Spatial Panel Modeling of Provincial Inflation in Indonesia to Mitigate Economic Impacts of Health Crises. Media Statistika, 17(1), 105–116. https://doi.org/10.14710/medstat.17.1.105-116

Astuti, A. M., Setiawan, Zain, I., & Purnomo, J. D. T. (2020). A Review of Panel Data on Spatial Econometrics Models. *Journal of Physics: Conference Series*, *1490*, 012032. https://doi.org/10.1088/1742-6596/1490/1/012032

Bland, A. R., Barraclough, M., Trender, W. R., Mehta, M. A., Hellyer, P. J., Hampshire, A., Penner, I. K., Elliott, R., & Harenwall, S. (2024). Profiles of Objective and Subjective Cognitive Function in Post-COVID Syndrome, COVID-19 Recovered, and COVID-19 Naïve Individuals. *Scientific Reports*, *14*(1), 13368. https://doi.org/10.1038/s41598-024-62050-x

Brata, A. G., Pramudya, E. P., Astuti, E. S., Rahayu, H. C., & Heron, H. (2021). COVID-19 and Socio-Economic Inequalities in Indonesia: A Subnational-Level Analysis (ERIA Discussion Paper Series No. 371). ERIA. https://www.eria.org/publications/covid-19-and-socio-economic-inequalities-in-indonesia-a-subnational-level-analysis/

Chicco, D., Warrens, M. J., & Jurman, G. (2021). The Coefficient of Determination R Squared is More Informative than SMAPE, MAE, MAPE, MSE and RMSE in Regression Analysis Evaluation. *PeerJ Computer Science*, *7*(e623). https://doi.org/DOI 10.7717/peerj-cs.623

Elhorst, J. P. (2014). *Spatial Econometrics: From Cross-Sectional Data to Spatial Panels*. Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-642-40340-8

Gao, J. (2021). *Fundamentals of Spatial Analysis and Modelling* (1st ed.). CRC Press. https://doi.org/10.1201/9781003220527

Hasiru, L. S., Djakaria, I., & Hasan, I. K. (2022). Penerapan Model Spasial Durbin dengan Uji Lanjutan Local Indicator of Spatial Autocorrelation untuk Melihat Penyebaran Stunting di Kabupaten Bone Bolango. *Jambura Journal of Probability and Statistics*, *3*(1), 19–28. https://doi.org/10.34312/jjps.v3i1.13083

Hidayati, N., Karuna, E. E., & Alwansyah, M. A. (2024). A Panel Data Spatial Regression Approach for Modeling Poverty Data In Southern Sumatra. *Journal Statistics and Data* Science, *03*(02).

Kibtiah, T. M. & Medeleine. (2023). Indonesia's Economic Recovery in a Post-Pandemic: Under the New Normal on Society 5.0. *E3S Web of Conferences*, *388*, 04016. https://doi.org/10.1051/e3sconf/202338804016

Mahendra, I. Y., Fannysa, F., Agusinta, L., & Suryobuwono, A. (2021). *T*he Impact of PPKM Implementation and Travel Behavior Characteristics on the Transportation Mode Selection in Bekasi. *Global Research on Sustainable Transport & Logistics*, *1*(1).

Maloney, P., Kompaniyets, L., Yusuf, H., Bonilla, L., Figueroa, C., & Garcia, M. (2023). The Effects of Policy Changes and Human Mobility on the COVID-19 Epidemic in the Dominican Republic, 2020–2021. *Preventive Medicine Reports*, *36*, 102459. https://doi.org/10.1016/j.pmedr.2023.102459

Nurjanah, N., Rinaldi, A., & Putri, R. (2023). Spatial Error Model pada Tingkat Kemiskinan Kabupaten/Kota di Provinsi Lampung. *Variance: Journal of Statistics and Its Applications*, *5*(1), 47–54. https://doi.org/10.30598/variancevol5iss1page47-54

O'Sullivan, D., & Unwin, D. J. (2010). *Geographic Information Analysis* (2nd ed). New York: John Wiley.

Ozili, & Peterson, K. (2022). The Acceptable R-square in Empirical Modelling for Social Science Research. *Social Research Methodology and Publishing Results*, *1*(1).

Pribadi, D. O., Saifullah, K., Putra, A. S., Nurdin, M., Iman, L. O. S., & Rustiadi, E. (2021). Spatial Analysis of COVID-19 Outbreak to Assess the Effectiveness of Social Restriction Policy in Dealing with the Pandemic in Jakarta. *Spatial and Spatio-Temporal Epidemiology*, *39*, 100454. https://doi.org/10.1016/j.sste.2021.100454

Romanillos, G., García-Palomares, J. C., Moya-Gómez, B., Gutiérrez, J., Torres, J., López, M., Cantú-Ros, O. G., & Herranz, R. (2021). The City Turned Off: Urban Dynamics During the COVID-19 Pandemic Based on Mobile Phone Data. *Applied Geography*, *134*, 102524. https://doi.org/10.1016/j.apgeog.2021.102524

Saputro, D. R. S., Muhsinin, R. Y., & Widyaningsih, P. (2018). 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*(012104).

Sari, D. A., & Pasaribu, R. A. (2023). Humans as Individual and Social Beings in Everyday Life. *International Journal of Students Education*, *1*(1): 78–80.

Serrao-Neumann, S., Crick, F., & Low Choy, D. (2018). Post-disaster Social Recovery: Disaster Governance Lessons Learnt from Tropical Cyclone Yasi. *Natural Hazards*, *93*(3), 1163–1180. https://doi.org/10.1007/s11069-018-3345-5

Setiadi, W., Rozi, I. E., Safari, D., Daningrat, W. O. D., Johar, E., Yohan, B., Yudhaputri, F. A., Lestari, K. D., Oktavianthi, S., Myint, K. S. A., Malik, S. G., Soebandrio, A., & on Behalf of the Wascove Team. (2022). Prevalence and Epidemiological Characteristics of COVID-19 After One Year of Pandemic in Jakarta and Neighbouring Areas, Indonesia: A Single Center Study. *PLOS ONE*, *17*(5), e0268241. https://doi.org/10.1371/journal.pone.0268241

Torres-Favela, M., & Luna, E. M. (2024). The Role of Informality in The Economic Growth, Employment, and Inflation During the COVID-19 Crisis. *Latin American Journal of Central Banking*, 100150. https://doi.org/10.1016/j.latcb.2024.100150

Utomo, D. S. C., & Nugroho, A. (2022). *Estimated Economic Growth Rate Based on Efek Decrease in PPKM Level Using Support Vector Regression Method*. *4*(2): 233–251.

Yang, M., Han, C., Cui, Y., & Zhao, Y. (2021). COVID-19 and Mobility in Tourism Cities: A Statistical Change-Point Detection Approach. *Journal of Hospitality and Tourism Management*, *47*, 256–261. https://doi.org/10.1016/j.jhtm.2021.03.014

Yang, T., Yu, N., Yang, T., & Hong, T. (2023). How Do Urban Socio-Economic Characteristics Shape A City's Social Recovery? An Empirical Study of COVID-19 Shocks in China. *International Journal of Disaster Risk Reduction*, *90*, 103643. https://doi.org/10.1016/j.ijdrr.2023.103643

Zhang, J., Zhao, S., Peng, C., & Gong, X. (2022). Spatial Heterogeneity of the Recovery of Road Traffic Volume from the Impact of COVID-19: Evidence from China. *Sustainability*, *14*(21), 14297. https://doi.org/10.3390/su142114297