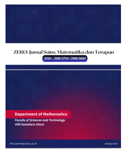
Zero : Jurnal Sains, Matematika, dan Terapan
E-ISSN : 2580-5754; P-ISSN : 2580-569X
Volume 9, Number 2, 2025
DOI: 10.30829/zero.v9i2.25933
Page: 430-442



A Spatio-Temporal Panel Data Model for Coronavirus Deaths: Evidence from Europe

[1] Asep Rusyana
Department of Statistics, Universitas Syiah Kuala, Banda Aceh, Indonesia

[2] Anwar Fitrianto
Study Program of Statistics and Data Science, IPB University, Bogor, Indonesia

[3] Marzuki
Department of Statistics, Universitas Syiah Kuala, Banda Aceh, Indonesia
Department of Mathematics, Universiti Malaysia Terengganu, Kuala Nerus, Malaysia

Article Info

Article history:

Accepted, 16 October 2025

Keywords:

Coronavirus;
General Spatial Model;
Lagrange Multiplier;
Spatial Autoregressive;
Spatial Error Model;
Spatial Weight Matrix;
Temporal Spatial Panel Data.

ABSTRACT

The spatial panel data model is a model in which the independent variables are estimated to be influenced by place, time, and explanatory variables. Almost all countries have been attacked by the coronavirus, starting in China in February 2020. Some of the affected residents died, some recovered, and some are still under surveillance. This study aims to identify the most accurate model with the response variable of population deaths due to coronavirus and the independent variables of the number of cases with coronavirus and the number of tests. This study uniquely compares multiple spatial models for pandemic analysis. The four models are SAR, SEM, GSM, and Temporal Spatial Panel Data. The fourth model enables more accurate analysis of COVID-19 mortality by capturing both spatial and temporal dependencies. As a result, the spatial temporal panel data model is the best model with a coefficient of determination of 63.9%.

Corresponding Author:

Asep Rusyana,
Department of Statistics,
Faculty of Mathematics and Natural Sciences,
Universitas Syiah Kuala, Banda Aceh, Indonesia
Email: asep.rusyana@usk.ac.id

1. INTRODUCTION

Panel data are a combination of cross-section and time-series data. Cross-sectional data can be location or firm, while time data can be daily, weekly, monthly, or annual [1]. If the cross-section is location and the effect of location on the response variable is considered, the model can be a spatial panel data model. In the spatial panel data model, the assumption of errors that must spread freely is not hampered by the presence of location effects because the location effect is included in the spatial weight matrix component (W) [2].

A panel data model in which the independent variable (y) is affected by the previous independent variable ($y_{t-1}, y_{t-2}, \ldots$) is called a dynamic or temporal panel data model [3]. If the panel data are modeled by including W components and independent variables, it is called a static spatial panel data model [4]. If the panel data model includes W, $y_{t-1}, y_{t-2}, \ldots$, and explanatory variables, the model is called a spatial temporal panel data model [5].

Corona Virus Disease 2019 (Covid-19) is a disease that began attacking China in early February 2020. The virus spread quickly to other countries, and almost all countries had residents who were affected by the disease, died from it, or recovered from it [6]. Data on patients commonly referred to as Patients under Surveillance (PwCP), the number of people who die, and the number of people who recover in each country can be seen on several websites, such as John Hopkins University 2020, Worldometer 2020, Virusncov 2020, and Our World in Data 2020, among others. Countries in continental Europe are among the worst affected by the coronavirus [7]. This study uses fourteen (14) countries, namely, Italy, England, Turkey, Russia, Belgium, Switzerland, Portugal, Austria, Poland, Romania, Denmark, Norway, Czechia, and Serbia. The reason for taking these fourteen (14) countries is because these countries are the countries with the most corona attacks in Europe and have complete data.

Several studies have been conducted using spatial panel data models. The Lagrange Multiplier (LM) tests on the Rook and Queen spatial weight matrices in a spatial temporaldata model were compered with varying lambda values (coefficients of the spatial weight matrix in the model), and the distribution of the response variables was Normal and Gamma [8]. This LM test identifies the effect of lambda, gamma (temporal coefficients), and rho (spatial coefficients and lags of independent variables). Berra involves both simulated and real data. A research was condacted to identify factors that affect the number of tourist visits in Indonesia using annual data from 2011-2015 [9]. Six models were compared, namely Spatial Autoregressive (SAR) with a spatial weighting matrix using the adjacency method (W1), SAR with a spatial weighting matrix using the inverse distance weighting (W2), Spatial Error Model (SEM) with W1, SEM with W2, Spatial Durbin Model (SDM) with W1 and SDM with W2. In this study, SEM with W1 was the best model because it had the largest R2, smallest RMSE, and smallest AIC. Although several studies have utilized spatial data models for various applications, there is a lack of research that systematically compares multiple spatial and spatio-temporal panel data models for modeling coronavirus deaths, particularly in the context of European countries over time. This study addresses this gap by evaluating the performance of four such models using real COVID-19 data.

This study aims to identify the best model among the four built models: SAR, SEM, General Spatial Model (GSM), and spatial-temporal panel data models. The response variable is the number of deaths due to corona while the explanatory variables are the number of cases of corona and the number of tests by country. The use of multiple models in this study is designed to ensure a thorough evaluation of spatial and temporal effects in the data. Statistical tests are applied to prevent overfitting and to select the most suitable model, with the final choice based on clear performance metrics. This approach provides robust and reliable conclusions about the factors influencing coronavirus deaths. Although detailed descriptions of panel data models are included, the main aim of this study is to advance understanding of COVID-19 deaths by identifying the most suitable modeling approach. Streamlining these sections allows the research to better emphasize the significance and real-world impact of accurately modeling pandemic outcomes.

The choice of SAR, SEM, GSM, and especially the spatio-temporal panel data model is justified by the need to capture both spatial and temporal dependencies inherent in COVID-19 mortality data. Unlike simpler models, the spatio-temporal approach accounts for how deaths in one country are influenced by neighboring countries and by previous time periods, making it particularly suitable for pandemic analysis. This comprehensive modeling strategy is supported by statistical tests and aligns with the complex, interconnected nature of COVID-19 spread and outcomes. The novelty of this study is its direct comparison of multiple spatial and spatio-temporal models for COVID-19 deaths, using rigorous statistical tests to select the best approach. By modeling both spatial and temporal interactions together, the study addresses gaps left by previous research and demonstrates the advantages of spatio-temporal analysis for understanding pandemic dynamics. While prior studies seldom directly compared several spatial and spatio-temporal panel data models for COVID-19 mortality, this research uniquely conducts a systematic comparison using European data, revealing the superior performance of the spatio-temporal model.

2. RESEARCH METHOD

2.1. Data

The data used in this study are the number of people infected with the coronavirus, the number of tests conducted by countries, and the number of deaths in 14 European countries from March 1, 2020, to May 2, 2020 [10]. These countries are Italy, the United Kingdom, Turkey, Russia, Belgium, Switzerland, Portugal, Austria, Poland, Romania, Denmark, Norway, Czechia, and Serbia. These countries were chosen because travel routes between European countries are very smooth, and they have been relatively severely affected by the coronavirus, making it interesting to examine the factors influencing the disaster [11]. Another reason is that these countries have comprehensive datasets. The data structure is shown in Table 1. Countries are indicated by index i; for example, i = 1 means Italy, ..., i = 14 means Serbia. Time is indicated by t; for example, t = 1 means week 1, ..., t = 9 means week 9. Y_{it} is the number of deaths due to coronavirus in country i at time t. Meanwhile, X_{1it} is the number of coronavirus cases in country i at time t. X_{2it} is the number of tests conducted by country i in week t [12]. The study uses data from March to May 2020, this period is crucial for understanding the initial dynamics of COVID-19 before later interventions and virus changes. Focusing on this early phase allows for clearer analysis of spatial and temporal effects on mortality, providing valuable insights for future outbreak preparedness and response. The study relies on reputable data sources, special care was taken to select countries with the most complete and consistent records, minimizing the risk of missing or inconsistent data. The use of robust statistical methods further ensures the reliability of the results, even in the presence of minor data quality issues. Focusing on cases and tests ensures the use of the most reliable and consistently reported data across all countries, which is critical for robust and unbiased modeling—especially during the chaotic early phase of the pandemic. This approach minimizes data quality issues and allows the model to accurately capture the core drivers of COVID-19 mortality, providing a strong, reliable foundation for both immediate insights and future research.

Table 1. Spatial-temporal panel data structure two lags of the dependent variables and one lag of the independent variables

i	*t*	Y_{it}	WY_{it}	WY_{it-1}	Y_{it-1}	Y_{it-2}	X_{it}
1	1	y1,1	$\sum_{i=j=1}^{n} w_{1,j}y_{i,1}$	$\sum_{i=j=1}^{n} w_{1,j}y_{i,0}$	y1,0	y1,0	x1,1
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
1	9	y1,9	$\sum_{i=j=1}^{n} w_{1,j}y_{i,9}$	$\sum_{i=j=1}^{n} w_{1,j}y_{i,8}$	y1,8	y1,7	x1,9
2	1	y2,1	$\sum_{i=j=1}^{n} w_{2j}y_{i1}$	$\sum_{i=j=1}^{n} w_{2,j}y_{i,0}$	y2,0	y2,0	x2,1
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
2	9	y2,9	$\sum_{i=j=1}^{n} w_{2j}y_{i,9}$	$\sum_{i=j=1}^{n} w_{2,j}y_{i,8}$	y2,8	y2,7	x2,9
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
n	1	y*n*,1	$\sum_{i=j=1}^{n} w_{14,j}y_{i,1}$	$\sum_{i=j=1}^{n} w_{14,j}y_{i,0}$	y*n*,0	y*n*,0	x*n*,1
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
n	9	y*n*,9	$\sum_{i=j=1}^{n} w_{14,j}y_{i,9}$	$\sum_{i=j=1}^{n} w_{14,j}y_{i,8}$	y*n*,8	y*n*,7	x*n*,9

Remark:

n	The number of countries	*Y*	Dependent variable
T	The *T*- Week	Y_{it-p}	*p* lags of the dependent variable
W	Spatial weight matrix	*X*	Independent Variable

2.2. Lagrange Multiplier (LM) Test to See the Effect of Time and Place

The LM test is also known as the Breusch-Pagan test. Three hypotheses will be tested: the influence of place and time, the influence of place, and the influence of time [13], [14].

Testing the influence of Place and Time

$H_0: \sigma_\mu^2 = \sigma_\lambda^2 = 0$ (There is no influence of place and time)

H_1: at least one that is not equal to zero (there is an influence of time or place)

The test statistic is in Equation (1).

$$LM = \frac{nt}{2(n-1)(t-1)}\left[(n-1)\left[1-\frac{\hat{\delta}'(I_n \otimes J_t)\hat{\delta}}{\hat{\delta}'\hat{\delta}}\right]^2 + (t-1)\left[1-\frac{\hat{\delta}'(J_n \otimes I_t)\hat{\delta}}{\hat{\delta}'\hat{\delta}}\right]^2\right] \tag{1}$$

where:

$\hat{\delta}$ is the regression model error,

In is the identity matrix of size n × n,

⊗ is the Kronecker product,

Jt is a matrix of size t × t.

The decision criteria were as follows: reject H0 if the LM test statistic value > χ_2^2 or the p < α

Testing the influence of place

$H_0: \sigma_\mu^2 = 0$ (There is no influence of place)
$H_1: \sigma_\mu^2 \neq 0$ (There is influence of place)
The test statistic is in Equation (2).

$$LM_1 = \frac{nt}{2(t-1)}\left[1 - \frac{\hat{\delta}'(I_n \otimes J_t)\hat{\delta}}{\hat{\delta}'\hat{\delta}}\right]^2 \tag{2}$$

The decision criteria are reject H0 if the LM1 test statistic value > χ_1^2 or the p value < α

Testing the influence of time

$H_0: \sigma_\lambda^2 = 0$ (There is no influence of time)
$H_1: \sigma_\lambda^2 \neq 0$ (There is influence of time)
The test statistic is in Equation (3).

$$LM_2 = \frac{nt}{2(n-1)}\left[1 - \frac{\hat{\delta}'(J_n \otimes I_t)\hat{\delta}}{\hat{\delta}'\hat{\delta}}\right]^2 \tag{3}$$

The decision criteria are reject H0 if the LM2 test statistic value > χ_1^2 or the p value < α

2.3. Panel Data Model

There are three (3) panel data models:

Mixed Effects Model

The mixed-effects model is similar to ordinary linear regression. Each observation was assumed to have an equal influence on the model. Parameter estimates can be found using the ordinary least squares method [15]. The model is in Equation (4).

$$y_{it} = \beta_0 + \sum_{j=1}^{k} \beta_j x_{jit} + \varepsilon_{it} \tag{4}$$

Fixed Effects Model

The fixed-effects model with n individuals focuses only on those individuals. This model has the following assumptions: (1) μi is fixed so that its estimate can be calculated, (2) $\varepsilon_{ij} \sim \mathrm{N}(0, \sigma_\varepsilon^2)$ and is identically stochastic, and xit and εit are independent for each i and t. The parameter estimation uses an ordinary least squares approach [16]. The model is in Equation (5).

$$y_{it} = \beta_0 + \sum_{j=1}^{k} \beta_j x_{jit} + \mu_i + \varepsilon_{it} \tag{5}$$

Random Effects Model

The random-effects model selects individuals randomly from a large population. The assumptions of this model are: (1) $\mu_i \sim N(0, \sigma^2)$ and is stochastically independent, (2) $\varepsilon_{ij} \sim \mathrm{N}(0, \sigma_\varepsilon^2)$ and is identically stochastically independent, and xit and εit are independent for each i and t. Parameter estimation uses a least-squares approach [17]. The model is in Equation (6).

$$y_{it} = \beta_0 + \sum_{j=1}^{k} \beta_j x_{jit} + \mu_i + \varepsilon_{it} \tag{6}$$

2.4. Chow Test and Hausman Test

Chow and Hausman tests were used to identify the most suitable panel data model for use with the research data. These tests, like hypothesis testing in general, are conducted in three (3) stages: determining the hypothesis, determining the test statistic, and determining the decision criteria [18].

Chow Test

$H_0: \mu_1 = \mu_2 = \cdots = \mu_{n-1} = 0$ (pooled model)
H_1: there is at least one $\mu_i \neq 0$ (fixed effect model)
Test Statistics is in Equation (7).

$$F = \frac{(JKG_{pooled} - JKG_{fixed})/(n-1)}{JKG_{fixed}/(nT-n-K)} \tag{7}$$

The decision criteria: Reject H0 if F > F(n-1)(nT-n-K) or p-value < α.

Hausman Test

$H_0: E(\varepsilon_{ij}|x_{it}) = 0$ (random effect model)
$H_1: E(\varepsilon_{ij}|x_{it}) \neq 0$ (fixed effect model)
The test statistics is in Equation (8).

$$\chi^2_{count} = \hat{q}[Var(\hat{q})]^{-1}\hat{q} \tag{8}$$

Where:
$\hat{q} = \hat{\beta}_{random} - \hat{\beta}_{fixed}$
The decision criteria: Reject H0 if $\chi^2_{count} > \chi^2_{(k,\alpha)}$ or p_value < α.

2.5. Spatial Weighting Matrix (W)

Approaches to managing contiguity relationships between locations can be implemented using Rook, Bishop, or Queen Contiguity [19].

1	2	3
4	5	6
7	8	9

(a)

1	2	3
4	5	6
7	8	9

(b)

1	2	3
4	5	6
7	8	9

(c)

Figure 1. Contiguity method for (a) rook, (b) bishop, and (c) queen contiguity

If number 5 in Figure 1 is a location, then locations 2, 6, 8, and 4 are connected locations for Rook contiguity, locations 1, 3, 7, and 9 are connected to 5 for Bishop contiguity, and locations 1, 2, 3, 4, 6, 7, 8, and 9 are connected to 5 for Queen contiguity. The contiguity matrix shows the intersecting relationships between a location and its neighbors. The contiguity matrix is symbolized by C, and cij is the element of matrix C that describes the value at location i that is adjacent to location j. cij = 1 if location i is directly adjacent to j, and cij = 0 if location i is not directly adjacent to j.

The spatial weighting matrix (W) is formed by normalizing the previous matrix [20]. A frequently used normalization method is to set the sum of the values of each row to zero. W of size n × n has elements wij where wij = 1 if locations i and j are adjacent, wij = 0 if i and j are not adjacent, and wii = 0. The relationship between wij and cij can be seen in Equation (9).

$$w_{ij} = \frac{c_{ij}}{\sum_{j=1}^{n} c_{ij}} \tag{9}$$

2.6. Lagrange Multiplier Test

This test was used to identify the presence of spatial influence in the data. The LM test consists of the Standard Spatial Lag and Standard Spatial Error tests [21].

Standard Spatial Lag Test

The standard spatial lag test was used to determine the influence of spatial lag on the models.
The hypothesis for the test is as follows.
$H_0 : \rho = 0$ (no spatial lag dependence)
$H_1 : \rho \neq 0$ (there is spatial lag dependence)
The test statistic is in Equation (10).

$$LM_{lag} = \frac{\left[\frac{\boldsymbol{\varepsilon' W y}}{\boldsymbol{\varepsilon'\varepsilon}/n}\right]^2}{\boldsymbol{D}} \tag{10}$$

Where, $\boldsymbol{D}$ can be seen in Equation (11).

$$D = \left[\frac{(WX\hat{\beta})'\left(I-(X'X)^{-1}X'(WX\hat{\beta})\right)}{\hat{\sigma}^2}\right] + tr(W'W + WW) \tag{11}$$

IT = T × T identity matrix,
W = spatial weighting matrix,
$\hat{\sigma}^2$ = mean square error.
The decision criterion was as follows:
Reject H_0 if $LM_{lag} > \chi^2_{(1)}$ or p-value < α.

Spatial Standard Error Test

A spatial standard error test was performed to identify the influence of spatial errors on the model.

$H_0 : \lambda = 0$ (no spatial error dependence)

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

The test statistic is in Equation (12).

$$LM_{error} = \frac{\left[\frac{\varepsilon' W \varepsilon}{\grave{\varepsilon} \varepsilon / n}\right]^2}{tr(W^2 + W'W)} \tag{12}$$

The decision criterion is

Reject H0 if $LM_{error} > \chi^2_{(1)}$ or p-value < α.

2.7. Spatial Panel Data Models

The spatial panel data models developed in this study were SAR, SEM, and the GSM.

SAR (Spatial Autoregressive) Panel Data Model

There are several forms of static spatial models, including the Spatial Autoregressive (SAR), Spatial Error Model (SEM), and Spatial Durbin Model (SDM) [1].

The SAR model is in Equation (13).

$$y_{it} = \rho \sum_{j=1}^{n} w_{ij} y_{ijt} + \grave{x}_{it}\beta + \mu_i + \mu_t + \varepsilon_{it} \tag{13}$$

The parameters were estimated using the maximum likelihood method (ML). The function is in Equation (14) and (15).

$$Log(L) = -\frac{nT}{2} log(2\pi\sigma^2) + T log|I_n - \rho W| \tag{14}$$

$$-\frac{1}{2\sigma^2} \sum_{i=1}^{n} \sum_{t=1}^{T} \left(y_{it} - \rho \sum_{j=1}^{n} w_{ij} y_{ijt} - \grave{x}_{it}\beta - \mu_i - \mu_t\right)^2 \tag{15}$$

SEM (Spatial Error Model) Panel Data Model

The SEM model is in Equation (16) and (17).

$$y_{it} = \grave{x}_{it}\beta + \mu_i + \mu_t + \Phi_{it} \tag{16}$$

$$\Phi_{it} = \lambda \sum_{j=1}^{n} w_{ij} \Phi_{ijt} + \varepsilon_{it} \tag{17}$$

where:

Φit is the spatial autocorrelation error,

λ is the spatial autocorrelation coefficient,

μi is the effect of location i, μt is the unobserved effect of time t,

The SEM model parameters are estimated using the maximum likelihood function in Equation (18) and (19) [21]:

$$Log(L) = -\frac{nT}{2} log(2\pi\sigma^2) + T log|I_n - \lambda W| \tag{18}$$

$$-\frac{1}{2\sigma^2} \sum_{i=1}^{n} \sum_{t=1}^{T} \left(y_{it} - \lambda \sum_{j=1}^{n} w_{ij} y_{ijt} - \left(\grave{x}_{it} - \lambda\left[\sum_{j=1}^{n} w_{ij} x'_{it}\right]\right)\beta\right)^2 \tag{19}$$

GSM (General Spatial Model) Panel Data Model

The GSM model is in Equation (20) and (21).

$$y = \rho W y + X\beta + u \tag{20}$$

$$u = \lambda W u + \varepsilon \tag{21}$$

Where:

$$\boldsymbol{\varepsilon} \sim N(0, \sigma^2 \boldsymbol{I})$$

The maximum likelihood method can be used to estimate GSM parameters. The maximum likelihood function is in Equation (22) and (23) [22]:

$$Log(L) = c(y) - \frac{n}{2} ln(\sigma^2) + ln|I_n - \lambda W| + ln|I_n - \rho W| \quad (22)$$

$$-\frac{1}{2\sigma^2}\{(I - \lambda[W])[(I - \rho[W])y - X\beta]\}^T\{(I - \lambda[W])[(I - \rho[W])y - X\beta]\}^T \quad (23)$$

2.8. Temporal Spasial Panel Data Model

The spatio-temporal panel data model contains both spatial and temporal elements. The model is in Equation (24).

$$Y_{i,t} = \lambda_0 W_n Y_{i,t} + \rho_0 W_n Y_{i,t-1} + \gamma_1 Y_{i,t-1} + \gamma_2 Y_{i,t-2} + \cdots + \beta_1 X_{1i,t} + \beta_2 X_{2,i,t} + \cdots + V_{i,t} \quad (24)$$

where:
$Y_{i,t}$ is the value of the dependent variable for the ith place and tth time,
$W_n Y_{i,t}$ is the n-dimensional spatial weighting matrix of the ith observation and time t,
$W_n Y_{i,t-1}$ is the n x n-dimensional spatial weighting matrix of the ith observation and time (t-1),
$Y_{i,t-1}$ is the response variable of the ith place at time (t-1),
$Y_{i,t-2}$ is the response variable of the ith place at time (t-2),
$X_{1i,t}$ is the value of the 1st independent variable for the ith place and time t,
$X_{2,i,t}$ is the value of the 2nd independent variable for the ith place and time t,
$V_{n,t}$ is the value of the disturbance [8], [23], [24]

2.9. Steps of the research

Data analysis was conducted using R Statistics software and other statistical packages. The research steps were as follows:

a. Panel data were collected in the form of the number of people who died from the coronavirus (Y), the number of people infected with the coronavirus (X1), and the number of tests conducted by country (X2) in European countries. Data sources were obtained from several websites.
b. The effects of time, place, and both were identified using the Lagrange Multiplier test. This step was performed to determine whether the 14 countries influenced coronavirus deaths. Furthermore, this step was used to determine whether the first week of March, the second week of March, and the fifth week of April 2020 influenced the number of coronavirus deaths. If there was an influence of time and country, the time and spatial components were included in the models. Focusing on 14 European countries with the most complete data during a critical early phase of the pandemic ensures high data quality and comparability. This targeted approach allows for robust analysis of COVID-19 mortality dynamics, providing reliable results that can inform future, broader studies.
c. Three panel data models were used: combined, fixed, and random [25]. This step was performed to select the most appropriate model for the research data analysis. Chow and Housman tests were used to select one of the three models.
d. Creation of the spatial weighting matrix (W).
e. A binary spatial weighting matrix was created based on the concept of queen-contiguity [26].
f. Identifying the presence of spatial lag and spatial error dependencies.
g. Spatial panel data models for SAR, SEM, GSM, and temporal data were built by incorporating the W component from Step 4.
h. The best model was determined based on the coefficient of determination (R2) values for the four models, see Equation (25).

$$R^2 = 1 - \frac{SSError}{SSTotal} = 1 - \frac{\Sigma(y_i - \hat{y}_i)^2}{\Sigma(y_i - \bar{y})^2} \quad (25)$$

The R^2 value ranges between 0 and 1, with values closer to 1 indicating better performance of the model. R^2 indicates the amount of variation in y that can be explained by the model [27].

3. RESULT AND ANALYSIS

3.1. Descriptive Statistics

The cumulative total number of coronavirus deaths for the fourteen (14) European countries observed can be classified into three (3) groups (Figure 2). Group 1, which had the highest number of deaths, included Italy and the United Kingdom. Group 2 is Belgium, which has fewer deaths than Italy and the United Kingdom but more than the other eleven (11) countries. Group 3 consisted of 11 countries with relatively low numbers of deaths.

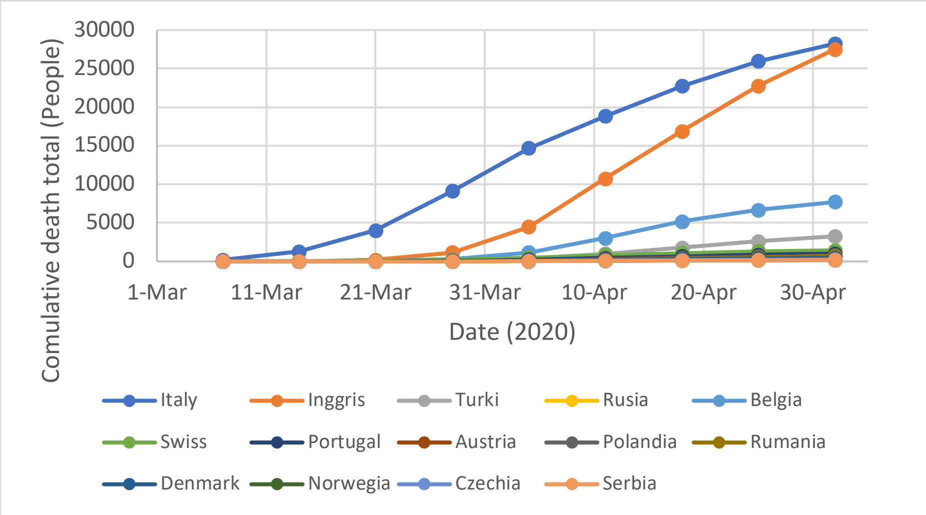


Figure 2. Cumulative total deaths due to coronavirus in 14 European countries from March 1 to May 2, 2020

3.2. Correlation between Variables

There was a correlation between the dependent (Y) and independent (X) variables, see Table 2. The correlation coefficient for Y and X1 was 0.791, with a p-value of <0.001. Because the correlation coefficient approaches +1 and p-value <0.05, Y and X1 are positively correlated at $\alpha = 0.05$. The correlation between X1 and X2 was 0.678. If the correlation between X1 and X2 is considered, they are positively correlated with rxy = 0.678 and a p-value of <0.001. The presence of a correlation between X1 and X2 indicates multicollinearity in the model if X1 and X2 are simultaneously entered into the model. Therefore, only X1 was entered into the model to avoid this assumption. The correlation method used in this study is highly valuable for public health because it provides robust, actionable insights for immediate decision-making, even when causality cannot be fully established. This approach ensures that policy responses are informed by real, statistically significant patterns in the data, which is crucial during rapidly evolving health crises.

Table 2. Pearson Correlation (rxy) for Dependent and Independent Variables

	Y	X1	X2
Y	1.000	0.791	0.687
		<.001	<.001
X1	0.791	1.000	0.255
	<.001		0.04
X2	0.678	0.255	1.000
	<.001	0.004	

3.3. Identification of the Influence of Time and Place

The influence of place, time, and time and place was measured using the Lagrange Multiplier (LM) test. This LM test is also called the Breusch-Pagan test. The results are shown in Table 3. Time, place, and place affected the number of coronavirus deaths at α=0.05. This can be seen from the p-value for the influence of time and place, which is less than 0.05.

Table 3. LM (Breusch-Pagan) Test for Identification of the Influence of Time and Place

Effects	Chi-Square	Degree of Fredom	p-value
Place and Time	42.668	2	<0.001
Place	40.963	1	<0.001
Time	1.705	1	0.192

3.4. Panel Data Model Selection

Chow and Hausman tests were used to select the most appropriate influence models. The null hypothesis of the Chow test is that the pooled model is more suitable than the fixed-effects model, while the alternative hypothesis is that the fixed-effects model is better than the pooled model. The null hypothesis of the Hausman test is that the random effects model is more suitable than the fixed effects model, and the alternative hypothesis is that the fixed effects model is better than the random effects model.

From Table 4, the p-value of the Chow test is less than 0.05, so H0 is rejected, meaning that the fixed effects model is more suitable. The Hausman test was then performed with a p-value of 0.128, which is less than 0.05; therefore, H0 is accepted, meaning that the random effects model is more suitable than the fixed effects model.

Table 4. Chow and Housman Tests for Model Selection

	Chow Test	Hausman Test
χ^2_{count}	-	4.119
F	5.601	-
Degree of freedom 1	13	2
Degree of freedom 2	110	-
p-value	<0.001	0.128

The selection results indicate that the random-effects panel model is the most suitable. A random-effects panel model was constructed, and the results are presented in Table 5.

Table 5. Parameter estimates for the random effects model

Variable	Estimation	Z	p-value
Intercept	25.576	0.159	0.874
X1	0.088	12.064	< 0.001
Coefficient of Determination (R2)		0.540	

The random model generated from Table 5 is in Equation (26).

$$y_{it} = 25.576 + 0.088x_{1it} + \varepsilon_{it} \tag{26}$$

3.5. Spatial Weighting Matrix (W)

The spatial weighting matrix plays a crucial role in the spatial panel model and is symbolized by W. The resulting W matrix is a 14 × 14 matrix because the countries involved in this study are the 14 European countries with the highest number of coronavirus cases and have data for Y, X1 and X2 (Figure 6). W is a component of spatial panel data.

Country i \ Neighbor j	1	2	3	4	5	6	7	8	9	10	11	12	13	14
1	0.00	0.00	0.00	0.00	0.00	0.50	0.00	0.50	0.00	0.00	0.00	0.00	0.00	0.00
2	0.00	0.00	0.00	0.00	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
3	0.00	0.00	0.00	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
4	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.50	0.00	0.00	0.50	0.00	0.00
5	0.00	0.50	0.00	0.00	0.00	0.00	0.50	0.00	0.00	0.00	0.00	0.00	0.00	0.00
6	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.50	0.00	0.00	0.00	0.00	0.00	0.50
7	0.00	0.00	0.00	0.00	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
8	0.33	0.00	0.00	0.00	0.00	0.33	0.00	0.00	0.00	0.00	0.00	0.00	0.33	0.00
9	0.00	0.00	0.00	0.33	0.00	0.00	0.00	0.00	0.00	0.33	0.00	0.00	0.33	0.00
10	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	1.00	0.00	0.00	0.00	0.00	0.00
11	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	1.00	0.00	0.00
12	0.00	0.00	0.00	0.50	0.00	0.00	0.00	0.00	0.00	0.00	0.50	0.00	0.00	0.00
13	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.50	0.50	0.00	0.00	0.00	0.00	0.00
14	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	1.00	0.00	0.00	0.00	0.00

Remarks:
1 = Italy, 2 = Inggris, 3 = Turki, 4 = Rusia, 5 = Belgia, 6 = Swiss, 7 = Portugal, 8 = Austria, 9 = Polandia, 10 = Rumania, 11 = Denmark, 12 = Norwegia, 13 = Czechia, 14 = Serbia

Figure 3. Spatial Weighting Matrix W

The procedure for creating the W matrix is as follows: (1) Create a 14 × 14 matrix C, where each row represents a country and each column represents its neighbors. (2) Enter 1 for neighboring countries; for example, Row 1 represents Italy. Italy's neighbors are Switzerland (number 6) and Austria (number 8); therefore, row 1 is assigned 1 in columns 6 and 8, respectively. This step was repeated for rows 2 through 14. For countries that do not have neighbors in direct contact, it is assumed that they have direct contact with the closest country. (3) C, where the total number in each row is made 1, is called W For example, row 1 for columns 6 and 8 is made 0.5 each so that the total is 1, do step 3 for rows 2 to row 14. The results are presented in Figure 3.

3.6. Spatial Influence Test

A Lagrange Multiplier (LM) test was conducted to examine the spatial dependence of the data. The p-values for the spatial lag and spatial error were less than 0.05 (Table 6), indicating a spatial relationship between the lag and error. Because of the relationship between lag and error, SAR, SEM, GSM, and spatiotemporal panel data models were constructed.

Tabel 6. Spatial Effect Test

LM test	LM score	p-value
Spatial lag	6.350	0.012*
Spatial error	8.360	0.004*

3.7. SAR, SEM, and GSM Models

The SAR spatial panel data model has spatially correlated dependent variables. In this study, the SAR model uses a random effects panel model in accordance with the Hausman test results in §4.4 and a two-way effect (time and place) in accordance with the LM test in §4.3. Parameter estimates are presented in Table 7.

Table 7. Panel Data Model using SAR

Variable	Parameter Est.	t	p-value
ρ	0.033	0.397	0.691
Intersep	15.345	0.093	0.926
X1	0.0873	12.116	<0.001
R2	0.611		

The SEM spatial panel data model contains spatial correlation error. In this study, the panel model uses a random effects model in accordance with the Hausman test results in Section 4.4, while the two-way effect, or place and time, is in accordance with the LM test in Section 4.3. Parameter estimates for this SEM can be seen in Table 8.

Table 8. Panel Data Model using SEM

Variable	Parameter Est.	t	p-value
λ	-0.072	-0.776	0.438
Intersep	27.130	0.167	0.868
X1	0.087	12.296	<0.001
R2	0.612		

The GSM spatial panel data model has spatially correlated dependent variables and error terms. In this study, the GSM model uses panel data with random effects according to the Hausman test in Section 4.4 and two-way effects according to the LM test in Section 4.3. The parameter estimates are presented in Table 9.

Table 9. Panel Data Model using GSM

Variable	Parameter Est.	t	p-value
ρ	0.092	0.093	0.323
λ	-0.130	-1.181	0.237
Intercept	-1.655	-0.010	0.992
X1	0.086	12.354	<0.001
R2	0.610		

From Tables 7, 8, and 9, SEM is the most accurate because it has the highest coefficient of determination. The SEM model is in Equation (27) and (28).

$$y_{it} = 27{,}130 + 0{,}087x_{1it} + \phi_{it} \tag{27}$$

$$\phi_{it} = -0{,}072\sum_{j=1}^{9} w_{ij}\phi_{it} + \varepsilon_{it} \tag{28}$$

where wij is the weight of country i and j is the neighbor.

3.8. Spatial-Temporal Panel Data Model

The spatiotemporal panel data model contains spatial influence components, autoregression, and independent variables (IVs). The results of the model estimation are shown in Table 10. Based on Table 10, the coronavirus death rate is influenced by the number of coronavirus cases in the country, and there is an autoregression of Yt-1 and Yt-2. The model is also influenced by spatial influences, as indicated by WY and Yt-1.

Table 10. Spatial Dynamic Panel Data Estimator

Variable	DF	Estimation	Standard Deviation	t	p-value
WY	1	-0.066	0.001	-120.44	<.0001
WYt-1	1	0.334	0.005	64.88	<.0001
Yt-1	1	0.749	0.002	323.57	<.0001
Yt-2	1	-0.409	0.001	-3616.80	<.0001
X1	1	1.042	0.012	85.53	<.0001
			R2 = 0.639		

The resulting spatiotemporal panel data model is in Equation (29).

$$Y_{it} = -0{,}066\ W_i Y_{i,t} + 0{,}334 WY_{i,t-1} + 0{,}749 Y_{i,t-1} - 0{,}409 Y_{i,t-2} + 1{,}042 X_{it} \quad (29)$$

3.9. Best Model

When comparing the coefficients of determination of the four (4) models, the spatial-temporal panel data model has the highest value (Table 11). Therefore, the spatial-temporal panel data model was selected as the best model for this study and is described below. Overfitting can be avoided by using strict statistical tests and objective criteria to select the best model, ensuring that only the most reliable and interpretable results are presented. This approach makes the findings robust and clearly identifies which model provides the most useful insights. An R-squared of 64% is a strong enough result for real-world pandemic modeling, showing that the model captures the main drivers of COVID-19 deaths with reliable, comparable data. This provides valuable, actionable insights for public health, even as some external factors remain outside the model's scope.

Table 11. Model Comparison

Models	R^2
Random effect model	0.540
Random effect SAR	0.612
Random effect SEM	0.612
Random effect GSM	0.609
Temporal Spatial Panel Data	0.639

4. CONCLUSION

The spatiotemporal panel data model was the best panel data model of the four models developed. The dependent variable and the dependent variable with lag one are spatially correlated. There is an autoregression with response variables at lags one and two. The variable influencing the number of new coronavirus deaths in a country is the number of new coronavirus cases. The coefficient of determination for this model was 63.9%.

The study's findings offer policymakers a robust, evidence-based model for predicting and managing COVID-19 mortality, enabling more effective and timely public health interventions. This approach supports better resource allocation and targeted responses, directly strengthening pandemic policy and preparedness. Future research could extend this approach to other infectious diseases or apply it to different regional datasets to enhance epidemic preparedness and response strategies.

5. REFERENCES

[1] S. Feng, T. Tong, and S. N. Chiu, “Statistical Inference for Partially Linear Varying Coefficient Spatial Autoregressive Panel Data Model,” Mathematics, vol. 11, Nov. 2023, doi: 10.3390/math11224606.

[2] K. Murayama and T. Gfrörer, “Thinking clearly about time-invariant confounders in cross-lagged panel models: A guide for choosing a statistical model from a causal inference perspective.,” Psychol Methods, Sep. 2024, doi: 10.1037/met0000647.

[3] J. Chen, Q. Feng, and D. Fan, “Vehicle Trajectory Prediction Based on Local Dynamic Graph Spatiotemporal–Long Short-Term Memory Model,” World Electric Vehicle Journal, vol. 15, Jan. 2024, doi: 10.3390/wevj15010028.

[4] B. Jin, Y. Wu, C. R. Rao, and L. Hou, “Estimation and model selection in general spatial dynamic panel data models,” Proceedings of the National Academy of Sciences, vol. 117, Feb. 2020, doi: 10.1073/pnas.1917411117.

[5] J. P. Elhorst, “The dynamic general nesting spatial econometric model for spatial panels with common factors: Further raising the bar.,” Review of Regional Research, vol. 42, Nov. 2021, doi: 10.1007/s10037-021-00163-w.

[6] R. H. El-Sokkary et al., “Characteristics and predicting factors of Corona Virus Disease-2019 (COVID-19) among healthcare providers in a developing country.,” PLoS One, vol. 16, Jan. 2021, doi: 10.1371/journal.pone.0245672.

[7] R. Lembo, L. Cianfanelli, G. Landoni, and A. Frontera, “Air pollutants and SARS-CoV-2 in 33 European countries.,” Acta Bio Medica : Atenei Parmensis, vol. 92, Jan. 2021, doi: 10.23750/abm.v92i1.11155.

[8] A. K. Bera, O. Doğan, S. Taşpınar, and Y. Leiluo, “Robust LM tests for spatial dynamic panel data models,” Reg Sci Urban Econ, vol. 76, Sep. 2018, doi: 10.1016/j.regsciurbeco.2018.08.001.

[9] D. Arifatin, “Analisis Spasial Data Panel untuk Mengkaji Faktor-faktor yang Memengaruhi Kunjungan Wisatawan Mancanegara di Indonesia Tahun 2011-2015,” IPB University, Bogor, 2018.

[10] Our World in Data, “Data on COVID-19 by Our World in Data,” https://github.com/owid/covid-19-data/tree/master.

[11] V. Kontis et al., “Lessons learned and lessons missed: impact of the coronavirus disease 2019 (COVID-19) pandemic on all-cause mortality in 40 industrialised countries and US states prior to mass vaccination.,” Wellcome Open Res, vol. 6, Feb. 2022, doi: 10.12688/wellcomeopenres.17253.2.

[12] WHO, “Europe Health Information Gateway ,” https://gateway.euro.who.int/en/indicators/hfa_471-5011-number-of-hospitals/#id=19525.

[13] M. H. Pesaran, “General Diagnostic Tests for Cross Section Dependence in Panels,” Empir Econ, vol. 60, May 2020, doi: 10.17863/cam.5113.

[14] D. Yuan, M. A. I. Gazi, I. Harymawan, B. K. Dhar, and A. I. Hossain, “Profitability determining factors of banking sector: Panel data analysis of commercial banks in South Asian countries,” Front Psychol, vol. 13, Nov. 2022, doi: 10.3389/fpsyg.2022.1000412.

[15] D. T. Nguyen, “Parametric Identification of Electric Drives Using the Ordinary Least Squares Method,” Jan. 2021, doi: 10.1109/elconrus51938.2021.9396566.

[16] Y. R. Lee and J. E. Pustejovsky, “Comparing random effects models, ordinary least squares, or fixed effects with cluster robust standard errors for cross-classified data.,” Psychol Methods, vol. 29, Dec. 2024, doi: 10.1037/met0000538.

[17] - Joshi and - Wooldridge, “Correlated Random Effects Models with Endogenous Explanatory Variables and Unbalanced Panels,” Ann Econ Stat, Jan. 2019, doi: 10.15609/annaeconstat2009.134.0243.

[18] J. Suteja, E. S. Alghifari, A. Lestari, A. A. Susiadi, A. Gunardi, and A. S. Yulianti, “Investment Decision and Firm Value: Moderating Effects of Corporate Social Responsibility and Profitability of Non-Financial Sector Companies on the Indonesia Stock Exchange,” Journal of Risk and Financial Management, vol. 16, Jan. 2023, doi: 10.3390/jrfm16010040.

[19] K. Suryowati, A. Faradila, and R. D. Bekti, “A Comparison of Weights Matrices on Computation of Dengue Spatial Autocorrelation,” IOP Conf Ser Mater Sci Eng, vol. 335, Apr. 2018, doi: 10.1088/1757-899x/335/1/012052.

[20] W. Sun et al., “Geospatial Analysis of Urban Expansion Using Remote Sensing Methods and Data: A Case Study of Yangtze River Delta, China,” Complexity, vol. 2020, Aug. 2020, doi: 10.1155/2020/3239471.

[21] Z. Zhang, Z. Li, and Y. Song, “On ignoring the heterogeneity in spatial autocorrelation: consequences and solutions,” International Journal of Geographical Information Science, vol. 38, Aug. 2024, doi: 10.1080/13658816.2024.2391981.

[22] F. R. Pinto, C. F. C. Marcellos, C. Manske, and A. Gomes Barreto, “Statistical analysis of parameters and adsorption isotherm models.,” Environ Sci Pollut Res Int, vol. 31, Feb. 2024, doi: 10.1007/s11356-023-31820-x.

[23] S. Cheng, P. Peng, and F. Lu, “A lightweight ensemble spatiotemporal interpolation model for geospatial data,” International Journal of Geographical Information Science, vol. 34, Feb. 2020, doi: 10.1080/13658816.2020.1725016.

[24] S. Xia et al., "GCN-LSTM Based Transient Angle Stability Assessment Method for Future Power Systems Considering Spatial-Temporal Disturbance Response Characteristics," Protection and Control of Modern Power Systems, vol. 9, Nov. 2024, doi: 10.23919/pcmp.2023.000116.

[25] M. T. Majeed and I. Ozturk, "Environmental degradation and population health outcomes: a global panel data analysis.," Environmental Science and Pollution Research, vol. 27, Feb. 2020, doi: 10.1007/s11356-020-08167-8.

[26] Y. Chen, "Spatial autocorrelation equation based on Moran's index," Sci Rep, vol. 13, Nov. 2023, doi: 10.1038/s41598-023-45947-x.

[27] D. Uzun Ozsahin, E. Precious Onakpojeruo, B. Bartholomew Duwa, A. G. Usman, S. Isah Abba, and B. Uzun, "COVID-19 Prediction Using Black-Box Based Pearson Correlation Approach," Diagnostics, vol. 13, Mar. 2023, doi: 10.3390/diagnostics13071264.