PAPER

Title: A spatial econometric analysis of COVID-19 infection rates and socio-economic factors: Evidence from Spain (*Un análisis econométrico espacial de las tasas de infección por COVID-19 y sus factores socioeconómicos: Evidencia para España)*

Authors and e-mails of them: Irene González Rodríguez (gonzalrodri@unican.es), Marta Pascual Sáez (marta.pascual@unican.es) y David Cantarero Prieto (david.cantarero@unican.es)

Department: Departamento de Economía

University: Universidad de Cantabria

Subject area: *(please, indicate the subject area which corresponds to the paper)*
2. Desigualdad y pobreza de los territorios

Abstract: *(maximum 300 words)*

The impact of the COVID-19 pandemic imposes large economic and social costs on many societies. However, its impact is not the same between countries and between regions. The aim of this study is to assess whether the infection rate depends on socioeconomic factors and whether there are spatial interactions between the Spanish regions. To this end, the Moran's test and spatial econometric models are used. The results suggest that the cumulative rate of confirmed COVID-19 cases shows a positive spatial autocorrelation and tends to form spatial clusters. In addition, the COVID-19 mortality rate, the COVID-19 intensive care unit admission rate, the population density, the illiteracy rate, the unemployment rate and the percentage of people with problems arising from pollution are found to influence the spread of the coronavirus. This study suggests that the links between regions matter. Our analysis can be considered in order to apply policy intervention measures to reduce health population inequalities.

Keywords: *(maximum 6 words) pandemic, COVID-19, spatial econometrics, socioeconomic inequalities.*
JEL codes: I14

1. Introduction

In January 2020, the World Health Organization declared the coronavirus outbreak an international public health emergency. Since then, many countries have been struggling to put an end to it. Although the impact of the pandemic has differed not only from country to country, but also from one region to another, it is imposing huge economic, social, employment and health costs on multiple societies. Even before this crisis, many authors were already stressing the enormous damage that a global pandemic could cause. The World Bank suggested that it could reduce the world's gross domestic product by 5% Jonas (2013), while the World Health Organization estimated the expected annual losses at 6% of global revenues (WHO, 2018). Despite these warnings, societies were not prepared in advance to deal with this COVID-19 health crisis. This has left, citizens in a constant search for solutions to minimize the impact of this unprecedented pandemic.

However, the crisis affects not only the economy, but also health, both physical and mental. On the one hand, people are afraid of the possible direct effects of a COVID-19 infection: worsening of chronic diseases causing of lasting disabilities and even death. On the other hand, all the measures taken to curb the coronavirus spread (quarantines, closure of businesses, social distancing) are disrupting people's lives (Saha et al., 2020). Uncertainty leads to anguish, stress and worry, which can develop into profound mental problems, such as anxiety or depression, or even sleep or eating disorders, and can affect individuals' physical health (McGinty et al., 2020).

Furthermore, the spread of the coronavirus is not random, but depends on many factors (Bhatnagar, 2020). This has led many researchers to study the spatial relationships involved in that spread and the factors that influence it. For instance, Shobande and Ogbeifun (2020) present maps showing that spatial changes in the spread of the coronavirus are positively related to its mortality rates in 79 different countries. Another example is that of Paez et al. (2020), who study the incidence of the virus in Spain and its links with different socio-economic indicators. There are also studies at regional level. De Cos et al. (2020) suggest that, in the Spanish region of Cantabria, the areas most vulnerable to pandemic outbreaks are those with the greatest numbers of people linked to one and other through social, economic and travel relationships.

The aim of this study is to present an empirical analysis to assess the confirmed case rates of COVID-19 and the socioeconomic factors that may influence it, taking into account possible spatial dependencies between Spanish regions. In this sense, we ask whether social inequalities affect COVID-19 results. The initial hypothesis is that there is spatial autocorrelation in the sample and that socioeconomic inequalities influence the spread of the virus.

The paper is organized as follows: Section 2 reviews the existing literature, Section 3 describes the data used in the empirical analysis, Section 4 presents the methodology, Section 5 shows the main results of the analysis, Section 6 discusses these results and Section 7 presents the study's findings.

2. Literature review

Even though the pandemic emerged relatively recently, there is already a large body of literature on it. Seeking to understand whether the spread of the virus depends on socio-economic factors and whether spatial interactions between regions play a role, we review papers that use spatial econometrics to answer their questions and hypotheses.

First, we review studies on the European Union (EU). Amdaoud et al. (2021) analyse the spatial heterogeneity of the spread of the coronavirus to understand its transmission channels. Exploratory analyses of deaths up to 31 May 2020 indicate the presence of high-high clusters in northeast France, northwest Spain, southeast and central England and Wales. These high-high clusters are geographical areas that exhibit positive spatial autocorrelation, i.e regions with high numbers of deaths tend to be surrounded by other regions with similar figures. However, low-low clusters are found in Denmark, eastern Austria and Germany, western France and southern Italy. On the one hand, results from spatial regression models suggest that high COVID-19 mortality rates are positively correlated with per capita Gross Domestic Product, unemployment, distance from attainment of EU targets and the rate of population ageing. However, social trust and the number of hospital beds and doctors are correlated with low mortality rates. Thus, wealth, income and public health policies can explain disparities between regions. Sannigrahi et al. (2020) also study the spatial relationship between COVID-19 outcomes and socio-demographic characteristics, but in this case in 31 European countries. They find a spatial association between COVID-19 deaths and selected socio-demographic variables (poverty, income and population), especially in Austria, Slovenia, Italy, Germany and Switzerland. They indicate as a suggested future line of research that it would be interesting to include some environmental factors in the analyses.

Focusing on single-country studies, Ehlert (2020) explores the spatial relationships between COVID-19 positive cases and death rates in 401 German counties up to mid-June 2020. He finds sufficient evidence that infections and deaths are positively and significantly related to median age, the number of people working in care of the elderly, early cases since the beginning of the pandemic and population density. Likewise, Sun et al. (2021) examine the spatial relationship between inequalities in COVID-19 mortality rates and disparities in environmental and socioeconomic factors in England. Their results suggest greater spatial inequality in COVID-19 mortality than in mortality from other causes. Relative humidity and accessibility to hospitals are negatively related to COVID-19 deaths, while the unemployment rate is positively related. Mena et al. (2021) find a strong association between socioeconomic status and COVID-19 mortality in Chile and show that mortality rates in young people are lower in high-income municipalities than in low-income municipalities. Their results indicate that socioeconomic inequalities have major consequences for health outcomes.

Next, we review studies on China. Guliyev (2020) examines the spatial factors and effects of the coronavirus, using spatial panel data models to show the relationship between COVID-19 positive cases with deaths and recoveries in 31 regions of China. Measuring the total effects, it is found that the rate of infection is positively related to the death rate and negatively related to the recovery rate. Focusing on the direct effects, a one percentage point increase in the death rate implies a 32% increase in the infection rate. Indirect effects suggest that the mortality rate in a given region has positive effects on neighboring regions. In this sense, a one percentage point increase in the mortality rate leads to a confirmed positive change of 1.7% points in neighboring regions. Li et al. (2021) study the spatial impact of public health spending on regional economic growth in 31 Chinese provinces during the COVID-19 pandemic. Estimates from three different spatial econometric models suggest that there is spatial clustering and that public health spending is correlated with regional economic growth.

Finally, we look at published studies on the United States (US). Baum and Henry (2020) analyse daily changes in COVID-19 positive cases at county level there to determine whether there are spatial dependencies between neighbors. They combine spatial econometric methodology with socioeconomic factors (gender, age, ethnicity, income, pollution, health insurance and complicated health conditions) and find evidence that these variables contribute to the spatial spread of the virus. Sun et al. (2020) study the spatial relationships of COVID-19 prevalence also in US counties. The authors argue that ordinary least squares estimates overestimate the virus prevalence in counties with small observed rates. They solve this problem by using spatial models that can assess some of the geographic disparities in the spread of the virus. Other authors, such as Yang et al. (2021), examine inequalities in confirmed COVID-19 rates in New York City. Bayesian spatial modelling estimation shows that confirmed COVID-19 case rates are positively related to racial minority groups, household size and the elderly population, and negatively related to the number of teleworkers. The areas with the highest infection rates are the Bronx and Queens, while the greatest spatial effects are found in Manhattan and Brooklyn.

Table 1: Studies reviewed.

Study	**Country**	**Data year**	**Spatial model**	**Results**
Amdaoud et al. (2021)	12 European countries	2014 - 2020	SAR	Spatial clusters exist. Income and public health policies are able to explain disparities across regions.
Baum and Henry (2020)	United States	2020	SAR	There are socio-economic factors that favour the spread of the virus
Ehlert (2020)	Germany	2016 - 2020	SAR, SEM and SAC	Infections and deaths are positively related to age, work in care for the eldery and population density.
Guliyev (2020)	China	2005 - 2019	SAR, SEM, SAC and SDM	An increase in mortality in a given region causes the mortality rate in neighbouring regions to rise.
Li et al. (2021)	China	2019 - 2020	SDM	There is spatial clustering and public health spending is correlated with regional economic growth.

Mena et al. (2021)	Chile	Not available	Bayesian spatial model	COVID-19 mortality rates are lower in high-income municipalities than in low-income municipalities.
Sannigrahiet al. (2020)	12 European countries	2014 – 2020	SAR and SEM	There is a spatial association between COVID-19 deaths and poverty, income and population.
Sun et al. (2020)	United States		SAR, SEM and SAC	There are spatial relationships in the prevalence and spread of the virus.
Sun et al. (2021)	United Kingdom	2014 - 2020	SAR and ESF	There is greater spatial inequality in COVID-19 mortality than in mortality from other causes.
Yang et al. (2021)	United States	2014 - 2020	Bayesian spatial model	Confirmed COVID-19 rates are positively related to racial minority groups and household size.

*SAR (spatial autoregression model), SER (spatial error model), SDM (spatial durbin model), ESF (eigenvector spatial filtering).

Source: Authors' own work

3. Data

According to the World Health Organization (WHO, 2021), Europe has the second highest number of COVID infections[1] in the world among large geographical areas (51,912,774). It is followed by South-East Asia (22,675,230), the Eastern Mediterranean (9,147,412) and Africa (3,316,851). At the top of the ranking are the Americas with 62,281,517 infections and at the bottom is the Western Pacific with 2,469,292. By countries, the top 3[2] in terms of cumulative numbers of positives cases are the United States (32,002,328), India (19,557,457) and Brazil (14,659,011). Spain is in ninth place with 3,514,942.

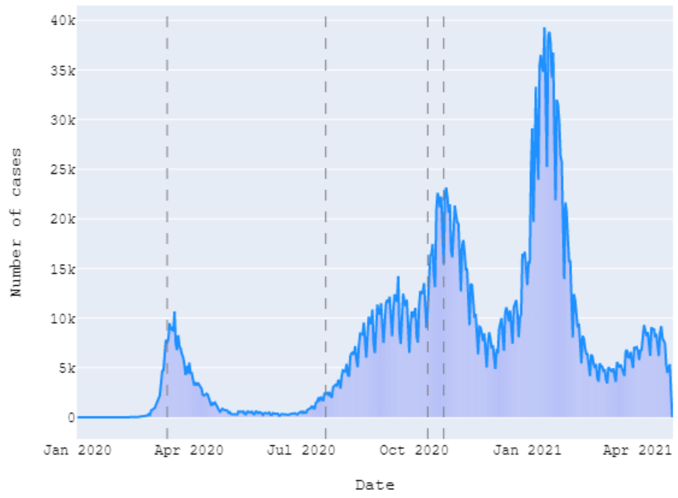
Given the large number of people infected in Spain, the geographical scope take for this study covers all 17 Spanish regions and the two autonomous cities. Given that the study is carried out on the COVID-19 pandemic, the time period for data varies between 2019

[1] Data as of 26 April 2021
[2] Data as of 3 May 2021

and 2021, depending on each variable. The spatial data are extracted from GADM (2021).

As in Guliyev (2020), Ehlert (2020) and Baum and Henry (2020), the dependent variable studied here is the cumulative rate of confirmed COVID-19 cases, measured as a percentage (*positiverate*). We calculated this rate from the total number of confirmed cases up to 9 April 2021[3], extracted from the Carlos III Health Institute (referred to by its Spanish acronym ISCIII) (ISCIII, 2021), and the total population in 2020 (measured in numbers of individual), taken from the National Statistical Institute (referred to by its Spanish acronym INE) (INE, 2021e). To provide an idea of the trend in infections in Spain, Figure 1 shows the new confirmed cases of COVID-19 in the country[4] from January 2020 to April 2021[5].

Figure 1: COVID-19 trend in new cases per day in Spain.



Source: Authors' own work based on ISCIII (2021)

Following the literature review, the independent variables used by other researchers have been compiled and adapted to this study. We have selected the following:

- The total COVID-19 mortality rate measured in percentage (*mortalityrate*), derived using total COVID-19 deaths up to 9 April 2021[6] drawn from ISCIII (2021), and of the total population in 2020 (measured in numbers of individuals), taken from INE

[3] Last available update when data were extracted.

[4] The first dashed vertical line corresponds to strict lockdown (first alarm stage) on 14 March 2020, the second to the end of the first alarm stage (21 June 2020), the third to the second alarm stage (25 October 2020) and the last one to the start of the COVID-19 vaccination (27 December 2020).

[5] Latest available update when data were extracted.

[6] Latest available update when data were extracted.

(2021e). It was decided to include this independent variable as done by Guliyev (2020), Amdaoud et al. (2021) and Maiti et al. (2021).

- The rate of admissions to Intensive Care Units (ICU) with COVID-19, measured in percentage (*icurate*) is calculated from the total number of cases admitted to ICUs up to 9 April 2021[7], drawn from ISCIII (2021), and of the total population in 2020 (measured in numbers of individuals), taken from INE (2021e). Flaxman et al. (2020) also include this in their analysis.

- The population density (*popden*) is the total population per square kilometer as a logarithm. It is obtained from the total population for 2020 (measured in numbers of individuals), taken from INE (2021e), and the surface area of each autonomous community (in square kilometers) taken from INE (2021b). This indicator is included in the same way that Amdaoud et al. (2021), Ehlert (2020) and Yang et al. (2021).

- The illiteracy rate (*illiteracyrate*) is the proportion of the total population of each community in the last quarter of 2020 aged 16 and over who are illiterate, taken from INE (2021b). Ehlert (2020) includes a similar variable, which characterizes the level of education of individuals.

- The life expectancy at birth (*leb*) (measured in years) in 2019[8] is drawn from INE (2021a). Ehlert (2020) also includes this variable in his study.

- Population with long-term illnesses or health problems, measured as a percentage of persons aged 16 and over (*longtermd*). Data are for 2019[9] and are from INE (2021c). Ehlert (2020) uses a similar variable in his analysis.

- Unemployment rate in 2019[10] (*unemploymentrate*), taken from INE (2021g). This indicator is included as in Amdaoud et al. (2021).

- Population suffering from pollution and other environmental problems (*pollution*) in 2019[11] measured in percentage and extracted from INE (2021f). This variable is included as recommended by Baum and Henry (2020).

[7] Latest available update when data were extracted.
[8] Latest year available.
[9] Latest year available.
[10] 2019 was chosen instead of 2020 so that the figure does not reflect unemployment caused by the COVID-19 pandemic.

Table 2: Variables selected for macroeconomic analysis.

Variables	**Description**	**Source**	**Year**
COVID-19 results			
positiverate	Cumulative COVID-19 positive cases as of 9 April 2021 out of total population in 2020	ISCIII and INE	2021
mortalityrate	Cumulative deaths from COVID-19 infection up to 9 April 2021 out of total population in 2020	ISCIII and INE	2021
icurate	Cumulative ICU admissions due to COVID-19 infection up to 9 April 2021 out of total population in 2020	ISCIII and INE	2021
Health			
leb	Average life expectancy in number of years from birth, if the mortality pattern for the period observed is maintained	INE	2019
longtemrd	Population with chronic diseases, measured as a percentage of people aged 16 years and over	INE	2019
Socioeconomic factors			
illiteracyrate	Illiterate population aged 16 and over as a proportion of total population	INE	2020
popden	Total population per square kilometer in logarithm	INE	2020
unemploymentrate	Ratio of the number of unemployed to the number of working-age persons	INE	2019
Environmental factor			
pollution	Population suffering from pollution and other environmental problems	INE	2019

Source: Authors' own work

[11] Latest year available.

4. Methods

This section describes the main tools used in the empirical analysis. We use a spatial econometric methodology, since it is important to take into account possible interactions between autonomous communities when using data on them (Serrano and Valcarce, 2000).

4.1 Spatial autocorrelation, weight matrix and Moran's Test

Although researchers detected a decay effect in distance (Zipf, 1949) as early as 1800, it was not until the 1960s that the term “spatial autocorrelation” was introduced by authors such as Paelinck (1967) and Cliff (1969). In 1979, Paelinck suggested the term “spatial econometrics” to refer to the union of the economic, statistical and mathematical theory that encompasses all regional and urban modelling problems (Paelinck et al., 1979).

Currently, spatial autocorrelation is the basis of the field of spatial econometrics. Getis (2007) outlines the many advantages of analysing this autocorrelation, which can be summarised as follows:

- It brings to light evidence of any model misspecification.
- It shows spatial dependence relationships between variables and establishes the strength of spatial effects.
- It is able to focus on a single spatial unit to study what effect it may have on neighbouring units, and vice versa.
- It tests stationarity and spatial heterogeneity.
- It enables hypotheses about spatial interactions to be tested.

Spatial autocorrelation can be understood as the dependence of observations on different geographical areas (Fotheringham, 2009). Griffith (1987) defines it as *“the correlation between the values of the same variable strictly attributable to their relatively close position on a two-dimensional surface”.* This spatial autocorrelation is positive if the values of the variable of interest tend to be similar in nearby geographical units (i.e low values are close to other low values, medium values are close to medium values, and high values are close to high values). However, spatial autocorrelation is negative if geographically close values of a variable are not similar on a map. Finally, it is zero if the phenomenon studied is completely random in space. It is worth noting that most indicators studied in the social sciences tend to be positively spatially autocorrelated (Griffith, 1987).

In this regard, an essential prerequisite for studying the autocorrelation of a variable in a specific geographical area is to know the distance and the geographical relationship between some spatial units and others. This geographical connectivity is expressed through the weights matrix, denoted by *W*. This is a matrix with *n* rows and *n* columns, whose elements indicate which areas are neighbors and which are not, and whose main diagonal always takes a value of zero to prevent an observation from being defined as a neighbor of itself. It is normalized by dividing all the elements of the matrix by the number of neighbors so that the sum of its rows is one (LeSage, 2008). The criteria for determining neighborhood between places are many:

- Geographical criteria: contiguity criteria on border sharing (linear, rook, bishop, queen), nearest-neighbor criteria to give each unit the same number of neighbors (k-nearest-neighbors), distance criteria (a unit is a neighbor if it is closer or further away than a given distance) and gravitational criteria.
- Economic criteria: similarities between economic indicators.
- Others: depending on the study.

For this analysis, the geographical contiguity criterion is chosen. Therefore, an element takes a value of 1 if autonomous communities share a geographical border and 0 if they do not. We denote the weight matrix *W* such that:

$$\mathbf{W} = \begin{pmatrix} w_{11} & w_{12} & \dots & w_{1j} \\ w_{21} & w_{22} & \dots & w_{2j} \\ \dots & \dots & \dots & \dots \\ w_{i1} & w_{i2} & \dots & w_{ij} \end{pmatrix} \tag{1}$$

where w_{ij} are the different weights between each pair of regions i and j. The main diagonal is replaced by zeros:

$$\mathbf{W} = \begin{pmatrix} 0 & w_{12} & \dots & w_{1j} \\ w_{21} & 0 & \dots & w_{2j} \\ \dots & \dots & \dots & \dots \\ w_{i1} & w_{i2} & \dots & 0 \end{pmatrix} \tag{2}$$

Once this is applied to the data under study, it must be standardised so that the sum of its rows is one.

There are several statistics that enable the spatial autocorrelation of the sample to be analysed. In this case, it is contrasted with Moran's test and its associated statistic *I*, to learn what type of econometric model should be used in the regression analysis (traditional model or spatial model). Moran (1948) asks whether the presence of some specific characteristic in a geographical unit makes its presence in neighboring units more or less probable or, on the other hand, whether the elements that cause a phenomenon are statistically independent.

In this sense, the null hypothesis of Moran's test is that the data are randomly distributed, as opposed to the alternative hypothesis that there is spatial autocorrelation. Based on Serrano and Valcarce (2000), we use the Moran's *I* statistic, which can be interpreted similarly to a covariance:

$$\mathrm{I} = \frac{\mathrm{N}}{\mathrm{S_0}} \cdot \frac{\Sigma_{\mathrm{ij}}^{\mathrm{N}} \mathrm{w_{ij}} \left(\mathrm{x_i} - \overline{\mathrm{x}}\right)\left(\mathrm{x_j} - \overline{\mathrm{x}}\right)}{\mathrm{S_i^N} \left(\mathrm{x_i} - \overline{\mathrm{x}}\right)^2} \quad \mathrm{i} \neq \mathrm{j} \tag{3}$$

where S_0 is the scale factor equal to the addition of the weights, N is the sample size, x_i is the value of the variable of interest x in the region i and $\bar{x}$ is the sample mean of the variable x. Since the weight matrix is standardised, N is equal to S_0.

Its interpretation is linked to its expected value (the negative value of the unit divided by the sample size minus one). Formally it is written as follows:

$$E(I) = -\frac{1}{N-1} \tag{4}$$

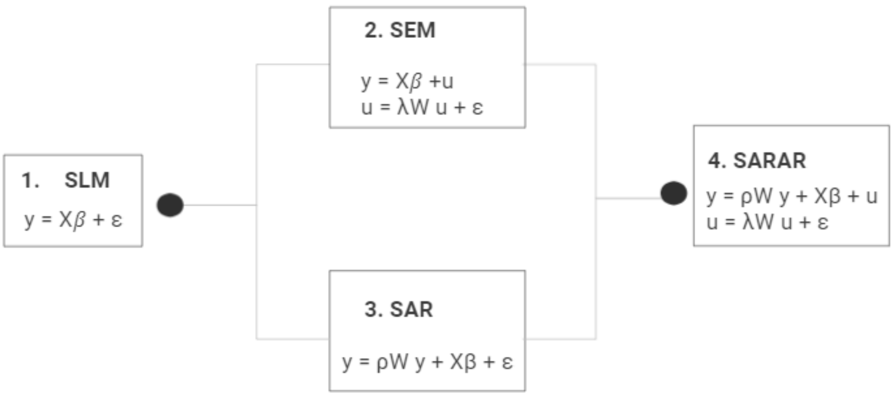
If the statistic *I* takes a value greater than its expected value *E(I)*, the variable of interest has positive spatial autocorrelation. However, if the opposite is true ($I < E(I)$), the spatial autocorrelation of the sample is negative.

<u>4.2 Spatial models</u>

Depending on where the spatial dependence of the data is assumed to lie, one spatial model or another can be used. Interpreting the spatial models coefficients as if they were partial derivatives (as in standard linear regression models, SLM) is correct only for the spatial error model (SEM) (LeSage and Dominguez, 2012). Additional nuances are required to interpret the coefficients of models with spatial dependence on the dependent variable or on the independent variables such the spatial autoregression model (SAR), the spatially-lagged X model (SLX), the mixed spatial autoregression spatial error model (Kelejian–Prucha, SAC or SARAR), the spatial Durbin models (SDM) and the spatial Durbin error model (SDEM).

This study focuses only on the SLM, SEM, SAR and SAC models. The transition from one model to other can be seen in Figure 2. Starting from an SLM with no spatial dependencies, it is possible to obtain an SAR model by including endogenous spatial lags (ρW_y) or an SEM by including spatially lagged errors (λW_u) is added in the error term). Finally, an SAC model can be obtained from the SAR model, which has both endogenous and error lags, by adding spatially lagged errors or from the SEM by introducing endogenous spatial lags.

Figure 2: Some spatial models.



Source: Authors' own work based on Golgher and Voss (2016)

Although the SLM does not include spatial interactions, its use is recommended to present data spatial patterns and their possible misspecification (Golgher and Voss, 2016). According to LeSage and Pace (2009), this model can be denoted as:

$$y = \alpha i_n + X\beta + \varepsilon,\ \varepsilon N(0, \sigma^2 I_n) \tag{5}$$

where y is the dependent variable vector, α represents the intercept coefficient with which an n-dimensional column vector of ones i_n is associated, X represent the matrix of independent variables, β is the regression slope coefficients and ε denotes the white noise error.

The spatial error model (SEM) assumes that the spatial dependence is in the spatial error (LeSage, 2008). This model is defined such that:

$$\begin{aligned} y &= X\beta + u; \\ u &= \lambda W u + \varepsilon \end{aligned} \tag{6}$$

where u is the error term, λ is the spatial coefficient and W is the weight matrix.

The *y* expected value and the partial derivatives of the two models above coincide, so SEM can be interpreted in the same way as SLM.

$$\mathrm{E}(y \mid X) = \alpha i_n + X\beta \tag{7}$$

$$\begin{pmatrix} \frac{\partial y_1}{\partial x_{1k}} & \cdots & \frac{\partial y_1}{\partial x_{nk}} \\ \cdots & \cdots & \cdots\cdot \\ \frac{\partial y_n}{\partial x_{1k}} & \cdots & \frac{\partial y_n}{\partial x_{nk}} \end{pmatrix} = \begin{pmatrix} \beta k & \cdots & 0 \\ \cdots & \cdots & \cdots \\ 0 & \cdots & \beta_k \end{pmatrix} = \beta_k I_n \tag{8}$$

Based on the Golgher and Voss (2016)} methodology, the spatial lag model (SAR) assumes that it is the endogenous variable that is spatially correlated. This can be formalised as:

$$y = \alpha i_n + \rho W y + X\beta + \varepsilon,\ \varepsilon N(0, \sigma^2 I_n) \tag{9}$$

where ρ is the coefficient for the endogenous variable W_y, a variable representing a function of neighboring values of the dependent variable.

Finally, the mixed spatial autoregression spatial error SAC (sometimes referred to as SARAR), which includes spatial autoregression proposed by Kelejian and Prucha (1998) is considered:

$$\begin{aligned} y &= \alpha i_n + \rho W y + X\beta + u; \\ u &= \lambda W u + \varepsilon,\ \varepsilon N(0, \sigma^2 I_n) \end{aligned} \tag{10}$$

Golgher and Voss (2016) show that, in contrast to SLM and SEM, the y expected value for SAR and SAC models is:

$$\mathrm{E}(y \mid X) = (I - \rho W)^{-1}(\alpha i_n + X\beta) \tag{11}$$

Similarly, the partial derivatives include the matrix $(I - \rho W)^{-1}$ which has non-zero elements on its main diagonal. This implies spillover effects. They can be divided into direct effects, which represent the expected average change in all observations of y in a given spatial unit for a one-unit increase of an independent variable in that spatial area, and indirect effects, which determine the change in y in a specific spatial unit for a one-unit increase of an independent variable in another region (Golgher and Voss, 2016).

$$\begin{pmatrix} \frac{\partial y_1}{\partial x_{1k}} & \dots & \frac{\partial y_1}{\partial x_{nk}} \\ \dots & \dots & \dots \\ \frac{\partial y_n}{\partial x_{1k}} & \dots & \frac{\partial y_n}{\partial x_{nk}} \end{pmatrix} = (I - \rho W)^{-1} \begin{pmatrix} \beta k & \dots & 0 \\ \dots & \dots & \dots \\ 0 & \dots & \beta_k \end{pmatrix} = \beta_k (I - \rho W)^{-1} \tag{12}$$

5. Results

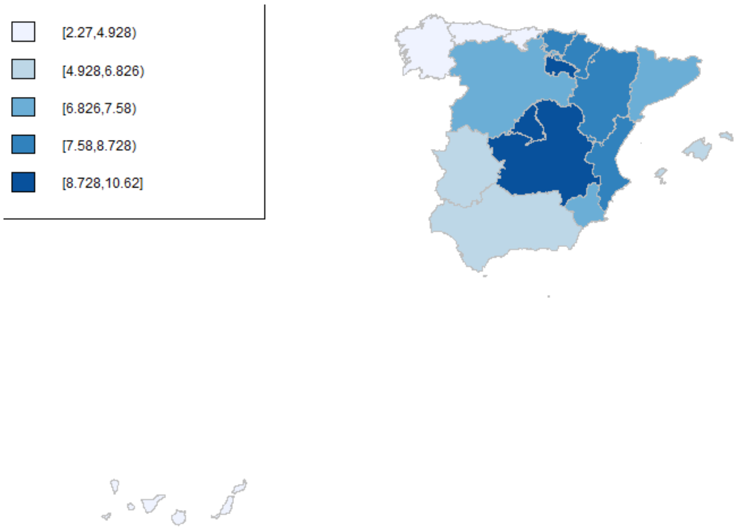
This section presents the results of our analysis in the form of several tables, as an aid to discussing the most important ideas in the following sections. It is divided into an initial exploratory analysis of the data and a subsequent regression analysis of different econometric models. All the empirical analysis in this section has been carried out with the R program.

5.1 Exploratory spatial data analysis

Spatial analysis requires an exploratory analysis of spatial data to enable spatial distributions to be described and spatial association patterns and/or atypical locations to be shown (Serrano and Valcarce, 2000). Maps are a good tool for showing data spatially and quickly. We rely on Figure 3, which shows the cumulative positive rate in Spain up to 9 April 2021[12] and can thus give us an intuitive idea of the potential correlation between neighboring communities. There clearly appear to be spatial clusters. The highest infection rates are in the center of Spain, while the lowest are in the north-west. Additionally, spatial autocorrelation can be anticipated: communities with higher positive rates tend to be surrounded by communities with high rates, those with intermediate rates are surrounded by neighbors with similar rates, and those with lower rates are close to communities with low rates.

[12] Latest data available at the time of extraction.

Figure 3: Cumulative confirmed cases up to 9 April 2021. Spain.



Source: Authors' own work based on ISCIII (2021) and INE (2021e)

The above figure leads us to believe that positive rates do have spatial dependencies between communities. We therefore show in Figure 4 how spatial interactions between regions are distributed. The stand-out point is that both *Castilla y León* and *Castilla La Mancha* are highly connected with their neighboring regions. Specifically, *Castilla y León* is directly connected with 8 communities, while *Castilla La Mancha* is directly connected with 6. *Cantabria* is directly connected only with *Asturias*, *País Vasco* and *Castilla y León*.

Figure 4: Spatial connections between Spain's autonomous communities.

Source: Authors' own work

It is therefore necessary to implement Moran's test, which reveals not only whether there is spatial autocorrelation in the positive rate but if so also what kind of autocorrelation it is. In this case, Moran's I statistic takes a value of 0.26 and is statistically significant at the 5% significance level. There is thus sufficient evidence to reject the null hypothesis of no spatial autocorrelation, so the positive rate observations in the sample can be said to have spatial dependencies. The expected value of the statistic is -0.06. The actual statistic is larger than its expected value (0.26 > -0.06), so the spatial autocorrelation is positive, i.e. the rates are similar across neighbouring regions: Autonomous Communities with high (low) rates will be surrounded by others with high (low) rates.

5.2 Spatial analysis

As indicated above, there is positive spatial autocorrelation in the dependent variable, so we use different spatial regression models to establish macro-level relationships between COVID-19 infections and various socio-economic factors.

First, the linear regression model is estimated as a starting point. Life expectancy at birth, the percentage of the population with long-term health problems and the percentage of people who have problems with pollution or other environmental problems are not significant, but the independent variables are significant overall (the F-statistic is statistically significant at the 1% level). As expected, the COVID-19 mortality rate, the COVID-19 ICU admission rate, population density and the illiteracy rate have a significant positive relationship with the cumulative confirmed case rate. However, the unemployment rate has an inverse relationship with the number of infections.

Second, the spatial error model is estimated by maximum likelihood. In this case, all variables are significant at least at the 10% significance level, except for long-term diseases. In addition, life expectancy at birth and people with environmental problems now influence the rate of confirmed COVID-19 cases. The sign of the relationships between the dependent and independent variables is the same as in the previous model. The impact of changes in the illiteracy rate on the rate of positives is greater when spatial dependencies are taken into account. In this sense, a 1% increase in the illiteracy increases the infected rate by 2.65%, ceteris paribus. On the other hand, with a 1% increase in the unemployment rate, the infected rate decreases by 0.39%, ceteris paribus.

This inverse relationship could be explained by commuting between the usual places of residence and work: the fewer people there are in work, the less they commute, so, the less contact they have with the rest of the population.

Table 3: Estimation of SLM, SEM, SAR and SARAR models.

Variables	**SLM**	**SEM**	**SAR**	**SARAR**
COVID-19 results				
mortalityrate	19.458***	18.759***	17.338***	17.403***
icurate	31.269**	31.724***	32.159***	31.957***
Health				
leb	0.512	0.391**	0.463*	0.423***
longtemrd	0.002	0.041	0.019	0.010
Socioeconomic factors				
illiteracyrate	1.183*	1.134***	1.294***	1.292***
popden	2.175***	2.649***	1.966***	2.341***
unemploymentrate	-0.278*	-0.389***	-0.220***	-0.312***
Environmental factor				
pollution	-0.075	-0.110***	-0.051**	-0.067
R^2	0.934			
R^2 adjusted	0.884			
F statistic	18.220***			

Note: *p<0.1;**p<0.05;***p<0.01

Source: Authors' own work

As discussed above, the estimated coefficients from the SLM and SEM can be interpreted directly, but those from the SAR and SARAR models cannot. The spillover effects are therefore presented in Table 4. The total effect is the sum of direct and indirect effects. The direct effect expresses the marginal effect of a 1% change in the independent variable on the positive COVID-19 rate in the same autonomous community, while the indirect effect is the marginal effect of a 1% change in the independent variable on the positive COVID-19 rate of all neighboring regions.

Table 4: Direct, indirect and total effects of SAR and SARAR models.

	SAR			**SARAR**		
Variables	**Direct effect**	**Indirect effect**	**Total effect**	**Direct effect**	**Indirect effect**	**Total effect**
COVID-19 results						
mortalityrate	17.410	2.203	19.613	17.453	1.816	19.268
icurate	32.291	4.087	36.377	32.048	3.334	35.382
Health						
leb	0.465	0.059	0.523	0.424	0.044	0.468
longtemrd	0.020	0.022	0.038	0.010	0.001	0.016
Socioeconomic factors						
illiteracyrate	1.299	0.164	1.463	1.295	0.134	1.430
popden	1.974	0.249	2.224	2.348	0.244	2.592
unemploymentrate	-0.221	-0.028	-0.249	-0.313	-0.032	-0.345
Environmental factor						
pollution	-0.050	-0.006	-0.058	-0.067	-0.007	-0.074

Source: Authors' own work

6. Discussion

The COVID-19 pandemic began in early 2020 and is still ongoing. To help policy makers to make the right decisions on containing and mitigating the spread of the virus, it is important to conduct studies to understand the explanatory factors and spatial patterns of the virus as well as its effects on people's health.

Spain is one of the countries hardest hit by COVID-19. We first studied the spatial interactions of the confirmed coronavirus rate among its regions, as well as its relationship with various socio-economic factors. The literature and our results both show that the spread of COVID-19 is uneven and depends on many factors (Amdaoud et al., 2021; Sannigrahi et al., 2020 and Mena et al., 2021). As in other countries, such as China (Guliyev, 2020), Germany (Ehlert, 2020) and the US (Baum and Henry, 2020 and Maiti et al., 2021), an analysis of the rate of positive cases of COVID-19 in Spain reveals spatial interactions between the regions, with spatial autocorrelation being positive. The results vary when these interactions are taken into account, so it is important to include them in the analyses. Thus, measures taken to curb the coronavirus spread (quarantines, closure of businesses, social distancing) must be taken as a whole, as what happens in one region will both directly and indirectly affect its neighbours.

As expected and as suggested by Guliyev (2020), we find that the coronavirus mortality rate is strongly positively related to the rate of confirmed cases. Another important finding is that, as in Germany (Ehlert, 2020), population density and life expectancy at birth in Spain are positively and significantly related to the rate of confirmed cases. This highlights the importance of taking precautionary measures in areas with high population densities and of paying special attention to the elderly during this crisis. We also find that unemployment rates are negatively related to infection rates.. Along these lines, Yang et al. (2021) finds an inverse relationship between the percentage of workers who work at home and confirmed cases of COVID-19 in New York, indicating that staying at home reduces the risk of infection.

Moreover, the direct and indirect effects of some regions on others can be quantified. The direct effects of the mortality rate on the positive rate in Spain are somewhat smaller than those found for China by Guliyev (2020) (17.41 and 32.49, respectively), while the indirect effects are slightly higher (2.20 vs. 1.70).

7. Conclusions

This study presents two empirical analyses of outcomes and consequences of COVID-19. We conduct a macroeconomic analysis of the links between cumulative confirmed coronavirus case rates and several socio-economic factors in Spain's autonomous communities, considering the spatial interactions between them.

The statistical results show spatial dependencies between autonomous communities. The cumulative rate of confirmed COVID-19 cases tends to form spatial clusters and shows a positive spatial autocorrelation: autonomous communities with a particular level are surrounded by others with similar levels. We also find that the COVID-19 mortality rate, the COVID-19 ICU admission rate, the population density, the illiteracy rate, the unemployment rate and the percentage of people with problems arising from pollution influence the infection rate. The socioeconomic factor that has the greatest influence on those infected is the illiteracy rate. In this sense, a 1% increase in this variable increases the positive COVID-19 rate by between 1.97% and 2.65%, ceteris paribus (depending on the model selected). The initial hypothesis that there is spatial autocorrelation in the sample and that socio-economic inequalities influence the spread of the virus can thus be accepted.

This study fosters the inclusion of spatial econometrics in the field of health economics research and the updating of the empirical literature on the regional transmission of COVID-19, specifically in Spain's autonomous communities. The results provide a better understanding of the spread of the virus from a socio-economic perspective, and this contribution may be of interest to policy makers and health agencies. Public health policies need to pay special attention to equity and accessibility to minimize all divergences in health between citizens and to promote flexibility in hospitals. Moreover, it is worth recalling the reflection of Díaz (2020), who indicates that the way out of this crisis necessarily implies global cooperation and both social and political innovation.

However, this is a static study. With the data available it not possible to dynamize the analysis, as it has only been possible to use data from a single time period. A second limitation is that to optimise the analysis data at the level of municipalities would be needed, so as to analyse the spatial interactions between territories in much greater depth. However, such data were not available to us.

As future lines of research and to oyangvercome the limitations listed above, a further study on this subject is proposed, but from a dynamic perspective. New data will become available over time and can be included in this study. Thus, these analyses can then be evaluated over a longer time horizon. We also suggest analyzing the relationships between COVID-19 infection or mortality rates and health determinants, not just socio-economic factors. Finally, these analyses could also be further explored in other areas to compare results and draw new conclusions.

Bibliography

Amdaoud, M., Arcuri, G., & Levratto, N. (2021). Are regions equal in adversity? A spatial analysis of spread and dynamics of COVID-19 in Europe. *The European Journal of Health Economics*, 1-14.

Baum, C. F., & Henry, M. (2020). Socioeconomic factors influencing the spatial spread of COVID-19 in the United States. *Miguel, Socioeconomic Factors influencing the Spatial Spread of COVID-19 in the United States (May 29, 2020)*.

Bhatnagar, M. R. (2020). A Statistical Model for the Spread of COVID19 in Clusters. *ResearchGate, DOI, 10*.

CLIFF, A. (1969). D. and J. K. ORD "The Problem of Spatial Autocorrelation". *Scott, A.*

Cos, O. D., Castillo, V., & Cantarero, D. (2020). Facing a Second Wave from a Regional View: Spatial Patterns of COVID-19 as a Key Determinant for Public Health and Geoprevention Plans. *International Journal of Environmental Research and Public Health*, *17*(22), 8468.

Díaz, D. (2020). The economic dimension of the pandemic. URL https://www.tni. org/en/article/the-economic-dimension-of-the-pandemic.

Ehlert, A. (2021). The socio-economic determinants of COVID-19: a spatial analysis of German county level data. *Socio-Economic Planning Sciences*, 101083.

Flaxman, S., Mishra, S., Gandy, A., Unwin, H. J. T., Mellan, T. A., Coupland, H., ... & Bhatt, S. (2020). Estimating the effects of non-pharmaceutical interventions on COVID-19 in Europe. *Nature*, *584*(7820), 257-261.

Fotheringham, A. S. (2009). “The problem of spatial autocorrelation” and local spatial statistics. *Geographical analysis*, *41*(4), 398-403.

Getis, A. (2007). Reflections on spatial autocorrelation. *Regional Science and Urban Economics*, *37*(4), 491-496.

Golgher, A. B., & Voss, P. R. (2016). How to interpret the coefficients of spatial models: Spillovers, direct and indirect effects. *Spatial Demography*, *4*(3), 175-205.

Griffith, D. A. (1987). Spatial autocorrelation. *A Primer (Washington, DC, Association of American Geographers).*

Guliyev, H. (2020). Determining the spatial effects of COVID-19 using the spatial panel data model. *Spatial statistics*, *38*, 100443.

INE (2021a). Instituto nacional de estadística. esperanza de vida al nacimiento por comunidad autónoma, según sexo. URL https://www.ine.es/jaxiT3/Datos.htm?t=1448.

INE (2021b). Instituto nacional de estadística. extensión superficial de las comunidades autónomas y provincias, por zonas altimétricas. URL https://www.ine.es/inebaseweb/pdfDispacher. do?td=33597.

INE (2021c). Instituto nacional de estadística. personas con enfermedades o problemas de salud de larga duración por ccaa y periodo. URL https://www.ine.es/jaxi/Datos.htm?path=/t00/ ICV/dim3/l0/&file=31401.px.

INE (2021d). Instituto nacional de estadística. Población de 16 y más años por nivel de formación alcanzado, sexo y comunidad autónoma. porcentajes respecto del total de cada comunidad. URL https://www.ine.es/jaxiT3/Datos.htm?t=9459.

INE (2021e). Instituto nacional de estadística. Población por comunidades y ciudades autónomas y tamaño de los municipios. URL https://www.ine.es/jaxiT3/Datos.htm?t=2915.

INE (2021f). Instituto nacional de estadística. Población que sufre problemas de contaminación y otros problemas ambientales. URL https://www.ine.es/ss/Satellite?L=es_ES&c=INESeccion_C&cid=1259944616263&p=1254735110672&pagename=ProductosYServicios% 2FPYSLayout&param1.

INE (2021g). Instituto nacional de estadística. Tasa de paro por ccaa y periodo. URL https: //www.ine.es/jaxi/Tabla.htm?path=/t00/ICV/dim2/l0/&file=21201.px&L=0.

ISCIII (2021). Instituto de Salud Carlos III. Covid-19. URL cnecovid.isciii.escovid

Jonas, O. B. (2013). Pandemic risk.

Kelejian, H. H., & Prucha, I. R. (1998). A generalized spatial two-stage least squares procedure for estimating a spatial autoregressive model with autoregressive disturbances. *The Journal of Real Estate Finance and Economics*, *17*(1), 99-121.

LeSage, J., & Pace, R. K. (2009). *Introduction to spatial econometrics*. Chapman and Hall/CRC.

LeSage, J. P., & Dominguez, M. (2012). The importance of modeling spatial spillovers in public choice analysis. *Public Choice*, *150*(3), 525-545.

Li, X., Chen, F., & Hu, S. (2021). Spatial Spillover Effect of Government Public Health Spending on Regional Economic Growth during the COVID-19 Pandemic: An Evidence from China. *Complexity*, *2021*.

Maiti, A., Zhang, Q., Sannigrahi, S., Pramanik, S., Chakraborti, S., Cerda, A., & Pilla, F. (2021). Exploring spatiotemporal effects of the driving factors on COVID-19 incidences in the contiguous United States. *Sustainable cities and society*, *68*, 102784.

McGinty, E. E., Presskreischer, R., Han, H., & Barry, C. L. (2020). Psychological distress and loneliness reported by US adults in 2018 and April 2020. *Jama*, *324*(1), 93-94.

Mena, G. E., Martinez, P. P., Mahmud, A. S., Marquet, P. A., Buckee, C. O., & Santillana, M. (2021). Socioeconomic status determines COVID-19 incidence and related mortality in Santiago, Chile. *Science*, *372*(6545).

Moran, P. A. (1948). The interpretation of statistical maps. *Journal of the Royal Statistical Society. Series B (Methodological)*, *10*(2), 243-251.

OECD (2021). Organisation for economic co-operation and development. life expectancy at birth. URL https://data.oecd.org/healthstat/life-expectancy-at-birth.htm

Paelinck, J. H. P. (1967). L’efficacité de la politique économique régionale. *Namur, Faculte des Sciences Economiques*, *58*.

Paelinck, J. H., Klaassen, L. H., Ancot, J.-P., and Verster, A. (1979). Spatial econometrics, vol. 1. Saxon House

Paez, A., Lopez, F. A., Menezes, T., Cavalcanti, R., & Pitta, M. G. D. R. (2020). A spatio-temporal analysis of the environmental correlates of COVID-19 incidence in Spain. *Geographical analysis*.

Saha, K., Torous, J., Caine, E. D., & De Choudhury, M. (2020). Psychosocial Effects of the COVID-19 Pandemic: Large-scale Quasi-Experimental Study on Social Media. *Journal of medical internet research*, *22*(11), e22600.

Sannigrahi, S., Pilla, F., Basu, B., Basu, A. S., & Molter, A. (2020). Examining the association between socio-demographic composition and COVID-19 fatalities in the European region using spatial regression approach. *Sustainable cities and society*, *62*, 102418.

Serrano, R. M., & Valcarce, E. V. (2000). *Técnicas econométricas para el tratamiento de datos espaciales: la econometría espacial* (Vol. 44). Edicions Universitat Barcelona.

Shobande, O., & Ogbeifun, L. (2020). A Spatial Econometric Analysis on the Impact of COVID-19 on Mortality Outcome. *Economics and Business*, *34*(1), 179-200.

Sun, F., Matthews, S. A., Yang, T. C., & Hu, M. H. (2020). A spatial analysis of the COVID-19 period prevalence in US counties through June 28, 2020: where geography matters?. *Annals of Epidemiology*, *52*, 54-59.

Sun, Y., Hu, X., & Xie, J. (2021). Spatial inequalities of COVID-19 mortality rate in relation to socioeconomic and environmental factors across England. *Science of The Total Environment*, *758*, 143595.

WHO (2018). Boletín de la organización mundial de la salud. volumen 96, número 2, febrero 2018, 77-144. URL https://www.who.int/bulletin/volumes/96/2/es/.

WHO (2021). World health organization. WHO coronavirus (covid-19) dashboard. URL https: //covid19.who.int/.

Yang, T. C., Kim, S., Zhao, Y., & Choi, S. W. E. (2021). Examining spatial inequality in COVID-19 positivity rates across New York City ZIP codes. *Health & place*, *69*, 102574.

Zipf, G. K. (1949). Human behavior and the principle of least effort: An introduction to human eoclogy.