How are corona infections in different countries related to infections in bordering countries?

By

Jin Dai
(1133713)

A Project thesis

for the degree of

Master of Science

in Department of Information and Computing Sciences

at

Utrecht University

July, 2022

How are corona infections in different countries related to infections in bordering countries?

Jin Dai

Applied Data Science

Abstract

It has been more than two years since the Covid-19 epidemic began in Europe, this long-term pandemic affecting all countries on the continent. This study aims to find the relationship between the number of new infections between countries by analyzing the data of new infections for nearly two years, that is, whether new infections in other countries will cause changes in new infections in their bordering countries. To answer this question, we compared the performance of panel models and spatial panel models to verify the correlation. To answer this question, we compare the performance of the panel model and the spatial panel model, find the most suitable model by comparison, and then verify the correlation. The results show that when there is a one-week lag, the number of new infections in the focal country, the number of new infections in neighboring countries, and the focal country's epidemic prevention policy index will all have a positive correlation with the number of new infections in the country this week. These results show that the spread of Covid-19 does have a spatial impact, so this impact factor should be taken into account in the prevention and control of the epidemic.

Keywords**:** Covid-19, adjacent countries, spatial panel, maximum likelihood, spatial weighted matrix, R.

Contents

Chapter 1

Introduction

COVID-19 was first reported in Wuhan, China, and subsequently spread worldwide. The virus was named by the WHO as COVID-19 because the first traceable case worldwide was found in December 2019. Since then, countries around the world have been engaged in a protracted war against COVID-19. The pandemic has become known as the fifth documented pandemic since the 1918 flu pandemic because of the extreme geographical spread of the disease and the sheer number of people affected [1].

Some people who have lived in Europe for a long time and have gone through the whole process of the epidemic believe that in Europe, the COVID-19 epidemic started in March, but the truth is that as early as January 24, 2020, the first confirmed case that can be traced back of COVID-19 were found in France. Since then, the unstoppable coronavirus has ravaged the European continent, and by April, the confirmed death toll had surpassed 100,000 [2].

Out of 7 continents, the European continent ranks third in the number of countries. At the same time, what can be observed from the map of Europe is that a large number of European countries have many "neighbors" [3]. For example, Russia, as a big country with a vast territory, has the most "neighbors" in the world [4]. In Europe, these neighboring countries provided the basis for the creation of the European Union. After World War II, Western European countries were the first to unite. Due to their geographical proximity and similar cultural heritage, these countries chose to unite to seek closer political, economic and social ties. This allows these relatively small countries to form a collective to increase their strength and resist risks [5]. In the following decades, the size of the EU gradually expanded, and more and more European countries joined the EU either for development or for asylum, which made the EU one of the three largest economies in the world [6].

The union of European countries not only creates benefits for these countries, but also brings invisible constraints and risks to themselves. The synergistic policy brought about by the union is a double-edged sword. This is particularly evident in the spread of the epidemic. Unlike countries on other continents, because of the existence of the EU, European countries tend to adopt more coordinated policies within a certain framework when faced with various affairs. This makes it not easier than other countries to make decisions such as closing

borders and suspending flights. Even during the high epidemic period, these countries tried their best to release the measures so as not to affect the communication link between countries. It can be observed that when the epidemic situation tends to improve, the epidemic prevention policies of most EU countries tend to be relaxed almost simultaneously [7]. On the one hand, this practice of maintaining "ties" allows countries to cooperate in preventing and controlling the epidemic; on the other hand, it leads to a more rapid spread of the virus from country to country.

During the epidemic, European neighbors helped each other, not only in the convergence of policies, but also in the mutual assistance and sharing of medical resources between countries [8]. This “tie” between neighbors is strengthened even more in the face of difficulties. Therefore, it is difficult to admit that European countries are dealing with the epidemic independently because they are being so much influenced by their neighbors. This means that the epidemic situation between countries might be related to a certain extent. A country's prevention and control policies will affect the changes in the number of sick people during this period, and the increase or decrease in the number of sick people in that country will affect the degree of infection and spread to neighboring countries.

In this research, what needed to be explored is the development of the epidemic in various European countries. It assumed that there is an infectious disease transmission network within Europe, and we will construct a new case network by comparing the new cases in various countries over time to observe whether the number of new cases in one country will be affected by the number of new cases in neighboring countries. Impact. At the same time, we will consider the policies of various countries during the same period to analyze the reasons for one of these situations. In addition to the impact of the bordering relationship between countries, the impact of the country's measure response index is also added to the analysis to reduce the impact of other factors on the research.

The aim of this paper is to gain an understanding of the number of new infections in neighboring countries and the impact of neighboring countries' epidemic prevention and control policies on the number of new infections in their own countries. The research method used was spatial panel regression. In methodology section, some other types of research that are helpful in understanding the research topics and methods used in this paper will be presented.

The corona virus epidemic is still spreading all over the world, and various research on it can still be called "current" research. In addition to comparing epidemic-related research, we are also able to predict our hypothesis from some other papers with similar topics or similar research methods.

The spread of epidemics has always been a research topic of concern. All we can know is that in a modern society with a highly developed economy and transportation, the infection paths of the virus have become more diverse and more difficult to accurately control. According to Young's research, in human cases of WNV infection, newly infected areas often border areas that were already infected in previous years, indicating that the virus spread from affected areas to unaffected adjacent areas [9]. In Heyman's study, hantavirus, a virus that causes haemorrhagic fever with renal syndrome, was transmitted via rodent populations that reproduced and migrated across the European plate, resulting in countries previously unreported cases of the virus are gradually seeing new infections [10]. When we focus on covid, what can be found is that even though SARS-CoV-2 infects a variety of animals, animal infections are still rare, and current cases of animal infections are often caused by an extremely large number of human infections [11]. Therefore, we can speculate that the virus is able to spread from an infected area to an uninfected area through humans, so the number of infections in one place will have an impact on the number of infected people in neighboring areas.

Meanwhile, Vandenbroucke-Grauls, in their Multiple resistant staphylococci prevalence study, mentions the Dutch case of infection rates, noting that the Netherlands is one of the country has the lowest MRSA infection rates in Europe, because the Netherlands implements strict infection control measures, including but not limited to isolating carriers and treating patients and carriers with antibiotics [12]. It can be seen from this paper that strict, reasonable and scientific control measures can also affect the spread of viruses or bacteria even between neighboring countries, resulting in very different infection rates among countries.

Our exploration of hypotheses about the subject will rely on the spatial panel model for regression analysis. Phang used the spatial panel model of six different regions to study the spatial pattern of dissemination at the regional level in Sarawak, Malaysia. Earlier, this study presented a weak positive spatial autocorrelation between confirmed cases and regions, but the emergence of high-valued spatial clusters over time suggests that the spread of COVID-19 in neighboring regions has a positive impact on the confirmed case rates in specific regions of Sarawak. have a significant impact. The study used spatial weighting schemes to analyze regions, suggesting that the regional spread of the virus can be reasonably fitted by spatial models [13]. Another study was from Mitze, starting from the specific work commuting aspect, introduced a spatial panel error correction model (SP-ECM). The results provide strong evidence for the spatial dependence of COVID-19 data and demonstrate that this spatial correlation is largely dependent on the prespecified work commute factor [14].

In this paper, the ultimate goal is to determine the spatial relationship between the number of infections in European countries and to analyze whether neighboring countries influence each other in the spread of the virus. We will use the spatial panel model to run the dataset and compare the obtained results with the predictions.

Chapter 2

Methodology

The research focuses on how corona infections in different European countries are related to those in bordering countries. We achieved this goal by selecting several of these variables and using spatial panel regression to analyze the impact of these variables on corona infections in various countries.

2.1 Data

2.1.1 Data collection

The data used in the research came from a data source called Data on COVID-19 (coronavirus) on GitHub, collected by Our World in Data. [15]. This dataset records the number of coronavirus infections in many countries around the world since February 2020, including a series of related data such as single-day new case numbers, daily cumulative new case numbers, population, population density, and stringency index.

2.1.1 Data extraction

For the topics explored in this research, we extracted data content for 108 weeks in Europe starting in March 2020 and ending in March 2022. This time period basically includes the time period when the corona epidemic has just begun to spread in Europe and the time period when the infection has decreased, and the societies of various countries have gradually opened up again.

We selected three factors from a large amount of information in the database as our research variables: new_cases_per_million_own, avg_neighbour_cases_per_million and ci_own, these are names they show in the data table. New_cases_per_million_own represents the number of new infections per million population, avg_neighbour_cases_per_million expresses the average number of new infections in all neighboring countries around a country. To reduce the impact of the size of the population of each country on the average, the weighted average method is used here, and the new cases are weighted according to the population of each country. Ci_own stands for containment index for own country.

The initial data set is composed of data from countries around the world. The

first thing to do is to filter out the data of countries in the European region. The obtained data contains a total of 49 countries. After sorting out the list of countries, an adjacency matrix is made according to the neighboring countries of these countries. In the process of this step, it can be found that some countries do not have bordering countries, so the data rows and columns containing these countries will be deleted.

The next steps that need to be taken are to reintegrate the number of new infections recorded by day, change such data for each country to weekly data according to the same time period, and register the number of these weeks in order from 1. In this step, we integrate daily data into weekly data in order to make the data perform better in the model. Daily data is often affected by public holidays, holidays or other specific events, skewing the data and reducing the accuracy of the analysis. But when we use weekly data, we can mitigate the impact of these special events on the data. For example, the weekly data will include two public holidays on weekends, which reduces the impact of public holidays on the number of infections recorded.

After the data of each country is sorted, what needs to be found is the new infections of bordering countries of each country, weighted and added up the number of new infections according to the different populations of these countries, so that each country can get weighted average of new infections in neighboring countries. What needs to be done after that is the column representing the containment index data, because some countries have missing data, we also removed the data for these 5 countries from the dataset and the adjacency matrix. After completing these steps, we replaced the country names with numbers, which facilitated the use of R language tools in the subsequent analysis process.

2.2 Spatial Panel Regression

Panel data refers to the indicator data of different objects at different times. The figure 1 below shows an example of panel data presented with this research data.

Figure 1.

iso_code	location	id	date	week	new_cases_per_million_own	avg_neighbour_cases_per_million	ci_own
ALB	Albania	1	06/03/2022	106	205,71	7117,57	50,00
ALB	Albania	1	13/03/2022	107	145,84	8509,34	50,00
ALB	Albania	1	20/03/2022	108	120,78	7864,57	50,00
ALB	Albania	1	27/03/2022	109	113,86	6205,03	50,00
AND	Andorra	2	02/03/2020	1	12,92	11,41	2,38
AND	Andorra	2	08/03/2020	2	0,00	82,65	12,92
AND	Andorra	2	15/03/2020	3	1124,70	253,43	31,97
AND	Andorra	2	22/03/2020	4	2844,06	622,46	42,52

The location/id column reflects different sections of the data, that is, different

countries in our case. The week column reflects the time series of the data, indicating that the data is divided into different weeks according to time. Panel data is a combination of the two. The panel model, also known as panel regression, is to use such panel data to study the relationship of regression and influence. Panel models can be divided into three types, namely FE model (fixed effect model), POOL model (mixed estimation model) and RE model (random effect model). In this paper, the data were analyzed using the FE model.

We use the FE model because the variables for which the relationship needs to be analyzed have been selected in this research. The full name of the FE model is the fixed effect model, and 'fixed' indicates that the analysis object of the model is fixed, not randomly selected. Therefore, we use the FE model in this study in order to fix the scope of the study within the study object, that is, the performance over time within the country. By importing country infection data into the model, it is possible to observe the number of infections within countries over time, rather than comparing changes between countries. There are many factors that might explain differences between countries, such as the extent to which the government's new cases reporting mechanism can be implemented, and the implementation progress of the policy to urge suspected cases that have been reported but not tested. In addition, we narrowed these changes in the epidemic to a smaller scale, that is, we only focused on the number of new infections rather than deaths and treated recoveries.

Chapter 3

Empirical results and conclusion

We use R as the language tool to do the analysis. To do the spatial anlysis in R, the library we use here are: plm, splm, sp, maps, maptools.

First, we divide the data into variables, and separate independent variables and dependent variables. Among them, the independent variables are the column new_cases_own_lag, column avg_neighbour_cases_lag, column ci_own_lag, and the dependent variable is the column new_cases_per_million_own.

The three independent variables here come from a one-week lag processing of the original data. We do not simply use the original data as independent variables but gave a lag processing, because the information feedback within this week often has a delayed effect. For the analysis of new cases in focal countries this week, we need to use the data of neighboring countries and focal countries in the previous week, and the parameter data of measurement in focal countries for the previous week as independent variables.

In order to facilitate reading and understanding, the above factors are simplified as follows: Simplify new_cases_per_million_own to cases_own; simplify new_cases_own_lag to cases_own_lag; simplify avg_neighbour_cases_lag to neighbour_cases_lag.

To get a better overall understanding of the data, we descriptive statistics for variables. From the result, It can be seen from the distribution that the variables on numbers of cases have strong outliers and are not normally distributed. To prevent outlier effects and make the distributions look more like a normal distribution we use a log-transformation for the three variables on infections cases.

Then we built an influencing factor equation like:
Log(cases_own+1) ~ log(cases_own_lag+1) + log(neighbour_cases_lag+1) + ci_own_lag

From table1 we can observe that the effect of outliers has been minimal.

Table 1. descriptive statistics for variables

	log(cases_own)	log(cases_own_lag)	log(neighbour_cases_lag)	ci_own_lag
Minimun	0.00995	0.00995	0.00995	0.01
1st Quartile	5.35551	5.33817	0.32930	48.30
Median	6.80501	6.79649	0.88377	57.14
Mean	6.52420	6.50978	2.03772	56.44
3rd Quartile	7.85687	7.84351	2.21811	64.59
Maximun	10.92387	10.92387	10.29598	90.00

3.1 Panel models

3.1.1 Fixed effects model and random effects model

We turn the raw data into panel data, which enables panel regression. We have mentioned before that panel regression has multiple models, and here we tested both the fixed-effect model and the random-effect model, in this step, the estimations are without consideration for spatial autocorrelation.

First, we obtained the results in Table 2 through the fixed effects model. The fixed effects model analyzes the fixed effects at the individual level, that is, without considering the time factor. The time factor has been reflected through lag.

Table 2. Fixed effects model Coefficients result

	Estimate	Std. Error	t-value	Pr(>|t|)
log(cases_own_lag)	0.9548	0.0060	160.4666	< 2e-16 ***
log(neighbour_cases_lag)	0.0152	0.0110	1.3833	0.1666
ci_own_lag	-0.0122	0.0007	-16.5101	< 2e-16 ***

(Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1)

From the results, it can be seen that the local cases and local policies in the previous week are related to the new local cases this week. Among them, for each additional unit of new local cases in the previous week, the average increase of local cases this week was about 0.95 units; for each additional unit of local policies in the previous week, the average decrease of local cases this week was about 0.01 units. And we also know that the R-Squared is 0.9348, means can explain 93% of the variation of new cases this week. Most importantly, we found that through the model analysis, the number of new cases in the adjacent

countries in the previous week was not related to the number of new cases in the focal countries this week.

Then, we obtained the results in Table 2 through the random effects model.

Table 3. Random effects model coefficients results

	Estimate	Std. Error	z-value	Pr(>|t|)
log(cases_own_lag)	0.9589	0.0042	225.9521	< 2e-16 ***
log(neighbour_cases_lag)	-0.0048	0.0029	-1.5925	0.1113
ci_own_lag	-0.0083	0.0006	-13.4884	< 2e-16 ***

In the results of this model, both the individual level and the quasi-demeaned variance are 0, suggesting that random effects are not appropriate.

The new cases in the focal country in the previous week and the local policy in the previous week are related to the new cases in this week. The average increase of local cases this week was about 0.96 units; for each additional unit of local policies in the previous week, the average decrease of local cases this week was about 0.01 units. The R-Squared is 0.9358, means can explain 94% of the variation of new cases this week.
Moreover, in the random effects model, the number of new cases in the adjacent area in the previous week is still not related to the number of new cases in the local area this week.

3.1.2 Hausman test

From the Hausman test, we found that the p value is less than 0.05, then we chosen the fixed-effect model to do further analysis. The result of the Hausman test tells that there is a significant difference in coefficient estimates between the fixed-effect model and the random-effect model, so the fixed-effect model is better than the random-effect model.

3.2 Spatial panel models

3.2.1 Spatial weight matrix

To do the spatial panel regression, we need to add spatial correlation to the model. According to the Tobler's First Law of Geography, " everything is related to everything else, but near things are more related than distant things". The spatial

weight matrix can describe the degree of correlation between things. According to the type, it can be divided into adjacency matrix and distance matrix. In this analysis, the spatial weighted matrix is an adjacent matrix of European countries.

According to the connection method, we can construct the following spatial weight matrix W to reflect the adjacency relationship between countries:

$$W=\begin{bmatrix} \omega_{11} & \cdots & \omega_{1n} \\ \vdots & \ddots & \vdots \\ \omega_{n1} & \cdots & \omega_{nn} \end{bmatrix}$$

where,

$$\omega_{ij} = \begin{cases} 1, & when\ country\ i\ is\ adjacent\ to\ country\ j \\ 0, & \text{when country i is not adjacent to country j} \end{cases}$$

Next, we use GeoDa to create a spatial weight matrix. GeoDa is an open-source software specially designed to deal with spatial autocorrelation analysis. Geoda supports a variety of spatial algorithms and graphs, and is often used for data visualization and graph analysis.
In GeoDa, we select the corresponding connection method as needed, because it is an adjacency matrix, and the spatial weight matrix created by GeoDa is a gal file.

We get the data description of the spatial weight matrix: FRA is the iso_code corresponding to the country France, it has seven neighboring countries, namely DEU- Deutschland, LUX- Luxembourg, BEL-Belgium, ESP-Spain, ITA-Italy, CHE-Switzerland, AND- Andorra; and so on, UKR- Ukraine also has seven neighboring countries, namely BLR- Belarus, RUS-Russia, ROU- Romania, HUN- Hungary, SVK-Slovak, POL-Poland, and MDA- Moldova.

3.2.2 LM-test

We next tested whether the spatial lag model or the spatial error model was more suitable than the model without spatial effects using the LM test and the Robust LM test.

The LM test is the Lagrange multiplier test, which is used to test whether there is a serial correlation in the residual series of the model. The null hypothesis is that there is no correlation; the alternative hypothesis is that there is a p-order autocorrelation.

Table 4. LM-tests Results

	lme	lml	rlme	rlml
LM	368.04	1044.3	105.14	781.35
p-value	< 2.2e-16	< 2.2e-16	< 2.2e-16	< 2.2e-16

(lme is short for LM test for spatial error dependence; lml is short for LM test for spatial lag dependence; rlme is short for Locally robust LM test for spatial error dependence sub spatial lag; rlml is short for Locally robust LM test for spatial lag dependence sub spatial error.)

From the result, all p-values are less than 0.05, so the null hypothesis should be rejected. In test lme, P<0.05, there is autocorrelation, it is appropriate to use the error model; in test lml, P<0.05, there is a lag correlation, it is appropriate to use the lag model; In test rlme, P<0.05, there is autocorrelation, it is appropriate to use the error model; in test rlml, P<0.05, there is a lag correlation, it is appropriate to use the lag model. Because there are no significant results in one respect of lag or error and not significant in the other, we finally choose the SARAR model that includes both lag and error.

3.2.3 Spatial Models

Spatial effects include spatial autocorrelation or spatial dependence and spatial heterogeneity [16]. The spatial autocorrelation means that the sample observations in one area are related to the observations in other areas, the observations lack spatial independence, and the degree of spatial correlation is determined by the absolute and relative positions. The spatial heterogeneity refers to the heterogeneity of spatial effects at the regional level due to the heterogeneity of spatial units. The spatial correlation here comes from two aspects: spatial measurement errors and connections between adjacent regions. Therefore, the two basic models we use are: spatial auto regression model, SAR and spatial error model, SEM.

spatial auto regression model, SAR:

$$y = \rho W y + x\beta + \varepsilon$$

spatial error model, SEM:

$$y = x\beta + \mu$$
$$\mu = \lambda W \mu + \varepsilon$$

In the formula, y is the dependent variable, x is the independent variable vector, β is the variable coefficient, W is the spatial weighted matrix, ε is the residual vector composed of independent residual random variables. ρ and λ are the spatial autoregression coefficient and the spatial autocorrelation

coefficient.

We generate a new equation for the spatial panel model, here we also use the logarithmic variables:

For SARAR model, the equation is:
log(new_cases_own+1)~log(neighbour_cases_lag+1)+ ci_own_lag ;

For SEM model, the equation is:
log(cases_own+1)~log(cases_own_lag+1)+log(neighbour_cases_lag+1)+ci_own_lag .

Function used in the model is:

$$y_N(t) = \lambda W_y + X_N(t)\beta + u_N(t)$$

$$u_N(t) = \rho W_N u_N(t) + \epsilon(t)$$
$$\epsilon_N = (e_T \otimes I_N)\mu_N + v_N$$

Different from the previous non-spatial panel model, in the analysis of this spatial panel model, the results show that both the number of new infections in the adjacent countries and the policies of the focus countries are positively correlated with the number of new infections in the focus countries. This gives a support to our conjecture for the first time.

The SARAR model contains lag and error, and the coefficients of them are all significant, which are p values of rho and lambda are all less than 0.05. The model has a good fit.

Table 5. SARAR model coefficients results

	Estimate	Std. Error	t-value	Pr(>|t|)
Error variance parameters - ρ	-0.559136	0.021753	-25.703	< 2.2e-16 ***
Autoregression parameter-λ	0.721804	0.011045	65.354	< 2.2e-16 ***
log(neighbour_cases_lag+1)	0.41692291	0.01783629	23.3750	< 2.2e-16 ***
ci_own_lag	0.00795422	0.00098107	8.1077	5.158e-16 ***

From the results of this model, for every one unit increase in the logarithmic value of new cases in focal countries last week, the logarithmic value of local new cases this week increased by 0.72 units. For every one unit increase in the logarithmic value of new cases in adjacent countries last week, the logarithmic value of local new cases this week increased by 0.41units. For every increase in the logarithmic value of local policies last week, the logarithmic value of local

new cases this week increased by 0.008 units.

And we also use SEM model to fit:

Table 6. SEM model coefficients results

	Estimate	Std. Error	t-value	Pr(>|t|)
Error variance parameters - ρ	0.232983	0.017518	13.299	< 2.2e-16 ***
log(new_cases_own_lag + 1)	0.94375189	0.00598062	157.8016	< 2.2e-16 ***
log(neighbour_cases_lag+1)	0.00678650	0.01063905	0.6379	0.5235
ci_own_lag	-0.01131397	0.00076149	-14.8577	<2e-16 ***

From the results of this model, for every one unit increase in the logarithmic value of new cases in focal countries last week, the logarithmic value of local new cases this week increased by 0.94 units. For every one unit increase in the logarithmic value of new cases in adjacent countries last week, the logarithmic value of local new cases this week increased by 0.0071units. For every increase in the logarithmic value of local policies last week, the logarithmic value of local new cases this week decreased by 0.011 units.

Chapter 4

Conclusion

This study explored the relationship between new infections in a country and new infections in its bordering countries during a pandemic Covid-19. First, we selected study variables and screened out the countries that needed to be analyzed based on the available data.

We chose to use different models to analyze the panel data. First, when we use the simple panel model without adding spatial factors, the coefficient of new infections in neighboring countries is negative, indicating that new infections in neighboring countries will not have a positive impact on new infections in focal countries. And it can also be observed that the coefficients for this term are not significant in both effect models. After confirming the existence of fixed effects and adding the spatial weight matrix to the panel model for analysis, the coefficient of this item in subsequent model appears significant. This shows that the spatial factor does have influence on the actual case analysis.When we use the spatial panel model, we can observe from the results that the two models fit the data differently. In the SARAR model, all variables are significant, but it can be observed that the coefficient of the containment index is positive, indicating that the result is the opposite of our prediction that a higher containment index would lead to more new cases in the country. In the SEM model, it can be observed that the coefficient of the containment index is negative, indicating that the result is consistent with our prediction, that is, the larger the containment index, the fewer new cases in the country. However, in this model, there are insignificant variables, that is, new cases in neighboring countries cannot affect new cases in the focal country.

In the process of research, we can confirm that both the lag part and the error part are all have significance for the fitting of the model. A one-week lag was designed for the data in the study. This value can be freely selected, and one or more weeks of lag can be selected. In the study, the lag of the data for one week can already show the correlation.

One limitation of the study was that the dataset used was only two years from the time Covid-19 was already on the radar of most people in Europe. Therefore, among the 38 selected countries, some countries with higher population density or with popular tourist cities in the world already had more cases at the initial time when the data started. Others, with underdeveloped tourism and sparsely

populated countries, had almost no infections at the start of the data. Such bias makes the data somewhat unbalanced. Another more important limitation is that in the two years of data expression, the new coronavirus has undergone multiple mutations [17]. These mutations make various coronaviruses with different names have different characteristics. For example, the original strain of the new coronavirus is highly virulent and difficult to cure, and the infected population has a greater chance of death. After nearly two years of transmission, the existing coronavirus strain has become milder, the lung damage of the infected person has been reduced, the infection symptoms have eased, and the fatality rate has also dropped significantly. On the contrary, the infection probability has greatly increased. This allows the virus to have different infection rates in different periods, and the policies issued by governments in different periods are also very different.

In addition to data limitations, there are limitations in model and variable selection in other aspects of the study. In terms of models, we have observed that different models tend to give different conclusions, which can further compare the fitting mechanisms of these two models to distinguish their preferences at different fitting levels. In terms of variables, in addition to the selection of other countries new cases and national policies, other variables can be added in subsequent research to conduct more detailed analysis of this topic. After all, at the moment, the new crown epidemic is still a long way from the complete end.

Reference

[1] Yen-Chin Liu, Rei-LinKuo, Shin-Ru Shih. COVID-19: The first documented coronavirus pandemic in history[D]. :Biomedical Journal, 2020. 328-333. https://doi.org/10.1016/j.bj.2020.04.007

[2] Sean Fleming. Europe surpasses 100,000 confirmed coronavirus deaths – how did we get there? WORLD ECONOMIC FORUM, 2020. https://www.weforum.org/agenda/2020/04/coronavirus-deaths-europe/

[3] Jason Shvili. Continents By Number Of Countries[Z]. : WorldAtlas,2021. https://www.worldatlas.com/geography/continents-by-number-of-countries.html

[4] Joseph Kiprop. European Countries With The Most Neighbors[Z]. : WorldAtlas,2018. https://www.worldatlas.com/articles/european-countries-bordering-most-other-countries.html

[5] Matthew J. Gabel. Creation of the European Economic Community[Z]. : Britannica, 2022. https://www.britannica.com/topic/European-Union/Creation-of-the-European-Economic-Community

[6] eurostat. Facts and figures on the European Union economy[Z]. : European Union, 2022. https://european-union.europa.eu/principles-countries-history/key-facts-and-figures/economy_en

[7] schengenvisa. Travel: List of EU Countries Keeping in Place COVID Restrictions in June & Entry Rules Explained[Z]. : schengenvisa, 2022. https://www.schengenvisainfo.com/news/travel-list-of-eu-countries-keeping-in-place-covid-restrictions-in-june-entry-rules-explained/

[8] Society. Solidarity: how EU countries help each other fight Covid-19[Z]. : News European Parliament, 2020. https://www.europarl.europa.eu/news/en/headlines/society/20200402STO76415/solidarity-how-eu-countries-help-each-other-fight-covid-19

[9] Young Johanna J, Haussig Joana M, Aberle Stephan W, Pervanidou Danai, Riccardo Flavia, Sekulić Nebojša, Bakonyi Tamás, Gossner Céline M. Epidemiology of human West Nile virus infections in the European Union and European Union enlargement countries, 2010 to 2018. Euro Surveill. 2021;26(19):pii=2001095. https://doi.org/10.2807/1560-7917.ES.2021.26.19.2001095

[10] Heyman P, Vaheri A, ENIVD members. Situation of hantavirus infections and haemorrhagic fever with renal syndrome in European countries as of December 2006. Euro Surveill. 2008;13(28):pii=18925. https://doi.org/10.2807/ese.13.28.18925-en

[11] Smriti Mallapaty. The search for animals harbouring coronavirus — and why it matters[J].nature,:,2021. 26-28.

https://www.nature.com/articles/d41586-021-00531-z
[12] Vandenbroucke-Grauls C. Epidemiology of staphylococcal infections--a European perspective. Journal of Chemotherapy (Florence, Italy). 1994 Apr;6 Suppl 2:67-70. PMID: 7799057.
https://europepmc.org/article/med/7799057
[13] Phang, P., Labadin, J., Abd Rahman, Se. (2022). Spatial Panel Data Modelling of COVID-19 Spreading at District Level in Sarawak, Malaysia. In: Alfred, R., Lim, Y. (eds) Proceedings of the 8th International Conference on Computational Science and Technology. Lecture Notes in Electrical Engineering, vol 835. Springer, Singapore.
https://doi.org/10.1007/978-981-16-8515-6_34
[14] Mitze, T., Kosfeld, R. The propagation effect of commuting to work in the spatial transmission of COVID-19. J Geogr Syst 24, 5–31 (2022). https://doi.org/10.1007/s10109-021-00349-3
[15] Data Source: covid-19-data/public/data at master · owid/covid-19-data · GitHub
[16] Anselin, L. (1988) Spatial Econometrics: Methods and Models. Kluwer Academic, Dordrecht.
https://doi.org/10.1007/978-94-015-7799-1
[17] Ewen Callaway.BEYOND OMICRON: WHAT'S NEXT FOR SARS-COV-2 EVOLUTION[J].Nature,:Springer Nature, 204-208.