01-07-2022

The impact of COVID-19 containment measures on COVID-19 cases in Europe: a spatial panel data analysis

Ibrahim El Salih
6867065

Organizations

Faculty of Social Sciences Utrecht University

Faculty of Information and Computer Science Utrecht University

Utrecht University

Acknowledgments

I would like to thank my supervisors Vincent Buskens and Mahdi Shafiee Kamalabad from Utrecht University. They both have supported and guided me throughout this process of writing my master thesis. In addition, I would like to thank Jin Dai, my fellow student who has been working with me on this topic.

Contents

List of tables

List of figures

List of abbreviations

GMM	Generalized method of moments
LM	Lagrange multiplier
ML	Maximum likelihood
SAR	Spatial autoregressive model
SARAR	Spatial autoregressive model with spatial disturbances
SEM	Spatial error model
SWM	Spatial weighs matrix
WHO	World health organization

Abstract

Background

The COVID-19 pandemic is the largest outbreak of an infectious disease in modern history. Europe has reported the most cases of COVID-19 as of July 2022 and every country has implemented some form of containment measures to mitigate the spread of COVID-19. Outbreaks of infectious diseases are however characterized by containment measures taken by one country having substantial consequences for others. Studying these spatial effects is crucial to understand the trajectory of COVID-19.

Methods

Spatial panel data models have been used to study the effects of containment measures on COVID-19 cases. Spatial panel data models have been chosen as this type of analysis encompasses both the temporal and spatial dimension. This study has used 38 European countries over a 109-week period. A fixed effects spatial lag and spatial error model has been conducted on this data.

Results

When a country implements stricter containment measures, this will decrease cases two weeks later in the same country. Furthermore, when countries implement stricter containment measures, this will decrease COVID-19 cases two weeks later in Europe. Spatial effects are present in containment measures on COVID-19 cases.

Conclusion

Understanding the spatial and temporal dimension of containment measures is crucial to design proper containment measures for COVID-19 and future infectious diseases with pandemic potential.

Chapter 1 Introduction

The COVID-19 pandemic is the largest outbreak of an infectious disease in modern history. Since early 2020, COVID-19 cases have spread throughout the world and have been going up and down in this period. At this moment of writing, 544,324,069 confirmed cases have been reported globally, including 6,332,963 deaths. For contact transmission, COVID-19 transmission occurs via human-to-human contact due to coughing, sneezing, speaking, singing, or breathing. In addition, people may become infected when touching their eyes, nose, or mouth after touching contaminated surfaces. The disease spread quickly across borders and continents, given the connected world we live in (World Health Organization, 2022). At the beginning of the pandemic, the World Health Organization (WHO) provided recommendations for governments on how to mitigate the spread of COVID-19, both pharmaceutical and non-pharmaceutical (World Health Organization, 2022). This included recommendations on an individual level such as face mask wearing but also recommendations on a group level such as mobility restrictions. Every country in the world implemented some form of these containment measures (Ritchie et al., 2022). Outbreaks of infectious diseases are characterized by containment measures taken by one country having substantial consequences for others (Gersovitz, 2014).

To tackle cross-border issues such as a pandemic, countries form multilateral agreements. In Europe, both small and large countries have benefited from the rise in multilateral agreements in the past decades. The COVID-19 pandemic has however showed weaknesses in the European multilateral system. The initial response was slow and inadequate, reflecting the urgency which was given on pandemic preparedness in previous years of European countries. Countries pursued their own self-interest in terms of containment measures to reduce cases (Jit et al., 2021)

Just as with any disease, prevalence results from individual characteristics. However, spatial dimensions and containment measures have a relation with the spread of COVID-19 (Wang et al., 2021). After two years, the world has learned more on the trajectory of COVID-19 and how to reduce cases. COVID-19 tends to increase exponentially (Vetter et al., 2020) (Palmer et al., 2021). Implementing strict containment measures at an early stage will limit the spread of COVID-19 through a country (Haug et al., 2020). The higher connectedness of countries implies higher virus circulation between the countries. The rate of COVID-19 cases has a spatial dimension just as other infectious diseases. Countries with a higher level of connectedness are prone to higher case rates. Furthermore, regions are more prone to higher cases when their neighbouring regions have a high incidence rate (Wang et al., 2021).

Knowledge is thus available on the impact that cases have on neighbouring countries. Knowledge on the effects that containment measures over time have on the cases of its neighbouring countries is however limited. Studying these temporal and spatial determinants is crucial to understand the entire magnitude of COVID-19 spread through a network. Europe is the most connected continent in the world in terms of human mobility, Therefore, using this continent as the region for this analysis is all the more relevant (Amdaoud et al., 2021).

1.1 Relevancy, urgency, objective

At this moment of writing, the COVID-19 pandemic is still ongoing. The latest Omicron variant has shown signs of weaker symptoms, reducing COVID-19 hospitalization rates all over Europe (WHO, 2022). European governments have lifted containment measures due to these lower COVID-19 hospitalization rates, giving governments time to reflect on the first two years of the pandemic (Ritchie et al., 2022).

Understanding the entire scope of the effects of containment measures on cases is urgent as the COVID-19 pandemic does not only encompass health related issues. As time progresses, more attention is focused on the socioeconomic effects due to COVID-19 measures. Just as the health system has been hit by COVID-19, so has the socioeconomic system. These systems may need years to recover, giving decision-makers more elements to consider in possible new COVID-19 waves (Amdaoud et al., 2021). In addition, new types of infectious disease with pandemic potential will appear in the future (Jit et al., 2021).

Given the connected world that we live in it is crucial to understand the spatial dimension between countries to design proper mitigation strategies for COVID-19 but also for future infectious diseases. Studying these temporal and spatial dimensions of COVID-19 is crucial, given that Europe is highly connected but the decision-making on containment measures is based on a national level. Uncovering the effects that containment measures taken on a national level have on its own cases but also on other countries paints a more complete picture of the impact of containment measures. The aim of this study is therefore to research the effects of COVID-19 containment measures of countries on COVID-19 cases in neighbouring countries in Europe. This will be done using a spatial panel data model to encompass both temporal and spatial effects.

1.2 Research question:

To what extent do COVID-19 containment measures of a country have an impact on COVID-19 cases of neighbouring countries?

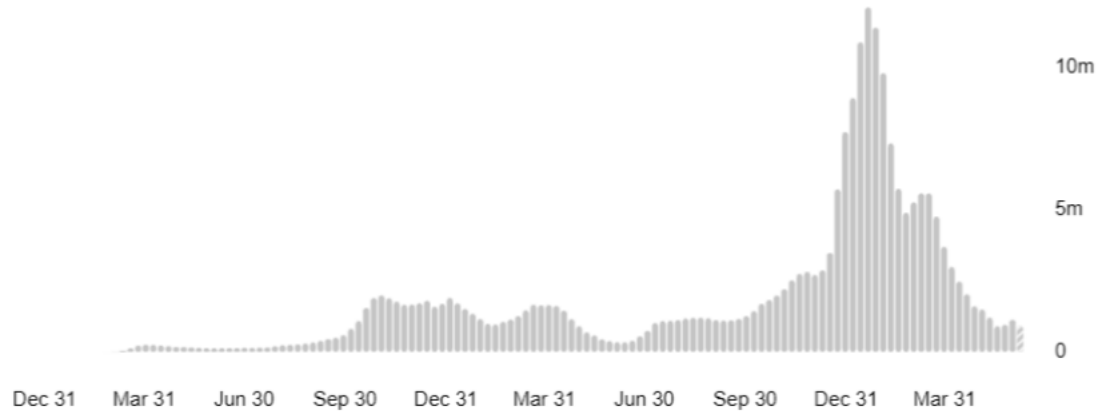
Chapter 2 Literature review

In this section a literature review will be conducted. Literature on the trajectory of cases of COVID-19 will be given. In addition, containment measures will be discussed and how these containment measures will be quantified during this study.

2.1. COVID-19 cases

At this moment of writing, 227,862,893 confirmed cases of COVID-19 have been reported in Europe, including 2,760,872 deaths. The true COVID-19 cases and deaths are expected to be higher. The first confirmed case was reported in France on the 24th of January 2020. By the 17th of March 2022, every country in Europe had confirmed cases. Europe experienced several peaks in cases. By far the largest peak in cases occurred in December 2021 due to the emergence of the Omicron variant. This peak lasted several weeks. As of July 2022, Europe is the most affected continent in the world in terms of confirmed cases (WHO, 2022).

Figure 1: COVID-19 cases in Europe 01-01-2020 until 01-07-2022 (WHO, 2022)



COVID-19 is a disease with strong spatial dimension in infection spread. The trajectory of COVID-19 throughout Europe did not go homogeneous. Significant differences occurred in infection rates between regions. Regional impacts were highly heterogenous and the effects on countries differed both on cases and on deaths. Differences did not only occur among countries but also within countries. The reason for this is the regional socioeconomic differences (OECD, 2020).

The relationship between infectious diseases and socioeconomic factors is not new. This topic has however raised more attention in the past decade. Multiple studies have been conducted which concluded that omitting the spatial element from an analysis on virus spread would lead to a partial understanding of infectious diseases (Linard et al., 2007) (Stanturf et al., 2015). These socioeconomic factors can be grouped into four categories: demographic determinants, income determinants, health care determinants and institutional determinants (Amdaoud et al., 2021).

Demographic determinants are particularly important for Europe as Europe is the oldest continent in the world and is expected to stay this for the foreseeable future. In addition, Europe is the most densely populated continent in the world (Eurostat, 2022). This is important as around 95% of people who die of COVID-19 are over 60 years, and 50% are over 80 years (WHO, 2022). The wealth and income in a region also play a role in the occurrence of COVID-19 just as in other diseases. In addition, the type of occupations in a region plays a role. Occupations which require one to be in close proximity of another will be more likely to be prone to infectious diseases. This was seen in for instance cleaners, healthcare workers, and teachers (Mongey et al., 2020). Furthermore, a significant relation exists between poverty and the prevalence of infectious diseases. This is because people in poverty are more likely to be less educated, are less likely to have health insurance, and work more often in close proximity occupation (Talavera & Perez, 2009) (Olson et al., 2012). Countries with a higher GDP per capita rate are however prone to higher infection rates, what may seem contradictive because of the previous statement. The reason behind this is that countries with higher levels of GDP per capita are also countries higher levels of connectedness between individuals. In addition, countries with higher levels of GDP per capita are more open to international trade and the movement of people. Clusters of COVID-19 for instance occurred at the centres of economic activity (Ascani et al., 2020). Furthermore, a country its health infrastructure plays a role in coping with pandemics. Well-structured health infrastructure positively affects a government its capacity to deal with rising hospitalization rates (Zanakis et al., 2007) (Gizelis et al., 2017). In addition, the country is better equipped in detecting, diagnosing, and reporting new cases (Hosseini et al., 2010) (Hogan et al., 2018). The institutional factor plays a role in the level of trust citizens have in its government. Government containment measures take time to enforce. Trust in public authorities is therefore of the essence to enforce these measures. Higher levels of trust in government institutions have a significant effect on mitigating the spread of COVID-19 (OECD, 2020).

The trajectory of COVID-19 in countries is therefore not solely dependent on the virus itself. Other factors in a country are at play which may influence the prevalence of COVID-19. It is important to be aware of these factors as they are mostly fixed over a relatively short time period. Understanding that demographic determinants, income determinants, health care determinants and institutional determinants al have an impact on cases is important to understand the spatial dimension of COVID-19 and design proper containment measures. This is because these determinants are challenging to influence to contain cases. These determinants are thus assumed to be time invariant and fixed when containing cases of COVID-19 (Terriau et al., 2021).

For the remainder of this study, COVID-19 cases will be considered new confirmed cases of COVID-19 per 1,000,000 people. Counts can include probable cases, where reported.

2.2. COVID-19 containment measures

Every European country has implemented some form of containment measure to control the spread of COVID-19, both pharmaceutical and non-pharmaceutical. Just as cases, the strictness of these containment measures has been going up and down in the past two years (Ritchie et al., 2022). The decision-making process was led by individual countries instead of on a multi-country level. At the beginning of the pandemic, European countries implemented relatively homogeneous types of non-pharmaceutical interventions to stop the spread of COVID-19. In general, some type of lockdown was implemented due to a lack of medical equipment and knowledge on the virus. Countries reintroduced border controls and social distancing measures were taken (Maurice et al., 2020). When cases dropped due to the seasonality of the virus and vaccines, countries lifted their restrictions. This process of rising cases leading to stricter measures was an ongoing process. The only difference with the beginning of the pandemic was that the types of measures differed more among countries as medical equipment was now more on hand and governments learned more about the virus. Countries weight independently which type of measure to take (Jit et al., 2021).

After two years, the world has learned from the impact the disease and containment measures had and still have on health and socioeconomic systems. In addition to the physical health risks, COVID-19 has had an impact on the mental health of people. Furthermore, socioeconomic systems have been disrupted due to containment measures, increasing economic hardship, food insecurity, lack of education, and lack of access to healthcare (Van Lancker & Parolin, 2020). Since early 2021, containment measures are not always received well across Europe. Reasons for this is economic loss, psychological burdens, inadequate communication, lack of long-term perspective, increasing vaccination coverage, and lack of trust in the government (Iftekhar et al., 2021).

Even though the measures are not always received well, the effect of stricter COVID-19 measures on lowering cases is significant. Countries which are found implementing fast and strict restrictions see a reduction in cases, with less and less new cases every day. When countries ease their restrictions, this increases cases. This however has a smaller effect than when countries tighten their measures. If population immunity is not achieved, measures are needed to mitigate the spread COVID-19 cases. The earlier these measures are implemented the more effective they are (Iftekhar et al., 2021). Countries therefore need to weigh whether they give more preference to the reduction of cases or the socioeconomic effect of stricter containment measures.

Knowledge on the impact of containment measures on COVID-19 cases in neighbouring countries is limited. One study by Ahmed and May (2021) mentions the impact of containment measures on cases in neighbouring regions. This study concludes that stricter containment measures in Northern Ireland have had a negative impact on COVID-19 cases in districts in Ireland which laid at the border of Northern Ireland.

The strictness of containment measures can be measured with the containment index. This is a measure of response metrics of governments. The index uses the following thirteen indicators: school closures, workplace closures, cancellation of public events, restrictions on public gatherings, closures of public transport, stay-at-home requirements, public information campaigns, restrictions on internal movements, international travel controls, testing policy, extent of contact tracing, face coverings, and vaccine policy. The index is calculated by taken the mean of the thirteen indexes. A higher score indicates a stricter containment measure. The containment index ranges from 0 to 100. When the decision-making process in not uniform across a country, the response level of the strictest region is taken (Ritchie et al., 2022).

Chapter 3 Theoretical framework

To study both the temporal and spatial dimension of COVID-19, a spatial panel data model will be used which encompasses both aspects. This section will describe the theory behind the use panel data and spatial panel data and their mathematical background. In addition, different models used in panel data models and spatial panel data models will be introduced.

3.1 Panel data model

Panel data has a different structure than ordinary data frames. Panel data refers to a cross section of observations which is repeated over multiple time periods. Examples of observations are individuals, groups, counties, regions, and countries. This analysis is often used in econometrics, epidemiology, and social science (Baltagi et al., 2013). The following equation displays a panel data regression model:

$$y_{it} = x'_{it}\beta + u_i + \varepsilon_{it}, \tag{1}$$

where *y* is the dependent variable, *i* is the index for observations, and *t* is the index for time. x'_{it} denotes the 1 x *k* vector of observations for the independent variables. β is the *k* x 1 vector of undetermined coefficients. u_i is the unobserved individual effects. ε is the error time of disturbance that varies with the individual and time. When is u_i is related to the x'_{it} , the panel data model is a fixed effects model, otherwise it is a random effects model (Fotheringham and Rogerson, 2008).

Fixed effects assume different intercepts in the regression equation. Fixed effects are variables that are constant across individuals. Examples are age, sex, and ethnicity. They either do not change or change at a constant rate over time. In theory, variables can change. These changes are however assumed to change at a slow rate and thus treated as constant.

The alternative is random effects. Random effects assume different disturbances in the regression equation. Examples are prices, costs, temperatures. These variables change faster over time and are therefore considered random.

Determining whether to use a fixed or random effects model can be done with tests. This can be done using with the F test and Wald test for fixed effects and the Breusch-Pagan Lagrange Multiplier (LM) test for random effects as can be seen in table 1. If both fixed and random turn out, the Hausman test can be conducted. The null hypothesis of the Hausman test is that either the fixed or random effects are not correlated with other independent variables. If the null hypothesis is rejected, random effect models will be suffering from the Gauss-Markov theory. This leads to biased estimates. Fixed effect models will remain unbiased. Therefore, if the null hypothesis is rejected, fixed effect models must be used (Bell & Jones, 2015).

Table 1. Determining the panel data model

Fixed effect (F test or Wald test)	**Random effect (Breusch-Pagan LM test)**	**Model**
H_0 is not rejected (No fixed effect)	H_0 is not rejected (No random effect)	Pooled
H_0 is rejected (Fixed effect)	H_0 is not rejected (No random effect)	Fixed effect model
H_0 is not rejected (No fixed effect)	H_0 is rejected (Random effect)	Random effect model
H_0 is rejected (Fixed effect)	H_0 is rejected (Random effect)	Hausman test

3.2 Spatial panel data model

Spatial data analysis deals with spatial interactions and spatial structures in regression models for cross-sectional and panel data (Anselin, 1998). In the past, spatial data analysis which uses geography was used in regional science and economics. More recently, it is used in social sciences and epidemiology. The focus has been put more on this type of analysis because of the interest in spatial effects, being spatial autocorrelation and spatial heterogeneity. Spatial autocorrelation concerns the degree of spatial dependence among observations in a geographic space. Spatial heterogeneity concerns the uneven distribution of various concentration of each observation in a geographic space (Katchova et al., 2002).

To perform spatial data analysis, the structure of spatial relationships among observations must be known. This can be captured using a spatial weights matrix (SWM). The SWM encompasses which observations are neighbours and how their values relate. The SWM can be defined as *W* where elements W_{ij} indicate whether observation *i* and *j* are spatially close. The dimensions of the SWM are *NxN* with *N* being the sample size. This can cause difficulties with large sample sizes. In an SWM, the diagonal elements are put to zero. The matrix is row standardized. This means that weights of each row need to add up to one. An SWM can be based on contiguity and on distance. Contiguity means whether a region is adjacent to another region. Distance means the inverse distance among regions up to a distance band. The SWM can be used to conduct spatial regression. Spatial regressions account for the spatial dependence in the data. This spatial dependence is added to the regression equation with spatial lag and spatial error (Lam & Souza, 2020).

The spatial lag model can be used when the spatial dependence revolves around the dependent variable. This model is a spatial autoregressive model, and it includes a spatially lagged dependent variable *y.* An example of this is that the number of COVID-19 cases in a region depends on the number of COVID-19 cases of its neighbouring regions. The independent variables in this equation explain the variation in the dependent variables which is not explained by the neighbours. The spatial lag is defined as *Wy.* This means that the dependent variable *y* is a weighted average of its neighbouring values. The model for spatial lag regression is called the spatial autoregression (SAR) model and is defined as:

$$y_{it} = \rho W y_t + x'_{it}\beta + u_i + \varepsilon_{it} \quad (2)$$

where *ρWyt* is the spatial lag of the dependent variable. *ρ* is the spatial autoregression coefficient. The *Wyt* is the contiguity based on the matrix and can be defined as:

$$Wy_t = \sum_{j=1}^{n} w_{ij} y_{it} \tag{3}$$

If *ρ* is statistically significant, there is spatial dependence in the dependent variable. The value for *ρ* is the degree of dependence. (Gelfand et al., 2010).

The spatial error model can be used when you want to correct for spatial autocorrelation because of the use spatial data. It does not matter here whether the model is spatial. The structure of the spatial relationship is not known. In this model you include spatially correlated errors due to the unobserved features which are related to the neighbour. The spatial error model (SEM) can be defined as:

$$y_{it} = x'_{it}\beta + u_i + \varepsilon_{it}$$
$$\varepsilon_{it} = \lambda W \varepsilon_t + v_{it} \tag{4}$$

where λWεt is the spatial error term and v_{it} is the random error term. This random error term is assumed to be independent. In this equation λ is the autoregressive factor. If λ is statistically significant, unobserved independent variables with spatial autocorrelation is present. This can be noticed with a trend in of spatial autocorrelation in the residuals (Elhorst, 2014).

3.2.1 Maximum Likelihood and Generalized Method of Moments

When the model is determined, both Maximum Likelihood (ML) and Generalized Method of Moments (GMM) estimators can be implemented in the model. The GMM approach is more robust than ML given that the ML approach uses assumptions about the entire distribution whereas GMM uses assumptions about specific moments. The GMM approach is therefore computationally less intensive and reduces running time. This makes the GMM method more suitable for large datasets and when both spatial lag and spatial error is present. Large datasets for spatial panel data models are considered to have 10,000 cross sectional observations. This is therefore not relevant for this study. Results are expected to be relatively similar when using either method (Millo & Piras, 2012).

Chapter 4 Methods and techniques

To answer the research question, a spatial panel data analysis will be conducted. This will be done by firstly conducting a panel data analysis. Secondary data will be used from Our World in Data. To conduct these analyses SQL, GeoDA, QGIS, and R will be used.

4.1 Data preparation

Data was extracted from Our World in Data (Ritchie et al., 2022). Both data files on cases and containment index were merged using SQL code. The values were aggregated into weekly data as they were presented in daily data with R code. This was done to remove daily fluctuation in the data. Panel data is a data structure which consists of a cross section of individuals (countries) repeated over several time periods (109 weeks) (Baltagi et al., 2013). For this study, the cross section of individuals is 38 countries on the European continent. The number of countries is limited to 38 for two conditions:

- The country needs at least one neighbouring country in terms of land borders.
- The country needs to have data available for COVID-19 cases and containment measures.

After going through these conditions, 38 countries remained. The time periods were taken from the 2nd of March 2020 until the 27th of March 2022, accounting for 109 weeks. Mean imputation was conducted as the number of missing values is less than one percentage being 0.56% .

The analysis has been conducted in R with the “pml” and “spml” package (Yves & Croissant, 2008) (Millo & Piras, 2012). The pml package will be used for the panel model data analysis. The spml package will be used for the spatial panel model analysis. These packages are designed based on the theory described in the theoretical framework. For the analysis, requirements for both packages are that data are transformed into panel data. To transform the data to panel data an index for the observations and an index for the time is needed. These indexes are listed below:

Indices

- **Observations**: Country_ID
- **Time:** Week

Both packages also require a dependent variable and independent variable(s). These are as follows:

Dependent variable

- **log(cases):** the country its confirmed cases of COVID-19 per 1,000,000 people. Counts can include probable cases, where reported.

Independent variables

- **log(cases_lag_2):** the country its confirmed cases of COVID-19 per 1,000,000 people with a two-week lag. Counts can include probable cases, where reported.
- **log(cases_nb_lag_2):** the average confirmed cases of COVID-19 per 1,000,000 people of the neighbouring countries with a two-week lag. Counts can include probable cases, where reported.
- **ci_lag_2:** the containment index of the country with a two-week lag.
- **ci_nb_lag_2:** the average containment index of the neighbouring countries with a two-week lag.

When the distributions of the dependent and four independent variables is visualized, it is displayed that a significant left skew is visible in the variables cases, cases_lag_2 and cases_nb_lag_2. As these variables are skewed, they are logarithmically transformed to acquire a distribution that reflects a normal distribution (Benoit, 2011), Therefore, the log for these variables is taken. The ci_lag_2 and ci_nb_lag_2 remain as is.

Figure 2. Distributions of variables

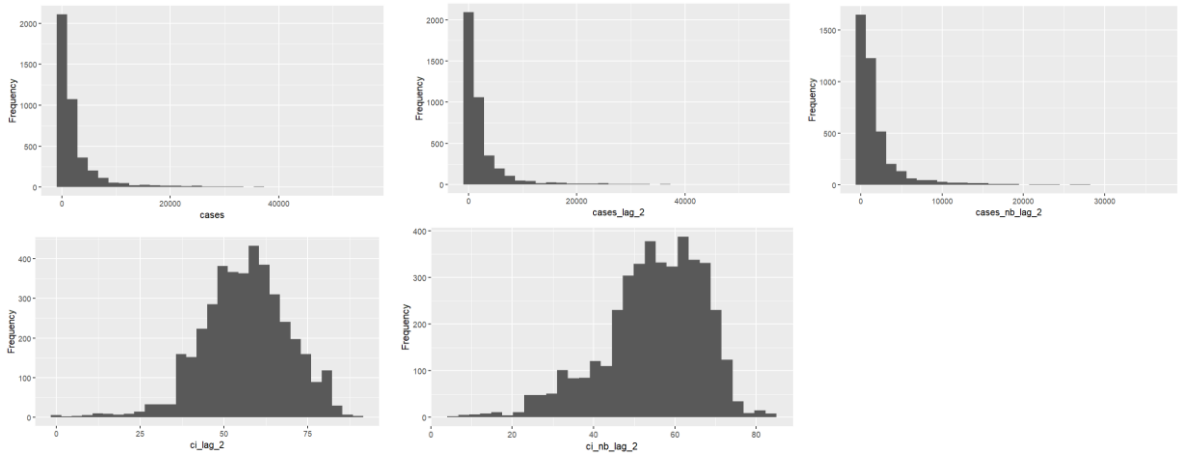
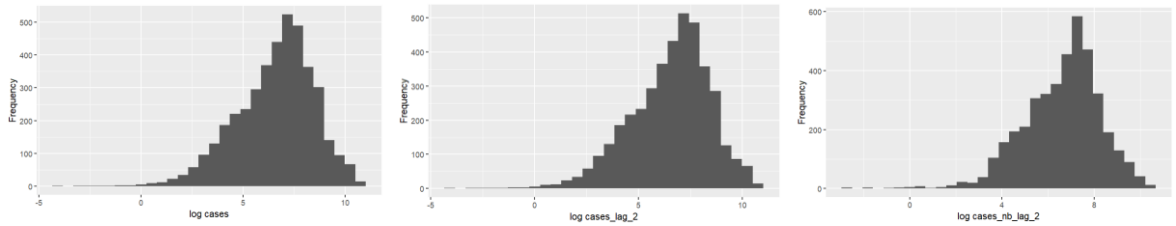


Figure 3. Distribution of variables after log transformation



The panel data frame consists of 38 countries with seven features over the same time period (week 1 to 109). Each country has values for every week, making it a balanced panel data set, a requirement for the plm and splm package in R.

The SWM was computed using QGIS, Geoda, and R. To construct an SWM, a panel data frame is needed. In addition, a Shapefile and a GAL file are needed. To construct the Shapefile and GAL file, data on the 38 countries was extracted from Natural Earth Data (Natural Earth Data, 2022). This data includes features of countries and how they relate. This datafile was imported into QGIS. In QGIS, the 38 countries from the QGIS data frame were merged with the 38 countries from the panel data frame. After selection and merging, a new data frame was extracted, and a Shapefile was created. In GeoDA the relationships whether countries are neighbours is defined based on contiguity, thus adjacent. These relationships are then extracted from GeoDA as a GAL file. Both the Shapefile and GAL file were imported in R. In R code was written to construct the SWM based on contiguity.

4.2 Data models

4.2.1 Panel data model

The plm package provides four functions. For this analysis only the plm function will be used. This function estimates the fixed and random effect models. (Croissant & Millo, 2008). The plm package requires that each row of the data corresponds to a specific observation (country) and time (week). The data argument must be as follows:

- NULL (the default value), it is then assumed that the first two columns contain the individual and the time index and that observations are ordered by individual and by time period.
- A character string, which should be the name of the individual index.
- A character vector of length two containing the names of the individual and the time index.
- An integer which is the number of individuals (only in case of a balanced panel with observations ordered by individual).
- A panel data frame where individual and time indexes need to be specified.

The display of the plm function is similar to the linear model lm() function. The first argument needs to include the formula and the second argument should be data as described in the above five possibilities. Additionally, three arguments are available:

- index: this argument enables the estimation functions to identify the structure of the data, such as the individual and the time period for each observation.
- effect: the kind of effects to include in the model, such as individual effects, time effects or both.
- model: the kind of model to be estimated, most of the time a model with fixed effects or a model with random effects.

4.2.2 Spatial panel data model

The splm package builds further on the plm package, adding the spatial dimension. Just as in the plm package, the spml package requires that each row of the data corresponds to a specific individual (country) and time (week) (Millo & Piras, 2012). The data argument must be as follows:

- A data frame whose first two variables are the individual and time indexes. The index argument should be left to the default value (i.e., NULL).
- A data frame and a character vector indicating the indexes variables.
- A panel data frame where individual and time indexes need to be specified.

The structure of the associations between the spatial unit is represented by an SWM. An SWM W is a $N \times N$ positive matrix (Bates & Maechler, 2012). The observations appear in both rows and columns of the matrix. The non-zero elements represent whether two locations (countries) are neighbours. The diagonal elements are thus all set to zero. The reason for this is to exclude self-neighbours. The spatial weights are row standardized. The SWM can be a matrix object or a listw object (Lam & Souza, 2020).

Chapter 5 Findings

In this section the results of the analysis will be discussed. The results of the panel data model and the spatial panel data model will be presented including the R output. The results will shed a light on the spatial dependence of COVID-19 cases and containment measures. Firstly, the analysis will start with the panel data model without spatial autocorrelation. Secondly, the Hausman test will determine whether to choose the fixed effects or random effects model. Thirdly, the LM tests will determine whether to choose a SAR or SEM model. Fourthly, the results of the spatial panel data models will be presented. In table 2 a summary of the variables can be seen.

Table 2. Summary statistics of dependent and independent variables

Statistic	N	Mean	St. Dev.	Min	Max
Log(cases)	4142	6.53	2.05	0.01	10.92
log(cases_lag_2)	4142	6.53	2.05	0.01	10.92
log(cases_nb_lag_2)	4142	6.62	1.66	0.07	10.50
ci_lag_2	4142	56.44	12.73	0.01	90.00
ci_nb_lag_2	4142	54.79	12.18	5.58	83.81

5.1 Panel data model

The panel data model can be used select the most suitable specification for the spatial panel data model. The model is without spatial autocorrelation and the Hausman test and LM test will be conducted to proceed with the spatial panel data model. Table 3 shows the results of the panel data model analysis. The pooled data model will be used as baseline to compare the fixed effects and random effects models. These models take into account the unobserved heterogeneity. In the fixed effects model, an increase in the country its cases will increase its cases two weeks later. A similar conclusion can be drawn of the of the cases of its neighbouring countries, although on a smaller scale. The measures of a country do not have a significant effect on its cases two weeks later. While the measures of its neighbouring countries lead to a decrease in cases two weeks later. The rounded adjusted R square is 0.31. For the random effects model all independent variables are significant. An increase in the country its cases and that of its neighbours increases the cases of the country after two weeks. While when a country and when neighbouring countries implement stricter measures, this leads to a reduction in the cases of the country after two weeks. Where the measures of the neighbouring countries have a bigger effect. The rounded adjusted R square is 0.31.

Table 3. Summary of panel data models

Model	pooled data	fixed effects	random effects
log(cases_lag_2)	0.8964***	0.8770***	0.8964***
log(cases_nb_lag_2)	-0.0318	0.0127	-0.0318
ci_lag_2	-0.0110***	-0.0097***	-0.0106***
ci_nb_lag_2	-0.0058***	-0.0215***	-0.0058***
constant	1.8018***		
observations	4142	4142	4142
R squared adjusted	0.7070	0.7015	0.7070

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

5.2 Tests

5.2.1 Determining the panel data model

Determining fixed effects or random effects can be done with the F test for fixed effects, LM test for random effects, and the Hausman test when both are considered.

The F test for fixed effects rejects the null hypothesis, indicating that fixed effects are present in the data:

- $F(37,4100) = 3.568$, $p < 0.001$.

The LM test for random effects rejects the null hypothesis as well, indicating that random effects are present in the data:

- 4.5778, $p < 0.001$.

As both tests reject the null hypothesis, the Hausman Test will be conducted. The Hausman Test rejects the null hypothesis of absence of correlation between individual effects and independent variables. Therefore, the analysis proceeds with the fixed effects model:

- $X^2(4) = 113.22$, $p < 0.001$

5.2.2 Determining the spatial panel data model

Determining the which spatial panel data model to use can be done with the LM tests. These tests include the LM test for spatial lag dependence, LM test for spatial error, and two extensions of these tests (RLMlag, RLMerr).

LM test for spatial lag dependence:

- LM(1) = 1385.10, $p < 0.001$

LM test for spatial error dependence:

- LM(1) = 1949.70, $p < 0.001$

The results of the LM tests conclude that for the spatial panel data model a spatial autoregressive model with spatial autoregressive disturbances is preferred. The LM test for spatial lag dependence and spatial error dependence both reject the null hypothesis, indicating that both spatial lag and spatial error dependence is present. Therefore, no choice can be made on whether to use a SAR or SEM model. Therefore, robust tests are conducted. Test 3 The rlml version tests for the absence of a spatial autoregressive term when the model contains a spatial autoregressive term in the errors. The rlme tests for the absence of a spatial autoregressive term in the errors when the models contain a spatial autoregressive term.

Locally robust LM test for spatial lag dependence sub spatial error (RLMlag):

- LM(1) = 24.608, $p < 0.001$

Locally robust LM test for spatial error dependence sub spatial lag (RLMerr):

- LM(1) = 589.17, $p < 0.001$

Both tests reject the null hypothesis and therefore it is not possible to make a choice using a SAR or SEM model based on the LM tests. Therefore, the procedure is to follow the SARAR model , which includes both the spatial dependencies. Nonetheless, the SEM model is performed. This is due to the SAR dimension being present in the cases_nb_lag_2 and ci_nb_lag_2 variables.

5.3 Spatial panel data model

The results of the spatial panel data analyses are displayed in table 4. Again, the pooled data model will be used as a baseline. Two models based on ML are displayed, the baltagi and KKP error model, and the GMM model is displayed.

Conclusions of all three models are similar. Firstly, an increase in cases leads to an increase in cases two weeks later. Secondly, an increase in strictness of containment measures decrease cases two weeks later. Thirdly, an increase in strictness of containment measures in neighbouring countries decreases cases two weeks later. All these three variables are significant, $p < 0.001$. The cases of the neighbouring countries decrease cases two weeks later. This variable is however not significant in all three models.

The results of the two models based on ML are similar. This is due to the KKP being an extension of baltagi and is used for particularly for spatial lag in SAR models (Baltagi & Liu, 2016). The results of GMM model differ from the ML models. This is due to that the ML approach uses assumptions about the entire distribution whereas GMM uses assumptions about specific moments (Millo & Piras, 2012).

To conclude, cases of a country have a positive effect on a country its cases two weeks later. Whereas containment measures of a country and its neighbouring countries have a negative effect on its cases two weeks later. There appears to be no relationship between cases of neighbouring countries and a country its own cases.

Table 4. Summary of spatial panel data models

Model	**pooled data model**	**fixed effects baltagi error model**	**fixed effects KKP error model**	**fixed effects GMM model**
log(cases_lag_2)	0.8690	0.8432***	0.8432***	0.8783***
log(cases_nb_lag_2)	-0.0411	-0.0151	-0.0151	-0.0326**.**
ci_lag_2	-0.0045	-0.0084***	-0.0084***	-0.0127***
ci_nb_lag_2	0.0007	-0.0145***	-0.0145***	-0.0086***
constant	1.1348			
Rho	0.6521	0.6539***	0.6539***	0.2475***
observations	4142	4142	4142	4142

Signif. codes: 0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

6 Discussion

Results from this study have shown that an increase in COVID-19 cases in a country result in higher COVID-19 cases two weeks later in the same country. A result which is already widely described in the literature (Vetter et al., 2020) (Palmer et al., 2021). Furthermore, this study has shown that when a country implements stricter containment measures, this will decrease cases two weeks later in the same country. A result which is also already widely described in the literature (Haug et al., 2020). A spatial effect which has also been described in the literature, is that regions are more prone to higher cases when their neighbouring regions have a high incidence rate (Wang et al., 2020). The results of this study however are that this relation is not significant. This could be explained by that the study of Wang et al. (2020) was conducted on regions in China instead of on a country level. In addition, different types of analyses were conducted.

The main aim of this study is to research the effects of COVID-19 containment measures of countries on COVID-19 cases in neighbouring countries in Europe over time. The results from this study show that when countries implement stricter containment measures, this will decrease COVID-19 cases two weeks later in Europe. Spatial effects are present in containment measures on COVID-19 cases. This is in line with the study of Ahmed and May (2021) which researched this topic but on a smaller scale in Ireland and Northern Ireland.

This negative relation can be explained by that the containment index incorporates mobility restrictions which affects cross border relations. The containment index for instance incorporates closures of public transport, stay-at-home requirements, restrictions on internal movement, and international travel controls. Where a decrease in mobility across regions is associated with lower COVID-19 cases (Warren & Skillman, 2020) (Nouvellet et al., 2021) These restrictions reduce the flow of people across borders and thus the connectedness of countries.

This study has several limitations. Firstly, this study has been focused on the impact of containment measures on cases. The registration of cases differs between country but also within countries over time. Countries in Europe have had different testing policies throughout the pandemic which could have an impact on results (Goniewicz et al.,2020). Secondly, this study has been solely focused on cases. An addition to this study could be to incorporate other elements such as COVID-19 hospitalizations or deaths to encompass the full scope of the pandemic. Furthermore, in some countries the priority in registration has been focused more on hospitalizations and deaths than on cases (Karanikolos & Mckee, 2020). Thirdly, this study has been conducted on a national level. An additional dimension could be to study this topic on a regional level. Containment measures in some countries have been conducted on a regional level instead of on a national level. In addition, when measures of the containment index vary within countries, the containment index takes the level of the strictest region (Ritchie et al., 2022).

A strength of this study is that it is the first study which focuses on the impact of containment measures on COVID-19 cases across borders over time. The use of the spatial panel data model across countries provides a more complete picture of the pandemic than studies focusing on one country or on one point in time. It highlights the spatial and the temporal dimension of COVID-19. Elements which are crucial to understand with infectious diseases and the connected world that we live in. .

To conclude, this study shows that just as country its own containment measures negatively influence cases, so does the containment measures of its neighbouring countries. These findings contribute to understanding the relationship between containment measures and cases on a multi-country level. Understanding the spatial and temporal dimension of containment measures is crucial to design proper containment measures for COVID-19 and future infectious diseases with pandemic potential.

Bibliography

Ahmed, R., & May, P. (2021). Does high COVID-19 spread impact neighbouring countries? Quasi-experimental evidence from the first year of the pandemic in Ireland. *HRB Open Research*.

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*, 629-642.

Anselin, L., & Bera, A. K. (1998). Spatial dependence in linear regression models with an introduction to spatial econometrics. *Statistics textbooks and monographs*, 237-290.

Ascani, A. F. (2021). The geography of COVID-19 and the structure of local economies: The case of Italy. *Journal of Regional Science*, 407-441.

Baltagi, B. H. (2013). A generalized spatial panel data model with random effects. *Econometric reviews*, 650-685.

Baltagi, B. H. (2016). Random effects, fixed effects and Hausman's test for the generalized mixed regressive spatial autoregressive panel data model. *Econometric Reviews*, 638-658.

Bell, A., & Jones, K. (2015). Explaining fixed effects: Random effects modeling of time-series cross-sectional and panel data. *Political Science Research and Methods*, 133-153.

Benoit, K. (2011). Linear regression models with logarithmic transformations. *London School of Economics*, 23-36.

Croissant, Y., & Millo, G. (2008). Panel data econometrics in R: The plm package. *Journal of statistical software*.

Eurostat. (2021, October 13). *Are you younger or older than the median age in your region?* . From Eurostat: https://ec.europa.eu/eurostat/web/products-eurostat-news/-/ddn-20211013-2#:~:text=In%202020%2C%20the%20median%20age,population%20stood%20at%2043.9%20years.

Gersovitz, M. (2014). Infectious disease externalities. *Encyclopedia of Health Economics*, 35.

Gizelis, T. I., Karim, S., Østby, G., & Urdal, H. (2017). Maternal health care in the time of Ebola: a mixed-method exploration of the impact of the epidemic on delivery services in Monrovia. *World Development*, 169-178.

Goniewicz, K., Khorram-Manesh, A., Hertelendy, A. J., Goniewicz, M., Naylor, K., & Burkle, F. M. (2020). Current response and management decisions of the European Union to the COVID-19 outbreak: a review. *Sustainability*.

Haug, N., Geyrhofer, L., Londei, A., Dervic, E., Desvars-Larrive, A., Loreto, V., & Klimek, P. (2020). Ranking the effectiveness of worldwide COVID-19 government interventions. *Nature human behaviour*, 1303-1312.

Iftekhar, E. N., Priesemann, V., Balling, R., Bauer, S., Beutels, P., Valdez, A. C., & ... & Willeit, P. (2021). A look into the future of the COVID-19 pandemic in Europe: an expert consultation. *The Lancet Regional Health-Europe*, 100185.

Jit, M., Ananthakrishnan, A., McKee, M., Wouters, O. J., Beutels, P., & Teerawattananon, Y. (2021). Multi-country collaboration in responding to global infectious disease threats: lessons for Europe from the COVID-19 pandemic. *The Lancet Regional Health-Europe*, 100221.

Karanikolos, M., & McKee, M. (2020). How comparable is COVID-19 mortality across countries? *Eurohealth*, 45-50.

Katchova, A. L., Sherrick, B. J., & Barry, P. J. (2002). The effects of risk on farmland values and returns.

Lam, C., & Souza, P. C. (2020). Estimation and selection of spatial weight matrix in a spatial lag model. *Journal of Business & Economic Statistics*, 693-710.

Linard, C. L. (2007). Determinants of. *Int. J. Health Geogr*, 15.

Maurice, E., Besnier, T., & Lazarovici, M. (2020). Restoring free movement in the Union. *Lett-Fond Robert Schuman*.

Millo, G., & Piras, G. (2012). splm: Spatial panel data models in R. *Journal of statistical software*, 1-38.

Mongey, S., Pilossoph, L., & Weinberg, A. (2020). Which workers bear the burden of social distancing? *CEPR Press*, 69-86.

Natural Earth Data. (2022). *Cultural vectors*. From Natural Earth Data: http://www.naturalearthdata.com/downloads/10m-cultural-vectors/

Nouvellet, P., Bhatia, S., Cori, A., Ainslie, K. E., Baguelin, M., Bhatt, S., & Donnelly, C. A. (2021). Reduction in mobility and COVID-19 transmission. *Nature* , 1-9.

OECD. (2020). *: The territorial impact of COVID-19: Managing the crisis.* OECD. From http://www.oecd.org/coronavirus/policy-responses/the-territorial-impact-of-COVID-19-managing-the-crisis-across-levels-of-government-d3e314e1/

Olson, N. A., Davidow, A. L., Winston, C. A., Chen, M. P., Gazmararian, J. A., & Katz, D. J. (2012). A national study of socioeconomic status and tuberculosis rates by country of birth. *BMC public health*, 1-7.

Palmer, S. C. (2021). COVID-19 hospitalization rates rise exponentially with age, inversely proportional to thymic T-cell production. *Journal of the Royal Society Interface*, 20200982.

Ritchie, H., Mathieu, E., Rodés-Guirao, L., Appel, C., Giattino, C., Ortiz-Ospina, E., . . . Roser, M. (2022). *Coronavirus Pandemic (COVID-19)*. From Our World in Data: https://ourworldindata.org/coronavirus

Stanturf, J. A., Goodrick, S. L., Warren Jr, M. L., Charnley, S., & Stegall, C. M. (2015). Social vulnerability and Ebola virus disease in rural Liberia. *PLoS One*, e0137208.

Talavera, A., & Perez, E. M. (2009). Is cholera disease associated with poverty? *The Journal of Infection in Developing Countries*, 408-411.

Terriau, A., Albertini, J., Montassier, E., Poirier, A., & Le Bastard, Q. (Scientific reports). Terriau, A., Albertini, J., Montassier, E., Poirier, A., & Le Bastard, Q. *Estimating the impact of virus testing strategies on the COVID-19 case fatality rate using fixed-effects models*, 1-8.

Van Lancker, W. &. (2020). COVID-19, school closures, and child poverty: a social crisis in the making. *The Lancet Public Health*, e243-e244.

Vetter, P., Vu, D. L., L'Huillier, A. G., Schibler, M., Kaiser, L., & Jacquerioz, F. (2020). Clinical features of covid-19. *Bmj*, 369.

Wang, Q., Dong, W., Yang, K., Ren, Z., Huang, D., Zhang, P., & Wang, J. (2021). emporal and spatial analysis of COVID-19 transmission in China and its influencing factors. *International Journal of Infectious Diseases*, 675-685.

Warren, M. S., & Skillman, S. W. (2020). Mobility changes in response to COVID-19. *arXiv preprint*.

WHO. (2022). *WHO Coronavirus (COVID-19) Dashboard*. From WHO: https://covid19.who.int/

Zanakis, S. H., Alvarez, C., & Li, V. (2007). Socio-economic determinants of HIV/AIDS pandemic and nations efficiencies. *European Journal of Operational Research*, 1811-1838.

7 Appendices

7.1 List of countries

1	Albania
2	Andorra
3	Austria
4	Belarus
5	Belgium
6	Bosnia and Herzegovina
7	Bulgaria
8	Croatia
9	Czechia
10	Denmark
11	Estonia
12	Finland
13	France
14	Germany
15	Greece
16	Hungary
17	Iceland
18	Ireland
19	Italy
20	Kosovo
21	Latvia
22	Liechtenstein
23	Lithuania
24	Luxembourg
25	Moldova
26	Netherlands
27	Norway
28	Poland
29	Portugal
30	Romania
31	Russia
32	Serbia
33	Slovakia
34	Slovenia
35	Spain
36	Sweden
37	Ukraine
38	United Kingdom

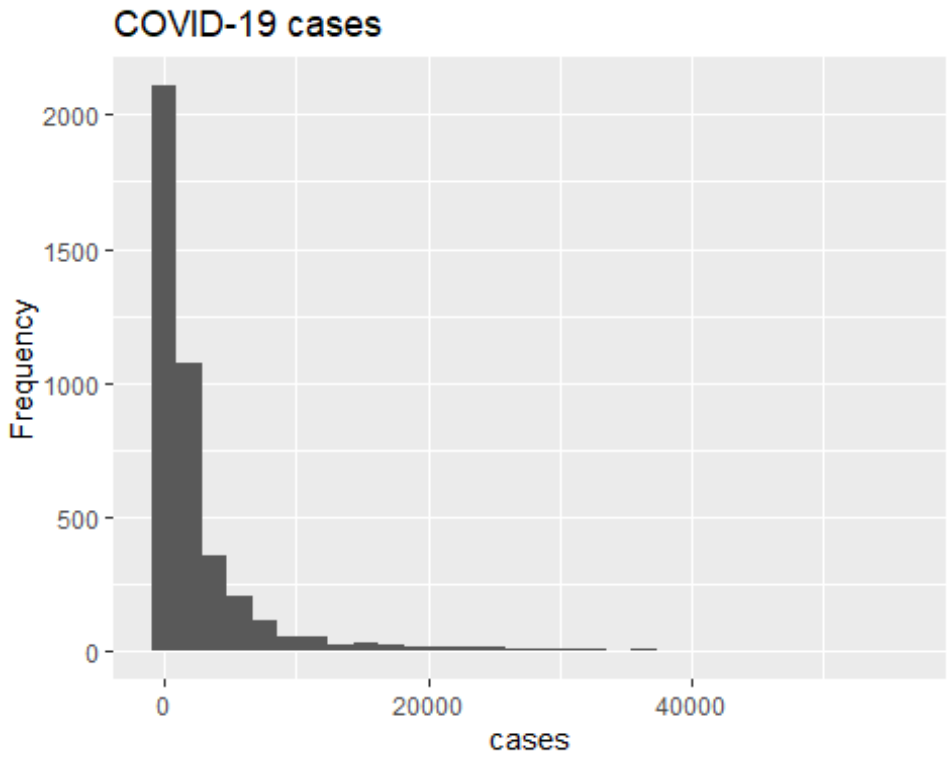
7.2 R code

```
#Loading libraries
library(splm)

library(plm)

library(dplyr)

library(spdep)

library(readr)

library(rgdal)

library(lmtest)

library(ggplot2)
```

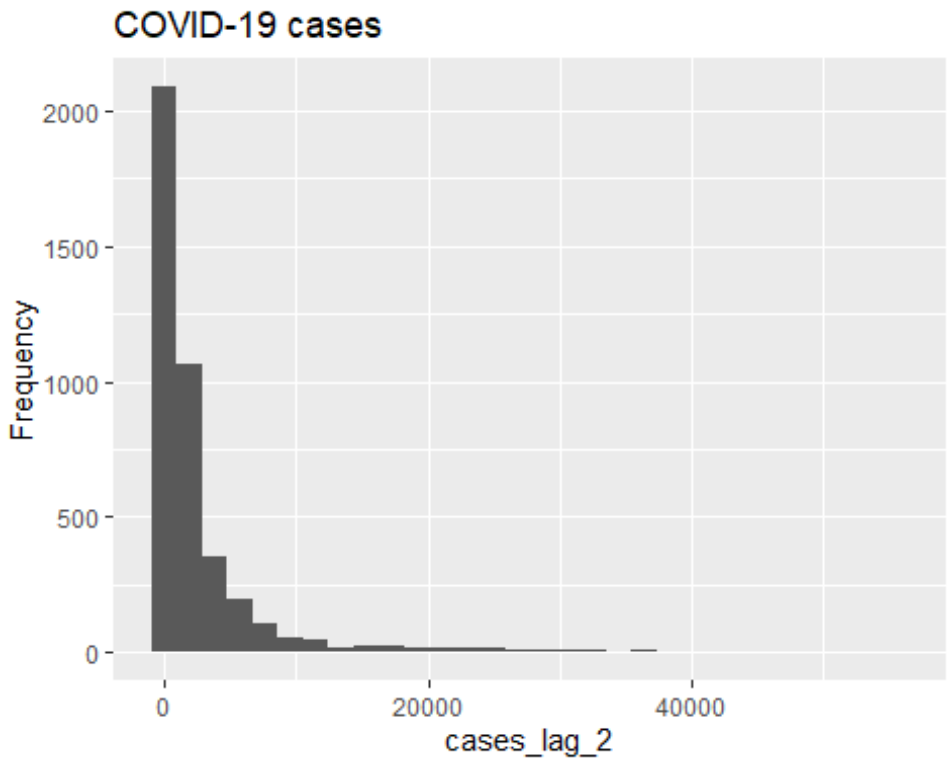
Data preparation

```
#Reading in Excel file
library(readxl)

CI <- read_excel("thesisproject.xlsx")

#Creating column: "cases_lag_2"
step1  <- CI %>%
  group_by(id) %>%
  dplyr::mutate(cases_lag_2 = dplyr::lag(cases, n = 2, default = NA)) %>%
  as.data.frame()
#Creating column: "cases_nb_lag_2"
step2 <- step1 %>%
  group_by(id) %>%
  dplyr::mutate(cases_nb_lag_2 = dplyr::lag(cases_nb, n = 2, default = NA)) %>%
  as.data.frame()
#Creating column: "ci_lag_2"
step3 <-step2 %>%
  group_by(id) %>%
  dplyr::mutate(ci_lag_2 = dplyr::lag(ci, n = 2, default = NA)) %>%
  as.data.frame()
#Creating column "ci_nb_lag_2"
step4 <- step3 %>%
  group_by(id) %>%
  dplyr::mutate(ci_nb_lag_2 = dplyr::lag(ci_nb
, n = 2, default = NA)) %>%
  as.data.frame()

#Creating panel data
pdata <- pdata.frame(step4, index=c("id", "week"))

#Calculating percentage of missing values
sum(is.na(pdata))/53846

## [1] 0.00564573

#Mean imputation
pdata$cases[is.na(pdata$cases)] <- mean(pdata$cases,na.rm = TRUE)
pdata$cases_lag_2[is.na(pdata$cases_lag_2)] <- mean(pdata$cases_lag_2,na.rm = TRUE)
pdata$cases_nb_lag_2[is.na(pdata$cases_nb_lag_2)] <- mean(pdata$cases_nb_lag_2,na.rm = TRUE
)
pdata$ci_lag_2[is.na(pdata$ci_lag_2)] <- mean(pdata$ci_lag_2,na.rm = TRUE)
pdata$ci_nb_lag_2[is.na(pdata$ci_nb_lag_2)] <- mean(pdata$ci_nb_lag_2,na.rm = TRUE)

#Setting 0's from the lagged variables to 0.001.
pdata[pdata == 0] <- 0.001
```

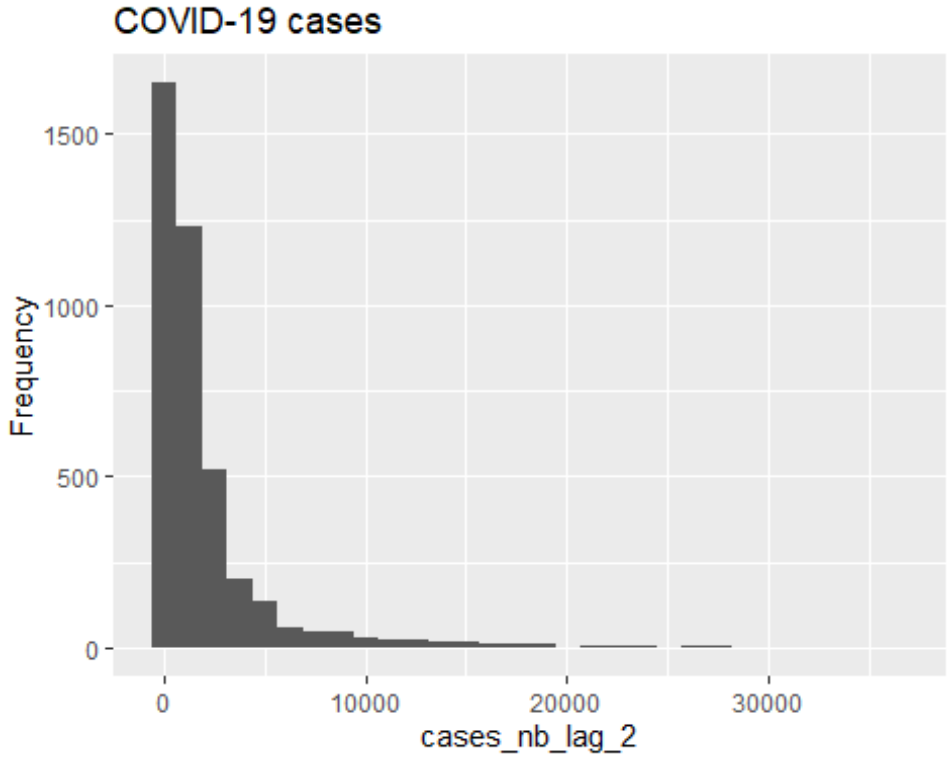
```
#Visualizing cases
ggplot(step4, aes(x = step4$cases)) +
  geom_histogram() +
  labs(title = "COVID-19 cases",
       x = "cases",
       y = "Frequency")
```




```
#Visualizing cases_lag_2
ggplot(step4, aes(x = step4$cases_lag_2)) +
  geom_histogram() +
  labs(title = "COVID-19 cases",
       x = "cases_lag_2",
       y = "Frequency")
```

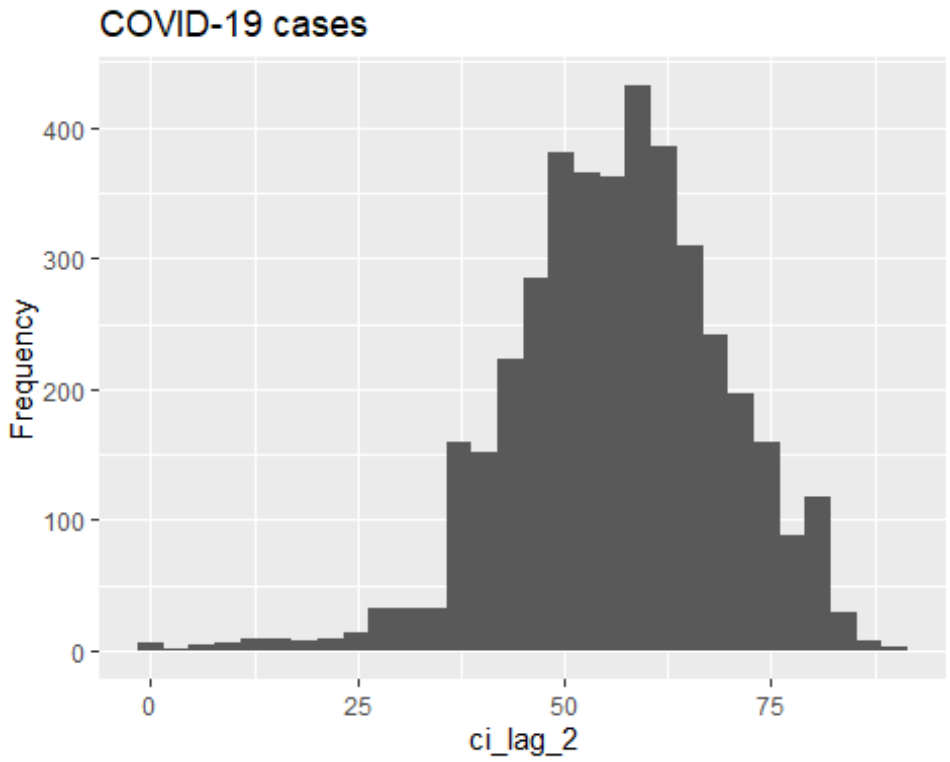



```
#Visualizing cases_nb_lag_2
ggplot(step4, aes(x = step4$cases_nb_lag_2)) +
  geom_histogram() +
  labs(title = "COVID-19 cases",
       x = "cases_nb_lag_2",
       y = "Frequency")
```

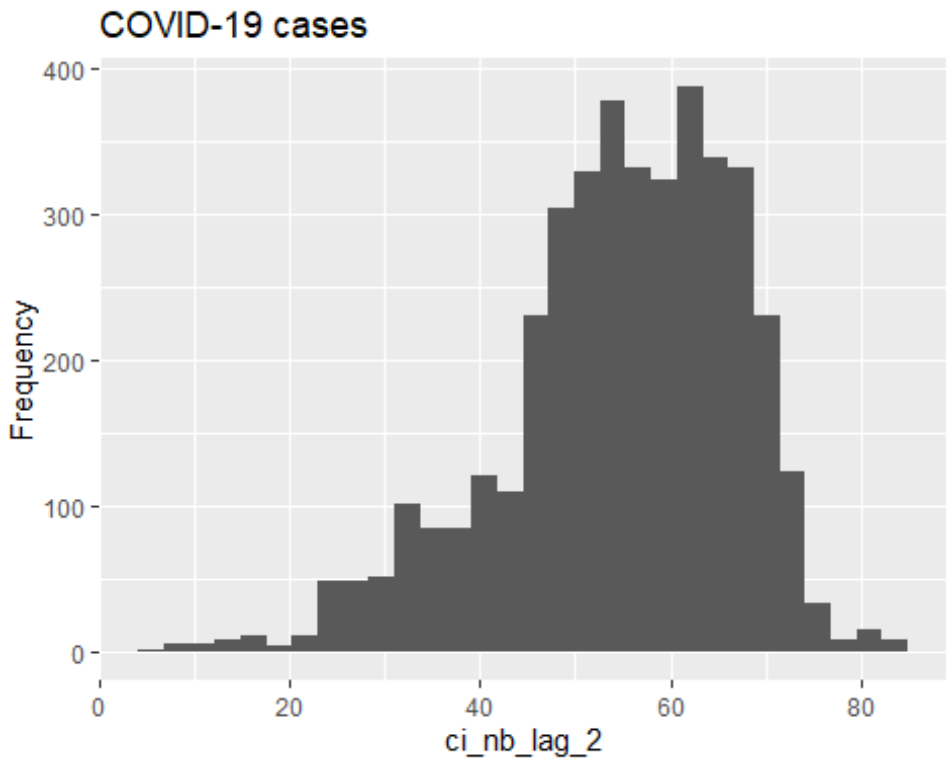



```
#Visualizing ci_lag_2
ggplot(step4, aes(x = step4$ci_lag_2)) +
```

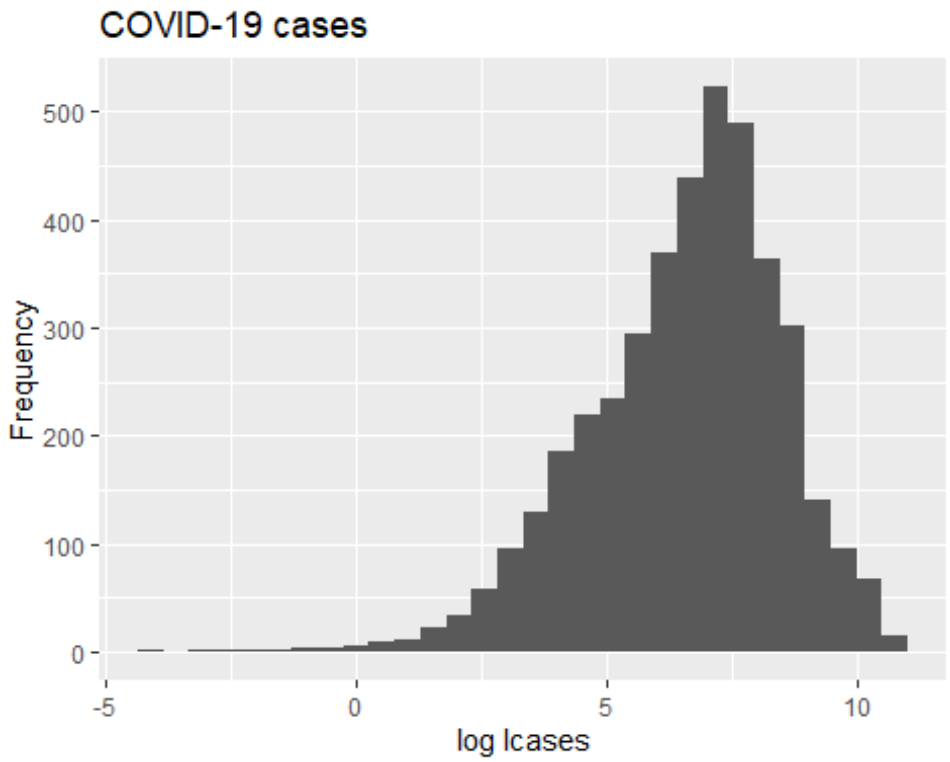
```
  geom_histogram() +
  labs(title = "COVID-19 cases",
       x = "ci_lag_2",
       y = "Frequency")
```



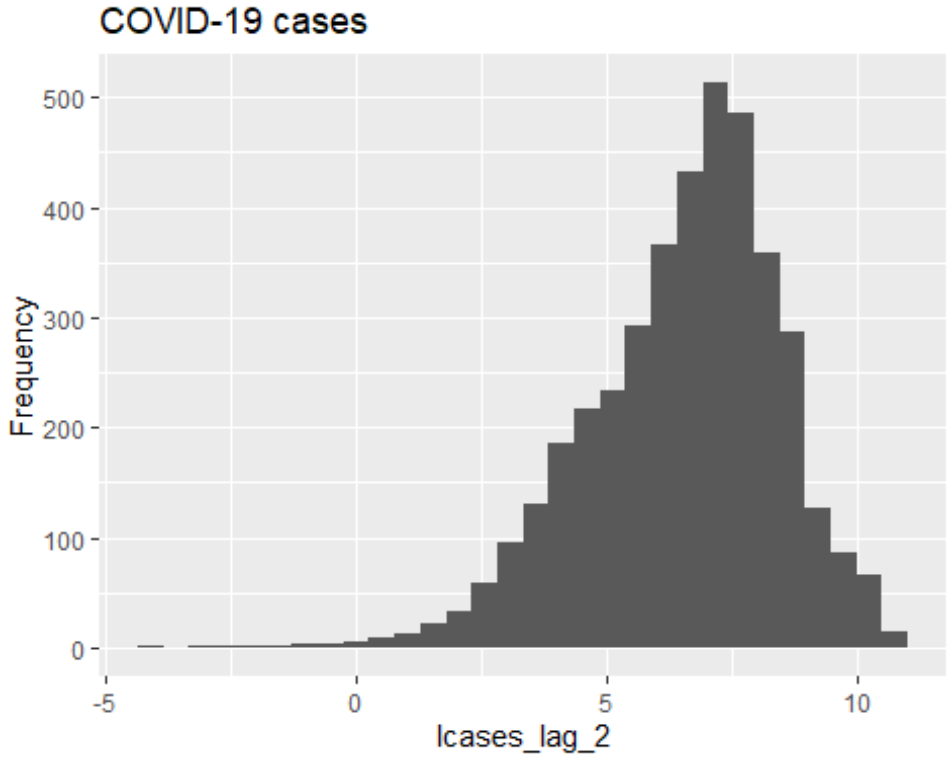

```
#Visualizing ci_nb_lag_2
ggplot(step4, aes(x = step4$ci_nb_lag_2)) +
  geom_histogram() +
  labs(title = "COVID-19 cases",
       x = "ci_nb_lag_2",
       y = "Frequency")
```




```
#Log transformation for visualizing
lcases <- log(step4$cases)
lcases_lag_2 <- log(step4$cases_lag_2)
lcases_nb_lag_2 <- log(step4$cases_nb_lag_2)

#Visualizing lcases
ggplot(step4, aes(x = lcases)) +
  geom_histogram() +
  labs(title = "COVID-19 cases",
       x = "log lcases",
       y = "Frequency")
```

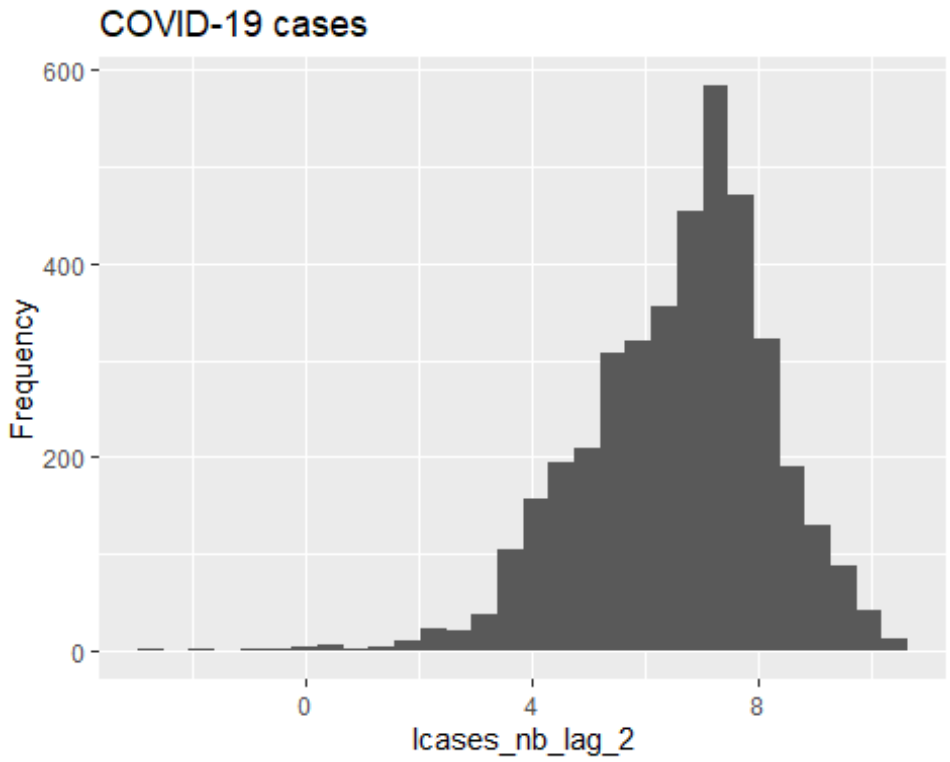



```
#Visualizing lcases_lag_2
ggplot(step4, aes(x = lcases_lag_2)) +
  geom_histogram() +
  labs(title = "COVID-19 cases",
       x = "lcases_lag_2",
       y = "Frequency")
```




```
#Visualizing lcases_nb_lag_2
ggplot(step4, aes(x = lcases_nb_lag_2)) +
```

```
  geom_histogram() +
  labs(title = "COVID-19 cases",
       x = "lcases_nb_lag_2",
       y = "Frequency")
```




```
#Creating y and x
y <- cbind(log(pdata$cases+1))
x <- cbind(log(pdata$cases_lag_2+1), log(pdata$cases_nb_lag_2+1), pdata$ci_lag_2, pdata$ci_
nb_lag_2)
summary(y)

##        V1
##  Min.   : 0.001
##  1st Qu.: 5.365
##  Median : 6.822
##  Mean   : 6.534
##  3rd Qu.: 7.868
##  Max.   :10.924

summary(x)

##        V1               V2                V3               V4
##  Min.   : 0.001   Min.   : 0.06766   Min.   : 0.001   Min.   : 5.58
##  1st Qu.: 5.376   1st Qu.: 5.58297   1st Qu.:48.693   1st Qu.:48.01
##  Median : 6.833   Median : 6.90230   Median :56.594   Median :55.84
##  Mean   : 6.535   Mean   : 6.62850   Mean   :56.439   Mean   :54.79
##  3rd Qu.: 7.860   3rd Qu.: 7.67288   3rd Qu.:64.400   3rd Qu.:63.69
##  Max.   :10.924   Max.   :10.50163   Max.   :90.000   Max.   :83.81

#Creating formula argument
fm <- y~x

#Creating spatial weights matrix based on contiguity
EU <- readOGR(dsn = ".", layer = "EUNOICELAND")

## OGR data source with driver: ESRI Shapefile
## Source: "C:\Users\ibrah\Documents", layer: "EUNOICELAND"
## with 38 features
```

```
## It has 177 fields
## Integer64 fields read as strings:  NE_ID CI_id CI_week

EU.nb=read.gal("EUGAL.gal", region.id=EU$BRK_A3)
eulist <- nb2listw(EU.nb)
summary(eulist)

## Characteristics of weights list object:
## Neighbour list object:
## Number of regions: 38
## Number of nonzero links: 146
## Percentage nonzero weights: 10.1108
## Average number of links: 3.842105
## Link number distribution:
##
##  1  2  3  4  5  6  7  8  9
##  4 10  5  7  4  1  4  2  1
## 4 least connected regions:
## DNK IRL GBR PR1 with 1 link
## 1 most connected region:
## DEU with 9 links
##
## Weights style: W
## Weights constants summary:
##    n   nn S0       S1       S2
## W 38 1444 38 25.48924 165.2069
```

Panel data models

```
#Pooling model
pooling <- plm(y~x, data = pdata, model = "pooling")
summary(pooling)

## Pooling Model
##
## Call:
## plm(formula = y ~ x, data = pdata, model = "pooling")
##
## Balanced Panel: n = 38, T = 109, N = 4142
##
## Residuals:
##      Min.   1st Qu.    Median   3rd Qu.      Max.
## -8.097897 -0.320251  0.095388  0.473884  6.648639
##
## Coefficients:
##               Estimate Std. Error t-value  Pr(>|t|)
## (Intercept)  1.8017614  0.0954685 18.8728 < 2.2e-16 ***
## x1           0.8963874  0.0151521 59.1594 < 2.2e-16 ***
## x2          -0.0317756  0.0177803 -1.7871 0.0739908 .
## x3          -0.0105924  0.0015476 -6.8446 8.799e-12 ***
## x4          -0.0057802  0.0016231 -3.5612 0.0003733 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Total Sum of Squares:    15537
## Residual Sum of Squares: 4548.1
## R-Squared:      0.70727
## Adj. R-Squared: 0.70699
## F-statistic: 2498.86 on 4 and 4137 DF, p-value: < 2.22e-16

#Fixed effects model
fixed <- plm(y~x, data = pdata, model = "within")
summary(fixed)
```

```
## Oneway (individual) effect Within Model
##
## Call:
## plm(formula = y ~ x, data = pdata, model = "within")
##
## Balanced Panel: n = 38, T = 109, N = 4142
##
## Residuals:
##      Min.   1st Qu.    Median   3rd Qu.      Max.
## -8.037351 -0.302965  0.098852  0.468844  6.162920
##
## Coefficients:
##      Estimate Std. Error t-value  Pr(>|t|)
## x1  0.8769909  0.0162512 53.9647 < 2.2e-16 ***
## x2  0.0127460  0.0191951  0.6640    0.5067
## x3 -0.0096971  0.0021157 -4.5834 4.710e-06 ***
## x4 -0.0214552  0.0027060 -7.9287 2.832e-15 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Total Sum of Squares:    14891
## Residual Sum of Squares: 4401.2
## R-Squared:      0.70443
## Adj. R-Squared: 0.70147
## F-statistic: 2442.85 on 4 and 4100 DF, p-value: < 2.22e-16

#Random effects model
random <- plm(y~x, data = pdata, model = "random")
summary(random)

## Oneway (individual) effect Random Effect Model
##    (Swamy-Arora's transformation)
##
## Call:
## plm(formula = y ~ x, data = pdata, model = "random")
##
## Balanced Panel: n = 38, T = 109, N = 4142
##
## Effects:
##                 var std.dev share
## idiosyncratic 1.073   1.036     1
## individual    0.000   0.000     0
## theta: 0
##
## Residuals:
##      Min.   1st Qu.    Median   3rd Qu.      Max.
## -8.097897 -0.320251  0.095388  0.473884  6.648639
##
## Coefficients:
##               Estimate Std. Error z-value  Pr(>|z|)
## (Intercept)  1.8017614  0.0954685 18.8728 < 2.2e-16 ***
## x1           0.8963874  0.0151521 59.1594 < 2.2e-16 ***
## x2          -0.0317756  0.0177803 -1.7871 0.0739177 .
## x3          -0.0105924  0.0015476 -6.8446  7.67e-12 ***
## x4          -0.0057802  0.0016231 -3.5612 0.0003692 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Total Sum of Squares:    15537
## Residual Sum of Squares: 4548.1
## R-Squared:      0.70727
## Adj. R-Squared: 0.70699
## Chisq: 9995.42 on 4 DF, p-value: < 2.22e-16
```

Tests

```
#F Test
pFtest(fixed, pooling)

##
##  F test for individual effects
##
## data:  y ~ x
## F = 3.6968, df1 = 37, df2 = 4100, p-value = 4.268e-13
## alternative hypothesis: significant effects

#LM Test
plmtest(pooling)

##
##  Lagrange Multiplier Test - (Honda) for balanced panels
##
## data:  y ~ x
## normal = 4.7508, p-value = 1.013e-06
## alternative hypothesis: significant effects

#Hausman Test
phtest(fm, pdata)

##
##  Hausman Test
##
## data:  fm
## chisq = 122.46, df = 4, p-value < 2.2e-16
## alternative hypothesis: one model is inconsistent

#Hausman Test robust to spatial autocorrelation
sphtest(fm,pdata,listw =eulist, spatial.model = "error", method="ML")

##
##  Hausman test for spatial models
##
## data:  x
## chisq = 29.815, df = 4, p-value = 5.338e-06
## alternative hypothesis: one model is inconsistent

#Hausman Test for spatial models
sphtest(fm,pdata,listw =eulist, spatial.model = "lag", method="ML")

##
##  Hausman test for spatial models
##
## data:  x
## chisq = 42.955, df = 4, p-value = 1.057e-08
## alternative hypothesis: one model is inconsistent

#LM test 1
slmtest(fm, pdata, listw = eulist, test = "lme", model="within")

##
##  LM test for spatial error dependence
##
## data:  formula (within transformation)
## LM = 2595.2, df = 1, p-value < 2.2e-16
## alternative hypothesis: spatial error dependence

#LM test 2
slmtest(fm, pdata, listw = eulist, test = "lml",model="within")
```

```
##
##  LM test for spatial lag dependence
##
## data:  formula (within transformation)
## LM = 1608.5, df = 1, p-value < 2.2e-16
## alternative hypothesis: spatial lag dependence

#LM test 3
slmtest(fm, pdata, listw = eulist, test = "rlme", model="within")

##
##  Locally robust LM test for spatial error dependence sub spatial lag
##
## data:  formula (within transformation)
## LM = 1002.1, df = 1, p-value < 2.2e-16
## alternative hypothesis: spatial error dependence

#LM test 4
slmtest(fm, pdata, listw = eulist, test = "rlml", model="within")

##
##  Locally robust LM test for spatial lag dependence sub spatial error
##
## data:  formula (within transformation)
## LM = 15.411, df = 1, p-value = 8.647e-05
## alternative hypothesis: spatial lag dependence
```

Spatial panel data models

```
spml(fm, pdata,
listw = eulist, lag=FALSE,model="pooling")

## Warning in if (class(covTheta) == "try-error") {: the condition has length > 1
## and only the first element will be used

##
## Call:
## spreml(formula = formula, data = data, index = index, w = listw2mat(listw),     w2 = lis
tw2mat(listw2), lag = lag, errors = errors, cl = cl)
##
## Coefficients:
## (Intercept)           x1           x2           x3           x4
##   1.3476212    0.8689291   -0.0410703   -0.0045199    0.0006763
##
## Error covariance parameters:
##    rho
## 0.6521

#Baltagi error
sem_b <- spml(fm, data = pdata,
listw = eulist, lag=FALSE,model="within", effect="individual", spatial.error="b")
summary(sem_b)

## Spatial panel fixed effects error model
##
##
## Call:
## spml(formula = fm, data = pdata, listw = eulist, model = "within",
##     effect = "individual", lag = FALSE, spatial.error = "b")
##
## Residuals:
##     Min.  1st Qu.   Median  3rd Qu.     Max.
## -7.88497 -0.30704  0.10713  0.48018  6.06617
##
## Spatial error parameter:
```

```
##     Estimate Std. Error t-value  Pr(>|t|)
## rho 0.653870   0.010924  59.855 < 2.2e-16 ***
##
## Coefficients:
##      Estimate Std. Error t-value  Pr(>|t|)
## x1  0.8431958  0.0114678 73.5272 < 2.2e-16 ***
## x2 -0.0151300  0.0148681 -1.0176    0.3089
## x3 -0.0083915  0.0013989 -5.9985 1.992e-09 ***
## x4 -0.0145063  0.0021239 -6.8300 8.490e-12 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
#KKP error
sem_b <- spml(fm, data = pdata,
listw = eulist, lag=FALSE,model="within", effect="individual", spatial.error="kkp")
summary(sem_b)
```

```
## Spatial panel fixed effects error model
##
##
## Call:
## spml(formula = fm, data = pdata, listw = eulist, model = "within",
##     effect = "individual", lag = FALSE, spatial.error = "kkp")
##
## Residuals:
##     Min.  1st Qu.   Median  3rd Qu.     Max.
## -7.88497 -0.30704  0.10713  0.48018  6.06617
##
## Spatial error parameter:
##     Estimate Std. Error t-value  Pr(>|t|)
## rho 0.653870   0.010924  59.855 < 2.2e-16 ***
##
## Coefficients:
##      Estimate Std. Error t-value  Pr(>|t|)
## x1  0.8431958  0.0114678 73.5272 < 2.2e-16 ***
## x2 -0.0151300  0.0148681 -1.0176    0.3089
## x3 -0.0083915  0.0013989 -5.9985 1.992e-09 ***
## x4 -0.0145063  0.0021239 -6.8300 8.490e-12 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
#GMM
sem_gm<- spgm(fm, data=pdata,
listw = eulist, model="within", moments="fullweights",
spatial.error = TRUE)
summary(sem_gm)
```

```
## Spatial fixed effects error model (GM estimation)
##
## Call:
## spgm(formula = fm, data = pdata, listw = eulist, model = "within",
##     spatial.error = TRUE, moments = "fullweights")
##
## Residuals:
##      Min.   1st Qu.    Median   3rd Qu.      Max.
## -7.870906 -0.296702  0.089798  0.461623  6.409945
##
## Estimated spatial coefficient, variance components and theta:
##           Estimate
## rho        0.24752
## sigma^2_v  1.03528
##
## Coefficients:
##      Estimate Std. Error t-value  Pr(>|t|)
## x1  0.8782547  0.0162055 54.1949 < 2.2e-16 ***
## x2 -0.0325944  0.0189711 -1.7181   0.08578 .
```

```
## x3 -0.0127190  0.0017362 -7.3257 2.377e-13 ***
## x4 -0.0085781  0.0018946 -4.5278 5.961e-06 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```