Received: 26 January 2021 | Revised: 24 October 2021 | Accepted: 25 October 2021
DOI: 10.1111/pirs.12647

FULL ARTICLE

Urbanization and COVID-19 incidence: A cross-country investigation

Rafael González-Val[1,2] | Fernando Sanz-Gracia[1]

[1]Universidad de Zaragoza, Zaragoza, Spain
[2]Institut d'Economia de Barcelona (IEB), Barcelona, Spain

Correspondence
Rafael González-Val, Universidad de Zaragoza, Departamento de Análisis Económico, Facultad de Economía y Empresa, Gran Vía 2, 50005, Zaragoza, Spain.
Email: rafaelg@unizar.es

Funding information
Ministerio de Ciencia e Innovación, Grant/Award Numbers: ECO2017-82246-P, PID2020-112773GB-I00, PID2020-114354RA-I00; Gobierno de Aragón, Grant/Award Number: ADETRE research group; European Regional Development Fund

Abstract
This paper investigates the determinants of the diffusion and intensity of the COVID-19 at the country level, focusing on the role played by urban agglomeration, measured using three urban variables: percentage of the urban population, population density, and primacy. We estimate the influence of urban agglomeration on two outcome variables: cumulative number of cases and deaths per 100,000 inhabitants up to 31 December 2020, using both parametric and semiparametric models. We also explore possible spatial effects. The non-linear effects of the urban variables on the intensity of the disease reveal non-monotonous relationships, suggesting that it is the size of the urban system that is linked to a stronger incidence.

KEYWORDS
COVID-19 incidence, parametric and semiparametric methods, urban variables

JEL CLASSIFICATION
C14; I12; R19

1 | INTRODUCTION

The disease known as COVID-19 appeared for the first time in the Chinese city of Wuhan in November 2019, spreading worldwide and being declared a universal pandemic by the World Health Organization on 11 March 2020.

In April 2020, half of the world's population was under some kind of lockdown. In addition to the very serious health consequences of this shock, the economic impact in most countries has also been catastrophic.

Throughout 2020, interesting literature about the economic effects of COVID-19 has emerged. Baker et al. (2020) define three uncertainty indices to estimate their impact on the real GDP of the United States. Two papers that were written in February and March 2020 (Maital & Barzani, 2020; McKibbin & Fernando, 2021) predict, under several scenarios, the worldwide macroeconomic effects of the shock, concluding that a global recession is coming. Topcu and Gulal (2020) analysed the impact of COVID-19 on emerging stock markets. The contraction of public fiscal revenues (income tax) in the US is quantified in Clemens and Veuger (2020). Other works have adopted a microeconomic approach. In this line, Martin et al. (2020) and Sumner et al. (2020) have studied the consequences of the virus on household income, consumption, and poverty. Following with microeconomic visions, Elenev et al. (2020) modelled the consequences of the pandemic as a drop in workers' productivity with the associated fall in firms' revenues; in addition, Bartik et al. (2020) estimated the effects of the shock on small business outcomes and expectations. Of course, the consequences of the COVID-19 have been analysed from many other points of view, beyond economic; see, to cite just a couple of works, Bashir et al. (2020) for the environmental impact and Rodríguez-Pose and Burlina (2021), who study which factors determine the geography of COVID-19-related excess mortality in Europe. However, the study of the economic effects of contagious diseases did not begin with the irruption of COVID-19. See, for example, the medieval age and the Black Death, Voigtländer and Voth (2013) and Jedwab et al. (2019), and Alfani and Percoco (2019) for the impact of the pandemic during 1629–1630 in Italy.

The objective of this paper is essentially different and complementary to the literature described in the previous paragraph. Instead of analysing the impact of the pandemic on a range of variables, we focus on how different socioeconomic and urban variables affect the incidence of the virus. Our main explanatory variables are urban primacy (a measure of the degree of concentration of the population in the largest city of a country), the percentage of the urban population, and population density. These three variables represent different measures of urban and population agglomeration. To our knowledge, only a few studies have adopted a similar approach, all of them using US data. Wheaton and Thompson (2020) regress the *per capita* contagion rate over GDP *per capita*, density, and percentage of industrial and commercial use for 351 nuclei of Massachusetts; Carozzi et al. (2020) estimate the link between population density and COVID-19 spread and severity considering US counties, and Geng et al. (2021) analyse the sparseness of COVID-19 infections and their time variations across the US at scales ranging from county to continental scale. Nevertheless, the spatial scale of these studies addressing the patterns of the spread of the disease within one unique country limits the scope of their conclusions compared to our cross-country analysis.

Why do we expect any relationship between urbanization or population agglomeration and COVID-19 intensity? From an epidemiological point of view, that concentration of people and, in general, the existence of large and dense urban areas is one of the key factors that facilitate the transmission of viruses (Alirol et al., 2011; Neiderud, 2015). This is the main reason for the worldwide implementation of lockdown practices. Therefore, the main working hypothesis we want to test is whether a direct and positive relationship exists between the three urban variables and the incidence of COVID-19 at the country level, considering both cases and deaths. To do so, we use a wide set of control variables that might also influence the intensity of the virus incidence. In this context, we can consider the contagion risk as a new urban congestion cost, apart from the traditional costs (e.g., commuting, pollution, traffic density, housing price), associated with the existence of urban agglomerations.

An important contribution of this work lies in the fact that we combine parametric and semiparametric specifications. The advantage of doing so is double. On the one hand, both approaches are complementary and give robustness to our results. On the other, the parametric specification assumes that the average relationship between COVID-19 incidence and the urban variables is linear, while the semiparametric approach is more flexible and specifically enables us to estimate functional forms that do not have to be necessarily linear. We expect, as the empirical exercise we carry out confirms, the relationship between the virus intensity and the urban variables to be not constant over the distribution of their values, showing concavities and convexities. This is for two main reasons: urban and new economic geography models show that congestion costs usually behave non-linearly. Second, infection curves across countries have shown that the virus spread is not linear.

The main results are the following. From the parametric approach, we conclude that the main explanatory variable is the percentage of the urban population, which exerts a positive influence on COVID-19 incidence. Moreover, we also test for possible spatial effects, finding weak evidence supporting spatial interactions between countries. From the semiparametric approach, we cannot find a systematic and monotonous positive relationship between the three urban variables and the intensity of the disease, as might be expected *a priori*. The link is positive only for intermediate and high values of the percentage of urban population, for low and very high values of density, and for very high values of the primacy.

The remainder of the paper is organized as follows. Section 2 presents the data used. In Section 3, we describe the methodology. Section 4 shows the main results, and Section 5 concludes.

2 | DATA

Our sample comprises 90 countries across five continents (see the list in the Appendix), and all variables are measured at the country level. The descriptive statistics and the sources of the variables used are described in Table 1. We consider two different measures of the COVID-19 incidence, updated up to 31 December 2020: the cumulative number of cases per 100,000 inhabitants and the cumulative number of deaths per 100,000 inhabitants. To make an easier interpretation, we normalize both of these, subtracting the mean and dividing by the standard deviation for our sample of countries. Hence, the mean is zero, and the standard deviation is equal to one.[1]

We are aware of some potential reporting bias in these statistics. Cases reported are the number of detected cases. In contrast, the true number of infected people might be larger because the detected cases depend crucially on the number of pharmaceutical tests carried out.[2] In the same way, mortality statistics are lower than the true number of deaths by COVID-19 in many countries, as the excess of mortality statistics are revealing. Some estimates of the real mortality are being released for some countries based on the excess of mortality, but, unfortunately, for most countries, the only reliable data source is the World Health Organization (WHO) data. Official statistics for cases and deaths probably underestimate the true incidence of the virus, so we can consider the official statistics by the WHO as a lower bound of the real effect of the COVID-19; if we can find any significant pattern (even with these data), the true underlying relationships are likely to be even stronger.

Our main explanatory variables have an urban dimension: percentage of urban population (over the total country population), population density (population per square kilometre), and urban primacy, defined as the population of the largest city of a country divided by its total population. The other explanatory variables can be grouped into three main categories. First, variables related to the level of development at the national level: a democracy index ranging from 0 for an absolute dictatorship to 10 for a complete democracy; the Human Development Index (HDI) by the United Nations; and GDP *per capita* (Wildman, 2021). Second, we include variables capturing the influence of globalization (Farzanegan et al., 2021), openness, and geography: trade (exports plus imports divided by GDP), the number of total air passengers divided by total population, and the geographical distance from the capital of each country to Wuhan in China (place of origin of the virus) or, in the case of European countries, distance to Milan (the main source of the virus in Europe in March 2020). The inclusion of this distance variable is debatable. We exclude distance to Wuhan/Milan in an alternative specification and use spatial models considering spatial matrices with bilateral distances between all countries.

Finally, we add three variables somehow directly related to COVID-19: the share of health expenditure over total GDP; the percentage (always negative) of change in the mobility to the workplace provided by Google,[3] and the COVID stringency index from the University of Oxford, which measures the severity of the control policies implemented by the governments ranging from a value of zero for January 2020 to 1 (the strictest control measures).[4]

TABLE 1 Descriptive statistics

Variable	Mean	Std. Dev.	Minimum	Maximum	Date	Source
Total population (thousands)	58,099.29	152,249.40	1,325.65	1,366,418	1 July 2019	United Nations
Cumulative cases	854,871.50	2,428,258	366	19,346,790	31 December 2020	World Health Organization
Cumulative deaths	18,730.68	45,978.42	0	335,789	31 December 2020	World Health Organization
Cumulative cases per 100,000 (normalized)	0.00	1.00	−1.14	2.61	31 December 2020	Own calculations
Cumulative deaths per 100,000 (normalized)	0.00	1.00	−1.01	3.18	31 December 2020	Own calculations
Primacy (%)	0.17	0.09	0.02	0.39	2020	United Nations
Urban population (%)	0.65	0.20	0.18	0.99	2020	United Nations
Population density	170.13	275.71	3.28	2,159.43	1 July 2019	United Nations
Democracy index	6.13	2.06	1.93	9.87	2019	The Economist
HDI	0.77	0.13	0.50	0.95	2019	United Nations
GDPpc	18,917.62	21,803.78	714.48	92,556.32	2019	The World Bank
Trade (% of GDP)	0.83	0.43	0.27	2.39	2019	The World Bank
Air passengers *per capita*	18.81	64.87	0.02	521.24	2018	International Air Transport Association
Distance to Wuhan/Milan	6,232.76	5,782.30	191.65	19,528.89	-	Own calculations
Health expenditure (% of GDP)	0.04	0.03	0.00	0.14	2017	World Health Organization
% Mobility change (workplaces)	−0.75	0.12	−0.91	−0.40	2020	Google
Covid stringency index	0.84	0.13	0.19	1.00	2020	University of Oxford

Notes: Data for 90 countries.

We want to emphasize the advantages of using country data over within-country data. Institutions, laws, health system, level of development, geography, and other characteristics change more intensively across countries than within countries. Therefore, the potential explanatory power of our variables is larger and can capture more properly the differences between countries with regards to the incidence of the COVID-19. For within-country studies, as all these variables are quite similar across sub-national units, the main (and only, in most cases) source of cross-sectional variation comes from the different non-pharmaceutical interventions adopted by the local and regional governments, limiting the scope of these analyses to the effectiveness of these lockdown measures.

3 | METHODOLOGY

A first way to test the relationships between our urban variables and the incidence of COVID-19 is to run a simple parametric OLS regression for our pool of 90 countries. We estimate the following parametric model:

$$\begin{aligned} Y_i = \alpha + \beta_1 URBAN\_POPULATION_i + \beta_2 DENSITY_i + \beta_3 PRIMACY_i + \\ + \sum_{i=1}^{m} \gamma_m X_{mi} + \delta_k + \varepsilon_i, \end{aligned} \tag{1}$$

where i indexes countries, Y_i is the measure of the incidence of COVID-19 (normalized cumulative cases per 100,000 and cumulative deaths per 100,000), our main variables of interest are $URBAN\_POPULATION_i$, $DENSITY_i$ and $PRIMACY_i$,[5] X_{mi} represents the rest control variables at the country level, δ_k is a vector of region dummies for the five major area categories (America, Africa, Asia, Eastern Europe, and Western Europe), and ε_i is the error term. The main coefficients of interest are the β_j, representing the effect on COVID-19 incidence of each one of our urban variables.

The full model includes all the control variables, but we also consider reduced versions of the model, including only one of the groups of the control variables. Furthermore, we also estimate a spatial Durbin model (Anselin, 1988; LeSage & Pace, 2009) with the aim of explicitly considering the impact of neighbouring countries on COVID-19 incidence.[6] The convenience of explicitly considering the spatial dimension in the analysis of the cross-country incidence of COVID-19 has been emphasized in Krisztin et al. (2020); spatial models have also been used to analyse the spatial diffusion of the disease within countries, see the studies by Paez et al. (2021) and Ehlert (2021) of the Spanish and German cases, respectively.

The spatial Durbin model includes three elements (LeSage & Pace, 2009): a spatial lagged dependent variable, a set of explanatory variables, and a set of spatial lagged explanatory variables. It can be expressed as:

$$\begin{aligned} Y_i = \alpha + \rho \sum_{j=1}^{n} W_{ij} Y_i + \beta_1 URBAN\_POPULATION_i + \beta_2 DENSITY_i + \beta_3 PRIMACY_i + \\ + \varphi_1 \sum_{j=1}^{n} W_{ij} URBAN\_POPULATION_i + \varphi_2 \sum_{j=1}^{n} W_{ij} DENSITY_i + \varphi_3 \sum_{j=1}^{n} W_{ij} PRIMACY_i + \\ + \sum_{i=1}^{m} \gamma_m X_{mi} + \sum_{i=1}^{m} \sum_{j=1}^{n} W_{ij} \varphi_m X_{mi} + \delta_k + \varepsilon_i, \end{aligned} \tag{1'}$$

with $|\rho| < 1$ being the spatial autoregressive parameter measuring the effect on the response variable of COVID-19 incidence in neighbouring countries (endogenous interaction relationships) and φ_i capturing the effects of the spatial lagged explanatory variables (exogenous interaction relationships). W_{ij} are the elements of the W spatial weight matrix, built using the k-nearest neighboring countries; following LeSage and Pace's (2009) suggestion, we compared the log-likelihood values of models with different weights matrices to set k to 5.[7] Distances between countries are calculated from the co-ordinates (longitude and latitude) of the country's capital city, obtained from the United Nations. In the spatial model the variable distance to Wuhan/Milan is excluded.

Besides the possible spatial effects, the functional relationship of the variables in Equation 1 can also be a problem. A parametric specification implies strong assumptions; it imposes a particular structure on the underlying relationship between the variables (that may not reflect the true underlying relationship), and the coefficients are not allowed to change over countries (the relationship is restricted to being stationary over the entire structure of the distribution of the urban variable).

To overcome these limitations, Durlauf (2001) suggested the use of semiparametric methods. This approach allows us to tackle the possible non-linear effect of the different urban variables on COVID-19 incidence in a more flexible way. For instance, the standard correlation index and the coefficients from parametric regressions give us only an aggregate average relationship between variables, and this relationship is restricted by the fact that it must remain unchanged through the entire distribution of the urban variables. In contrast, the semiparametric estimate allows Y_i to vary with these variables over the entire distribution, allowing for the linear effects of other conditioning variables. In related literature, Barrios and Strobl (2009), Lessmann (2014), and Díez-Minguela et al. (2020) have applied this methodology to the study of regional inequalities.

We perform a semiparametric analysis using the kernel regression estimator (Robinson, 1988). This consists of taking the following specification:

$$Y = \alpha + f(URBAN\_VARIABLE) + \gamma X + \varepsilon, \tag{2}$$

in which, for the sake of clarity, we drop the subscript *i*. X is a set of explanatory variables that are assumed to have a linear effect on Y, $f(.)$ is a smooth and continuous, possibly non-linear, unknown function of the corresponding *URBAN VARIABLE* (percentage of urban population, population density and primacy), and ε is a random error term. Thus, the model has a parametric (γX) and a non-parametric $f(URBAN\_VARIABLE)$ part. Only one non-parametric variable is allowed at a time, so we run a semi-parametric regression for each of the urban variables. In each case, the other two urban variables are included in the linear part of the model. Robinson's approach is a two-step methodology. First, $\widehat{\gamma}$ is estimated by applying a procedure similar to that whereby variables can be partialled out of an OLS regression (but using nonparametric regressions). Second, a kernel regression of $Y - \widehat{\gamma}X$ on each *URBAN_VARIABLE* is carried out. In both stages, a Gaussian kernel-weighted local polynomial fit is used for kernel regressions.

4 | RESULTS

4.1 | Parametric analysis

Table 2 shows the results for the OLS estimations of the parametric model (Equation 1) when the endogenous variable is the normalized relative number of cases. Each table column corresponds to a different set of explanatory variables grouped by category (see Section 2). Column (5) shows the results for the full specification of the model, including all the regressors and a set of regional dummies. The mean value of the VIF statistic shows tolerable values for all the models (including the full specification), which allows us to conclude that multicollinearity is not a problem.

The coefficient of the percentage of urban population is systematically significant and has the expected sign: more urbanized countries show a higher incidence of the disease. Also, population density exerts a certain positive influence on the incidence of COVID-19, although not so intense as that of the percentage of urban population. On the other hand, the coefficient of primacy is not significant in any case and even changes its sign. Therefore, it seems that the size of the urban system (percentage of urban population), and not the preeminence of the largest city (primacy), is the key urban variable that explains the incidence of the virus on the number of cases. From our country-level perspective, these results indicate that having a large city does not imply a higher incidence of the virus.

TABLE 2 Parametric estimates: Cases per 100,000 population

	(1)	(2)	(3)	(4)	(5)
Urban population (%)	2.926*** (0.477)	1.292** (0.611)	2.489*** (0.528)	1.479** (0.583)	1.325** (0.621)
Ln (Population density)	0.107 (0.083)	0.149* (0.087)	0.023 (0.083)	0.142 (0.093)	0.196* (0.100)
Primacy	−1.649 (1.277)	−1.114 (1.205)	−0.802 (1.477)	−0.269 (1.305)	0.286 (1.181)
Democracy index		−0.042 (0.050)			−0.135** (0.054)
HDI		5.108** (1.958)			−1.894 (2.470)
Ln (GDPpc)		−0.113 (0.204)			0.278 (0.234)
Trade (% of GDP)			0.338 (0.274)		0.129 (0.224)
Air passengers *per capita*			0.000 (0.001)		0.001 (0.001)
Ln (distance to Wuhan/ Milan)			−0.269*** (0.094)		−0.123 (0.236)
Health expenditure (% of GDP)				12.855** (5.076)	14.851** (6.862)
% Mobility change (workplaces)				−1.334 (0.951)	−0.626 (0.916)
COVID stringency index				0.862 (0.705)	1.401** (0.689)
Regional dummies	N	N	N	N	Y
Observations	90	90	90	90	90
Degrees of freedom	86	83	83	83	73
Adjusted R^2	0.247	0.341	0.379	0.363	0.556
Mean VIF	1.31	5.39	1.44	1.70	5.76

Notes: Dependent variable: normalized number of cumulative cases per 100,000 population. All regressions include a constant. Coefficient (robust standard errors). Regional dummies are dummies at five major area categories: America, Africa, Asia, Eastern Europe, and Western Europe (Asia is the base category). Significant at the *10%, **5%, ***1% level.

Although the main city could be severely hit at the beginning of the pandemic, ultimately, the intensity of the disease spread in the whole country depends on the urbanization level rather than in the population agglomeration in only one particular city.

Regarding the rest of the control variables, focusing on the last column, when all the regressors and geographical dummies are considered, only three controls significantly affect on the relative number of cases. The coefficient of the democracy index variable is significant and has the expected negative sign: democratic countries reported a significantly lower number of relative cases. The effect of the COVID stringency index is significant and positive, indicating that the larger the number of cases, the more the health and political authorities are forced to implement restrictive measures. Finally, the percentage of health expenditure is positively correlated with the dependent variable. An analysis of the data can shed some light on this result: among the top 15 countries ordered by the relative number of cases, you can find countries such as Belgium, Czechia, the US, Switzerland, Spain, and France; the last five (out of 90) are New Zealand, Thailand, Cambodia, Vietnam, and Tanzania.[8] As the number of detected cases depends directly on the number of pharmaceutical tests carried out, even if the relative number of infected people were the same in all countries, those countries that perform more tests would detect more cases, and the country's capacity to perform tests depends on the health expenditure. This explains why countries with high health expenditure rank highly for relative cases, while many underdeveloped countries with low health expenditure rank low.

The lack of statistical significance of the other control variables might indicate the wide worldwide diffusion of the virus across countries, despite the important heterogeneity across countries in those variables. Although these characteristics may have affected the impact of the virus, our results indicate that, ultimately, they do not have a significant effect on the cumulative number of cases after almost 1 year of pandemic, confirming that the diffusion of the virus is space-filling (Geng et al., 2021).

Table 3 reports the results of the parametric model for the normalized relative number of deaths. The VIF indicates that, again, there are no problems of multicollinearity. The coefficient of the percentage of urban population is significant in all cases and shows the expected sign: more urbanized countries also suffer a larger relative number of deaths. Regarding the other two urban variables, this time, population density is no longer significant, and urban

TABLE 3 Parametric estimates: deaths per 100,000 population

	(1)	(2)	(3)	(4)	(5)
Urban population (%)	2.784*** (0.548)	2.278*** (0.733)	2.434*** (0.534)	1.361** (0.620)	2.421*** (0.789)
Ln (Population density)	−0.038 (0.087)	−0.007 (0.091)	−0.095 (0.086)	−0.006 (0.084)	−0.055 (0.075)
Primacy	−3.487*** (1.219)	−3.106** (1.205)	−2.462* (1.320)	−2.089* (1.184)	−1.182 (1.062)
Democracy index		0.100** (0.044)			−0.038 (0.048)
HDI		5.351*** (1.678)			1.889 (2.289)
Ln (GDPpc)		−0.461** (0.197)			−0.295 (0.195)
Trade (% of GDP)			−0.019 (0.243)		−0.136 (0.237)
Air passengers *per capita*			0.001 (0.001)		0.001 (0.001)
Ln (distance to Wuhan/ Milan)			−0.254** (0.098)		−0.651*** (0.190)
Health expenditure (% of GDP)				13.133** (5.477)	6.534 (5.469)
% Mobility change (workplaces)				−1.448 (0.924)	0.244 (0.987)
COVID stringency index				1.371** (0.669)	1.805*** (0.587)
Regional dummies	N	N	N	N	Y
Observations	90	90	90	90	90
Degrees of freedom	86	83	83	83	73
Adjusted R^2	0.207	0.272	0.255	0.352	0.539
Mean VIF	1.31	5.39	1.44	1.70	5.76

Notes: Dependent variable: normalized number of cumulative deaths per 100,000 population. All regressions include a constant. Coefficient (robust standard errors). Regional dummies are dummies at five major area categories: America, Africa, Asia, Eastern Europe, and Western Europe (Asia is the base category). Significant at the *10%, **5%, ***1% level.

primacy exerts a negative influence on the relative number of deaths. The interpretation of these results seems straightforward: the existence of a very big city or a very large agglomeration not only does not increase the number of deaths but reduces it, and it is the complete urban structure of the country (including cities of all sizes: large, intermediate and small) that most contributes to the magnitude of the dependent variable in Table 3.

Besides the percentage of the urban population, in the last column, when all controls are added, another two variables help explain the relative number of deaths: the distance to Wuhan/Milan and the COVID stringency index. Both show the expected sign. The coefficient of the COVID stringency index is significant and positive, indicating that the larger the number of deaths, the more the health and political authorities are forced to implement restrictive measures, and the negative effect of distance suggests that the geographically farther away a country is from one of the two main sources of infection, the lower the number of deaths.

Table 4 shows the results of the spatial Durbin model (Eq. 1′) for the two dependent variables, estimated by maximum likelihood. For both dependent variables the spatial lag parameter (ρ) is not significant, indicating that the endogenous interaction relationships are not relevant either for the relative number of deaths or for the relative number of cases.

Therefore, the evidence supporting spatial effects is weaker than that found in previous research. Krisztin et al. (2020) found notable spatial spillover mechanisms in the early stages of the virus, considering annual country data from 23 January to 28 March 2020, for a set of 99 countries. However, here we are considering the cumulative incidence for almost 1 year (up to 31 December 2020), instead of focusing only on the initial outbreak of the disease. After all the international restrictions to people movements between countries and all the internal non-pharmaceutical interventions, we expect that spatial interactions between and within countries would have decreased. Krisztin et al. (2020) also observed a sharp drop in the intensity of spatial spillovers in later stages of the spread of the disease due to national travel bans, indicating that travel restrictions led to a reduction of cross-country spillovers.

Nevertheless, we find that some spatial lagged explanatory variables are significant in some cases, which confirms some kind of exogenous interaction relationships. However, the coefficient lag estimates reported in Table 4 must be interpreted with caution, because in spatially lagged models the estimated coefficient of an independent variable does not directly reflect its marginal effect on the dependent variable (Golgher & Voss, 2016). LeSage and Pace (2009) pointed out that using the point estimation method of the spatial regression model to test the spatial spillover effect leads to bias because the coefficient estimate of the explanatory variable does not represent the true partial regression coefficient and suggested that the direct, indirect, and total effects of a change in an independent variable should be calculated. Then, according to Elhorst (2014), the effects have the following interpretation: if an explanatory variable in a particular country changes, not only the relative number of cases and/or deaths in that country changes, but also relative number of cases and/or deaths in other countries change via the spatial spillover. Therefore, a change in one of the independent variables in a particular country has a direct effect on that country, but also an indirect effect on neighbouring countries, and the sum of the direct and indirect effects is the total effect of a change in that explanatory variable.

Table 4 reports the direct, indirect and total effects for the spatial Durbin model, estimated using equation 2.46 in LeSage and Pace (2009). The main conclusion from this Table is that spatial effects are limited, because just a few variables show a significant effect on the relative number of cases or deaths. Paez et al. (2021) obtained a similar result for the Spanish case, and their explanation was that under a situation of lockdown, inter-regional contagion is reduced. In our context, the same explanation applies, because international restrictions to people movements between countries were stronger than within countries.

For the relative number of cases, only the democracy and the stringency indexes have a significant total effect. Interestingly, population density has no significant total effect because direct and indirect effects have opposite signs and cancel out each other. Regarding the relative number of deaths, only the stringency index has a significant and positive direct effect, whereas the development index and the number of air passengers show a significant total effect driven by the indirect effect. What these results reveal compared to the OLS estimations shown previously is

TABLE 4 Coefficient estimates of the spatial Durbin model and decomposition estimates of the direct and indirect effects

	Cases per 100,000 population					Deaths per 100,000 population				
Dependent variable:	**Coefficient estimate**	**Coefficient lag estimate**	**Direct effect**	**Indirect effect**	**Total Effect**	**Coefficient Estimate**	**Coefficient Lag Estimate**	**Direct Effect**	**Indirect Effect**	**Total Effect**
Urban population (%)	−0.051 (0.633)	0.512 (1.557)	−0.071 (0.723)	0.421 (1.409)	0.349 (1.612)	0.426 (0.667)	−4.772*** (1.660)	0.417 (0.695)	−4.818** (2.186)	−4.400* (2.476)
Ln (Population density)	0.199*** (0.060)	−0.529*** (0.195)	0.223*** (0.059)	−0.472*** (0.173)	−0.250 (0.173)	−0.001 (0.063)	−0.118 (0.206)	−0.001 (0.061)	−0.119 (0.237)	−0.120 (0.248)
Primacy	−1.027 (0.861)	0.708 (2.537)	−1.068 (0.947)	0.827 (2.021)	−0.242 (2.398)	−1.693* (0.908)	2.892 (2.712)	−1.688 (1.036)	2.902 (3.012)	1.214 (3.519)
Democracy index	−0.019 (0.057)	−0.325*** (0.121)	−0.007 (0.060)	−0.254*** (0.096)	−0.261*** (0.097)	0.045 (0.059)	−0.101 (0.122)	0.045 (0.057)	−0.101 (0.132)	−0.057 (0.137)
HDI	−2.754 (2.228)	−0.822 (7.920)	−2.757 (2.369)	0.490 (6.506)	−2.708 (6.505)	3.545 (2.348)	17.112** (8.415)	3.576 (2.409)	17.337** (8.627)	20.913** (9.331)
Ln (GDPpc)	0.415** (0.191)	0.315 (0.480)	0.408** (0.204)	0.145 (0.427)	0.553 (0.430)	−0.175 (0.202)	−0.449 (0.513)	−0.176 (0.200)	−0.455 (0.559)	−0.631 (0.596)
Trade (% of GDP)	0.112 (0.200)	0.268 (0.634)	0.103 (0.195)	0.185 (0.532)	0.288 (0.624)	−0.260 (0.212)	−0.904 (0.670)	−0.262 (0.218)	−0.917 (0.770)	−1.179 (0.902)
Air passengers *per capita*	0.001 (0.001)	−0.001 (0.003)	0.001 (0.001)	−0.001 (0.002)	0.001 (0.003)	−0.001 (0.001)	−0.010*** (0.003)	−0.001 (0.001)	−0.010*** (0.004)	−0.011*** (0.004)
Health expenditure (% of GDP)	13.730*** (4.428)	0.507 (18.394)	13.885*** (4.715)	−3.105 (13.604)	10.780 (13.998)	3.675 (4.667)	−15.403 (19.150)	3.648 (5.006)	−15.520 (18.490)	−11.872 (20.280)
% Mobility change (workplaces)	−1.006 (0.715)	−1.261 (2.444)	−0.969 (0.734)	−0.748 (2.086)	−1.717 (2.336)	1.170 (0.753)	1.920 (2.574)	1.173 (0.725)	1.955 (2.924)	3.128 (3.168)
COVID stringency index	0.636 (0.531)	4.298*** (1.214)	0.473 (0.509)	3.262*** (0.929)	3.736*** (1.025)	1.317** (0.569)	5.178*** (1.395)	1.327** (0.593)	5.249*** (1.304)	6.576*** (1.515)
ρ	−0.321 (0.531)					0.012 (0.165)				
Regional dummies	Y					Y				
Observations	90					90				
LR test (p-value)	0.138					0.946				
Wald test (p-value)	0.091					0.941				

Notes: Dependent variables: normalized number of cumulative cases and deaths per 100,000 population. Direct and indirect effects computed from the Spatial Durbin Model, using equation 2.46 in LeSage and Pace (2009). Significant at the *10%, **5%, ***1% level.

that population density and the democracy index have a significant negative indirect effect on the relative number of cases, while urban population and the number of air passengers have a significant and negative indirect effect on the relative number of deaths; the development index show a large positive and significant indirect effect on relative deaths. The positive and significant effects found for the stringency index on both dependent variables are consistent with the OLS results.

To sum up, if we focus on the results of the full models, including all the regressors (column (5) in Tables 2 and 3), we find two variables always significant, and that show the expected sign: the COVID stringency index and, among our three urban variables, the percentage of the urban population of the country. The results of the estimated spatial Durbin model confirm the effect of the stringency index, whereas the influence of the urban population variable is found weaker than in the OLS results. However, as explained above, a parametric growth regression is not the best way to address non-linear relationships because of the strong assumptions implied. Therefore, we next estimate semiparametric models, allowing the effect of the urban variables to vary across their distributions and including the rest control variables in a linear way.

As explained above, congestion costs of urban agglomeration usually behave in a non-linear manner. Furthermore, besides econometric and flexibility reasons, the consideration of non-linear influences between the variables is also recommendable from an epidemiological perspective. In the words of Hu et al. (2013, Abstract): "Empirical data of human and wildlife diseases indicate that a non-linear function may work better when looking at the full spectrum of densities".

4.2 | Semiparametric analysis

Table 5 reports the results of the linear part of the model for each COVID-19 outcome variable. We only show results for the specification including all control variables. Results are consistent with those reported in column (5) in Tables 2 and 3, although there are slight differences. The first three numerical columns show the results for the relative number of cases and, respectively, including each of our urban variables in a nonparametric way. The explanatory variables that appear significant at least in two of the three columns are the percentage of urban population, the democracy index, the health expenditure as a percentage of GDP, the percentage of mobility change, and the COVID stringency index. All of these, except the percentage of mobility change, are also systematically significant in the last two columns of Table 2. The signs are as expected, except for the health expenditure variable, something explained above.

Regarding the relative cumulative deaths, the last three columns in Table 5, the variables with more explicative power (see also Table 3) are the percentage of urban population (this is, perhaps, the more pervasive outcome of this paper), the distance to Wuhan/Milan and the COVID stringency index, all of them with the expected sign.

Table 5 also reports the Hardle and Mammen's (1993) specification test; the null hypothesis is that parametric and non-parametric fits are not different from a statistical perspective. In the case of the relative number of cases, this null hypothesis cannot be rejected, thus rejecting any significant non-linear relationship for two out of three urban variables: percentage of urban population and primacy. For the relative number of deaths, we reject the null of the test at the 5% significance level for density and primacy, confirming the non-linear relationship between the relative number of deaths and these two urban variables. To sum up, in half of the six possible cases the non-linear approach is more suitable.

Figure 1 shows the non-parametric part of the estimation, including the 95% confidence bands. These graphs show the non-linear relationship between our three urban variables (percentage of urban population, density, and primacy), separately considered, and each of the two dependent variables used to measure the COVID-19 incidence: relative number of cases and the relative number of deaths. Moreover, as the two incidence rates are normalized, the value zero indicates the mean value across our sample of countries, and values significantly different from zero point to a higher or lower than average incidence. A first and important outcome is that the six figures show relevant

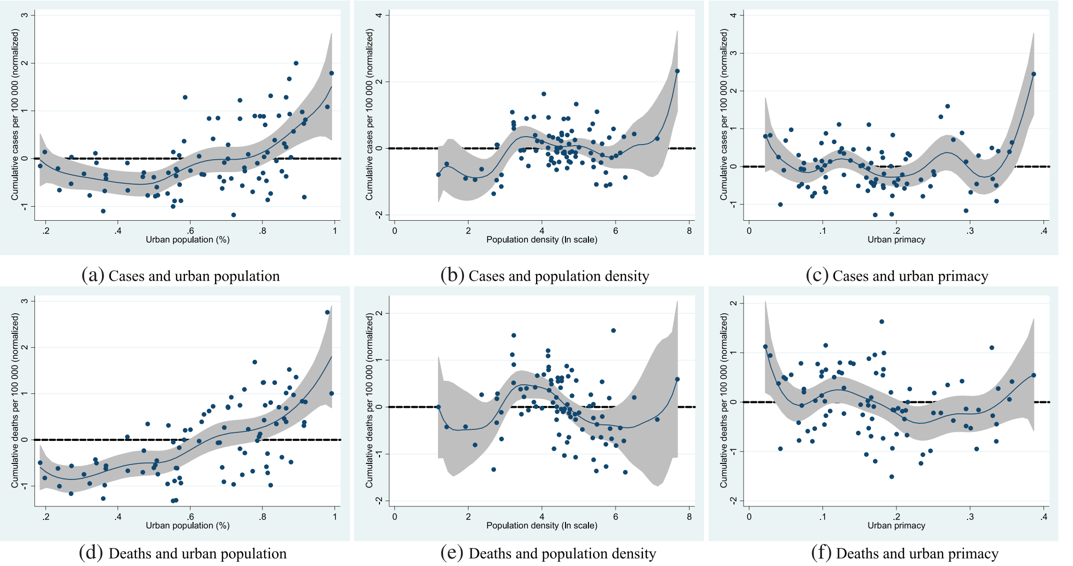
TABLE 5 Semi-parametric estimated, linear part of the model

	Cumulative cases per 100,000			Cumulative deaths per 100,000		
Dependent variable:	**(1)**	**(2)**	**(3)**	**(4)**	**(5)**	**(6)**
Urban population (%)	Nonpar. variable	1.721*** (0.584)	2.094*** (0.628)	Nonpar. variable	2.654*** (0.834)	2.767*** (0.810)
Ln (Population density)	0.157 (0.102)	Nonpar. variable	0.008 (0.085)	−0.094 (0.083)	Nonpar. variable	−0.107 (0.085)
Primacy	0.401 (1.153)	−0.808 (1.190)	Nonpar. variable	−1.136 (1.058)	−2.291** (0.980)	Nonpar. variable
Democracy index	−0.106** (0.048)	−0.146** (0.059)	−0.134** (0.055)	−0.010 (0.048)	−0.055 (0.053)	−0.052 (0.047)
HDI	−2.578 (2.215)	−2.488 (2.206)	−3.792* (2.146)	1.179 (2.364)	1.306 (2.129)	0.634 (2.066)
Ln (GDPpc)	0.299 (0.200)	0.289 (0.241)	0.259 (0.242)	−0.260 (0.235)	−0.189 (0.196)	−0.241 (0.188)
Trade (% of GDP)	0.052 (0.223)	0.028 (0.192)	0.090 (0.199)	−0.188 (0.210)	−0.255 (0.225)	−0.077 (0.235)
Air passengers *per capita*	0.000 (0.001)	0.001 (0.001)	−0.000 (0.001)	0.000 (0.001)	0.001 (0.001)	−0.001 (0.002)
Ln (distance to Wuhan/ Milan)	−0.276 (0.246)	−0.284 (0.211)	−0.453** (0.207)	−0.702*** (0.192)	−0.743*** (0.185)	−0.712*** (0.205)
Health expenditure (% of GDP)	12.496 (7.884)	13.207** (5.847)	15.007** (7.237)	5.847 (5.966)	4.687 (4.871)	7.055 (5.883)
% Mobility change (workplaces)	−0.164 (0.846)	−1.269* (0.714)	−1.549* (0.785)	0.805 (0.974)	−0.069 (0.931)	−0.257 (0.943)
Covid stringency index	1.562** (0.705)	1.821*** (0.604)	1.461** (0.663)	1.897*** (0.548)	2.196*** (0.549)	1.653*** (0.541)
Regional dummies	Y	Y	Y	Y	Y	Y
Hardle and Mammen's (1993) test, p-value	0.59	0.04	0.19	0.53	0.01	0.03
Observations	90	90	90	90	90	90
Effective degrees of freedom	72.5	69.8	62.8	72.8	71	72.6
Adjusted R^2	0.440	0.620	0.602	0.449	0.603	0.541

Notes: All dependant variables are normalized. All regressions include a constant. Coefficient (robust standard errors). Regional dummies are dummies at five major área categories: America, Africa, Asia, Eastern Europe, and Western Europe (Asia is the base category). Nonpar. Variable indicates the variable included in a non-parametric way in each regression. Significant at the *10%, **5%, ***1% level.

parts of the curves with concave and convex behaviours, far from being purely linear. However, in Figures 1(a), 1(c) and I(d), these non-linearities are not significantly different from a linear approach.

The influence of the percentage of urban population on the virus incidence (Figures 1(a) and 1(d)) is similar for our two dependent variables: first, for low percentages of the urban population, the effect can be well proxied by a horizontal straight line with incidence values lower than average; once the threshold of about 50% of urban population is reached then the relationship between both variables in the two graphs is clearly increasing, and for very high percentages the growth is at increasing rates, that is to say, convex. The urban primacy variable (Figures 1(c) and



(a) Cases and urban population (b) Cases and population density (c) Cases and urban primacy

(d) Deaths and urban population (e) Deaths and population density (f) Deaths and urban primacy

FIGURE 1 Semi-parametric estimates

1(f)) confirms this result at least to a certain extent. In effect, for low and intermediate primacies, the curve is, especially for the number of cases, a sinusoid around the zero value, and the relationship with the COVID-19 incidence variable is positive only for very high primacies (about 33% for the number of cases and about 25% for the deaths); therefore, for the majority of values of primacy, an increase in the relative importance of the largest city does not necessarily imply a growth in the number of cases or deaths.

Finally, Figures 1(b) and 1(e) define the behaviour of the population density variable. In the body of both distributions, especially for the number of deaths, an inverted U-shape relationship is found. For low (up to 3.8 in log scale in both graphs) and very high densities (beyond 6.2), the link between density and COVID-19 incidence is positive; for the rest of the distribution (which includes the majority of the observations in Figures 1(b) and 1(e)) a higher population density implies a lower virus incidence in terms of cases and deaths.

5 | CONCLUSIONS

The COVID-19 crisis has spurred the scientific community to search for the determinants and consequences of the pandemic from many points of view. In this paper, we have not analysed the impact of COVID-19 on a set of variables of interest, as is becoming usual in this literature. Our approach is different: we have studied how some variables can explain the diffusion and intensity of the virus in a sample of 90 countries. Two endogenous variables are defined: the cumulative number of cases and deaths per 100,000 inhabitants up to 31 December 2020. Our main explanatory variables are related to the urban dimension of each country: percentage of the urban population, population density (concentration of population per square kilometre), and urban primacy (percentage of the population of the largest city), to test whether higher values of these three variables imply or not a stronger incidence of the disease. A secondary working hypothesis we adopt is that the relationship between primacy and COVID-19 incidence might be non-linear, varying across the distribution of these urban variables, as agglomeration can generate a new non-linear congestion cost: the risk of infection cost.

From the parametric approach, we concluded that the main explanatory variable is the percentage of the urban population, in such a way that countries more urbanized suffer a higher incidence of the virus in terms of the relative number of cases and relative number of deaths. We also found a weaker positive influence of the density on the relative number of cases. Regarding primacy, it does not affect the number of cases and, with respect to relative deaths, in some models, the existence of a very large agglomeration not only does not increase the number of deaths but makes it decrease. Therefore, it seems that it is the complete urban structure of the country (with cities of all the sizes: large, intermediate, and small) that mainly helps explain the magnitude of the endogenous variables. Finally, we also obtain some evidence of spatial effects, although weaker than in previous research (Krisztin, Piribauer, & Wögerer, 2020).

From the semiparametric approach, we deduce that in three out of six possible cases (two endogenous variables and three urban explanatory variables), the non-linearities are statistically significant. Furthermore, we have not detected a systematic and monotonous positive relationship between the three urban variables and the intensity of the disease, as might be expected a priori. A positive relationship was detected only for intermediate and high values of the percentage of urban population; for low and very high values of density; and for very high values of primacy. At other points of the distribution, the relationship between COVID-19 incidence and the urban variables is constant or even negative. These non-linear behaviours have strong implications for policymakers and health authorities, and the causal link between urbanization and COVID-19 intensity clearly deserves further research.

ACKNOWLEDGEMENTS

The authors thank Cristian Liarte Blasco for valuable assistance with data. An earlier version of this paper was presented at the I International on-line Conference on Regional Science, with all the comments made by participants being appreciated. Finally, comments received from three anonymous referees have improved the version originally submitted. All remaining errors are ours.

FUNDING INFORMATION

Ministerio de Ciencia e Innovación, MCIN/AEI/10.13039/501100011033, Grants number: ECO2017-82246-P, PID2020-114354RA-I00 and PID2020-112773GB-I00; Gobierno de Aragón, Grant number: ADETRE research group; European Regional Development Fund (ERDF).

ORCID

Rafael González-Val https://orcid.org/0000-0002-2023-5726
Fernando Sanz-Gracia https://orcid.org/0000-0003-3725-0022

ENDNOTES

[1] This normalisation only implies a change in the scale of the variables. Results are robust to this re-scaling.

[2] Using data about the number of tests *per capita* by country from Hasell et al. (2020), we regressed our two variables measuring COVID-19 incidence on the cumulative tests per 100,000 inhabitants, finding that the tests' coefficient is significant explaining cases but not in the case of deaths. These coefficients are positive but very small, and they become even smaller when regional dummies are included. However, as the number of tests is not available for all the countries in our sample (we lose 25 observations), this variable is not included in the analysis, although these results are available from the authors upon request.

[3] Google provides daily data of the percent change in mobility for several categories: retail and recreation, grocery and pharmacy, parks, transit stations, workplaces, and residential. In all cases, the references to calculate the change are the immediate pre-pandemic values in January and February 2020. After several proofs we selected only the percentage change in mobility to workplace.

[4] Both the Google mobility measures and the stringency index are provided on a daily basis; to get a single value, we consider all the Spring values by country (in 2020, the highest peak of cases were reported around March and April for most countries) and then we compute the average for the mobility measure, and the maximum value for the stringency index.

[5] Although the three variables represent population agglomeration, we can include the three at the same time because they are not highly correlated. Bilateral correlations are −0.19 for urban population-density (log scale), 0.55 for urban population-primacy and −0.13 for primacy-density (log scale).

[6] As a preliminary analysis, we apply the Lagrange multiplier and Moran's *I* tests to the residuals of the model in Equation 1, using the full version of the model, including all controls and regional dummies, except the distance to Wuhan/Milan. The null hypothesis in all tests is that there is zero spatial autocorrelation. Results provide mixed evidence: whereas the robust Lagrange multiplier test rejects the null hypothesis of no spatial effects using both the standard spatial lag and errors models for the relative number of cases, for the relative number of deaths, we obtain some weak evidence supporting the spatial lag model, as the null cannot be rejected at the 10%. These results are available from the authors upon request.

[7] The log-likelihood values for different values of *k* are available from the authors upon request.

[8] Ranking of the cumulative number of cases per 100,000 inhabitants on 31 December, 2020.

REFERENCES

Alfani, G., & Percoco, M. (2019). Plague and long-term development: The lasting effects of the 1629–30 epidemic on the Italian cities. *The Economic History Review*, *72*(4), 1175–1201. https://doi.org/10.1111/ehr.12652

Alirol, E., Getaz, L., Stoll, B., Chappuis, F., & Loutan, L. (2011). Urbanisation and infectious diseases in a globalized world. *Lancet Infectious Diseases*, *11*, 131–141. https://doi.org/10.1016/S1473-3099(10)70223-1

Anselin, L. (1988). Lagrange multiplier test diagnostics for spatial dependence and spatial heterogeneity. *Geographical Analysis*, *20*, 1–17. https://doi.org/10.1111/j.1538-4632.1988.tb00159.x

Baker, S. R., Bloom, N., Davis, S. J., & Terry, S. J. (2020). COVID-Induced economic uncertainty. NBER Working Paper 26983. Available at: http://www.nber.org/papers/w26983

Barrios, S., & Strobl, E. (2009). The dynamics of regional inequalities. *Regional Science and Urban Economics*, *39*(5), 575–591. https://doi.org/10.1016/j.regsciurbeco.2009.03.008

Bartik, A. W., Bertrand, M., Cullen, Z., Glaeser, E. L., Luca, M., & Stanton, C. (2020). The impact of COVID-19 on small business outcomes and expectations. *PNAS*, *117*(30), 17656–17666. https://doi.org/10.1073/pnas.2006991117

Bashir, M. F., Benjiang, M. A., & Shahzad, L. (2020). A brief review of socio-economic and environmental impact of Covid-19. *Air Quality*, *Atmosphere and Health*, *13*, 1403–1409. https://doi.org/10.1007/s11869-020-00894-8

Carozzi, F., Provenzano, S., & Roth, S. (2020). Urban density and COVID-19. IZA Discussion Paper 13440. Available at: https://www.iza.org/publications/dp/13440/urban-density-and-covid-19

Clemens, J., & Veuger, S. (2020). Implications of the Covid-19 pandemic for state government tax revenues. NBER Working Paper 27426. Available at: https://www.nber.org/papers/w27426

Díez-Minguela, A., González-Val, R., Martinez-Galarraga, J., Sanchis, M. T., & Tirado, D. A. (2020). The long-term relationship between economic development and regional inequality: South-West Europe, 1860–2010. *Papers in Regional Science*, *99*(3), 479–508. https://doi.org/10.1111/pirs.12489

Durlauf, S. N. (2001). Manifesto for a growth econometrics. *Journal of Econometrics*, *100*(1), 65–69. https://doi.org/10.1016/S0304-4076(00)00055-5

Ehlert, A. (2021). The socio-economic determinants of COVID-19: A spatial analysis of German county level data. *Socio-Economic Planning Sciences*, 101083. https://doi.org/10.1016/j.seps.2021.101083

Elenev, V., Landvoigt, T., & Van Nieuwerburgh, S. (2020). Can the Covid bailouts save the economy? NBER Working Paper 27207. Available at: https://www.nber.org/papers/w27207

Elhorst, J. P. (2014). *Spatial econometrics: From cross-sectional data to spatial panels*. Springer. https://doi.org/10.1007/978-3-642-40340-8

Farzanegan, M. R., Feizi, M., & Gholipour, H. F. (2021). Globalization and outbreak of Covid-19: An empirical analysis. *Journal of Risk and Financial Management*, *14*(3), 105. https://doi.org/10.3390/jrfm14030105

Geng, X., Gerges, F., Katul, G. G., Bou-Zeid, E., Nassif, H., & Boufadel, M. C. (2021). Population agglomeration is a harbinger of the spatial complexity of COVID-19. *Chemical Engineering Journal*, *420*, 127702. https://doi.org/10.1016/j.cej.2020.127702

Golgher, A. B., & Voss, P. R. (2016). How to interpret the coefficients of spatial models: Spillovers, direct and indirect effects. *Spatial Demography*, *4*, 175–205. https://doi.org/10.1007/s40980-015-0016-y

Hardle, W., & Mammen, E. (1993). Comparing nonparametric versus parametric regression fits. *Annals of Statistics*, *21*, 1926–1947. https://doi.org/10.1214/aos/1176349403

Hasell, J., Mathieu, E., Beltekian, D., Macdonald, B., Giattino, C., Ortiz-Ospina, E., Roser, M., & Ritchie, H. (2020). A cross-country database of COVID-19 testing. *Scientific Data*, *7*, 345. https://doi.org/10.1038/s41597-020-00688-8

Hu, H., Nigmatulina, K., & Eckhoff, P. (2013). The scaling of contact rates with population density for the infectious disease models. *Mathematical Biosciences*, *244*, 125–134. https://doi.org/10.1016/j.mbs.2013.04.013

Jedwab, R., Johnson, N. D., & Koyama, M. (2019). Pandemics, places, and populations: Evidence from the Black Death. CESIFO Working Paper 7524. Available at: https://ssrn.com/abstract=3343855

Krisztin, T., Piribauer, P., & Wögerer, M. (2020). The spatial econometrics of the coronavirus pandemic. *Letters in Spatial and Resource Sciences*, *13*, 209–218. https://doi.org/10.1007/s12076-020-00254-1

LeSage, J., & Pace, R. K. (2009). *Introduction to spatial econometrics*. Chapman Hall/CRC Press. https://doi.org/10.1201/9781420064254

Lessmann, C. (2014). Spatial inequality and development. Is there an inverted-U relationship. *Journal of Development Economics*, *106*, 35–51. https://doi.org/10.1016/j.jdeveco.2013.08.011

Maital, S., & Barzani, E. (2020). The global economic impact of COVID-19: A summary of research. Samuel Neaman Institute for National Policy Research. Available at: https://www.neaman.org.il/EN/The-Global-Economic-Impact-of-COVID-19-A-Summary-of-Research

Martin, A., Markhvida, M., Hallegatte, S., & Walsh, B. (2020). Socio-economic impacts of COVID-19 on household consumption and poverty. *Economics of Disasters and Climate Change*, *4*, 453–479. https://doi.org/10.1007/s41885-020-00070-3

McKibbin, W., & Fernando, R. (2021). The global macroeconomic impacts of COVID-19: Seven scenarios. *Asian Economic Papers*, *20*(2), 1–30. https://doi.org/10.1162/asep_a_00796

Neiderud, C.-J. (2015). How urbanization affects the epidemiology of emerging infectious diseases. *Infection Ecology & Epidemiology*, *5*(1), 27060. https://doi.org/10.3402/iee.v5.27060

Paez, A., Lopez, F. A., Menezes, T., Cavalcanti, R., & Pitta, M. G.d. R. (2021). A spatio-temporal analysis of the environmental correlates of COVID-19 incidence in Spain. *Geographical Analysis*, *53*, 397–421. https://doi.org/10.1111/gean.12241

Robinson, P. M. (1988). Root-N consistent semiparametric regression. *Econometrica*, *56*(4), 931–954. https://doi.org/10.2307/1912705

Rodríguez-Pose, A., & Burlina, C. (2021). Institutions and the uneven geography of the first wave of the COVID-19 pandemic. *Journal of Regional Science*, *61*(4), 728–752. https://doi.org/10.1111/jors.12541

Sumner, A., Hoy, C., & Ortiz-Juárez, E. (2020). Estimates of the impact of COVID-19 on global poverty. United Nations University World Institute for Development Economics Research Working Paper 43/2020/. Available at: https://www.wider.unu.edu/publication/estimates-impact-covid-19-global-poverty

Topcu, M., & Gulal, O. S. (2020). The impact of COVID-19 on emerging stock markets. *Finance Research Letters*, *36*, 1–4. https://doi.org/10.1016/j.frl.2020.101691

Voigtländer, N., & Voth, H. J. (2013). The three horsemen of riches: Plague, war, and urbanization in early modern Europe. *The Review of Economic Studies*, *80*, 774–811. https://doi.org/10.1093/restud/rds034

Wheaton, W. C., & Thompson, A. K. (2020). Doubts about density: COVID-19 across cities and towns. Available at: https://ssrn.com/abstract=3586081

Wildman, J. (2021). COVID-19 and income inequality in OECD countries. *The European Journal of Health Economics*, *22*, 455–462. https://doi.org/10.1007/s10198-021-01266-4

How to cite this article: González-Val, R., & Sanz-Gracia, F. (2022). Urbanization and COVID-19 incidence: A cross-country investigation. *Papers in Regional Science*, *101*(2), 399–415. https://doi.org/10.1111/pirs.12647

APPENDIX A

LIST OF COUNTRIES

Argentina	Kenya
Australia	Kyrgyzstan
Austria	Latvia
Bahrain	Lebanon
Bangladesh	Lithuania
Belarus	Malaysia
Belgium	Mexico
Bolivia	Morocco
Bosnia and Herzegovina	Myanmar

Brazil	Nepal
Bulgaria	Netherlands
Cambodia	New Zealand
Cameroon	Nigeria
Canada	Norway
Chile	Oman
Colombia	Pakistan
Costa Rica	Peru
Croatia	Philippines
Czechia	Poland
Denmark	Portugal
Dominican Republic	Qatar
Ecuador	Republic of Korea
Egypt	Republic of Moldova
El Salvador	Romania
Estonia	Russian Federation
Finland	Saudi Arabia
France	Senegal
Georgia	Serbia
Germany	Slovakia
Ghana	South Africa
Greece	Spain
Guatemala	Sri Lanka
Haiti	Sweden
Honduras	Switzerland
Hungary	Tajikistan
India	Thailand
Indonesia	Turkey
Iraq	Uganda
Ireland	Ukraine
Italy	United Arab Emirates
Ivory Coast	United Kingdom
Jamaica	United Republic of Tanzania
Japan	United States of America
Jordan	Vietnam
Kazakhstan	Zimbabwe

DOI: 10.1111/pirs.12647

Resumen. Este artículo investiga los determinantes de la difusión y la intensidad de COVID-19 a nivel de país, centrándose en el papel que desempeña la aglomeración urbana, medida a través de tres variables urbanas: el porcentaje de población urbana, la densidad de población y la primacía. Se utilizaron modelos paramétricos y semiparamétricos para estimar la influencia de la aglomeración urbana en dos variables de resultado: el número acumulado de casos y de muertes por 100.000 habitantes hasta el 31 de diciembre de 2020. También se exploraron los posibles efectos espaciales. Los efectos no lineales de las variables urbanas sobre la intensidad de la enfermedad revelaron relaciones no monótonas, lo que sugiere que el tamaño del sistema urbano es lo que está vinculado a una mayor incidencia.

抄録: 本稿では、都市集積が果たす役割に焦点を当てて、3つの都市変数〔urban variable:都市人口の割合(パーセンテージ)、人口密度、首座都市性(primacy)〕を用いて測定して、国レベルでのCOVID - 19の拡散とその強度の決定要因を調査する。パラメトリックモデル及びセミパラメトリックモデルの両方を用いて、2020年12月31日までの住民10万人当たりの累積症例数と死亡数の2つの結果変数に対する都市集積の影響を推定した。また、空間効果も推定した。疾患の強度に対する都市変数の非線形効果は非単調関係を示したことから、発生率をより大きくするのは都市システムのサイズであることが示唆された。