www.cya.unam.mx/index.php/cya

Contaduría y Administración 65 (5), Especial COVID-19, 2020, 1-28

The effect of socioeconomic factors and health conditions on COVID-19 infection in the States of Mexico

Efecto de factores socio-económicos y condiciones de salud en el contagio de COVID-19 en los estados de México

Pablo Mejía Reyes*, Annel Hurtado Jaramillo, Liliana Rendón Rojas

Universidad Autónoma del Estado de México, México

Received October 28, 2020; accepted October 18, 2020
Available online November 26, 2022

Abstract

The aim of this document is to analyze the effects of a large battery of demographic, social, health and economic factors on the magnitude and intensity of SARS-CoV-2 contagion in the states of Mexico. To reach so, an extreme-bounds analysis in cross-section econometric models, with possible spatial dependence, is carried out. Our findings suggest that a greater population density (that impedes social distancing), the suffering of obesity and/or chronic degenerative diseases (diabetes and hypertension), and the lack of respect for health regulations have favored the spread of COVID-19. Social conditions of the population and economic characteristics seem to be not relevant. The public policy implications from our results are straightforward.

JEL Code: C21, I18, J11, K00, R15
Keywords: COVID-19; population density; health conditions; health regulations; spatial econometrics

*Corresponding author.
E-mail address: pmejiare@uaemex.mx (P. Mejía Reyes).
Peer Review under the responsibility of Universidad Nacional Autónoma de México.

http://dx.doi.org/10.22201/fca.24488410e.2020.3127

Resumen

El objetivo de este documento es analizar el impacto de múltiples factores demográficos, sociales, de salud y económicos en la magnitud e intensidad del contagio de SARS-CoV-2 en los estados mexicanos. Para ello se desarrolla un análisis de límites extremos (extreme-bounds analysis) en modelos de regresión de corte transversal, que pueden incluir efectos espaciales. Los resultados sugieren que una mayor densidad de población (que dificulta el distanciamiento social), el padecimiento de obesidad y/o enfermedades crónico-degenerativas (diabetes e hipertensión) y el no respeto a las disposiciones sanitarias han favorecido el contagio de COVID-19. Las condiciones sociales de la población y las características económicas de los estados no resultaron relevantes. Las implicaciones de política pública que se derivan de este resultado son directas.

Código JEL: C21, I18, J11, K00, R15
Palabras clave: COVID-19; densidad de población; condiciones de salud; regulación sanitaria; econometría espacial

Introduction

The disease named COVID-19 (Coronavirus Disease-2019) is caused by the new Severe Acute Respiratory Syndrome Type 2 Coronavirus (SARS-CoV-2), which is taxonomically placed in the family Coronaviridae (Gorbalenya et al., 2020). This family is composed of different viruses which cause diseases in humans and animals, ranging from the common cold to severe respiratory illnesses. COVID-19 is a transmissible disease that spreads at high speed (Gupta et al., 2020) as it is estimated that one infected person would spread the disease between 2.2 and 3.5 people on average (Díaz-Castrillón & Toro-Montoya, 2020; Callaway et al., 2020).

The first cases of COVID-19 were confirmed on December 2019 in Wuhan city, Hubei province of China, as an outbreak of pneumonia of unknown cause.[1] The virus spread rapidly to the rest of China and other countries. On mid-January, patients were identified in other Asian countries (Indonesia, Japan, Thailand and South Korea), and since the last week of January, in several European countries (France, Germany, Spain, United Kingdom and Italy). As in previous cases, the virus arrived in America via international travelers. The first case was reportedly confirmed in the United States (U.S.) on January 19, 2020, in Washington State (Holshue et al., 2020). With the accelerated spread of the virus to several countries in virtually every continent, the World Health Organization (WHO) declared the SARS-CoV-2 outbreak a health emergency of global significance on January 30 (WHO, 2020a; ECDC, 2020). In turn, the new coronavirus appeared in Latin America almost a month later: on February 26, the first case was

[1]The identification of the first patients in China and other countries was very uncertain, as it was an unknown virus that was only identified in January 12 in China. See WHO (2020a) for a sequence of events since the first case occurred.

confirmed in Sao Paulo, Brazil, while in Mexico, it appeared on February 28 in Mexico City (PAHO/WHO, 2020). Both cases involved people who had recently traveled from Italy. Due the situation, WHO officially declared the existence of a COVID-19 pandemic on March 11, 2020 (WHO, 2020a).

Since COVID-19 is a new disease, there is no vaccine or proven treatment for its cure or effective control. Therefore, the vast majority of countries in the world have taken physical distancing and the practice of thorough hygiene as general prevention measures (Lakshmi-Priyadarsini & Suresh, 2020). These measures required people to reduce their outdoor activities, gatherings, and geographic travels. As a result, school activities were suspended and tourism, recreation, entertainment, and other activities have been significantly reduced or canceled virtually all over the world. To varying degrees and with varying effectiveness, home confinement and the use of masks have been imposed.

However, the virulence of COVID-19, non-compliance with sanitary measures, and other social, health, and demographic factors have led to high levels of infection worldwide, especially when appropriate measures have not been implemented. For example, on July 31, the number of infections reached 4 388 566 in the USA, 3 300 000 in European countries and 770 412 in Africa. Only in countries where strict social distancing measures have been taken were the infections controlled relatively quickly, as in China, where only 87 956 cases had been found by the same date (WHO, 2020b).

In Mexico, 178 days after the first case was identified, 454 322 people have been infected, and 46 688 died. Disaggregated information, however, shows that infections have been distributed very heterogeneously among the country's states. In particular, up to July 31, the three states with the highest number of infections were Mexico City, the State of Mexico, and Tabasco, with 76 804, 55 726, and 22 961, respectively. In contrast, the States with the lowest number of positive cases were Colima (1 985), Zacatecas (2 905), and Nayarit (3 306). Of course, among other factors, this may be due to the population size of the States themselves (scale effect), so a more appropriate measure would be the prevalence rate. According to this, the states most affected up to July 31 were Tabasco, which had 8.9 cases per 1 000 inhabitants, Mexico City, 8.5, and Sonora, 6.0, respectively. In contrast, Chiapas reported only 1.0, Chihuahua 1.5 and Jalisco 1.6 (Gobierno de México,2020).

The world is undoubtedly facing an unprecedented health crisis caused by the new SARS-CoV-2 virus. Because it is a different variant, there has been a lot of speculation about the main factors which lead to its spread. However, relatively little research has been made on the role of various factors may played in its spread, which has also been contributed to the fact that it is an ongoing process. Therefore, this paper aims to analyze the impact of social, demographic, and economic factors, and prevailing health conditions on the cumulative levels of infection and the prevalence rate of COVID-19 on the states of Mexico. To this end, an extreme bounds analysis is performed in different econometric models to identify, within a wide range, the variables that can explain them. The results indicate that, in general, population density, compliance with health recommendations, and pre-existing diseases determine their magnitude and intensity.

The rest of the document consists of five sections. The first section briefly presents the evolution of SARS-CoV-2 infections at the international level until its arrival in Mexico to analyze infections at the state level. The second section reviews the relevant literature on the subject. The third section specifies the econometric model and defines the methodology used to identify the factors that can explain the cumulative level and prevalence rate of COVID-19. Section four presents and discusses the main results, and finally, the conclusions are drawn.

COVID-19 infection in Mexico and worldwide

The end of 2019 marks the beginning of the worst health crisis the world has experienced in the last 100 years, due to the spread of a novel virus called SARS-CoV-2. According to the South China Morning Post, the first case occurred in China on November 17, 2019, and after that, cases rapidly increased throughout the country. On December 27, Zhang Jixian, a physician at the Hubei Provincial Hospital of Chinese and Western Integrated Medicine, reported that the disease had been caused by a new coronavirus (Ma, 2020).

According to the World Health Organization (WHO, 2020a), the first death caused by the virus occurred on January 9, 2020, and on January 13, 2020, a case was reported in Thailand, the first outside China. In America, the U.S. reported the first patient on January 19. France, for its part, reported its first two patients on January 24, thus initiating the epidemic in Europe. One day later, on January 25, Australia reported 4 cases, bringing the disease to Oceania (Reuters, 2020).

On February 11, the WHO (2020a) announced that the official name of the disease would be "Coronavirus Disease," abbreviated as "COVID-19." In the following days, the first cases were confirmed in Egypt (the first in Africa), followed by Israel, and then in Brazil (on February 26), becoming the first country in Latin America with carriers of the new virus.

To control contagion, governments worldwide have adopted a wide range of measures. In the absence of proven medical procedures or a vaccine to eliminate the virus, they decided to impose restrictions on people's mobility in their immediate environment (home confinement) and their geographical movement (national and international travel). Thus, for example, on January 22, South Korea announced the temporary suspension of the entry of tourists throughout its territory; the Italian government, for its part, ordered the total isolation of the country, in addition to imposing stricter measures, such as a travel ban.[2] Spain decreed a State of Alarm so that the population was confined, while non-essential workers stopped working or did so from their homes, and schools, colleges and universities closed (Carrión et al., 2020). In America, Peru and Chile went into a state of emergency and closed their borders on March 15 (WHO, 2020a).

Despite these measures and others taken in different countries, the number of infections increased worldwide, albeit at different rates, leading the WHO to declare COVID-19 a global pandemic on March 11 (WHO, 2020a). The rapid growth in infections has led to many cumulative cases in several countries. For example, as of July 31, there have been a total of 4 495 014 in the USA; 288 522 in Spain; 186 573 in France; 247 158 in Italy; 2 610 102 in Brazil; 407 492 in Peru; 353 536 in Chile; and 454 322 in Mexico. Apparently, the countries where there were fewer numbers were those where rapid and strict control measures were adopted, such as China and South Korea, where a total of 87 489 and 14 305 people, respectively, were affected up to the same date (ECDC, 2020).

In the case of Mexico, the first patient with COVID-19 traveled from Italy. It was reported on February 28, 2020, which led the federal government to announce the beginning of Phase 1 of the disease —only with imported cases— on the 29th of the same month. The exponential expansion of the virus in the first weeks led Mexican authorities to implement the Jornada Nacional de Sana Distancia (National Healthy Distance Campaign) from March 23 until April 19. Among the measures adopted were voluntary confinement, the suspension of non-essential activities, the carrying out of educational and work activities from home as far as possible, the suspension of mass events (of more than 5 000 people) and special care for adults over 60 years. On the same day, the WHO indicated that local transmission of the disease had already occurred in Mexico, which meant that the country was now in Phase 2 of the pandemic. Therefore, on March 30, with 1 094 confirmed cases and 28 deaths, the federal government decreed a health emergency and decided on the home protection of people over 60 years of age and those with diabetes, hypertension, and heart diseases (Gobierno de México,2020).

On April 20, after 857 deaths and 9 501 confirmed cases, the beginning of Phase 3 was announced. Among the measures approved were the extension of the National Healthy Distance Campaign until May 30 and the extension of the suspension of non-essential activities. At the end of the National Healthy Distance Campaign, 87 517 confirmed cases and 9 779 deaths were reported (Gobierno de

[2]Italy imposed the most severe blockade outside of mainland China, restricting the mobility of more than 10 million people in the northern part of the country. Places affected by the closure included Milan's financial center and the tourist destination of Venice, while measures such as the closure of schools and universities and the cancellation of all public events, sporting events, and civil and religious ceremonies were imposed (Cartenì *et al.*, 2020). Italians remained under quarantine until May 3.

México,2020). Then, starting in early June, Mexico, like other countries, prepared for the reopening of the economy by transitioning to the so-called "New Normal" phase.[3]

With the implementation of a system based on the "traffic light" concept,[4] the federal administration ended the sanitary contingency decreed at the end of March and transferred to the governments of each state the authority to decide how to carry out the reopening (Gobierno de México,2020).

The Mexican government's strategy has not been successful from the point of view of the number of daily infections, which continued to increase at increasing rates until the last days of July. Only recently has there been a stagnation above 50 daily confirmed cases per million inhabitants, a very high figure compared to those of other countries in about the same period. Moreover, as of July 31, Mexico had a total of 454 322 infected people, corresponding to a prevalence rate[5] of 3 800.9 persons per million inhabitants (Gobierno de México,2020).

At the state level, the situation has been very heterogeneous due to their individual socioeconomic characteristics, the population's reactions to health measures, and the specific policies adopted by each state.[6] In particular, Figure 1 shows the cumulative infections in the states with the highest levels. Exponential growth can be observed in most of them, at least during the first 100 days. The same trend was maintained in states such as Tabasco, Veracruz, Guanajuato, Puebla and Nuevo Leon. Mexico City and the State of Mexico have shown linear growth in recent weeks, while in Sonora and Baja California, and to some extent in Tamaulipas, the pace of growth seems to be slowing down. In any given case, the levels of contagion have been very high.

[3]At that time, the Mexican economy was mired in the deepest recession of the last nine decades. According to figures from the National Institute of Statistics and Geography, INEGI (2020), production had decreased by 17.3% at an annual rate, investment had fallen 31.7% and 12.8 million formal and informal jobs had been lost.
[4]The concept of the epidemiological risk traffic light to move towards a new normal is a monitoring system to regulate the use of public space according to the risk of transmission of COVID-19. This traffic light is statewide and has four colors: Red, only essential economic activities are allowed. Orange, in addition, non-core economic activities are allowed to function with 30% of their personnel, while considering the maximum care measures for people with a higher risk of presenting a serious case of COVID-19, open public spaces are opened with a reduced capacity (number of people). Yellow, all work activities are allowed, taking care of people at highest risk. Open public space is opened on a regular basis, and closed public spaces are opened with reduced capacity. Green, all activities are allowed, including school activities (Gobierno de México,2020).
[5]The prevalence rate weights transmissions by a certain number of inhabitants, so it can be seen as a measure of pandemic intensity.
[6]States also adopted very heterogeneous measures. A summary is available upon request.

Figure 2, in turn, shows the evolution of infections per 1 000 inhabitants in the states where the highest levels have been reached; note that these are different from those shown in Figure 1, although some are repeated. Again, there are different patterns. The trend up to the end of July continues to be increasing in Coahuila, Yucatan, Tabasco and Baja California Sur, while in Mexico City, Quintana Roo and Campeche, there seems to be a stable linear growth. Only in Tamaulipas, Sonora and Sinaloa is there a slowdown in the rate of infection.

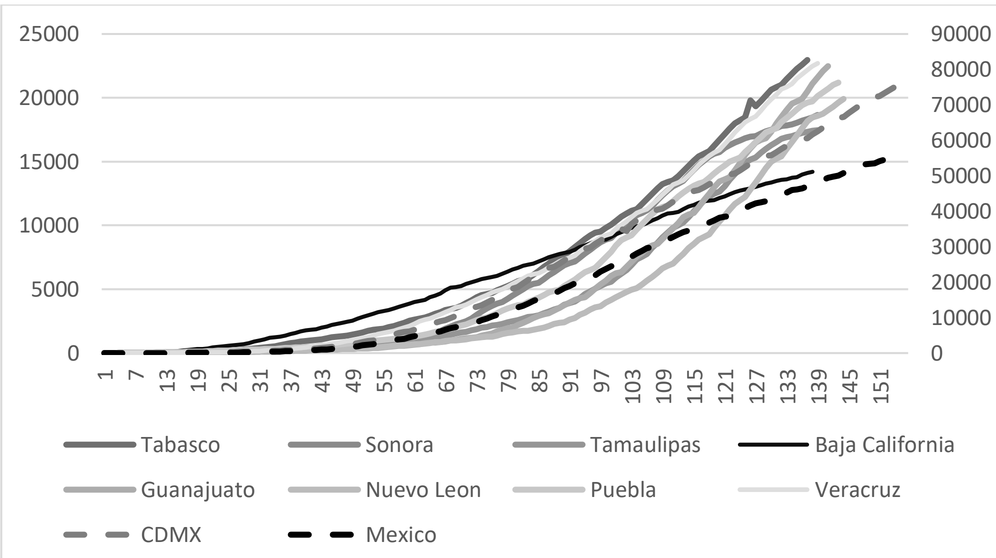

Figure 1. Cumulative COVID-19 infections in the states of Mexico with the highest levels (Number of persons)
Data for Mexico City and the State of Mexico are measured on the right axis.
Source: created by the authors based on information fromhttps://coronavirus.gob.mx/datos/#COMNac

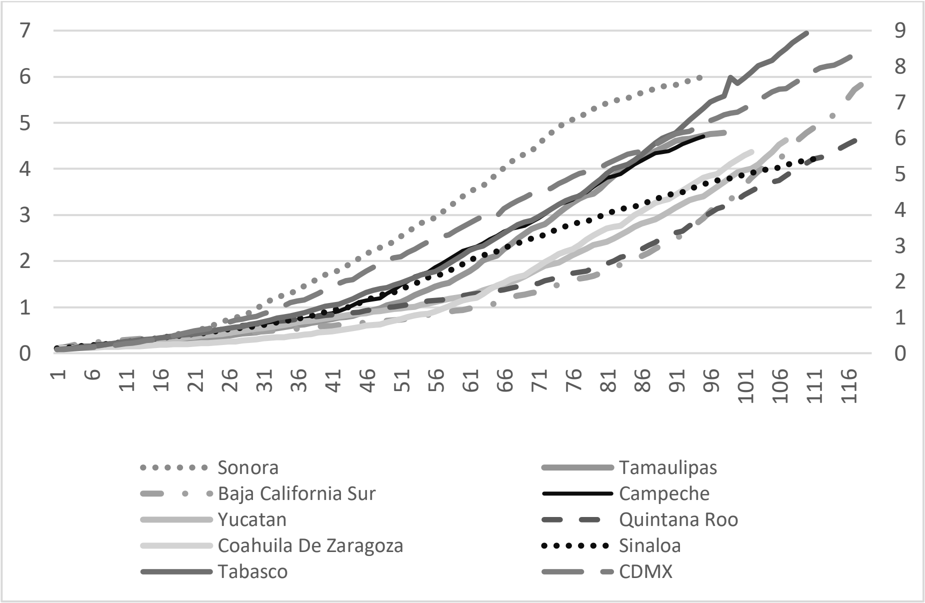

Figure 2. Prevalence rate of COVID-19 in the states with the highest levels (Contagions per 1000 inhabitants)
The horizontal axis measures the number of days since an average of 0.1 cases per 1 000 population. Cases in Mexico City and Tabasco are measured on the right axis.
Source: created by the authors based on information from https://coronavirus.gob.mx/datos/#COMNac

The spatial distribution of the cumulative number of infections among the states is shown in the map in Figure 3. As of July 31, 2020, the state with the highest number of infections was Mexico City (76 804), followed by the State of Mexico (55 726), Tabasco (22 961), Veracruz (22 703), Guanajuato (22 479), Puebla (21 192), Nuevo León (19 910) and Sonora (18 657), several of them with high population levels. In contrast, states with lower population volumes, large territorial areas, or predominantly rural environments have had a lower number of infections, such as Chihuahua (5 913), Durango (4 365), Zacatecas (2 905), Baja California Sur (4 692), Chiapas (5 481) or Campeche (4 706).

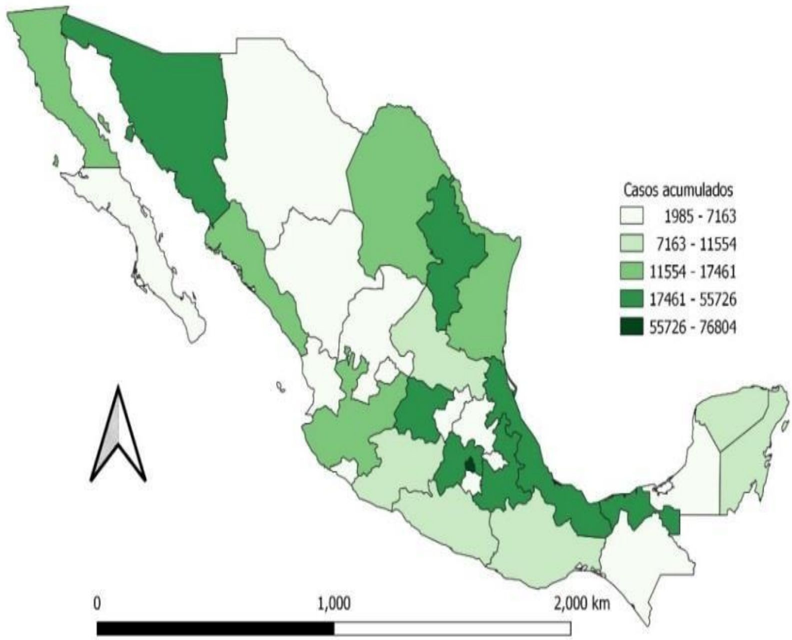

Figure 3. Cumulative COVID-19 infections in the states of Mexico (Persons)
Source: created by the authors based on information from https://coronavirus.gob.mx/datos/#COMNac

The map in Figure 4 shows the prevalence rate of COVID-19 (number of infections per 1 000 inhabitants) in Mexico. Although there are differences with respect to Figure 3, some patterns are repeated. In particular, Tabasco (8.9) and Mexico City (8.5) have the highest levels, followed by Sonora (6.4), Baja California Sur (5.8), Tamaulipas (4.7), Campeche (4.7), Quintana Roo (4.6) and Yucatan (4.6). At the other end of the scale are the states of Chihuahua (1.5), Zacatecas (1.7), Jalisco (1.6), Querétaro (1.6) and Chiapas (1.0), with the lowest number of infections per 1 000 inhabitants.

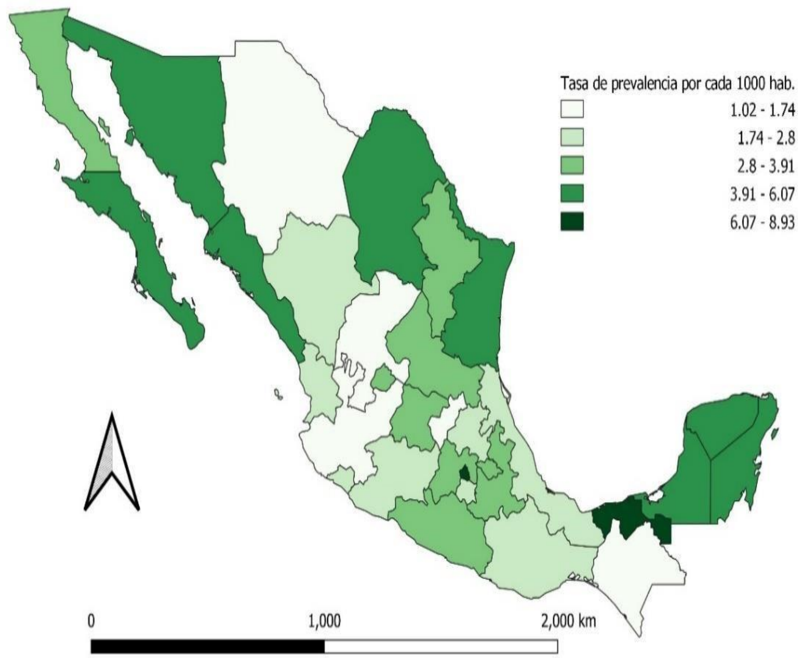

Figure 4. Prevalence rate of COVID-19 in the states of Mexico (Contagions per 1 000 inhabitants)
Source: created by the authors based on information from https://coronavirus.gob.mx/datos/#COMNac

As it can be observed in the maps in Figures 3 and 4, the spatial patterns of the magnitude and rate of COVID-19 infections change significantly. As expected, the former occurs to a greater extent on states with larger populations simply because of a scale factor, although not in all cases. The picture changes when these infections are measured as a proportion of the state population, which gives a measurement of the intensity of infection. In any case, the information sought in the following sections is what factors can explain these differences in magnitude and intensity of COVID-19 among Mexican states, for which an econometric model is used.

http://dx.doi.org/10.22201/fca.24488410e.2020.3127

Brief literature review

COVID-19 is a very virulent disease that took scientists and policymakers around the world by surprise, especially during its first few months. Gradually, however, more has been learned about its nature and the measures to control or reduce the rate of its spread. In this regard, although the measures to be taken are evident from a clinical point of view, the diagnosis is less clear with respect to the factors that may explain the magnitude and intensity of contagion between countries and regions of the same country. Given the seriousness of the situation, and with the phenomenon still ongoing, different studies have appeared that try to explain it for regions in China (Li et al., 2020; Xie et al., 2020), the USA (Papageorge et al., 2020; Gerritse, 2020; Mollalo et al., 2020; Yanga et al., 2020) or Europe (Onder et al., 2020; Cartenì et al., 2020), among others.

First, Lakshmi-Priyadarsini and Suresh (2020) employ the TISM (Total Interpretive Structural Modelling)[7] methodology to identify factors influencing the epidemiological characteristics of COVID-19 worldwide. Among their main findings are that social distancing and community awareness, age, air temperature, airflow and ventilation, population density, and humidity constitute key factors in this model. In addition, they find that host behavior, number of contacts, and personal hygiene practices are the linking factors.

At a country level, Xie et al. (2020) perform an exploratory spatial data analysis to examine the spatial and temporal differences in the spread of the pandemic in China. In addition, using the geo-detector method, which allows analysis of stratified spatial heterogeneity, they identify factors influencing the spread of COVID-19. They find that the population influx of Wuhan and the intensity of regional economic connection were the main factors influencing the spread of the epidemic, along with population distribution, transportation accessibility, average temperature, and medical facilities. Meanwhile, Yanga et al. (2020) evaluate the prevalence of comorbidities in patients infected with SARS-CoV-2 by performing a meta-analysis of seven hospital-based studies in China. They use odds ratios, random effects models and 95% confidence intervals (CI) to reach their conclusions, in which they highlight that underlying conditions, including cardiovascular and respiratory system diseases and hypertension, may be risk factors for severe cases compared to non-severe cases.

Cartenì et al. (2020) analyze the effect of mobility habits on the spread of coronavirus in Italy through a multiple linear regression model. The estimation links the number of daily cases with socioeconomic, environmental, health care, and mobility habits variables. The estimation results showed

[7]This methodology is used to analyze the interrelationship between factors that influence a certain phenomenon, establishing direct or transitive relationships between factors.

that mobility habits constituted the main variable explaining the number of COVID-19 infections. Environmental variables, number of tests per day, and proximity to the first Italian outbreak were also significant, especially in the early stage of infection. Onder et al. (2020) seek to explain the high COVID-19 mortality rates in Italian regions compared to those in other countries. They note that there was a high proportion of older patients with confirmed COVID-19 infection, so this factor could partly explain the differences and that deaths occurred mainly among older male patients with multi-morbidity.

In the case of Spain, Carrión et al. (2020) study the origin of spatial differences in the spread of COVID in the Basic Health Areas (ABS) of Catalonia, using different multiple regression models and quantile regressions. They conclude that virus transmission occurs rapidly in areas that concentrate high volumes of population and, although there is no conclusive relationship, they suggest that low socioeconomic status could also have a positive effect on the spread of the virus.

In the case of the U.S., Mollalo et al. (2020) use five different models: ordinary least squares (OLS), spatial lag model (SLM), spatial error model (SEM), geographically weighted regression (GWR), and multiscale GWR (MGWRX), to explain variations in COVID-19 incidence rate at county level. They consider 35 socioeconomic, environmental, topographic, behavioral, demographic, and topographical factors as the explanatory variables. Among their results, they highlight that the most important explanatory factors are income inequality, average family income, the proportion of black women, and the proportion of practicing nurses.

Gerritse (2020) finds, by estimating epidemiological regression equations, that transmission and incidence of COVID-19 infections differ markedly between different areas of the U.S.A. and that population density is associated with higher transmission rates, particularly at the onset of outbreaks. In turn, Dietz and Santos-Burgoa (2020) analyze the disproportionate impact of H1N1 and now COVID-19 in patients with obesity and severe obesity. They state that obesity is associated with a decrease in expiratory reserve volume, functional capacity, and respiratory system distension, which aggravates infections caused by SARS-CoV-2.

In the case of Mexico, Padilla-Santamaría et al. (2020) conducted an analytical retrospective cross-sectional study, analyzing the official number of confirmed cases and deaths due to COVID-19 in Mexico up to May 9, 2020. They state that males are infected more frequently than females, that case fatality is also higher in males, and that both deaths and infections are related to population density.

This literature review shows that relatively few analysis, especially in the case of Mexico, seek to understand the nature of COVID-19 contagion, possibly because it is an ongoing phenomenon of which many aspects are still unknown. Nevertheless, these studies constitute significant advances in identifying some of the main socioeconomic factors to explain its magnitude and intensity. Based on the results of these studies, the following section formulates an econometric model to explain contagion in the case of Mexican states.

Specification of the econometric model

The variables that have been used in the studies reviewed in the previous section to explain SARS-CoV2 (c) infection can be organized into different groups associated with demographic ($\boldsymbol{X}_1$), social ($\boldsymbol{X_2}$), health conditions ($\boldsymbol{X}_3$) and economic ($\boldsymbol{X}_4$) factors. To analyze the effect of these variables on the number of cumulative cases (magnitude) and the prevalence rate (intensity) of SARS-CoV-2 infection in the states of Mexico, the following general linear model is proposed:

$$\boldsymbol{c} = \boldsymbol{X_1 b_1} + \boldsymbol{X_2 b_2} + \boldsymbol{X_3 b_3} + \boldsymbol{X_4 b_4} + \boldsymbol{u} \tag{1}$$

where $\boldsymbol{c}$ is the variable to be explained and corresponds to an order vector ($n\ x$ 1) (with n = 32 states) containing the number of cumulative cases of contagion (CA) or the prevalence rate (TC); $\boldsymbol{X_i}$ (for i = 1, 2, 3, 4) denotes an order matrix ($n\ x\ k_i$) of explanatory variables, with associated parameters contained in the vector $\boldsymbol{b_i}$ of order ($k_i\ x$ 1), while $\boldsymbol{u}$ represents the vector of disturbances of order ($n\ x$ 1) following a white noise process, i.e. $\boldsymbol{u} \sim iid\ N(0, \sigma^2 I)$.

In particular, $\boldsymbol{X_1}$ contains demographic factors, such as population density (DP) and the percentage of male population (TH).[8] The former is associated with higher transmission rates, since the more people concentrated in smaller areas the greater the physical proximity, which raises the possibility of contagion (Kumar, 2015; Padilla-Santamaría et al., 2020; Hoehl et al., 2020; Gilbert et al., 2020; ECDC, 2020; Carrión et al., 2020). Statistics have shown a higher incidence of COVID-19 in males than in females (Onder et al., 2020; Padilla-Santamaría et al., 2020) at a ratio of 52.6% to 47.4% in the case of Mexico (Gobierno de México,2020), so it is expected that there will be more infections in states with a higher percentage of males. Thus, the matrix $\boldsymbol{X_1}$ would be composed of the vectors corresponding to these variables, $\boldsymbol{X_1} = [\boldsymbol{DP\ TP\ TH}]$.[9]

On the other hand, $\boldsymbol{X_2}$ collects variables referred to as social, mainly referring to people's attitudes toward health recommendations. In fact, numerous reports in the national and international media

[8]The definitions and sources of the variables are presented in Table A1 of the appendix.
[9]In addition, the number of persons per household and the number of rooms per person were considered, as well as the mobility index of the COVID-19 Observatory (2020), but none of these had statistically significant effects on *CA* or *TC*.
[10] The variables that were not significant in this category were the marginalization index, the percentage of population living in poverty, the social backwardness index, and the Gini coefficient.

attribute the increase in contagions to the lack of respect for social distancing (e. g. Flournoy and Morell, 2020). One problem with this variable is that it is not directly observable, so different proxy indicators are used that attempt to measure it. First, the Culture of Legality Index (ICL) is used, which combines variables that provide data on whether the people surveyed declare that they respect, are aware of, and comply with laws and civic practices (see MUCD, 2014). Second, a set of variables is used to measure the perception or respect for the rule of law ex post. Here variables such as perception of security ($PINS$), perception of state corruption ($PCORE$), costs of crime ($CDEL$), reported crime ($DDEN$) and crime incidence ($INDEL$) are taken into account. The matrix $\boldsymbol{X_2}$ = [$\boldsymbol{ICL\ PINS\ PCORE\ CDEL\ DEN\ INDEL}$].[10]

In the $\boldsymbol{X_3}$ matrix, variables associated with pre-existing diseases are considered, as there is evidence suggesting that the immune system of people with obesity and chronic degenerative diseases may have respiratory or cardiovascular complications, which would make them more prone to contract SARS-CoV-2 (Onder et al., 2020; Gupta et al., 2020; Dietz & Santos-Burgoa, 2020; Yanga et al., 2020). To capture these effects, variables such as the proportion of the total population with diabetes ($TDIIA$), hypertension ($THIP$) and obesity (TOB) are considered, as well as those with multi-morbidity associated with hypertension and obesity ($THIPOB$) and diabetes, hypertension and obesity ($THIPOBDIA$).[10]

Finally, COVID-19 infections may be subject to economic factors that prevent people from being confined to their homes or working from home, such as the case of workers in the informal sector or those with precarious jobs. Moreover, the population in conditions of poverty or marginalization may be more vulnerable to infection because they do not have the financial means, physical health conditions, or appropriate hygiene to cope with the virus (Flournoy & Morell, 2020). Other variables were also taken into account in this category, but only the labor gap (BL), defined as BL = (Unemployed + Underemployed + Available) / (Economically Active Population + Available)[11] was significant in the explanation; its values are shown in the matrix $\boldsymbol{X_4}$.[12]

Moreover, the transit of people from one place to another can cause contagion to spread between countries and regions of the same country (as in fact happened). To address this possibility, model (1) is extended to capture the spatial interaction of COVID-19 infections among Mexican states. Then, grouping all explanatory variables and the corresponding parameters in the matrix $\boldsymbol{X}$ = [$\boldsymbol{X_1\ X_2\ X_3\ X_4}$] <and the vector $\boldsymbol{b}$ = [$\boldsymbol{b_1\ b_2\ b_3\ b_4}$], model (1) would be:

[10]Alternative measures such as the number of physicians in public institutions and the health index, based on life expectancy at birth, were also considered (PNUD, 2015). However, none had statistically significant effects.
[11]The labor gap seeks to reflect the incidence of underemployment (people working part-time who want full-time employment) and the extent of hidden unemployment (people who are not actively seeking but would re-enter the labor force if the market allowed) (Blanchflower and Levin, 2015).
[12]The labor informality rate, the rate of critical employment conditions, and the extended unemployment rate were also considered, but none of them showed significant effects.

http://dx.doi.org/10.22201/fca.24488410e.2020.3127

$$\boldsymbol{c} = \rho \boldsymbol{W}\boldsymbol{c} + \boldsymbol{X}\boldsymbol{b} + \boldsymbol{u}$$

$$\boldsymbol{u} = \lambda \boldsymbol{W}\boldsymbol{u} + \boldsymbol{\varepsilon}$$

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

(2)

where $\boldsymbol{W}$ is the standardized spatial weights matrix per row, with entries equal to 1 if the states are neighbors and 0 otherwise (queen-like contiguity). Therefore, $\boldsymbol{Wc}$ is the autoregressive term of the model, so that ρ is the parameter measuring the substantive spatial dependence between each state and its neighbors; $\boldsymbol{Wu}$ and λ are defined analogously in relation to the part not explained by the variables suggested by the theory; $\boldsymbol{\varepsilon}$ is an order vector ($n \ x$ 1) following a white noise process. From this general model different alternatives can be defined: if $\lambda = 0$, the spatial lag model is obtained, while if $\rho = 0$, a spatial error model is obtained; in the case that both coefficients differ from 0, the general model named $SARMAA$ (Spatial autoregressive and moving average) by Anselin and Bera (1998) would be obtained.

The choice of the particular spatial model to be estimated is based on the widely used LM statistics (Lagrange multipliers). In their two versions, LM_ρ or LM_λ, they evaluate the null hypothesis of no substantive and residual spatial dependence, respectively. If the null hypothesis is rejected with either of them, the corresponding model is estimated, but if both test statistics are significant, the best specification is chosen; for example, if $LM_\rho > LM_\lambda$, the spatial lag model is estimated, and vice versa. Additionally, and according to the same logic, robust versions of the same statistics were used, which are valid even in the presence of specification errors. Generally, conventional cross-sectional models are estimated if the null hypothesis is not rejected.

Explanatory factors of SARS-CoV-2 transmission

The analysis of the determinants of COVID-19 transmissions in Mexican states is based on the general model specification and econometric techniques described in the previous section. In particular, the specification of the estimated models consists of two steps. First, the variables that have robust effects on the cumulative number of transmissions (magnitude) or on the prevalence rate (intensity) are selected by developing an extreme bounds analysis (EBA) proposed by Leamer (1983) and popularized by Levine and Renelt (1992). Second, given that transmissions may have moved from one state to another, evidence of spatial dependence is sought, and the corresponding econometric models are estimated.

EBA consists of estimating robust relationships between the explained variable and those suggested by the relevant literature. The starting point is a small model that is gradually expanded if the added variables are statistically significant, of the expected sign, and do not affect the estimates of the previously included variables. This approach could be seen as a "particular to general" approach, with clear advantages when there are few degrees of freedom, as in this case (32 states), and a wide variety of possible explanatory variables.

Then, the objective is to define the order of entry of the explanatory variables into the model. To do this, the starting point is a bivariate regression that includes as an explanatory variable only the one expected to have a robust relationship with the transmissions (based on theory or empirical evidence from other studies). In the case of the dependent variable prevalence rate (TC), the robust explanatory variable turned out to be population density (DP), while for cumulative transmissions (CA), it was one of the groups of social variables that measure respect for norms; specifically, ICL, CDEL, PCORE, DEN or INDEL, which leads to several initial models.[13] Tables 1 and 2 highlight these variables in bold.

Once this first variable is identified, each resulting model (several) is extended with a second one from the different groups, with the requirements mentioned above regarding statistical significance, sign, and no effect on the estimates of the first variable. In this second stage, the effects of variables from the set of health conditions and population density (DP), among others, were robust.[14]

This procedure was used to build models in which up to three explanatory variables were finally incorporated. Those that were added later did not show statistically significant effects. Thus, there are fourteen linear models containing variables with robust effects on transmission. From there, in a second stage, an analysis is done to determine whether it is necessary to incorporate spatial effects. The test statistics corresponding to Moran's I (Moran, 1948) and Lagrange multipliers, LM (Burridge, 1980; Anselin, 1988a, 1988b) are presented in Table A2 in the appendix. Models in which no spatial effects need to be incorporated are shown in Table 1, and spatial models in Table 2.[15] In both tables the normality and homoscedasticity specification tests are also reported; in the second one, the test for spatial residual dependence is added.[16]

Overall, Moran's I is statistically significant only in model 7 (with a p-value equal to 0.050). However, the LM statistics are significant in seven models: they suggest the estimation of a spatial lag

[13]An extensive analysis was also carried out using practically all the variables mentioned in the previous section as initial variables in the bivariate models.
[14]In the subsequent steps, the variable that was most significant to the least significant in previously estimated bivariate models is entered.
[15]These are the heteroscedasticity (Breusch and Pagan, 1980) and normality (Jarque and Bera, 1987) tests.
[16]Likelihood ratio test that contrasts whether there is any residual spatial autocorrelation remaining

model in cases 1, 2, 3, 4, 8 and 9 and of spatial error in 13. In the rest there is no evidence of spatial dependence, so they are estimated by MCO (see Table A2 in the appendix).

In the regressions presented in Table 1, none show normality or heteroscedasticity problems, since the corresponding p-values suggest not rejecting the null hypotheses at 5% significance. In turn, the explanatory power of the models is relatively high, with coefficients of determination above 0.7 in one case and around 0.9 in the rest.

It is noteworthy that all of them include a positive effect of population density (DP) on the magnitude of transmission (models 5, 6, 7, 10, 11 and 12) and the prevalence rate (model 14). The latter also includes the other demographic variable considered, TH, which has a positive and statistically significant coefficient, so it can be concluded that the higher the proportion of men in a state, the higher the transmission rate.[17] The third variable that explains the prevalence rate is the proportion of the state population that suffers from hypertension, which proves the hypothesis that previous illnesses that weaken the immune system favor the acquisition of the virus.

Interestingly, the first variable (in bold) with robust effects on cumulative transmissions is from the group of indicators that seek to measure compliance with regulations. The idea of incorporate these variables is that they serve as a proxy for the population's compliance with health regulations. In this sense, the Cost of crime (CDEL) is significant and has a positive effect on the level of transmission in models 5, 6 and 7, as does the proportion of the population with hypertension (THIP), obesity (TOB) or both (THIPOB). In turn, the variable Reported crimes (DEN) has similar effects in model 10, along with multi-morbidity associated with obesity, hypertension and diabetes (THIPOBDIA). Finally, the Incidence of crime (INDEL) is combined with diabetes (TDIA) or obesity (TOB) to explain the accumulation of COVID-19 transmissions in models 11 and 12, respectively.

[17]In addition, it has been observed that COVID-19 deaths in Mexico have occurred mainly among older male patients who also have multiple comorbidities (Padilla-Santamaría et al. (2020).

http://dx.doi.org/10.22201/fca.24488410e.2020.3127

Table 1
Factors influencing the magnitude and prevalence rate of COVID-19 transmission in Mexico

Variables	Dependent variable: CA						Dependent variable: TC
	(5)	(6)	(7)	(10)	(11)	(12)	(14)*
Intercept	-9100.3	-5087.1	- 11495.2	1140.26	11015.6	- 8744.36	-47.000
	(0.093)	(0.135)	(0.064)	(0.697)	(0.081)	(0.211)	(0.003)
DP	7.214	7.381	7.424	6.567	6.818	7.684	0.001
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.001)
TH							0.971
							(0.003)
CDEL	0.882	0.883	0.895				
	(0.000)	(0.000)	(0.000)				
DEN				0.069			
				(0.000)			
INDEL					0.289	0.280	
					(0.000)	(0.000)	
TDIA					1436.417		
					(0.015)		
THIP	709.173						0.131
	(0.014)						(0.043)
TOB			412.129			342.442	
			(0.012)			(0.060)	
THIPOB		835.286					
		(0.005)					
THIPOBDIA				1104.767			
				(0.033)			
R^2	0.910	0.916	0.911	0.898	0.891	0.881	0.719
Normal	1.212	0.839	2.024	0.576	0.297	1.553	1.906
	(0.546)	(0.657)	(0.363)	(0.749)	(0.862)	(0.460)	(0.385)
Heteroscedasticity	1.360	3.732	1.374	4.411	2.222	2.408	2.106
	(0.715)	(0.292)	(0.712)	(0.220)	(0.528)	(0.492)	(0.834)

The coefficients in bold, italics, and regular type correspond to the variables identified in first, second, and third place in the different models. The p-values are presented in parentheses. * The model is estimated with two binary variables that are 1 in Chihuahua and Tabasco and 0 in the other states, respectively, in order to capture outliers in the residuals. The estimated coefficients are 5.140 and -2.583, with p-values of 0.000 and 0.029, respectively.
Source: created by the authors.

Regarding the factors that explain the transmissions, the social variables that seek to measure respect for sanitary norms again stand out. Models 1 and 2 show a robust effect of the Culture of Legality Index (ICL), constructed by variables that evaluate how respondents report respect, knowledge, and compliance with laws and civic practices. The estimated coefficients are negative and statistically significant, suggesting that a greater culture of legality leads to a lower level of transmission and vice versa. Interestingly, these models also include variables that seek to measure people's behavior with regard to regulations: the higher the crime incidence (INDEL), the more transmissions were reported on that State. Interestingly, the population's perception of security (PINS) has the opposite sign to the expected, since logic would suggest that the higher the value of this variable, the more respect will exists for the rules in general and for health regulations in particular, so there should be fewer transmissions. Finally, as in the linear models, suffering from chronic degenerative diseases, in this case diabetes (TDIA), has contributed to more cases of COVID-19.

On the other hand, the spatial regression models are shown in Table 2. According to the implications of the LM statistics, first the spatial lag models and finally the spatial error models are shown. In general, the coefficients of substantive spatial autocorrelation (ρ), in the models of cumulative transmissions, and residual (λ), in those of prevalence rate are statistically significant at least at 5% (only in model 4, ρ is significant at 10% and in model 8 it is not). It is striking, however, that they are all negative, suggesting the non-existence of clusters in which states with high levels of transmission are neighbors of others with similar patterns and vice versa. The maps in Figures 3 and 4 suggest that this is indeed the case: there is a central corridor in which states with low levels of transmission are located adjacent to others on the Pacific Ocean coast and in central Mexico that have been severely affected by the pandemic. In sum, this evidence suggests that there is no interstate transmission.

Table 2
Factors influencing the magnitude and prevalence rate of COVID-19 transmission in Mexico (Spatial regression models)

Variables	Spatial lag. Dependent variable: CA						Spatial error Dependent variable: TC
	(1)*	(2)**	(3)	(4)	(8)	(9)	(13)
Intercept	26546.4	41785.6	-10680	-1987.14	-5757.87	-11877.4	-5.699
	(0.019)	(0.001)	(0.058)	(0.479)	(0.333)	(0.041)	(0.000)
DP			7.453	7.723	6.348	6.706	0.001
			(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
ICL	-92758.4	-122759					
	(0.001)	(0.000)					
CDEL			0.898	0.873			
			(0.000)	(0.000)			
PCORE					0.009		
					(0.000)		
DEN						0.071	
						(0.000)	
INDEL	0.295						
	(0.000)						
PINS		0.009					
		(0.010)					
TDIA	1838.04	1777.99	149892		1191.08	162722	

	(0.000)	(0.000)	(0.000)		(0.014)	(0.000)	
TOB							0.200
							(0.000)
THIPOBDIA				107859			
				(0.012)			
BL							7.873
							(0.025)
ρ	-0.293	-0.356	-0.275	-0.209	-0.206	-0.232	
	(0.008)	(0.007)	(0.009)	(0.078)	(0.108)	(0.035)	
λ							-0.437
							(0.039)
R^2	0.925	0.906	0.930	0.919	0.903	0.922	0.637
Normal	0.183	5.157	0.186	0.518	0.333	0.106	0.281
	(0.913)	(0.076)	(0.911)	(0.771)	(0.285)	(0.948)	(0.869)
Heteroscedasticity	0.713	2.159	2.677	1.596	1.817	4.859	***
	(0.950)	(0.827)	(0.444)	(0.660)	(0.611)	(0.182)	
Spatial dependence	5.841	6.663	6.490	3.413	2.875	4.349	
	(0.016)	(0.010)	(0.011)	(0.065)	(0.090)	(0.037)	

The coefficients in bold, italics, and regular type correspond to the variables identified in first, second, and third place in the different models. The p-values are presented in parentheses. * The model is estimated with a binary variable with a value of 1 in CDMX and 0 in the other states; the value of its coefficient is 40497.9, and the respective p-value is equal to 0.000. ** The model is estimated with two binary variables with a value of 1 in CDMX and State of Mexico and 0 in the other states; the values of the estimated coefficients are 60825.3 and 37200.9, with p-values both equal to 0.000. *** The test proposed by Kelejian and Prucha (2010) was used to obtain robust estimators with heteroscedasticity and autocorrelation.
Source: created by the authors.

As in the linear models, population density (DP) contributes positively to cumulative transmission (models 3, 4, 8 and 9) and to the prevalence rate (model 13). Similarly, variables associated with compliance with the norm are significant in explaining cumulative transmissions, such as the cost of crime (CDEL) in models 3 and 4; perception of corruption (PCORE) in model 8, and reported crime (DEN). As anticipated, these variables have positive and statistically significant effects. On the other hand, the diabetes rate (TDIA) enters in models 3, 8 and 9, while the multi-morbidity of diabetes, hypertension, and obesity (THIPOBDIA) enter in model 4.

In addition to DP, the prevalence rate (model 13) is explained by the obesity rate (TOB) and by the only economic variable that enters the specifications, the BL labor gap. The latter has a positive and significant effect, suggesting that the greater the proportion of the population that does not have a job, the more difficult it is for them to maintain social distance because they have to look for a means of subsistence, which exposes them more to transmission.

In general, the evidence reported in this paper is consistent with some of the findings of the early international literature on the subject. In particular, it is consistent with reports on the positive effects of population density on transmission, as the more people are concentrated in smaller spaces, the greater the physical proximity, which raises the likelihood of transmission (Kumar, 2015; Hoehl et al., 2020; Gilbert et al., 2020; and Padilla-Santamaría et al., 2020). Therefore, densely populated locations are more prone to more rapid viral spread (Gerritse, 2020; Lakshmi-Priyadarsini & Suresh, 2020). Similarly, it supports the approaches of Yanga et al. (2020) and Dietz and Santos-Burgoa (2020), who find that underlying diseases, including hypertension, diabetes, diseases of the respiratory system, cardiovascular diseases, and obesity may be risk factors for severe cases compared to non-severe cases. Finally, it gives empirical support to the conjecture that the lack of compliance with sanitary regulations dictated by the authorities causes a greater magnitude and intensity of COVID-19 transmission.

Conclusions

This document analyzes the impact of demographic, social, health, and economic factors on the cumulative levels of transmission and the prevalence rate of COVID-19 in Mexican states by means of an extreme bounds analysis developed in cross-sectional models in which the existence of spatial effects is also determined. Although several model specifications are reached, there are several results consistent with the findings in international literature that are worth highlighting.

First, the variables used as proxies to measure compliance with regulations in general and, indirectly, with health regulations issued by the authorities to contain the transmission of the coronavirus have significant effects on the magnitude and intensity of COVID-19 in the states. In particular, non-compliance with sanitary regulations appears to be an important factor in the spread of the virus. Second, SARS-CoV-2 transmission has been more frequent in states where the proportion of the population suffering from obesity or chronic degenerative diseases is higher due to the deterioration of the immune system that they cause. Third, the greater physical proximity between people in highly populated cities has been one of the central factors in transmission. Finally, it is important to mention that the explanatory power of the models is relatively high, suggesting that these factors and their combination can explain both the magnitude and intensity of transmission in the states of Mexico.

The main policy implications that can be derived directly from our results are related to the need to protect the population with obesity or hypertension and diabetes, since, as has been said, they are a highly vulnerable group. To maintain and reinforce sanitary measures, especially physical distance, the use of masks and personal hygiene, particularly in densely populated localities, is still necessary.

On the other hand, it is striking that some variables considered important by policymakers and public opinion were not relevant. For example, public policies dictated by the states, physical mobility, as well as different indicators of the population's living conditions (poverty, precarious employment and inequality) were not significant in our models. Nor do economic variables such as the proportion of activities highly sensitive to confinement (tourism and entertainment) or the size of the informal sector seem to be important in this process. Moreover, according to the results of the estimated spatial models, there is also no evidence of interstate coronavirus transmission.

The COVID-19 pandemic is a complex process, which is still ongoing and has multiple aspects that should be analyzed in the near future to understand better its nature and implications. The speed and timing of transmission and deaths are some aspects that could be analyzed with methods similar to those used here. The analysis, however, can be extended to look for possible non-linear relationships between the explanatory variables at more disaggregated levels, such as cities or neighborhoods. Some intuitively important variables may then be relevant.

References

Anselin, L. (1988a). Spatial Econometrics: Methods and Models. Kluwer Academic Publishers, Dordrecht, The Netherlands.

Anselin, L. (1988b). A test for spatial autocorrelation in seemingly unrelated regressions. Economics Letters, 28 (4), 335– 341. https://doi.org/10.1016/0165-1765(88)90009-2

Anselin, L. & Bera, A. K. (1998). Spatial Dependence in Linear Regression Models with an Introduction to Spatial Econometrics. In A. Ullah y D. Giles (eds.), Handbook of Applied Economic Statistics. (pp. 237-289). New York, NY. https:/doi.org/10.1201/9781482269901-36

Blanchflower, D. & Levin, A. (2015). Labor market slack and monetary policy. NBER Working Paper, No. 21094. Cambridge, MA. https://doi.org/10.3386/w21094

Breusch, T. & Pagan, A. (1980). The Lagrange multiplier and its applications to model specification in econometrics. Review of Economics Studies. 47 (1), 239-253. https://doi.org/ 10.2307/2297111

Burridge, P. (1980). On the Cliff-Ord test for spatial correlation. Journal of the Royal Statistical Society, 42 (1), 107–108. https://doi.org/10.1111/j.2517-6161.1980.tb01108.x

Callaway, E., Cyranoski, D., Mallapaty, S., Stoye, E. & Tollefson, J. (2020). The coronavirus pandemic in five powerful charts. Nature, 579, 482-483. http:/doi.org/10.1038/d41586-020-00758-2

Carrión, J. Ll., Di Paolo, A., García, A., López-Bazo, E., López-Tamayo, J., Matano, A., Moreno, R., Pons, E., Ramos, R., Royuela, V. & Suriñach, J. (2020). La propagación del COVID y su distribución geográfica en áreas pequeñas. Un análisis para Cataluña (I) Densidad y nivel socioeconómico. La Riqueza de las Naciones. Asociación Española de Ciencia Regional. (Disponible en: https://aecr.org/es/la-propagacion-del-covid-y-su-distribuciongeografica-en-areas-pequenas-un-analisis-para-cataluna-i-densidad-y-nivel-socioeconomico/) (Consultado: 25/08/2020).

Cartenì, A., Di Francesco, L. & Martino, M. (2020). How mobility habits influenced the spread of the COVID-19 pandemic: Results from the Italian case study. Science of the Total Environment, 741, 1-9. https://doi.org/10.1016/j.scitotenv.2020.140489

Díaz-Castrillón, F. J. & Toro-Montoya, A. I. (2020). SARS-CoV-2/COVID-19: el virus, la enfermedad y la pandemia. Medicina & laboratorio, 24 (3), 183-205. https://doi.org/10.36384/01232576.268

Dietz, W. & Santos-Burgoa C. (2020). Obesity and its implications for COVID-19 Mortality. Obesity, 28 (6). https://doi.org/10.1002/oby.22818

ECDC (2020). Data on the geographic distribution of COVID-19 cases worldwide. European Centre for Disease Prevention and Control. (Disponible en: https://www.ecdc.europa.eu/en/publications-data/download-todaysdata-geographic-distribution-covid-19-cases-worldwide)(Consultado: 03/08/2020).

Flournoy, M. & Morell, M. (2020, April 19). The 6 factors that determine coronavirus containment or devastation. The Washington Post. (Disponible en: https:/www.washingtonpost.com/opinios/2020/0419/6-factors-that-determinecoronovirus-containment-or-devastation/) (Consultado: 26/08/2020).

Gerritse, M. (2020). Cities and COVID-19 infections: Population density, transmission speeds and sheltering responses. Covid Economics vetted at real-time papers, Centre for Economic Policy Research (CEPR), 37, 1-26.

Gilbert, M., Pullano, G. & Pinotti, F. (2020). Preparedness and vulnerability of African Countries against importations of COVID-19: a modelling study. THE LANCET Regional Health, 395 (10227), 871-877. http/doi.org/10.1016/S0140-6736(20)30411-6

Gobierno de México (2020). Covid-19 México. (Disponible en: https://coronavirus.gob.mx/datos/#COMNac) (Consultado: 18/08/2020).

Gorbalenya, A. E., Barker, S.C., Baric, R.S., de Groot, R.J., Drosten, C., Gulyaeva, A.A., Haagmans, B.L., Lauber, C., Leontovich, A. M., Neuman, B. W., Penzar, D., Perlman, S., Poon, L. L. M., Samborskiy, D. V., Sidorov, I. A., Sola, I. & Ziebuhr, J. (2020). The species Severe acute respiratory syndrome–related coronavirus: classifying 2019-nCoV and naming it SARS-CoV-2. Nature Microbiology, 5, 536-544. https://doi.org/10.1038/s41564-020- 0695-z.

Gupta, R., Ghosh, A., Singh, A. & Misra, A. (2020). Clinical considerations for patients with diabetes in times of COVID19 epidemic. Diabetes & Metabolic Syndrome: Clinical Research & Reviews, 14 (3), 211-212. https://doi.org/10.1016/j.dsx.2020.03.002

Hoehl, S., Rabenau, H., Berger, A., Kortenbusch, M., Cinatl, J. & Bojka, S. (2020). Evidence of SARS-CoV-2 Infection in Returning Travelers from Wuhan, China. The New England Journal of Medicine, 328, 1278-1280. http//doi.org/ 10.1056/NEJMc2001899

Holshue, M.L., DeBolt, C., Lindguist, S., Lofy, K.H., Wiesman, J., Bruce, H. & Díaz, G. (2020). First case of 2019 novel coronavirus in the United States. New England Journal of Medicine, 382, 929-936. https://doi.org/10.1056/NEJMoa2001191

INEGI (2020). Banco de Información Económica. (Disponible en: https://www.inegi.org.mx/sistemas/bie/) (Consultado: 15/08/2020).

Jarque, C. M. & Bera, A. K. (1987). A test for normality of observations and regression residuals. International Statistical Review, 55 (2), 163-172. https://doi.org/10.2307/1403192.

Kelejian, H. H. & Prucha, I. R. (2010). Specification and estimation of spatial autoregressive models with autoregressive and heteroskedastic disturbances. Journal of Econometrics, 157 (1), 53-67. https://doi.org/10.1016/j.jeconom.2009.10.025

Kumar, M. (2015). A study of population distribution. International Journal of Latest Technology, Engineering Management and Applied Science, 4 (3), 24-28.

Lakshmi-Priyadarsini, S. & Suresh, M. (2020.) Factors influencing the epidemiological characteristics of pandemic COVID 19: A TISM approach. International Journal of Healthcare Managment, 13 (2), 89-98. http:/doi.org/10.1080/20479700.2020.1755804

Leamer, E. E. (1983). Let’s take the con out of econometrics. The American Economic Review, 73 (1), 31-43. (Disponible en: https://www.jstor.org/stable/1803924) (Consultado: 17/08/2020).

Levine, R. & Renelt, D. (1992). A sensitivity analysis of cross-country growth regressions. The American Economic Review, 82 (4), 942-963. (Disponible en: https://www.jstor.org/stable/2117352) (Consultado: 17/08/2020).

Li, B., Yang, J. Zhao, F., Zhi, L., Wang, X., Liu, L., Bi, Z. & Zhao, Y. (2020). Prevalence and impact of cardiovascular metabolic diseases on COVID-19 in China. Clinical Research in Cardiology, 109 (5), 531–538. http://doi.org/10.1007/s00392-020-01626-9

Ma, J. (2020). Coronavirus: China´s first confirmed Covid-19 case traced back to november 17. South China Morning Post. (Disponible en: https://www.scmp.com/news/china/society/article/3074991/coronavirus-chinas-firstconfirmed-covid-19-case-traced-back) (Consultado: 15/08/2020).

Mollalo, A., Vahedi, B., & Rivera, K. M. (2020). GIS-based spatial modeling of COVID19 incidence rate in the continental United States. Science of the Total Environment, 728, 1-8. https://doi.org/10.1016/j.scitotenv.2020.138884

Moran, P. A. P. (1948). The Interpretation of Statistical Maps. Journal of the Royal Statistical Society B (Methodological) 10 (2), 243-251. https://doi.org/10.1111/j.2517-6161.1948.tb00012.x

MUCD (2014). Primer Índice de Cultura de la Legalidad en México. (Disponible en: https://www.mucd.org.mx) (Consultado: 10/08/2020).

Observatorio COVID-19 (2020). Observatorio para la Contención del COVID-19 en América Latina. (Disponible en: www.ObsContenCovid.INFO) (Consultado: 08/08/2020).

WHO (2020a). Rolling updates on coronavirus disease (COVID-19). (Disponible en: https://www.who.int/emergencies/diseases/novel-coronavirus-2019/events-as-they-happen) (Consultado: 15/08/2020).

WHO (2020b). WHO Coronavirus (COVID-19) Dashboard. (Disponible en: https://covid19.who.int/table) (Consultado: 15/08/2020).

Onder, G., Rezza, G. & Brusaferro, S (2020). Case Fatality Rate and Characteristics of Patients Dying in Relation to COVID-19 in Italy. American Medical Association, 323 (18), 1775-1776. http://doi.org/10.1001/jama.2020.4683

Padilla-Santamaría, F., Maya-Franco, L. & Ferman-Cano, F. (2020). COVID-19 en México: Panorama Epidemiológico. Revista Cadena de Cerebros, 4 (1), 31-42. https://doi.org/10.5281/zenodo.3926806

PAHO/WHO (2020). Novel coronavirus (COVID-19). Epidemiological Update. Pan American Health Organization. (Disponible en: https://www.paho.org/en/documents/epidemiological-update-novel-coronavirus-covid-19-28- february-2020) (Consultado: 30/08/2020).

Papageorge, N., Zahn, M. V., Belot, M., van den Broek-Altenburg, E., Choi, S., Jamison, J. C. & Tripodi, E. (2020). Socio-demographic factors associated with self-protecting behavior during the Covid19 pandemic. NBER Working Paper No. 27378. https://doi.org/10.3386/w27378

PNUD (2015). Índice de Desarrollo Humano para las entidades federativas, México 2015. (Disponible en: https://www.mx.undp.org/content/mexico/es/home/library/poverty/indice-de-desarrollo-humano-para-lasentidades-federativas--mexi.html) (Consultado: 14/08/2020).

Reuters (2020). Coronavirus. Francia confirma dos casos, los primeros en Europa. El Universal. (Disponible en: https://www.eluniversal.com.mx/mundo/coronavirus-francia-confirma-dos-casos-los-primero-en-europa) (Consultado: 10/08/2020).

Xie, Z., Qin, Y., Li, Y., Shen, W., Zheng, Z. & Liu, S. (2020). Spatial and temporal differentiation of COVID19 epidemic spread in mainland China and its influencing factors. Science of the Total Environment, 744, 1-9. https://doi.org/10.1016/j.scitotenv.2020.140929

Yanga, J., Zheng, Y., Gou, X., Pu, K., Chen, Z., Guo, Q., Ji, R., Wang, H., Wang, Y. & Zhou, Y. (2020). Prevalence of comorbidities and its effects in patients infected with SARS-CoV-2: a systematic review and meta-analysis. International of Journal of Infectious Diseases, 94, 91-95. https://doi.org/10.1016/j.ijid.2020.03.017

Annex

Table A1
Definition of variables

Variables	Nomenclature	Units	Date	Data source
Cumulative COVID-19 cases by state	CA	Persons	July 31, 2020	https://coronavirus.gob.mx/datos/#COMNac
Prevalence rate*.	TC	Rate	July 31, 2020	https://coronavirus.gob.mx/datos/#COMNac
Population density	DP	Inhab/km^2	2020	CONAPO and INEGI
% of male population	TH	Percentage	2015	CONAPO (2020 estimates)
Culture of Legality Index**	ICL	Index	2014	MUCD https://www.mucd.org.mx
Costs of crime	CDEL	Pesos per person	2019	
Reported crimes	DEN	Crimes	2019	INEGI National Survey on Victimization and Perception of Security
Perception of safety	PINS	Percentage	2019	
Incidence of corruption***	INCOR		2019	
Perception of government corruption	PCORE	Percentage	2019	
Incidence of crime⁺	INDEL	Crimes /1000 inhabitants	2020	Ministry of Government
Percentage of population with obesity	TOB	Percentage	2018	
Percentage of the population previously diagnosed with hypertension	THIP TDIA	Percentage	2018	

http://dx.doi.org/10.22201/fca.24488410e.2020.3127

Percentage of the population previously diagnosed with diabetes		Percentage	2018	National Health and Nutrition Survey (ENSANUT) 2018; Ministry of Health, the National Institute of Public Health (INSP); and the National Institute of Statistics and Geography (INEGI).
Percentage of population with diagnosed hypertension and obesity	THIPOB	Percentage	2018	
Percentage of the population diagnosed with hypertension, obesity, and diabetes	THIPOBDIA	Percentage	2018	
Labor gap#.	BL		2020.I	INEGI https://www.inegi.org.mx/sistemas/Infoenoe/Default

* $TC = \frac{CA}{Población} \times 1000$. ** Provides data on whether the surveyed persons declare that they respect, are aware of, and comply with the laws and civic practices. *** Incidence of corruption per 100 000 inhabitants. + Common crimes per thousand inhabitants. # Percentage of the potential economically active population.
Source: created by the authors.

Table A2
Diagnostic tests for spatial dependence

Tests	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)
Moran's I	-0.655	-1.051	0.275	-0.300	-0.064	0.519	1.956	-0.464	-0.545	-0.135	-0.015	1.391	-1.436	1.146
	(0.512)	(0.293)	(0.783)	(0.764)	(0.949)	(0.603)	(0.050)	(0.642)	(0.586)	(0.892)	(0.988)	(0.164)	(0.151)	(0.252)
LM lag	4.460	5.651	5.165	3.381	2.668	1.766	1.009	2.880	3.749	1.940	2.304	0.219	2.630	0.014
	(0.034)	(0.017)	(0.023)	(0.066)	(0.102)	(0.184)	(0.315)	(0.090)	(0.053)	(0.164)	(0.129)	(0.640)	(0.105)	(0.907)
LM error	0.787	1.486	0.000	0.306	0.115	0.012	1.895	0.463	0.566	0.153	0.065	0.763	3.212	0.365
	(0.374)	(0.223)	(1.000)	(0.580)	(0.734)	(0.912)	(0.169)	(0.496)	(0.452)	(0.695)	(0.798)	(0.382)	(0.073)	(0.545)
Robust LM lag	3.699	4.218	6.458	3.251	2.869	2.561	3.852	2.489	3.211	1.938	2.638	1.517	0.106	0.327
	(0.054)	(0.040)	(0.011)	(0.071)	(0.090)	(0.110)	(0.050)	(0.115)	(0.073)	(0.164)	(0.104)	(0.218)	(0.745)	(0.567)
Robust LM error	0.027	0.053	1.293	0.176	0.317	0.808	4.738	0.073	0.028	0.151	0.400	2.061	0.687	0.678
	(0.869)	(0.818)	(0.255)	(0.675)	(0.573)	(0.369)	(0.030)	(0.788)	(0.868)	(0.697)	(0.527)	(0.151)	(0.407)	(0.410)
LM SARMA	4.487	5.704	6.458	3.557	2.985	2.573	5.747	2.952	3.777	2.091	2.704	2.280	3.317	0.692
	(0.106)	(0.057)	(0.040)	(0.169)	(0.225)	(0.276)	(0.057)	(0.229)	(0.151)	(0.351)	(0.259)	(0.320)	(0.190)	(0.708)

Source: created by the authors.