Since January 2020 Elsevier has created a COVID-19 resource centre with free information in English and Mandarin on the novel coronavirus COVID-19. The COVID-19 resource centre is hosted on Elsevier Connect, the company's public news and information website.

Elsevier hereby grants permission to make all its COVID-19-related research that is available on the COVID-19 resource centre - including this research content - immediately available in PubMed Central and other publicly funded repositories, such as the WHO COVID database with rights for unrestricted research re-use and analyses in any form or by any means with acknowledgement of the original source. These permissions are granted for free by Elsevier for as long as the COVID-19 resource centre remains active.
Research article

The impact of built and socio-economic environment factors on Covid-19 transmission at the ZIP-code level in Florida

Emre Tepe *

Department of Urban and Regional Planning, University of Florida, 444 Architectural Building P.O. Box 115706, Gainesville, FL, 32611, USA

A R T I C L E   I N F O

Keywords:
Covid-19 outbreak
Spatial regression modeling
Built environment
Socio-economic environment

A B S T R A C T

Most studies have explored the Covid-19 outbreak by mainly focusing on restrictive public policies, human health, and behaviors at the macro level. However, the impacts of built and socio-economic environments, accounting for spatial effects on the spread at the local levels, have not been thoroughly studied. In this study, the relationships between the spatial spread of the virus and various indicators of the built and socio-economic environments are investigated, using Florida ZIP-code data on accumulated cases before large-scale vaccination campaigns began in 2021. Spatial regression models are used to account for the spatial dependencies and interactions that are core factors in Covid-19 spread. This study reveals both the spillover dynamics of the coronavirus spread at the ZIP code level and the existence of spatial dependencies among the unobserved variables represented by the error term. In addition, the findings show a positive association between the expected number of Covid-19 cases and specific land uses, such as education facilities and retail densities. Finally, the study highlights critical socio-economic characteristics causing a substantial increase in Covid-19 spread. Such results could help policymakers, public health experts, and urban planners design strategies to mitigate the spread of future Covid-19-like diseases.

1. Introduction

The recent Covid-19 outbreak has affected human life at unprecedented levels since December 2019. The World Health Organization (WHO) declared this outbreak a pandemic on March 11, 2020. Many non-clinical researchers began investigating the effects of this outbreak in a broader geographical spectrum, while health scientists focused on the clinical and public safety aspects. Because of various unknowns in this disease, the WHO and the U.S. Centers for Disease Control and Prevention (CDC) warned the public about the highly contagious virus and suggested “social distancing” measures. Dense urban areas became hot spots for the illness because more people close to each other imply more virus transmission. The factors leading to the community spread of Covid-19 in urban and rural areas are related to their demographic, socio-economic, and physical characteristics and must be investigated.

In the early period of the pandemic, dense cities were faced with dramatic increases in positive cases. In response to this public health emergency, unconventional measures were taken to slow down the disease’s spread, such as asking people to continue their daily work and education virtually from their homes. Sudden changes in everyday life routines substantially reduced the demand for transportation. Google mobility data reports for the U.S. between February 15 and December 31, 2020, show that fewer people visited retail stores & recreational activities (−9.8%), transit stations (−12.16%), and workplaces (−23.31%), while more people visited groceries & pharmacies (1.31%), parks (28.64%), and residential locations (7.76%) as compared to the previous year (Google, 2020).

The pandemic also impacted various aspects of the communities, such as mental health, environmental quality, energy consumption, and economic development. Drastic measurements to reduce the spread have worsened mental health conditions worldwide (Aqed et al., 2022a, b). Limiting individuals’ mobility during lockdowns has forced shoppers to adopt online purchasing (al Halbusi et al., 2022). The pandemic conditions also changed the structure of energy demand during the lockdowns. Energy demand from residential areas has increased because of remote working, while fewer people commute and use workspaces. As such changes, environmental quality has been slightly improved in certain regions. Economic activities might also have been affected by such changes in environmental conditions due to the correlation between ecological indicators and economic development (Shah et al., 2020).

* Department of Urban and Regional Planning, University of Florida, 1480 Inner Road, Gainesville, FL 32611, USA.

E-mail address: emretpe@ufl.edu.
The pandemic conditions disrupted the existing human behaviors and urban activities. Lockdown measures have widened the inequalities in society and accelerated the transition from working and shopping in-person to virtual options for some groups of people, while others were left behind. In the post-pandemic period, cities are expected to adapt to these changes, and such adaptations might provide opportunities to implement approaches to achieve sustainable, inclusive, and equitable cities.

The relationship between population distribution and coronavirus spread must be investigated at a fine spatial scale to understand disease spread dynamics better. In the U.S., cities with dense populations and international airports were severely impacted in the early days. Shortly thereafter, rural areas started experiencing a rapid increase in positive cases and deaths. The share of fatalities in small towns and rural areas among the initial 100,000 deaths was 20%, but this share grew to 50% among the next 100,000 deaths (McMinn, 2020). Understanding the relationships between Covid-19 spread and the density structure and the socio-economic environment of urban and rural areas is necessary for proper policy interventions. However, the relationships between the spread and several built-environment and socio-economic factors have not been fully investigated by accounting for spatial dependencies at the local geographical levels.

This study investigates the spread of Covid-19 by considering various indicators of the built and socio-economic environments in Florida’s urban and rural areas before large-scale vaccination campaigns began. Covid-19 data was obtained from the Florida Department of Health (FDOH) daily reports, which provide information about the numbers of positive cases and deaths at the county, city, and ZIP code levels (Florida Department of Health, 2020). Socio-economic indicators at the block group level were obtained from the American Community Survey (ACS). Indicators of the built environment at the parcel level were derived from the Florida Statewide Parcel Database of the University of Florida (UF) GeoPlan Centers. Since these data sets are provided at different spatial levels, data at the parcel and block group levels were aggregated to the ZIP code level. The relationship between the numbers of positive cases at the ZIP code level and built and socio-economic environment indicators is explored using spatial regression models. Given the spatial nature of the spread of this disease, spatial dependency in the number of positive cases is very likely.

The rest of the paper is organized as follows. Section 2 presents a literature review. The modeling method is described in Section 3. The study area and the data are described in Section 4. The statistical analysis and a discussion of model results are presented in Section 5. Section 6 discusses policy recommendations. Section 7 concludes and presents areas for further research.

2. Literature

This literature review focuses on (1) research investigating the demographic, socio-economic, built & energy environments, and natural environmental factors in Covid-19 spread, (2) methods for statistical modeling of this spread, and (3) proposed solutions for providing sustainable, healthy, safe, and inclusive conditions in urban places.

Several studies have investigated the relationship between Covid-19 spread dynamics and demographic and socio-economic characteristics. Sharifi and Khavarian-Garmsir (2020) review the pandemic’s impacts on cities, showing that environmental factors and income inequality play a significant role in spreading Covid-19. Singh and Adhikari (2020) provide insight into the transmission patterns of Covid-19 among different age groups in India, China, and Italy, demonstrating that young people are more susceptible to the disease through higher transmission risks, while mortality rates are higher among the elderly population. Mollalo et al. (2020) examine the relationship between Covid-19 spread at the U.S. county level and median household income, income inequality, and percentages of nurses and black women in the county population over the period between January 22, 2020 and April 9, 2020. Their results indicate that median household income and income inequality are positively associated with the spread. Plümper and Neumayer (2020) conclude that the second pandemic wave in Germany disproportionately affected low-income neighborhoods. Drastic measures to slow down the spread exacerbated social inequality globally (Global reflections on covid-19 and urban inequalities, 2021). Sannigrahi et al. (2020) investigate the relationship between Covid-19 cases and related deaths and income and poverty factors in European countries, controlling for spatial dependencies. Li et al. (2021a) found a higher mortality risk for infected members of low-income communities in Brazil. Paez et al. (2021) show that population density and the share of people over 65 negatively impact the disease spread, while GDP per capita positively impacts it. All these model results show that poverty and low income are positively associated with the number of positive cases and deaths.

Other studies have investigated the impacts of population density and built environment characteristics on Covid-19 spread. Hamidi et al. (2020) compare Covid-19 infection and mortality rates for multiple U.S. counties, focusing on their density structure, and indicate that dense metropolitan areas have lower mortality rates because of superior healthcare systems and strict social distancing measures. Hamidi and Zandiastashbhar (2021) also show the importance of parks in dense urban areas during the pandemic, with residents using these parks with higher frequency. Andersen et al. (2021) analyze the relationship between Covid-19 spread in the U.S. and a set of socio-economic and physical indicators, accounting for spatial dependencies. Their findings show that denser urban areas increase the number of positive cases compared with less dense urban and rural areas. Bharda et al. (2021) report a moderate correlation between Covid-19 spread and district population density in India. Florida et al. (2021) predict potential transformations of cities in the post-pandemic period in three main ways: (1) the transition from in-person working and shopping to online will accelerate; (2) cities will improve cultural and civic activities while shopping and working places will be reduced; and (3) the economic dominance of large cities will remain while most mid-size cities and rural areas located far from the dynamic centers will shrink in size.

Recent studies have examined the impacts of environmental factors on Covid-19 spread. Paez et al. (2021) consider the effects of temperature, humidity, and sunshine on disease spread at the Spain province level, based on 30-day cross-sectional data, and show the statistically significant impacts of these factors while controlling for socio-economic factors (GDP per capita, share of population over 65, and population density). Moore et al. (2021) analyze the spatial pattern of 10,000 severe Covid-19 cases in the U.K. and their relationship with built-environment, locational, and air quality factors. They show that Covid-19 cases tend to be clustered around bars & restaurants, open spaces, and areas with high nitrogen oxide concentrations. Using a Random Forest model, Zhang et al. (2021b) show that weather and socio-economic factors are associated with Covid-19 cases at the county level. The results show that population, population density, and a social distance index (a combination of six mobility metrics) have the highest impacts on the number of Covid-19 cases. At the same time, daily temperature, humidity, shortwave radiation, precipitation, and wind speed are also essential factors. Magazzino et al. (2021) demonstrate, with a Deep Machine Learning method, that air pollution (PM2.5 and NO2) contributes to the deaths caused by Covid-19 in New York. A summary of the salient results from the studies investigating relationships between Covid-19 spread and various factors is presented in Table S1 in the supplementary document. Besides the above demographic, socio-economic, built environment, and environmental factors, the impacts of social distancing and lockdown measures have also been investigated. Sarkar et al. (2020) introduce a mathematical model to predict the disease spread and evaluate the possible outcomes of imposed social distancing and lockdown policies. Lancastle (2020) compares the effectiveness of social distancing measures in various countries and shows the positive impacts of such
measures in reducing Covid-19 spread. Lockdowns were implemented in many countries as a prevention method to reduce the spread of Covid-19. However, restricting social interactions causes an increase in mental health problems. According to a study conducted with university students in Pakistan, both full and partial lockdowns negatively impacted students’ mental health and quality of life. However, partial lockdown measures have comparatively lesser impacts (Aqeel et al., 2022a,b).

Pandemic conditions also impact energy consumption and economic development. The change in consumer demand and productions are likely to impact low-income households more than other groups (Kan-siime et al., 2021). Prevention measures during the pandemic have caused a significant decline in households’ food and energy consumption (Geng et al., 2022). Ge et al. (2022) investigate the contributions of female entrepreneurs to family income using innovative technologies to overcome financial challenges caused by the pandemic. Changes in Foreign Direct Investment (FDI) dynamics due to pandemic conditions may also affect sustainable environments. Zhang et al. (2021a) highlight the causal relationship between FDI and well-being and a sustainable environment with outward FDI. Zhang et al. (2022) highlight the importance of investment in green energy technologies to maintain sustainable growth while the pandemic adversely affects the transition to green energy use.

Various statistical methods have been used to model the spread of Covid-19 and other infectious diseases. Hu et al. (2013) show the importance of non-linear population density functions in such models, as previous studies indicate that linear functions may not adequately represent infectious disease dynamics across various populations. Several researchers have also explored the spatiotemporal patterns of Covid-19 spread (Franch-Pardo et al., 2020). Desjardins et al. (2020) study spatial clusters of Covid-19 cases at the county level in the U.S., using a prospective space-time scan statistic method, identifying 26 emerging space-time clusters between January 22 and March 27, 2020. Wang et al. (2010) introduce an agent-based approach to model influenza outbreaks in urban areas using Poisson space-time scan statistics. Yang et al. (2020) discuss spatiotemporal methods to analyze Covid-19 spread as a function of socio-economic, human mobility, and environmental factors. They underscore the importance of accounting for spatial and temporal dependencies in the modeling of Covid-19 spread. Araujo-Cardot et al. (2021) investigate the spatial socio-economic determinants of Covid-19 spread at the neighborhood level during Barcelona’s first and second waves, highlighting the importance of econometric and spatial analysis to understand the spatio-temporal dynamics of the spread. They show that population density, the share of school facilities, and the share of the young population in a given neighborhood are positively associated with the number of positive cases, with income having the opposite effect. Other studies have used temporal datasets to model coronavirus cases in different countries (He et al., 2020). The spatiotemporal Poisson distribution has been considered a suitable approach to model the spread of malaria cases in the Brazilian Amazon Forest between 1999 and 2008.

Spatial dependencies at different spatial resolutions have also been investigated. Using spatial econometrics methods, Ehlerl (2021) explores the relationships between socio-economic, demographic, and health-related variables and Covid-19-related cases and deaths in Germany. In this research, spatial spillover effects among districts are captured using Spatial Autoregressive (SAR), Spatial Error Model (SEM), and Spatial Two-stage least squares (2SLS) models. Similarly, Huang et al. (2020) and Li et al. (2021b) confirm the spatially heterogeneous relationships between Covid-19 spread and built environment characteristics in Hong Kong and China. Frank and Wali (2021) explore the benefits of physical activity in reducing the severity of Covid-19 infections and mortality risks and focus on built-environment characteristics, such as population density, design, and accessibility. Their findings indicate that highly aggregated data can be misleading due to significant spatial heterogeneity. Long and Ren (2022) highlight the importance of accounting for fine-scale spatial and temporal components in investigating the relationship between the Covid-19 spread and individuals’ mobility patterns and socio-economic conditions, indicating that these relationships vary over time. Yang et al. (2021) examine the first wave of Covid-19 spread in New York City at the ZIP code level, incorporating multiple socio-economic, locational, and built-environment factors. Their study is the most detailed geographical analysis of Covid-19 cases. However, they do not account for spatial dependencies.

Several researchers have also explored the spatiotemporal patterns of Covid-19 spread (Avery et al., 2020):}

\[ \frac{\partial S(t)}{\partial t} = -S(t)I(t)R_0 \text{ and } \frac{\partial I(t)}{\partial t} = S(t)I(t)R_0 - \gamma I(t) \]

In Eq. (1), \( R_0 \) is the transmission rate, and \( \gamma \) is the recovery rate. A person moves from one state to another by (1) contracting a virus from an infected individual or (2) recovering from the disease. At the initial stage of the disease spread, the number of cases increases slowly, but then exponentially due to the interactions of susceptible individuals
with infected people. Over time, the growth rate of infected people declines because of fewer susceptible people. The critical factor in the spread is \( R_0 \), which depends upon how spatially close susceptible and infected people are, and also upon the undertaken mitigation measures (masking, social distancing). However, this aggregate model does not explicitly account for space. As the virus can only be transmitted from one person to another if a person is in close contact with someone infected by Covid-19, the number of positive cases in one geographical unit can affect the number of cases in the surrounding areas and vice versa.

3.3. Spatial regression modeling

OLS regression modeling cannot handle spatial dependency in the transmission dynamics empirically. However, spatial regression models can account for such dependencies. Spatial effects like spatial autocorrelations (SA) and spatial heterogeneity (SH) in OLS models may cause biased estimation of error variance, where the estimated OLS parameters might remain unbiased, and the significance tests become misleading (Anselin and Griffith, 2005; Dale and Fortin, 2002). The Spatial Autoregressive Model (SAR) by Cliff and Ord (1988), and the General Spatial Model (GSM), also known as Spatial Autoregressive with additional Autoregressive error structure (SARAR) by (Kelejian and Prucha, 1998), are to be used to account for spatial dependency in the data.

Equation (2) represents a first-order SAR model (also named Spatial Lag Model), where the spatial autoregressive coefficient \( \rho \) measures the SA among the dependent variable (Anselin, 1989; Cliff and Ord, 1980). The SAR model is widely used to account for spatial dependency in the data (Anselin, 2001; Zhang et al., 2009).

\[
Y = X\beta + \rho WY + \varepsilon
\]

where \( Y \) is the vector of the dependent variable; \( X \) is the matrix of the explanatory variables; \( \beta \) is the vector of regression coefficients; \( W \) is the row-standardized spatial weight matrix based on a predefined spatial neighborhood structure; \( \varepsilon \) is the residuals vector.

It is, however, possible that there is further SA across the error term \( \varepsilon \), which represents the unobserved variables impacting \( Y \). The first-order GSM accounts for spatial dependencies in both the dependent variable and the error term, with

\[
Y = X\beta + \rho WY + u & u = \lambda Wu + \varepsilon
\]

where \( \lambda \) is the lag coefficient for the error term \( u \), and \( \varepsilon \) is an error term satisfying OLS conditions. Components in Eq. (3) can be combined as follows:

\[
Y = (I - \rho W)^{-1} X\beta + (I - \rho W)^{-1}(I - \lambda W)^{-1} \varepsilon
\]

Equation (4) clearly demonstrates the non-linear nature of the model, with non-linear interactions between the variables \( X \), \( \rho \), and \( \lambda \). In this research, a standard OLS model is first estimated, and the existence of SA in the error term is tested with the Moran I’s test. If SA is detected, then a SAR model is next estimated. A GSM model is estimated if SA is still detected in residual (Farzadfar et al., 2022; Schmidt et al., 2022).

4. Study area and data

As of December 23, 2020, the total number of positive cases in Florida was slightly over 1.2 million, and the total death toll was 20,874. The number of positive cases began rapidly increasing after the Federal Government’s guidelines for social distancing expired on April 30, 2020. The rate of increase in positive cases slowed down by mid-July 2020, but began increasing after September 2020 across the state. The numbers of identified positive cases per 100K population in the counties, including the cities of Jacksonville, Miami, and Tampa, were higher than in areas with fewer people. However, these conditions changed recently, with some counties composed mainly of rural areas having higher positive rates per capita. A similar pattern has also been observed for death rates.

Fig. 1 illustrates the distributions of positive cases in ZIP code areas as of December 23, 2020, with log-transformed positive cases. The distribution of positive cases is heavily right-skewed, with 48 ZIP code areas (out of 1102) reporting no positive cases. These 48 ZIP code areas are mainly uninhabited areas, such as forests and lakes, and are excluded from further analysis. Because of this skewed distribution, the number of positive cases is log-transformed (see Fig. S1 in the supplementary).

Built environment indicators, including land use, year-built, and most recent sale records, are derived from the 2019 statewide parcel database. The built environment composition of each ZIP code area is also derived from this database. Table 1 presents the variables and their sources. The densities of residential, commercial, and non-residential activities show where people gather. ZIP-code demographic and socio-economic compositions are derived from the 2019–2015 ACS 5-year estimates. ACS data provides comprehensive information at the block group level, and these data are then aggregated to the ZIP code level. Demographic and socio-economic indicators are essential for understanding the human aspects of Covid-19 spread. The mobility of Floridians during the pandemic is also included in this research. ACS provides data on workforce mode choices. Finally, Google offers mobility data for Florida at the county level. Pre-pandemic travel behavior provides clues about the current transportation demands.

Besides the above variables, the mask rule in a county is used as a control variable. Counties in Florida have implemented different approaches to enforcing mask rules: 24 out of 67 counties have issued some mask rules to keep their residents safe (Ogles, 2020). All these data are further explored with descriptive statistics (see Table S2 in the supplementary document).

5. Results

5.1. Spatial autocorrelation of Covid-19 infections

Spatial dependency in the number of positive cases at the ZIP code level is first investigated using the Global Moran’s I test. Fig. 2 displays the number of positive cases per 100K population at the ZIP code level in Florida on December 23, 2020. There is no firm rule for conceptualizing spatial relationships in the spatial econometrics and statistics literature. As a standard approach, various rules are tested to find the one that best fits the data. Due to the nature of this pandemic, the disease is likely to spread within close proximity. K-nearest neighbors and fixed distance-band rules are used to test various spatial conceptualizations (see Fig. S2 in the supplementary document). The inverse-distance-weight (IDW) method also accounts for potential distance-related spatial spillover effects. Row-standardization method is applied to all spatial weight matrices. Fig. S3 in the supplementary document displays the Moran’s I statistics with 99% confidence intervals based on the K-nearest neighbors rule. All test results are statistically significant at the 1% level. The SA index declines as the kth degree increases when the spatial effects are assumed fixed among the predefined number of K-nearest neighbors. Fig. S4 in the supplementary document illustrates SA tests using the fixed-distance-band rule. Moran’s I indices for both the fixed effect and the IDW approach decrease as the fixed distance increases. The K-nearest neighbors rule better captures the spatial dependency, with a comparatively smaller standard deviation, and is therefore used in the following analyses.

5.2. Exploratory regressions

We tested multiple model specifications to identify the best-fit SAR model using an IDW-applied spatial weight matrix based on 20 nearest neighbors. The best model is obtained when the p-value of the Moran’s I test for SA in the dependent variable reaches the smallest level. All combinations of independent variables were evaluated using SAR
(65,535 model configurations were estimated). AIC is a main goodness-of-fit measure for spatial regression models. Models’ AIC values range between 2938 and 4291, while the median is 3353 and the mean is 3243. Because the difference between the minimum and 1st quartile AIC values is 47, the variables included in the 1st quartile of tested models are further investigated to identify the best-fit model.

Table S3 in the supplementary document presents additional statistics for all the potential explanatory variables across the 65,535 models. For each variable, the following were computed: (1) percentage of time the sign is positive; (2) percentage of time the sign is negative; (3) percentage of time the variable is statistically significant at the 10% significance level. This analysis provides information on the stability of the estimated coefficients’ sign and significance. Except for the RESIDENT and MASKRULE variables, all the independent variables were statistically significant in more than half of the tested models. When the sign consistency of the estimated variables is analyzed, the construction age and single-family residential variables can be considered unstable, with almost equal numbers of positive and negative coefficients.

Pairwise Pearson’s correlation coefficients were also computed between pairs of explanatory variables. High correlations may cause multicollinearity issues. These coefficients are illustrated in Fig. S5 in the supplementary document. The average household size correlates with labor, student, and driving-alone ratios. Driving alone is also linearly associated with median age, labor ratio, and household size. These strong correlations will be eliminated from the final regression models to avoid multicollinearity issues.

Potential outliers in variables are also further investigated. Excess Kurtosis and Skewness statistics are presented in Table S2 in the supplementary document to provide more information about the distribution of these variables. Excess Kurtosis values give a clearer picture of the right and left tails’ size. Negative Excess Kurtosis values signal a higher probability of obtaining outliers (Yu et al., 2022). Boxplots are also presented in Fig. S6 in the supplementary document, where potential outliers are highlighted with circles. The CASES variable has many potential outliers. However, a logarithmic transformation significantly reduces the number of potential outliers. In the following statistical analysis, the log-transformed version is used. Among the independent variables, RHR, RETAIL, INDUSTRY, EDUCATION, INCOME, DROVEALONE, and PUBLICTRANSIT have potential outliers.
5.3. Selected models

After identifying the best out of the 65,535 models based on the AIC, the variables highly correlated with other variables were removed from this model. The square of the income variable was also added to the best model to test the non-linearity of the median household income effect. Table 2 presents the estimation results for the OLS, SAR, and GSM models. A significant SA pattern in the OLS residuals can cause biased estimators and inefficient test statistics (Anselin, 1988a). Also, spatial heterogeneity in the data causes nonstationary spatial relationships (Stewart Fotheringham, Brunsdon and Charlton, 2002). Similarly, the impacts of nonstationary conditions on the model results are also discussed in studies that estimate temporal models (Abbasi et al., 2021a,b; Raza Abbasi et al., 2021). The SAR approach can account for spatial heterogeneity through the dependent variable and reduce the SA level in the residuals. However, it is still possible to observe a significant level of SA in the SAR residuals. The GSM accounts for spatial dependence among the unobserved variables represented by SA in the SAR residuals. The GSM is superior to the SAR in reducing the SA level in the residuals and provides an identically independent distribution (Zhang et al., 2009). The standard errors are presented in brackets. The spatial lag coefficients (\( \rho \)) in the SAR and GSM models are highly significant, underscoring the importance of controlling for spatial dependency.

The Moran’s I test for the OLS residuals is highly significant, which indicates SA and supports the need for spatial regression. The Lagrange Multiplier (LM) test shows that SA is still present in the SAR model (LM = 31.36, p < 0.01) (Anselin, 1988b). The GSM is superior to the SAR in terms of AIC, but it still retains some degree of SA (LM = 48.75, p < 0.01). The GSM model is further investigated in the remainder of this paper. The positive spatial lag coefficient (\( \rho \)) of 0.14 indicates that an increase in positive cases in a given ZIP code area leads to a rise in the number of positive cases in neighboring ZIP codes. Thus, the model successfully captures the spillover dynamics of the virus spread. Also, the spatial autoregressive coefficient (\( \lambda \)) of 0.27 indicates that there are spatial dependencies among the unobserved variables represented by the error term.

Table 2
Results of regression models.

| Variable         | OLS     | SAR     | GSM     |
|------------------|---------|---------|---------|
| Intercept        | 2.6978  | 1.2413  | 1.9152  |
| INCOME           |         |         |         |
| POPULATION       | 0.019***| 0.004** | 0.023***|
| EDUCATION        | 0.2254  | 0.1802  | 0.1935  |
| RETAIL           | 0.0208  | 0.0183  | 0.0177  |
| DROVEALONE       | 1.3669  | 1.6278  | 1.6068  |
| SFR              |         |         |         |
| STUDENT          | 1.8042  | 1.6253  | 1.6244  |
| INCOME\(^2\)     | -0.0118 | -0.0111 | -0.0113 |
| DROVEALONE\(^2\) | 1.3669  | 1.6278  | 1.6068  |
| \( \lambda \)    | -       | -       | 0.2747  |
| \( \rho \)       | -       | 0.2067  | 0.1417  |

The standard errors are presented in brackets. The spatial lag coefficients (\( \rho \)) in the SAR and GSM models are highly significant, underscoring the importance of controlling for spatial dependency. The Moran’s I test for the OLS residuals is highly significant, which indicates SA and supports the need for spatial regression. The Lagrange Multiplier (LM) test shows that SA is still present in the SAR model (LM = 31.36, p < 0.01) (Anselin, 1988b). The GSM is superior to the SAR in reducing the SA level in the residuals and provides an identically independent distribution (Zhang et al., 2009). The standard errors are presented in brackets. The spatial lag coefficients (\( \rho \)) in the SAR and GSM models are highly significant, underscoring the importance of controlling for spatial dependency.

The Moran’s I test for the OLS residuals is highly significant, which indicates SA and supports the need for spatial regression. The Lagrange Multiplier (LM) test shows that SA is still present in the SAR model (LM = 31.36, p < 0.01) (Anselin, 1988b). The GSM is superior to the SAR in reducing the SA level in the residuals and provides an identically independent distribution (Zhang et al., 2009). The standard errors are presented in brackets. The spatial lag coefficients (\( \rho \)) in the SAR and GSM models are highly significant, underscoring the importance of controlling for spatial dependency.

The Moran’s I test for the OLS residuals is highly significant, which indicates SA and supports the need for spatial regression. The Lagrange Multiplier (LM) test shows that SA is still present in the SAR model (LM = 31.36, p < 0.01) (Anselin, 1988b). The GSM is superior to the SAR in reducing the SA level in the residuals and provides an identically independent distribution (Zhang et al., 2009). The standard errors are presented in brackets. The spatial lag coefficients (\( \rho \)) in the SAR and GSM models are highly significant, underscoring the importance of controlling for spatial dependency.

5.4. Direct and indirect effects

Because of the spatial lag term, the GSM must be considered as a non-linear model. Therefore, the estimated coefficients cannot be interpreted as marginal effects. When the spatial lag of the dependent variable is introduced into a model, it is necessary to compute direct and indirect effects. The GSM can be re-written by isolating the dependent variable (see Eq. (4)). The marginal effects can be obtained by using the biased derivative of the dependent variable with respect to a given explanatory variable. Because the dependent variable has been log-transformed, this derivative is:

\[
\frac{\partial Y}{\partial X_i} = (I - \rho W)^{-1} \beta_i, \quad Y
\]

where \( X_i \) is the \( k \)th column of the covariate matrix \( X \) and \( \beta_k \) is the \( k \)th coefficient.

Using Eq. (5), average direct, total, and indirect effects can be computed (LeSage and Pace, 2009; Park et al., 2021). Equations (6)–(8) represent the average direct (ADE), total (ATE), and indirect (ANE) effects:

\[
ADE_i = \bar{y}_i (I - \rho W)^{-1} \beta_i, \quad \frac{\partial Y}{\partial N_i} = \frac{\partial Y}{\partial X_i} \frac{\partial X_i}{\partial N_i}
\]

where \( N_i \) is the number of observations; \( I \) is a \( N \times N \) identity matrix; \( \bar{y}_i \) is a \( N \times 1 \) vector of ones; \( \bar{y}_i \) is the transpose of \( \bar{y}_i \) and \( 1_N \) is the trace operator. Table 3 presents the average elasticities of the direct and indirect effects of the explanatory variables of the GSM, with their significance levels.

As single-family residential density (SFR) increases by 1%, the expected number of cases in the same area decreases by 0.003%. We can expect this result when we consider human-interactions in single-family residential areas. Residents tend to stay home and reduce their interactions with other people. Therefore, such conditions limit the spread of Covid-19. Commercial land-use density (RETAIL) is another critical factor in spreading the disease. A 1% increase in commercial density results in a 0.012% increase in the expected number of positive cases in the same area. The same increase in RETAIL in neighboring ZIP codes raises the expected positive cases by 0.002%. Because of higher face-to-face interactions in commercial spaces, such a positive relationship is expected. The density of education facilities (EDUCATION) is also essential in community spread. A 1% increase in the density of education facilities increases the expected number of positive cases by 30% in the same area, while the same increase in surrounding ZIP codes increases the expected number of positive cases by 5.4%. Considering the activities in education facilities, it is understandable that people close to them have a higher risk of contracting the virus due to higher interactions through school-related activities.

Demographic and socio-economic indicators are incorporated into the model to gather a complete understanding of the Covid-19 spread.

Table 3
Average elasticities of the direct, indirect, and total impacts based on the GSM model.

| Variable      | Direct  | Indirect | Total  |
|---------------|---------|----------|--------|
| SFR           | -0.003**| -0.001   | -0.004*|
| RETAIL        | 0.012***| 0.002*   | 0.014***|
| EDUCATION     | 0.370***| 5.444*   | 35.723**|
| POPULATION    | 0.019***| 0.004**  | 0.023***|
| STUDENT       | 0.311***| 0.052*   | 0.363***|
| INCOME        | 4.123***| 0.708*   | 4.831***|
| INCOME\(^2\) | -0.060**| -0.010*  | -0.070**|
| DROVEALONE    | 58.198***| 9.387*   | 67.585***|

Significance levels: % *** 0.1% ** 1% * 5% . 10%.
Population is positively associated with the spread. More people mean a high number of positive cases. A 1% increase in a given ZIP code area population results in a 0.019% increase in the expected number of positive cases. The same increase in neighboring ZIP code areas also increases the expected positive cases by 0.004%. The student-population ratio (STUDENT) is another significant factor in the spread of the disease. A 1% increase in this ratio in a given ZIP code results in an almost 0.3% increase in the expected number of cases in the same area, while the same increase in surrounding ZIP codes increases the expected number of positive cases by 0.05%. These results suggest that the student population might be one of the main transmitters of the virus. Young students are known to be the primary transmitters of influenza-like illnesses (Cauchemez et al., 2008). Even though many schools were closed in 2020, young people might have had difficulty adjusting to social-distancing measures. The median household income (INCOME) is also positively associated with the number of positive cases. However, the relationship between the number of Covid-19 cases and median household income is not constant over different income levels. As shown in Fig. S7 in the supplementary document, there is a non-linear relationship, where areas with a median household income of less than 96,060 are more susceptible to contracting the virus than areas with a median household income higher than 96,060. A statistically significant indirect effect is only observed in the median household income. Finally, a 1% increase in the share of the driving-alone-to-work variable (DROVEALONE) in a given ZIP code area results in a 58.2% increase in positive cases in the same area. In comparison, the exact change in neighboring ZIP codes also increases the expected number of cases by 9.2%. This result suggests that people driving their cars pose more threats to others nearby. Based on our initial correlation analysis, the labor ratio and the driving-alone-to-workplace variables are strongly correlated. Therefore, a higher ratio of people driving alone to work in a given ZIP code area also indicates a higher concentration of employed populations who might be unable to maintain social distancing in their workplaces.

6. Discussion and policy-recommendations

This research should help policymakers and planners improve existing planning approaches to mitigate future outbreaks. The study results highlight the potential urban places where a virus can be easily transmitted. Schools, retail facilities, and workplaces are identified as the activity places where people congregate and allow a virus to spread. We should reevaluate current building design practices to prevent future spreads. School, retail, and workplace buildings should allow more natural air to penetrate indoors. Increased air movements may even improve energy efficiency.

This study also shows that low-income households, students, and essential workers have a higher risk of contracting the coronavirus. This disproportionate risk has widened inequalities. New urban planning approaches to reshaping cities will probably reduce the risk of future spread. However, policymakers should work on new policies to mitigate future risks. Improving Internet access and current online education systems can support self-isolation practices. Companies should continue investing in online technologies to allow their workforce to continue working virtually without losing efficiency. Finally, policymakers should consider a funding source for people who cannot continue working during future lockdowns.

7. Conclusions

The relationship between the Covid-19 spread and indicators of Florida’s built and socio-economic environments has been investigated at the ZIP code level, controlling for mobility and mask rules in counties. The availability of Covid-19 data at the ZIP code level has allowed for an exploration of the spatial dynamics of the spread at a fine spatial scale. Given the nature of the spread, spatial dependency in the number of positive cases is very likely, and spatial regression models have been used to capture these dependencies and provide unbiased and efficient results.

The model results provide detailed relationships between the built environment, socio-economic indicators, and disease spread. Some of these indicators play significant roles in the virus spread. This study contributes to the modeling of Covid-19 spread literature by: (1) capturing the spillover dynamics of the spread at the ZIP code level; (2) revealing the existence of spatial dependencies among the unobserved variables represented by the error term; (3) exploring positive association between the expected number of Covid-19 cases and specific land uses, such as education facilities and retail densities; (4) revealing critical socio-economic characteristics causing a substantial increase in Covid-19 spread.

There are several areas where to expand this research by eliminating existing limitations. Temporal dynamics of the spread can be incorporated into the model using FDOH’s available historical records of the number of cases at the ZIP code level. In addition, testing capacity can be included in the model as a control variable when this information is available. Furthermore, a similar investigation can be conducted at the block group level, where socio-economic and built environment variables are already available. Also, the impacts of built and socio-economic variables on the spread can be investigated by focusing on the period after the vast vaccination campaign. Finally, other states in different regions could also be analyzed using the same methodology to compare results with those for Florida.

Credit author statement

Emre Tepe: Conceptualization, Methodology, Software, Validation, Formal analysis, Investigation, Data curation, Writing- Original draft preparation, Writing- Reviewing and Editing.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

Data will be made available on request.

Acknowledgements

I would like to thank Dr. Jean-Michel Guldmann at Ohio State University for his valuable comments on this work. I also would like to thank anonymous reviewers for helpful comments.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.jenvman.2022.116806.

References

Abbasi, K.R., Abbas, J., Tufail, M., 2021a. Revisiting electricity consumption, price, and real GDP: a modified sectoral level analysis from Pakistan. Energy Pol. 149, 112087.
Abbasi, K.R., Adedoyin, F.F., Abbas, J., Hussain, K., 2021b. The impact of energy depletion and renewable energy on CO2 emissions in Thailand: fresh evidence from the novel dynamic ARDL simulation. Renew. Energy 180, 1439-1450.
al Halbusi, H., Al-Sulaiti, K., Abbas, J., Al-Sulaiti, I., 2022. Assessing factors influencing technology adoption for online purchasing amid COVID-19 in Qatar: moderating role of word of mouth. Front. Environ. Sci. 10.
Alexandrini, G., Janoschka, M., 2020. ‘Post-pandemic’ transnational gentrifications: a critical outlook. Urban Stud. 57 (15), 3202-3214.
Allam, Z., Bibri, S.E., Chabaud, D., Moreno, C., 2022. The ‘15-Minute City’ concept can shape a net-zero urban future. Human. Soc. Sci. Commun. 9 (1), 126.
Zhang, L., Ma, Z., Guo, L., 2009. An evaluation of spatial autocorrelation and heterogeneity in the residuals of six regression models. For. Sci. 55 (6), 533–548. https://doi.org/10.1093/forestscience/55.6.533. Retrieved from.

Zhang, Q., Naqvi, S.A.A., Shah, S.A.R., 2021a. The contribution of outward Foreign direct investment, human well-being, and technology toward a sustainable environment. Sustainability 13 (20), 11430.

Zhang, X., Maggioni, V., Houser, P., Xue, Y., Mei, Y., 2021b. The impact of weather condition and social activity on COVID-19 transmission in the United States. J. Environ. Manag., 114085

Zhang, Q., Shah, S.A.R., Yang, L., 2022. An appreciated response of disaggregated energies consumption towards the sustainable growth: a debate on G-10 economies. Energy 254, 124377.