Original article

A spatial analysis of the COVID-19 period prevalence in U.S. counties through June 28, 2020: where geography matters?

Feinuo Sun, PhD Candidate [a], [*], Stephen A. Matthews, PhD [b], Tse-Chuan Yang, PhD [a], Ming-Hsiao Hu, PhD [c]

[a] Department of Sociology, University at Albany, State University of New York, Albany, NY
[b] Department of Sociology & Criminology, and Department of Anthropology, The Pennsylvania State University, University Park, PA
[c] Department of Orthopedic Surgery, College of Medicine, National Taiwan University, Taipei, Taiwan

A B S T R A C T

Purpose: This study aims to understand how spatial structures, the interconnections between counties, matter in understanding the coronavirus disease 2019 (COVID-19) period prevalence across the United States.
Methods: We assemble a county-level data set that contains COVID-19–confirmed cases through June 28, 2020, and various sociodemographic measures from multiple sources. In addition to an aspatial regression model, we conduct spatial lag, spatial error, and spatial autoregressive combined models to systematically examine the role of spatial structure in shaping geographical disparities in the COVID-19 period prevalence.
Results: The aspatial ordinary least squares regression model tends to overestimate the COVID-19 period prevalence among counties with low observed rates, but this issue can be effectively addressed by spatial modeling. Spatial models can better estimate the period prevalence for counties, especially along the Atlantic coasts and through the Black Belt. Overall, the model fit among counties along both coasts is generally good with little variability evident, but in the Plain states, the model fit is conspicuous in its heterogeneity across counties.
Conclusions: Spatial models can help partially explain the geographic disparities in the COVID-19 period prevalence. These models reveal spatial variability in the model fit including identifying regions of the country where the fit is heterogeneous and worth closer attention in the immediate short term.

© 2020 Elsevier Inc. All rights reserved.

Introduction

Geography, referring to both an absolute location (i.e., specific place) and relative locations, matters in the outbreak of coronavirus disease 2019 (COVID-19). The dynamic data dashboards and news feeds clearly demonstrate great within-country spatial variations in the confirmed cases and deaths attributed to COVID-19 [1,2]. However, little formal research has used a spatial perspective to investigate the geographical disparities in the COVID-19 pandemic in the United States. This study, based on data through June 28,

2020, aims to show how spatial analysis may shed light on this issue.

Supplementing the long-standing focus on person and time of epidemiological research, the *place* has been recognized as an essential dimension of disease processes [3,4]. Previous studies show that the spatial heterogeneities of infectious diseases can result from either intrinsic population processes, including spatially aggregation of infected individuals and their nonrandom social interactions, or environmental influences acting across different spatial locations [5,6]. Public health scholars have called for attention on not only the role of place-based characteristics in the spread of diseases but also on the spatial relationships or interconnections between places [7,8]. Doing so allows a comprehensive understanding of the potential determinants of the novel disease [9–11].

The embeddedness and connectedness of the place is evident on a daily basis as much of the news cycle is driven by health, political,

* Corresponding author. Department of Sociology, University at Albany, State University of New York, 351 Arts & Sciences Building, 1400 Washington Ave, Albany, NY, 12222. Tel.: (+1) 518-730-6522.
E-mail address: fsun4@albany.edu (F. Sun).

economic, and social issues based on the different geographies of the interdependent processes including data reporting, decision-making, and policy enactment (both the imposition of stay-at-home orders and their relaxation). Decision makers at all levels—mayors, state representatives and governors—must adapt to directives or guidelines from higher up the hierarchy because what matters to them is what is going on in their 'local' constituency and the surrounding areas. These local decision makers have internalized the importance of absolute and relative location as well as first-hand knowledge of population composition and other contextual variables about where they live. The present study is motivated by the need to understand local COVID-19 conditions within regional and national contexts.

The purpose of this study is to first examine how the COVID-19 period prevalence distributes as of June 28, 2020, with thematic mapping and then to investigate how different spatial econometrics modeling approaches can inform understanding of the possible outliers (in terms of model fit or residuals), which may shed light on the transmission pattern. In the following sections, we describe our data and measures, the methods used, and our findings. The conclusion and implications of our study as well as a discussion of measurement and modeling issues related to our findings are summarized.

Materials and methods

For this study, we assembled a county-level data set for the contiguous United States ($n = 3106$ counties) using the Coronavirus Live Map [12], County Health Ranking and Roadmaps (CHRR) [13], U.S. Health Maps from the Institute for Health Metrics and Evaluation [14], the Area Health Resources Files [15], and Census Bureau Geographic Information System data [16].

Dependent variable

The dependent variable is the *COVID-19 period prevalence* (the number of cumulative confirmed cases per 100,000 population) in a county as of June 28, 2020. The data are provided by the Coronavirus Live Map that aggregates data from the Centers for Disease Control and Prevention and state- and local-level public health agencies. As the period prevalence is skewed, we log transform this variable as the Yeo-Johnson transformation [17] suggests.

Independent variables

Time is measured by the number of days since the first confirmed case in a county until June 28, 2020. To consider the nonlinear nature of infectious disease, we include the *square term of time* to capture an acceleration rate. In light of the racial/ethnic disparities in confirmed cases and deaths [18], we include racial/ethnic composition variables: the *percentage of non-Hispanic blacks* (hereafter blacks), *non-Hispanic Asians* (hereafter Asians), *Hispanics,* and *American Indian and Alaska Natives* (hereafter Native Americans). In addition, we consider the *percentage of older adults* (people who are older than 65 years), *unemployment rate,* and the *logged median income* to capture the age structure and socioeconomic conditions of a county. Furthermore, we consider the *nonwhite/white residential segregation index* (i.e. dissimilarity index), the *percentage of the uninsured*, the *percentage of households with at least one severe housing problem* (e.g., overcrowding or lacking major facilities), the *percentage of people who work outside the county of residence,* and *life expectancy*. These contextual variables are frequently used in social science research to capture fundamental conditions of, and inequalities in, society and the economy within an area.

The availability of medical resources in a county is measured by the Health Professional Shortage Area (*HPSA*) code. We capture health provisioning through two dummy variables identifying "the whole county is at shortage" and "part of the county is at shortage," respectively, with counties that are "not at any shortage" as the reference group. *Population density* is calculated by dividing the total county population by the land area of a county (Census Bureau GIS data), and this variable is log transformed. Population density has been known to be a factor for the transmission of infectious disease [19]. Most of the independent variables are drawn from 2018 to 2020 CHRR, except for the life expectancy (U.S. Health Map), the percentage of people who work outside the county of residence (American Community Survey 5-year estimates, 2014–2018), and the HPSA code (the Area Health Resources Files).

We compare the ordinary least squares (OLS) model and three spatial econometric models [20]. A spatial lag model is a model that examines how the infection burden in a county is influenced by the infection burden in adjacent counties. The spatial lag parameter (ρ) refers to the estimate of how the average logged period prevalence in neighboring counties is associated with the logged period prevalence of a focal county. By contrast, a spatial error model estimates the extent to which the OLS residual of a county is correlated with that in its adjacent counties. The spatial error parameter (λ) measures the strength of the relationship between the average residuals/errors in neighboring counties and the residual/error of a given county. Finally, a spatial autoregressive combined (SAC) model is a combination of the previous two models, which simultaneously considers the spatial lag and spatial error parameters. In the analysis presented, all spatial models are based on a first-order Queen spatial weight matrix, which defines a neighboring relationship between two counties when they share a common boundary or vertex (corner). The maps of the residuals generated for all four models are presented. These residual maps can inform our understanding of the spatial patterning of model fit predicting the COVID-19 period prevalence.

Results

Table 1 presents the descriptive statistics of the variables, and the last column includes the variance inflation factors. In an average U.S. county, there were 493.07 COVID-19−confirmed cases per 100,000 population, and not surprisingly, the distribution is positively skewed. The average number of days since the first confirmed case in a county was 88.30, with the maximum of 159 days (King county, WA). Regarding racial/ethnic composition, on average, 9.08% were blacks, 1.48% Asians, 9.69% Hispanics, and 2.08% Native Americans. The average percentage of the population older than 65 years was 19.31. The unemployment rate was slightly more than 4%, and on average, 11% of county population were uninsured. On average, 14.35% of households have at least one severe housing problem (e.g., no kitchen or plumbing facilities). As if to underline the spatial relationships between counties, on average, 30.80% of the adult population work outside their county of residence. The average life expectancy was 77.74 years, and only 10% of contiguous counties have no shortage of health professional shortage. We emphasize that multicollinearity is not a concern in our analysis as the variance inflation factors are all less than 4.

Figure 1 shows the spatial distribution of the logged COVID-19 period prevalence (by quintiles). Counties with high rates are clustered along the Boston−Washington corridor, in parts of the Rust Belt, the Black Belt with scattered high values found in the Mountain, Mexico/U.S. border, and West. By contrast, counties with a low period prevalence are concentrated in the Upper Great Plains, in Montana and Idaho, in west Texas, and in parts of central Appalachia.

Table 1
Descriptive statistics of variables used in this study, as of June 28, 2020 ($n = 3106$)*

Variable	Mean	SD	Minimum	Maximum	VIF
Confirmed cases per 100,000 (logged)	4.94	3.87	−16.12	9.50	—
Confirmed cases per 100,000	493.07	751.27	0	13,403.56	—
Time since the first confirmed case (day)	88.30	24.45	0	159	—
% Blacks	9.08	14.36	0	85.41	1.81
% Asians	1.48	2.51	0	38.31	1.84
% Hispanics	9.69	13.91	0.61	96.36	2.03
% Native Americans	2.08	6.69	0	92.52	1.39
% older than 65 y	19.31	4.65	4.83	57.59	1.83
% unemployed	4.09	1.40	1.30	18.09	1.78
Median income (logged)	10.84	0.24	10.14	11.85	3.74
Nonwhite-white segregation index	30.81	12.43	0.07	90.42	1.20
% of uninsured	11.42	5.11	2.26	33.75	2.14
% households with severe housing problems	14.35	4.35	0	39	1.86
% people work outside the county of residence	30.80	17.81	0	87.45	1.35
Life expectancy	77.74	2.37	66.81	86.83	3.24
HPSA					
% no shortage (reference group)	10.56				
% whole county is at shortage	26.50				3.10
% part of the county is at shortage	62.94				2.82
Population density (logged)	3.82	1.75	−1.48	11.18	3.19

* We show the original descriptive statistics in this table and emphasize that all continuous variables except for population density are standardized in our regression models.

The OLS and spatial modeling results are summarized in Table 2, and several findings are notable. First, the number of days since the first confirmed case and its square term follow the expectation. Specifically, the negative association between the square term and COVID-19 period prevalence ($\beta = -0.002$) suggests that the acceleration rate decreases with time, yet the total number of confirmed cases continues to grow ($\beta = 0.349$) since the first case. Second, the racial/ethnic composition of a county is important in determining the period prevalence, although the magnitude of the coefficients across models varies. The coefficients in the spatial error model are closer to that in the OLS model, whereas spatial lag and SAC models tend to yield comparable estimates. For example, the OLS model estimates that every 1% increase in the percentage of blacks is associated with 0.543-unit increase in the logged period prevalence and the magnitude of this relationship is 0.51 in spatial error model; however, the value drops by 15% to roughly 0.45 in the spatial lag and SAC models. The same pattern is observed for the percentage of Asians, Hispanics, and Native Americans, respectively. Third, the nonwhite-white segregation index, life expectancy, and population density are positively associated with the period prevalence. These three associations are consistent across all models and robust to the specification of spatial dependence.

We test if the residuals of each model are spatially correlated with the Moran's I statistic. As shown in Table 2, the OLS model has a Moran's I of 0.110, which is significant at the 0.001 level. The spatial lag model reduces Moran's I to 0.023, but it remains statistically significant. That being said, even after considering the average period prevalence of neighboring counties (i.e., the lag term), the spatially correlated errors suggest that this model omits variables that are not only related to the COVID-19 period prevalence but also spatially correlated. This finding is confirmed as the Moran's I is nonsignificant when spatial error terms are included in the analysis.

The residuals of all four models are visualized in Figure 2. Although these figures look similar, two major findings are worth noting. One is that spatial models improve the predicted values for counties with a low period prevalence, especially those in the Upper Great Plains (e.g., North/South Dakota, Wyoming, and Nebraska). Specifically, counties in the Upper Great Plains tend to have large and negative residuals in OLS, suggesting that the OLS model overestimates the period prevalence in these mostly sparsely populated counties. The other notable finding is that counties in the Great Plains (from Montana, North Dakota to southwestern Texas) show greater spatial heterogeneity (i.e. the pattern that neighboring counties have dissimilar values) in fit than those counties found along both coasts, even after considering the potential spatial autocorrelation in the analysis.

Discussion

Our findings suggest that there is great variability across the United States in the COVID-19 period prevalence and spatial models improve model fits especially for counties with a low period prevalence. Our model specification seems to explain reasonably well why counties have *high* levels of period prevalence; that is, most counties with a high period prevalence have relatively small residuals in our analysis. Nonetheless, it should be emphasized that some of the explanatory variables—such as the time since the first confirmed case, demographic composition, nonwhite-white segregation index, life expectancy, and population density—have significant impacts on the period prevalence of COVID-19, but they

Logged Prevalence
- -16.118- 4.488 (Q1)
- 4.488 - 5.250 (Q2)
- 5.250 - 5.867 (Q3)
- 5.867 - 6.552 (Q4)
- 6.552 - 9.503 (Q5)
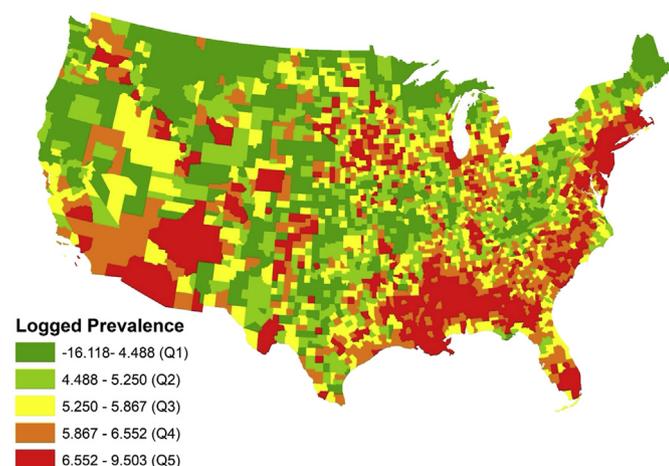
Fig. 1. Spatial distribution of the logged COVID-19 period prevalence by quintiles, as of June 28, 2020.

Table 2
OLS, spatial lag, spatial error, and SAC model for the period prevalence (logged), as of June 28, 2020

Variable	OLS		Spatial lag model		Spatial error model		SAC model	
	Estimate	SE	Estimate	SE	Estimate	SE	Estimate	SE
(Intercept)	−9.094***	0.244	−9.203***	0.238	−8.930***	0.257	−9.210***	0.248
Time	0.349***	0.006	0.338***	0.006	0.345***	0.006	0.342***	0.006
Time square	−0.002***	0.000	−0.002***	0.000	−0.002***	0.000	−0.002***	0.000
% Blacks	0.543***	0.052	0.448***	0.051	0.513***	0.061	0.465***	0.056
% Asians	0.227***	0.053	0.243***	0.051	0.214***	0.056	0.234***	0.054
% Hispanic	0.321***	0.055	0.280***	0.053	0.328***	0.064	0.298***	0.058
% Native Americans	0.216***	0.046	0.227***	0.045	0.221***	0.050	0.227***	0.047
% older than 65 y	−0.087	0.050	−0.119*	0.049	−0.079	0.054	−0.107*	0.051
% unemployed	−0.099	0.051	−0.105*	0.050	−0.087	0.056	−0.099	0.052
Median income (logged)	0.028	0.073	0.000	0.071	0.025	0.078	0.006	0.075
Nonwhite-white segregation index	0.141***	0.042	0.139***	0.041	0.106*	0.043	0.125**	0.042
% uninsured	0.074	0.055	0.078	0.054	0.101	0.064	0.085	0.058
% severe housing problems	−0.003	0.053	−0.004	0.052	−0.004	0.056	−0.004	0.054
% work outside the county of residence	0.046	0.045	−0.024	0.044	0.065	0.045	0.004	0.045
Life expectancy	0.176**	0.060	0.185**	0.058	0.181**	0.062	0.187**	0.060
HPSA (ref: no shortage)								
The whole county is at shortage	−0.180	0.154	−0.223	0.151	−0.153	0.153	−0.197	0.152
Part of the county is at shortage	−0.051	0.134	−0.049	0.131	−0.019	0.134	−0.034	0.133
Population density (logged)	0.168***	0.040	0.089*	0.040	0.195***	0.044	0.121**	0.043
ρ (spatial lag parameter)			0.192***				0.139***	
λ (spatial error parameter)					0.269***		0.121**	
AIC	13,608		13,501		13,516		13,494	
Observed Moran's I for residuals	0.110***		0.023*		−0.008		−0.003	

Level of significance: *$P < .05$, **$P < .01$, ***$P < .001$.

cannot fully account for the spatial pattern. Even with this caveat, the residual maps of spatial error and SAC models do help shed some light on identifying other potential explanatory variables for use in future research.

In addition, taking into account spatial structure improves the predicted values for the counties with a *low* period prevalence. From the comparison across different spatial regression models, the variability or heterogeneity in residuals is interesting. Here, the model may fit well in one county but fit poorly in neighboring counties. Similarly, we find counties where the COVID-19 period prevalence is severely underpredicted are often adjacent to counties where the model overpredicts. This checkerboard-like pattern is most visible across the Plain States and offers a stark contrast with the good model fit across much of the Atlantic seaboard and South East. We also note that a recent study [21] on COVID-19 suggests that spatial heterogeneity is fairly common in U.S. counties, which is supported by our findings. It should also be emphasized that while some scholars have noted the importance of spatial autocorrelation [21—23], they did not consider the potential impact of time on the pandemic and little research has considered the spatial lag and error term simultaneously. Our study advances the rapidly evolving literature by filling these gaps.

Given the array of demographic, social, economic, and health service—related variables in our models coupled with controls for population density and time since the first COVID case and incorporating spatial dependence into the model, the heterogeneity in the model fit underscores the complexity of COVID-19 period prevalence. On the one hand, there is the behavior and mutations of the virus itself, but there are also a host of modifiable social factors determined by federal, state, and local governments and institutions and the actions and behaviors of businesses and individuals vis-à-vis service provisioning, social distancing, and protection of the most vulnerable members of society. The complexity and levels of decision-making that have influenced the spread and intensity of COVID-19 have yet to be unpacked. We do not yet fully understand the reasons behind the high variability in testing availability and the rates of testing per capita. In this study, we looked at reported cases and that measure is subject to wide variability, driven by both the need to test in identified disease hotspots and in clusters of high-risk populations but also by the lack of testing and/or delays in testing. It may be a coincidence or a measurement concern that our model fit is most variable in the Great Plain states, an area including many of the states not to implement stay-at-home orders, and there have been fewer tests for COVID-19 in these states [24]. Future research is necessary.

We conducted several sensitivity analyses to assure the robustness of our findings. For example, we replaced the HPCA codes with other continuous measures, such as the number of hospital beds or physicians per 1000 population, but the results were similar. We also applied the principal component analysis (PCA) to a set of socioeconomic variables and created a PCA score to indicate the level of socioeconomic status of a county. Using this PCA score did not alter our conclusions or findings. These results are shown in Appendix Table A1. Furthermore, we implemented spatial regime models with different definitions of regime (e.g., stay-at-home vs. no stay-at-home order; metropolitan vs. nonmetropolitan) and found that the results were not changed. The results of these models are presented in Appendix Table A2. Finally, we visualized the residuals with different legend classifications (e.g., standard deviations, natural breaks, and quantiles), but the main visual patterns were consistent with the interpretation reported here.

This study is subject to some limitations. First, treating the period prevalence as a linear dependent variable may mask the great variations across counties. To our knowledge, there is no readily available software program that allows us to conduct a spatial lag model with a dependent variable that follows a Poisson or Binomial distribution. As such, we chose to use the log-transformed dependent variable with approximate normality. Second, our analysis relies on secondary data sources such as Coronavirus Live Map and CHRR, which also use the secondary data from federal agencies. Owing to the lack of data, we are unable to incorporate the county-level testing rates into our models, which in an ideal scenario ought to be associated with the period prevalence.

OLS

Spatial Lag Model

Spatial Error Model
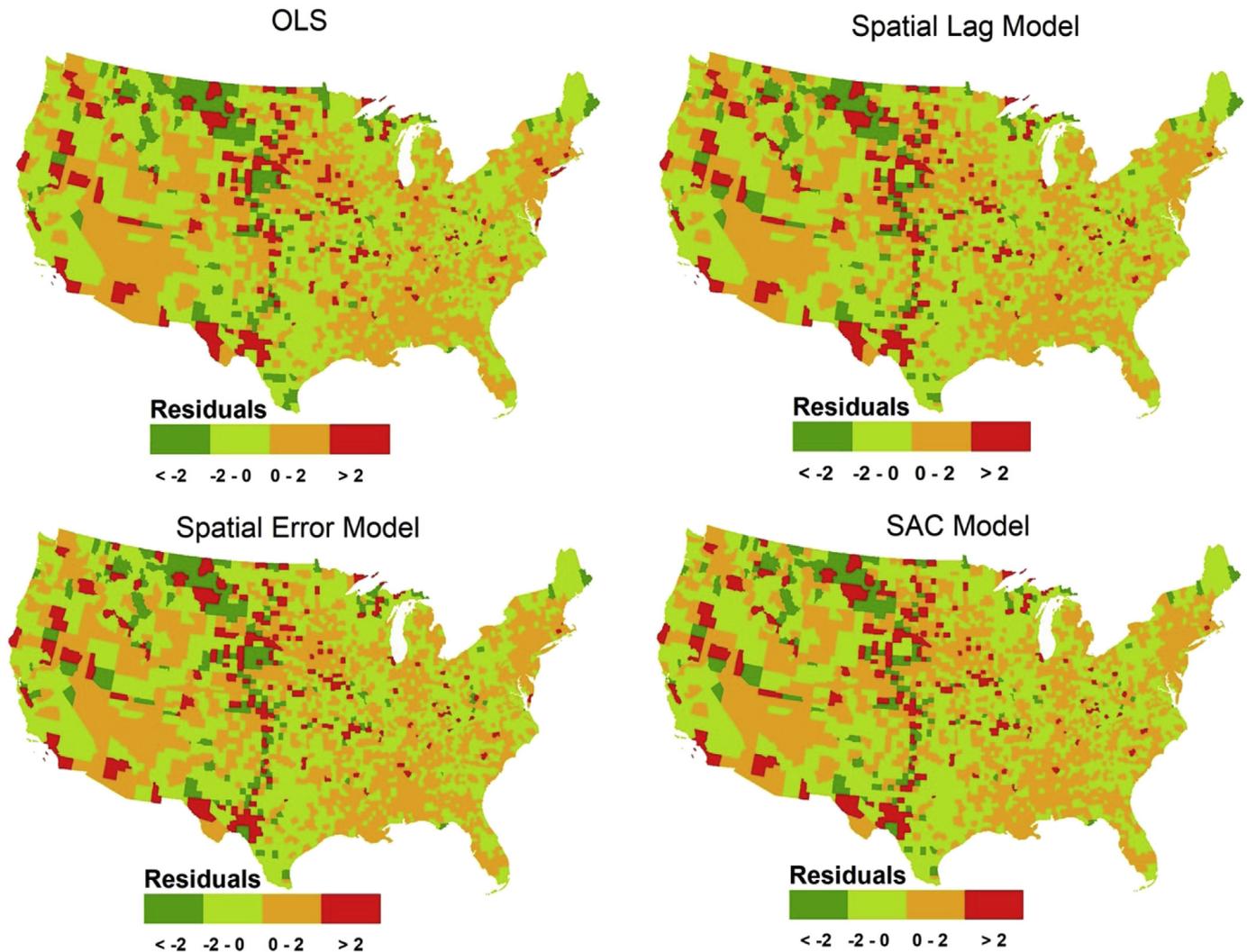
SAC Model

Fig. 2. Thematic maps for residuals of OLS, spatial lag, spatial error, and SAC model, as of June 28, 2020[†].[†]: With respect to the values, "< −2" means less than −2 and "> 2" means greater than 2.

We can only partially address the issue via a proxy for medical service provisioning and infrastructure using HPSA codes. In addition, while we consider the number of days since the first case in the analysis, this study remains cross-sectional and this design may mask temporal trends of the ongoing COVID-19 pandemic. All COVID-19 researchers are working in a dynamic environment, and we are all aware that findings may need to be revised by new data. Finally, there is a growing concern about asymptomatic cases [25], which cannot be included in our data. Future research is warranted to understand the impact of undercounted cases on geographical disparity of the COVID-19 crisis.

Conclusion

We believe in the old adage that "some models are useful." This study goes beyond data dashboards and description to contribute to emergent research using spatial models to look at the correlates of COVID-19 cases. Our results are consistent with expectations and a spatially informed study. For example, the variables with statistically significant associations with county-level COVID-19 cases include demographic variables (i.e., race/ethnicity), socioeconomic factors (i.e., income and housing conditions), and population mobility (i.e., the level of commuting ties between counties). The county-level analysis provides evidence on the embeddedness and connectedness of places and the importance of relative locations to local decision makers. What matters in the spread of COVID-19 is not only the contextual factors of a specific place but also the latent features of its neighbors. With additional data, which will inevitably be furnished, rigorous spatiotemporal analysis will play an important role. Our findings call for further efforts to help explain the spatial distribution and dynamics of this new infectious disease for prediction and prevention purposes.

CRediT authorship contribution statement

Feinuo Sun: Writing - original draft, Conceptualization, Formal analysis, Software. **Stephen A. Matthews:** Writing - review & editing, Conceptualization, Methodology. **Tse-Chuan Yang:** Conceptualization, Writing - review & editing, Visualization. **Ming-Hsiao Hu:** Writing - review & editing, Validation.

References

[1] Sharkey P. The US has a collective action problem that's larger than the coronavirus crisis 2020. https://www.vox.com/2020/4/10/21216216/coronavirus-social-distancing-texas-unacast-climate-change. [Accessed 20 May 2020].

[2] Danon L, Brooks-Pollock E, Bailey M, Keeling MJ. A spatial model of CoVID-19 transmission in England and Wales: early spread and peak timing. MedRxiv 2020:1–10.

[3] Glass GE. Update: Spatial aspects of epidemiology: The interface with medical geography. Epidemiol Rev 2000;2:136–9.

[4] Snow J. On the mode of communication of cholera. 1855. Salud Publica Mex 1991.

[5] Real LA, Biek R. Spatial dynamics and genetics of infectious diseases on heterogeneous landscapes. J R Soc Interface 2007;4:935–48.

[6] Grenfell BT, Bjørnstad ON, Kappey J. Travelling waves and spatial hierarchies in measles epidemics. Nature 2001;414:716–23.

[7] Rezaeian M, Dunn G, Leger St S, Appleby L. Geographical epidemiology, spatial analysis and geographical information systems: A multidisciplinary glossary. J Epidemiol Community Health 2007;61:98–102.

[8] Auchincloss AH, Gebreab SY, Mair C, Diez Roux AV. A Review of Spatial Methods in Epidemiology, 2000–2010. Annu Rev Public Health 2012;33:107–22.

[9] Ostfeld RS, Glass GE, Keesing F. Spatial epidemiology: An emerging (or re-emerging) discipline. Trends Ecol Evol 2005;20:328–36.

[10] Elliott P, Wartenberg D. Spatial epidemiology: Current approaches and future challenges. Environ Health Perspect 2004;112:998–1006.

[11] Jerrett M, Burnett RT, Goldberg MS, Sears M, Krewski D, Catalan R, et al. Spatial analysis for environmental health research: Concepts, methods, and examples. J Toxicol Environ Health A 2003;66:1783–810.

[12] USAFacts. Coronavirus Locations: COVID-19 Map by County and State. 2020. https://usafacts.org/visualizations/coronavirus-covid-19-spread-map/. [Accessed 20 May 2020].

[13] University of Wisconsin Population Health. County Health Rankings and Roadmaps. 2020. https://www.countyhealthrankings.org/. [Accessed 20 May 2020].

[14] IHME. Life expectancy at birth, both sexes, 2014. https://vizhub.healthdata.org/subnational/usa. [Accessed 1 July 2020].

[15] DHHS. Health. Resources and Servics Administration Health Professions. Area Health Resource File; 2019. https://data.hrsa.gov/topics/health-workforce/ahrf. [Accessed 1 May 2020].

[16] U.S. Department of Commerce USCBGD. TIGER/Line Shapefile, 2010, 2010 entity, Enterprise Rancheria, 2010 Census Tribal Census Tract AIA-based. US Census Bur TIGER/Line Shapefiles; 2010. https://www.census.gov/geographies/mapping-files/time-series/geo/tiger-line-file.html. [Accessed 1 May 2020].

[17] Weisberg S. Yeo-Johnson Power Transformations. Dep Appl Stat Univ Minnesota; 2001. https://www.stat.umn.edu/arc/yjpower.pdf. [Accessed 1 May 2020].

[18] CDC. COVID-19 in Racial and Ethnic Minority Groups. 2020. https://www.hsdl.org/?view&did=837299. [Accessed 20 May 2020].

[19] Chandra S, Kassens-Noor E, Kuljanin G, Vertalka J. A geographic analysis of population density thresholds in the influenza pandemic of 1918-19. Int J Health Geogr 2013;12:1–10.

[20] LeSage J, Pace RK. Introduction to spatial econometrics. Boca Raton, FL: CRC Press; 2009.

[21] Mollalo A, Vahedi B, Rivera KM. GIS-based spatial modeling of COVID-19 incidence rate in the continental United States. Sci Total Environ 2020;728:138884.

[22] Cordes J, Castro MC. Spatial analysis of COVID-19 clusters and contextual factors in New York City. Spat Spatiotemporal Epidemiol 2020;34.

[23] Zhang CH, Schwartz GG. Spatial Disparities in Coronavirus Incidence and Mortality in the United States: An Ecological Analysis as of May 2020. J Rural Health 2020;00:1–13.

[24] Jin B. Live tracker: How many coronavirus cases have been reported in each U.S. state?. https://www.politico.com/interactives/2020/coronavirus-testing-by-state-chart-of-new-cases/. [Accessed 9 May 2020].

[25] Furukawa N, JT B, Sobel J. Evidence Supporting Transmission of Severe Acute Respiratory Syndrome Coronavirus 2 While Presymptomatic or Asymptomatic, 26; 2020.

Appendix

Appendix 1. Sensitivity analysis with the SAC model using different covariates

Table A1
Sensitivity analysis for the period prevalence (logged)

Variable	Model A1		Model A2	
	Estimate	SE	Estimate	SE
(Intercept)	−9.461***	0.216	−9.146***	0.248
Time	0.343***	0.006	0.339***	0.006
Time square	−0.002***	0.000	−0.002***	0.000
% Blacks	0.434***	0.055	0.418***	0.057
% Asians	0.228***	0.054	0.251***	0.053
% Hispanic	0.295***	0.057	0.250***	0.058
% Native Americans	0.226***	0.047	0.210***	0.046
% older than 65 y	−0.106***	0.051	−0.150***	0.052
% unemployed	−0.093***	0.052	—	—
Median income (logged)	0.025	0.074	—	—
Nonwhite-white segregation index	0.123**	0.042	−0.088**	0.072
% uninsured	0.082	0.058	0.116*	0.055
% severe housing problems	0.013	0.053	−0.045	0.054
% work outside the county of residence	0.007	0.046	0.003	0.045
Life expectancy	0.191**	0.060	0.265***	0.061
Number of physicians per 1000 people	−0.026	0.049	—	—
Number of hospital beds per 1000 people	0.141***	0.041	—	—
HPSA (ref: no shortage)				
The whole county is at shortage	—	—	−0.255	0.153
Part of the county is at shortage	—	—	−0.075	0.133
Population density (logged)	0.139**	0.043	0.135**	0.043
SES Score (PCA)	—	—	0.118	0.042
ρ (spatial lag parameter)	0.144***		0.134***	
λ (spatial error parameter)	0.115**		0.131***	

Level of significance: *P < .05, **P < .01, ***P < .001.

Appendix 2. Sensitivity analysis with the spatial regime model

Table A2
Spatial regime models by the stay-at-home order and metropolitan status

Variable	Model A3				Model A4			
	Stay-at-home order		No stay-at-home order		Metropolitan		Nonmetropolitan	
	Estimate	SE	Estimate	SE	Estimate	SE	Estimate	SE
(Intercept)	−9.220***	0.232	−10.016***	1.240	−9.977***	0.212	−3.726**	1.178
Time	0.322***	0.007	0.451***	0.015	0.389***	0.007	0.152***	0.023
Time square	−0.002***	0.000	−0.003***	0.000	−0.002***	0.000	−0.001***	0.000
% Blacks	0.406***	0.076	2.679	1.840	0.522***	0.063	0.298**	0.093
% Asians	0.187***	0.053	0.533*	0.249	0.169	0.119	−0.001	0.003
% Hispanic	0.266**	0.092	0.552	0.369	0.260***	0.061	0.187*	0.091
% Native Americans	0.063	0.071	0.432***	0.123	0.234***	0.046	−0.102	0.422
% older than 65 y	−0.137	0.119	0.421	0.227	−0.139*	0.058	−0.115	0.137
% unemployed	−0.009	0.101	−0.484*	0.221	−0.099	0.054	−0.024	0.021
Median income (logged)	0.006	0.107	0.379	0.262	0.058	0.099	0.007	0.014
Nonwhite-white segregation index	0.151***	0.043	−0.218	0.127	0.113*	0.047	0.163	0.087
% uninsured	0.149*	0.063	0.445	0.267	0.111	0.059	0.074	0.084
Life expectancy	0.175	0.132	−0.065	0.153	0.226***	0.065	0.143	0.087
Population density (logged)	0.121**	0.042	0.421*	0.197	0.202***	0.051	0.029	0.084
ρ (spatial lag parameter)	0.144***				0.138***			
λ (spatial error parameter)	0.079*				0.109**			

Level of significance: *P < .05, **P < .01, ***P < .001.