Socioeconomic Factors Influencing the Spatial Spread of COVID-19 in the United States

Christopher F Baum (Boston College, DIW Berlin & CESIS)
Miguel Henry (Greylock McKinnon Associates)

London Stata Conference, September 2020

Motivation for the study

- The spread of COVID-19 has been heavily influenced by geography and proximity to areas with high concentrations of those infected
- A number of socioeconomic factors have also emerged as predictors of the pandemic's severity
- This paper analyzes US county-level daily data on confirmed cases through late May, taking both geographic and socioeconomic factors into account

Motivation for the study

- The spread of COVID-19 has been heavily influenced by geography and proximity to areas with high concentrations of those infected
- A number of socioeconomic factors have also emerged as predictors of the pandemic's severity
- This paper analyzes US county-level daily data on confirmed cases through late May, taking both geographic and socioeconomic factors into account

Motivation for the study

- The spread of COVID-19 has been heavily influenced by geography and proximity to areas with high concentrations of those infected
- A number of socioeconomic factors have also emerged as predictors of the pandemic's severity
- This paper analyzes US county-level daily data on confirmed cases through late May, taking both geographic and socioeconomic factors into account

Existing work

We located several recent studies related to the spatial geographic aspects of COVID-19 spread, including:

- Bailey et al., WP, Stern School, NYU
- Coven & Gupta, WP, Stern School, NYU
- Guliyev, *Spatial Statistics*, 2020
- Kang et al., *International Journal of Infectious Diseases*, 2020
- Kuchler et al., WP, CESifo
- Mollalo et al., *Science of the Total Environment*, 2020

Full references are provided in our working paper,
https://ideas.repec.org/p/boc/bocoec/1009.html

Existing work

We located several recent studies related to the spatial geographic aspects of COVID-19 spread, including:

- Bailey et al., WP, Stern School, NYU
- Coven & Gupta, WP, Stern School, NYU
- Guliyev, *Spatial Statistics*, 2020
- Kang et al., *International Journal of Infectious Diseases*, 2020
- Kuchler et al., WP, CESifo
- Mollalo et al., *Science of the Total Environment*, 2020

Full references are provided in our working paper, https://ideas.repec.org/p/boc/bocoec/1009.html

Existing work

We located several recent studies related to the spatial geographic aspects of COVID-19 spread, including:

- Bailey et al., WP, Stern School, NYU
- Coven & Gupta, WP, Stern School, NYU
- Guliyev, *Spatial Statistics*, 2020
- Kang et al., *International Journal of Infectious Diseases*, 2020
- Kuchler et al., WP, CESifo
- Mollalo et al., *Science of the Total Environment*, 2020

Full references are provided in our working paper,
https://ideas.repec.org/p/boc/bocoec/1009.html

Existing work

We located several recent studies related to the spatial geographic aspects of COVID-19 spread, including:

- Bailey et al., WP, Stern School, NYU
- Coven & Gupta, WP, Stern School, NYU
- Guliyev, *Spatial Statistics*, 2020
- Kang et al., *International Journal of Infectious Diseases*, 2020
- Kuchler et al., WP, CESifo
- Mollalo et al., *Science of the Total Environment*, 2020

Full references are provided in our working paper, https://ideas.repec.org/p/boc/bocoec/1009.html

Existing work

We located several recent studies related to the spatial geographic aspects of COVID-19 spread, including:

- Bailey et al., WP, Stern School, NYU
- Coven & Gupta, WP, Stern School, NYU
- Guliyev, *Spatial Statistics*, 2020
- Kang et al., *International Journal of Infectious Diseases*, 2020
- Kuchler et al., WP, CESifo
- Mollalo et al., *Science of the Total Environment*, 2020

Full references are provided in our working paper, https://ideas.repec.org/p/boc/bocoec/1009.html

Existing work

We located several recent studies related to the spatial geographic aspects of COVID-19 spread, including:

- Bailey et al., WP, Stern School, NYU
- Coven & Gupta, WP, Stern School, NYU
- Guliyev, *Spatial Statistics*, 2020
- Kang et al., *International Journal of Infectious Diseases*, 2020
- Kuchler et al., WP, CESifo
- Mollalo et al., *Science of the Total Environment*, 2020

Full references are provided in our working paper,
https://ideas.repec.org/p/boc/bocoec/1009.html

Description of the data

- Daily data on COVID-19 confirmed cases for counties in the 48 contiguous US states were accessed from the Johns Hopkins Center for Systems Science and Engineering for March 1, 2020 through May 23, 2020
- Data on county populations, including information on gender, race and age categories, were accessed from the US Census Bureau
- County-level socioeconomic factors were accessed from the County Health Rankings database
- The Policy Map/CDC BRFSS database provided data on health risk factors
- Spatial data were obtained from the US Census Bureau's MAF/TIGER database

Description of the data

- Daily data on COVID-19 confirmed cases for counties in the 48 contiguous US states were accessed from the Johns Hopkins Center for Systems Science and Engineering for March 1, 2020 through May 23, 2020
- Data on county populations, including information on gender, race and age categories, were accessed from the US Census Bureau
- County-level socioeconomic factors were accessed from the County Health Rankings database
- The Policy Map/CDC BRFSS database provided data on health risk factors
- Spatial data were obtained from the US Census Bureau's MAF/TIGER database

Description of the data

- Daily data on COVID-19 confirmed cases for counties in the 48 contiguous US states were accessed from the Johns Hopkins Center for Systems Science and Engineering for March 1, 2020 through May 23, 2020
- Data on county populations, including information on gender, race and age categories, were accessed from the US Census Bureau
- County-level socioeconomic factors were accessed from the County Health Rankings database
- The Policy Map/CDC BRFSS database provided data on health risk factors
- Spatial data were obtained from the US Census Bureau's MAF/TIGER database

Description of the data

- Daily data on COVID-19 confirmed cases for counties in the 48 contiguous US states were accessed from the Johns Hopkins Center for Systems Science and Engineering for March 1, 2020 through May 23, 2020
- Data on county populations, including information on gender, race and age categories, were accessed from the US Census Bureau
- County-level socioeconomic factors were accessed from the County Health Rankings database
- The Policy Map/CDC BRFSS database provided data on health risk factors
- Spatial data were obtained from the US Census Bureau's MAF/TIGER database

Description of the data

- Daily data on COVID-19 confirmed cases for counties in the 48 contiguous US states were accessed from the Johns Hopkins Center for Systems Science and Engineering for March 1, 2020 through May 23, 2020
- Data on county populations, including information on gender, race and age categories, were accessed from the US Census Bureau
- County-level socioeconomic factors were accessed from the County Health Rankings database
- The Policy Map/CDC BRFSS database provided data on health risk factors
- Spatial data were obtained from the US Census Bureau's MAF/TIGER database

- Race was categorized as White, Black and Hispanic, where Hispanic can be combined with any of the Census Bureau's racial categories
- County-level age distributions were aggregated into five categories. Ultimately, the only age category used in our analysis is the percentage of those 20–39 years of age
- socioeconomic factors included the county median income, prevalence of PM2.5 (fine inhalable particle) air pollution and the percentage of residents lacking health insurance
- The COVID-19 Health Risk Index combines the prevalence of five health conditions associated with the severity of infection: obesity, diabetes, high blood pressure, heart disease and chronic obstructive pulmonary disease (COPD)

- Race was categorized as White, Black and Hispanic, where Hispanic can be combined with any of the Census Bureau's racial categories
- County-level age distributions were aggregated into five categories. Ultimately, the only age category used in our analysis is the percentage of those 20–39 years of age
- socioeconomic factors included the county median income, prevalence of PM2.5 (fine inhalable particle) air pollution and the percentage of residents lacking health insurance
- The COVID-19 Health Risk Index combines the prevalence of five health conditions associated with the severity of infection: obesity, diabetes, high blood pressure, heart disease and chronic obstructive pulmonary disease (COPD)

- Race was categorized as White, Black and Hispanic, where Hispanic can be combined with any of the Census Bureau's racial categories
- County-level age distributions were aggregated into five categories. Ultimately, the only age category used in our analysis is the percentage of those 20–39 years of age
- socioeconomic factors included the county median income, prevalence of PM2.5 (fine inhalable particle) air pollution and the percentage of residents lacking health insurance
- The COVID-19 Health Risk Index combines the prevalence of five health conditions associated with the severity of infection: obesity, diabetes, high blood pressure, heart disease and chronic obstructive pulmonary disease (COPD)

- Race was categorized as White, Black and Hispanic, where Hispanic can be combined with any of the Census Bureau's racial categories
- County-level age distributions were aggregated into five categories. Ultimately, the only age category used in our analysis is the percentage of those 20–39 years of age
- socioeconomic factors included the county median income, prevalence of PM2.5 (fine inhalable particle) air pollution and the percentage of residents lacking health insurance
- The COVID-19 Health Risk Index combines the prevalence of five health conditions associated with the severity of infection: obesity, diabetes, high blood pressure, heart disease and chronic obstructive pulmonary disease (COPD)

Incorporation of spatial factors

- Our working hypothesis regarding spatial spread led to the development of a spatial autoregressive model
- The spatial weighting matrix was computed using the characteristic of *rook contiguity*, where two counties are considered as first-order neighbors if they share a border (and not merely a point)
- For robustness, we also tested an inverse spatial distance matrix, with less satisfactory results

Incorporation of spatial factors

- Our working hypothesis regarding spatial spread led to the development of a spatial autoregressive model
- The spatial weighting matrix was computed using the characteristic of *rook contiguity*, where two counties are considered as first-order neighbors if they share a border (and not merely a point)
- For robustness, we also tested an inverse spatial distance matrix, with less satisfactory results

Incorporation of spatial factors

- Our working hypothesis regarding spatial spread led to the development of a spatial autoregressive model
- The spatial weighting matrix was computed using the characteristic of *rook contiguity*, where two counties are considered as first-order neighbors if they share a border (and not merely a point)
- For robustness, we also tested an inverse spatial distance matrix, with less satisfactory results

The spatial autoregressive (SAR) model

- The spatial autoregressive model (LeSage and Pace, *Introduction to Spatial Econometrics*, 2009) evaluates the importance of contiguity via the spatial weighting matrix and allows for the inclusion of standard regressors
- Our initial SAR model includes only values of current confirmed cases, $C_{j,t}$, for 14- and 28-day lags, given the incubation period of the virus.
- These factors are added to a spatial lag of the dependent variable as well as a spatially lagged error term
- The latter term evaluates the correlation of the error term $\varepsilon_{j,t}$ with errors $\varepsilon_{k,t}$ in order to capture common shocks driven by contiguity
- The SAR model can be estimated in recent versions of Stata with the `spregress` command, using the maximum likelihood option

The spatial autoregressive (SAR) model

- The spatial autoregressive model (LeSage and Pace, *Introduction to Spatial Econometrics*, 2009) evaluates the importance of contiguity via the spatial weighting matrix and allows for the inclusion of standard regressors
- Our initial SAR model includes only values of current confirmed cases, $C_{j,t}$, for 14- and 28-day lags, given the incubation period of the virus.
- These factors are added to a spatial lag of the dependent variable as well as a spatially lagged error term
- The latter term evaluates the correlation of the error term $\varepsilon_{j,t}$ with errors $\varepsilon_{k,t}$ in order to capture common shocks driven by contiguity
- The SAR model can be estimated in recent versions of Stata with the `spregress` command, using the maximum likelihood option

The spatial autoregressive (SAR) model

- The spatial autoregressive model (LeSage and Pace, *Introduction to Spatial Econometrics*, 2009) evaluates the importance of contiguity via the spatial weighting matrix and allows for the inclusion of standard regressors
- Our initial SAR model includes only values of current confirmed cases, $C_{j,t}$, for 14- and 28-day lags, given the incubation period of the virus.
- These factors are added to a spatial lag of the dependent variable as well as a spatially lagged error term
- The latter term evaluates the correlation of the error term $\varepsilon_{j,t}$ with errors $\varepsilon_{k,t}$ in order to capture common shocks driven by contiguity
- The SAR model can be estimated in recent versions of Stata with the `spregress` command, using the maximum likelihood option

The spatial autoregressive (SAR) model

- The spatial autoregressive model (LeSage and Pace, *Introduction to Spatial Econometrics*, 2009) evaluates the importance of contiguity via the spatial weighting matrix and allows for the inclusion of standard regressors
- Our initial SAR model includes only values of current confirmed cases, $C_{j,t}$, for 14- and 28-day lags, given the incubation period of the virus.
- These factors are added to a spatial lag of the dependent variable as well as a spatially lagged error term
- The latter term evaluates the correlation of the error term $\varepsilon_{j,t}$ with errors $\varepsilon_{k,t}$ in order to capture common shocks driven by contiguity
- The SAR model can be estimated in recent versions of Stata with the `spregress` command, using the maximum likelihood option

The spatial autoregressive (SAR) model

- The spatial autoregressive model (LeSage and Pace, *Introduction to Spatial Econometrics*, 2009) evaluates the importance of contiguity via the spatial weighting matrix and allows for the inclusion of standard regressors
- Our initial SAR model includes only values of current confirmed cases, $C_{j,t}$, for 14- and 28-day lags, given the incubation period of the virus.
- These factors are added to a spatial lag of the dependent variable as well as a spatially lagged error term
- The latter term evaluates the correlation of the error term $\varepsilon_{j,t}$ with errors $\varepsilon_{k,t}$ in order to capture common shocks driven by contiguity
- The SAR model can be estimated in recent versions of Stata with the `spregress` command, using the maximum likelihood option

Incorporation of socioeconomic factors

We consider these stylized facts:

- Males have been more resistant to observing public health interventions (PHI) such as wearing masks and obeying social distancing
- Minority workers are more likely to have essential jobs that cannot be performed at home
- Although the aged are at greater risk if infected, younger individuals are more likely to ignore PHIs and congregate, triggering the recent reimposition of restrictions in Florida, Texas and Arizona
- Residents of counties with lower median income are less likely to be able to work from home, and more likely to be minorities.

Incorporation of socioeconomic factors

We consider these stylized facts:

- Males have been more resistant to observing public health interventions (PHI) such as wearing masks and obeying social distancing
- Minority workers are more likely to have essential jobs that cannot be performed at home
- Although the aged are at greater risk if infected, younger individuals are more likely to ignore PHIs and congregate, triggering the recent reimposition of restrictions in Florida, Texas and Arizona
- Residents of counties with lower median income are less likely to be able to work from home, and more likely to be minorities.

Incorporation of socioeconomic factors

We consider these stylized facts:

- Males have been more resistant to observing public health interventions (PHI) such as wearing masks and obeying social distancing
- Minority workers are more likely to have essential jobs that cannot be performed at home
- Although the aged are at greater risk if infected, younger individuals are more likely to ignore PHIs and congregate, triggering the recent reimposition of restrictions in Florida, Texas and Arizona
- Residents of counties with lower median income are less likely to be able to work from home, and more likely to be minorities.

Incorporation of socioeconomic factors

We consider these stylized facts:

- Males have been more resistant to observing public health interventions (PHI) such as wearing masks and obeying social distancing
- Minority workers are more likely to have essential jobs that cannot be performed at home
- Although the aged are at greater risk if infected, younger individuals are more likely to ignore PHIs and congregate, triggering the recent reimposition of restrictions in Florida, Texas and Arizona
- Residents of counties with lower median income are less likely to be able to work from home, and more likely to be minorities.

- Residents of counties with lower median incomes, particularly in rural areas, may have less access to high-quality health care facilities
- Residents of counties with higher levels of air pollution are more likely to be susceptible to airborne infection, and to adverse respiratory conditions
- Residents of counties with a larger fraction lacking health insurance are more likely to have poor access to quality health care
- Residents of counties with a high prevalence of complicating health risk factors are more susceptible to COVID-19

- Residents of counties with lower median incomes, particularly in rural areas, may have less access to high-quality health care facilities
- Residents of counties with higher levels of air pollution are more likely to be susceptible to airborne infection, and to adverse respiratory conditions
- Residents of counties with a larger fraction lacking health insurance are more likely to have poor access to quality health care
- Residents of counties with a high prevalence of complicating health risk factors are more susceptible to COVID-19

- Residents of counties with lower median incomes, particularly in rural areas, may have less access to high-quality health care facilities
- Residents of counties with higher levels of air pollution are more likely to be susceptible to airborne infection, and to adverse respiratory conditions
- Residents of counties with a larger fraction lacking health insurance are more likely to have poor access to quality health care
- Residents of counties with a high prevalence of complicating health risk factors are more susceptible to COVID-19

- Residents of counties with lower median incomes, particularly in rural areas, may have less access to high-quality health care facilities
- Residents of counties with higher levels of air pollution are more likely to be susceptible to airborne infection, and to adverse respiratory conditions
- Residents of counties with a larger fraction lacking health insurance are more likely to have poor access to quality health care
- Residents of counties with a high prevalence of complicating health risk factors are more susceptible to COVID-19

- The inclusion of socioeconomic factors and evaluation of their effects is complicated by their intercorrelations. For instance, most states rejecting Medicaid expansion are in the South, with lower percentages of health coverage, lower median incomes and a higher prevalence of adverse health conditions
- Although we did not consider political affiliation in our analysis, it has become evident in recent weeks that 'red' (Republican) states have been less likely to impose PHIs (such as lockdowns and mandatory use of masks) and more likely to reopen their economies too rapidly than the 'blue' (Democratic) states in which the initial waves of the virus were concentrated

- The inclusion of socioeconomic factors and evaluation of their effects is complicated by their intercorrelations. For instance, most states rejecting Medicaid expansion are in the South, with lower percentages of health coverage, lower median incomes and a higher prevalence of adverse health conditions
- Although we did not consider political affiliation in our analysis, it has become evident in recent weeks that 'red' (Republican) states have been less likely to impose PHIs (such as lockdowns and mandatory use of masks) and more likely to reopen their economies too rapidly than the 'blue' (Democratic) states in which the initial waves of the virus were concentrated

- The trajectory of US confirmed cases has now shifted, with New York, New Jersey, Massachusetts, Connecticut and Michigan reaping the benefits of strong restrictions following devastating levels of infection and widespread deaths, particularly among the elderly
- Rising caseloads have shifted to the 'Sun Belt' states, which are now epicenters of the pandemic in the US, forced to reimpose restrictions on economic activity to reduce the spread of the virus
- The median age of confirmed cases in Florida, a state with a large population of retirees, has dropped from 65 to 35 in recent weeks

- The trajectory of US confirmed cases has now shifted, with New York, New Jersey, Massachusetts, Connecticut and Michigan reaping the benefits of strong restrictions following devastating levels of infection and widespread deaths, particularly among the elderly
- Rising caseloads have shifted to the 'Sun Belt' states, which are now epicenters of the pandemic in the US, forced to reimpose restrictions on economic activity to reduce the spread of the virus
- The median age of confirmed cases in Florida, a state with a large population of retirees, has dropped from 65 to 35 in recent weeks

- The trajectory of US confirmed cases has now shifted, with New York, New Jersey, Massachusetts, Connecticut and Michigan reaping the benefits of strong restrictions following devastating levels of infection and widespread deaths, particularly among the elderly
- Rising caseloads have shifted to the 'Sun Belt' states, which are now epicenters of the pandemic in the US, forced to reimpose restrictions on economic activity to reduce the spread of the virus
- The median age of confirmed cases in Florida, a state with a large population of retirees, has dropped from 65 to 35 in recent weeks

Descriptive statistics

Table: Confirmed COVID-19 Counts for Lag Variables ($N = 3{,}107$)

Variable Name	Min	0.25	Median	0.75	0.99	Max	Date
$C_{j,t}$	0	7	33	157	8,316	197,266	May 23
$C_{j,t-14}$	0	5	23	107	6,887	183,289	May 9
$C_{j,t-28}$	0	3	14	68	4,916	155,113	April 25

These raw counts are converted to confirmed cases per 100,000 residents in the empirical analysis.

Table: Descriptive Statistics ($N = 3{,}107$)

Variable Name	Mean	Min	Max	Description
pct_male	50.08	43.13	73.16	Percentage of Males
pct_black	10.30	0.097	86.61	Percentage of Blacks
pct_hisp	9.69	0.61	96.36	Percentage of Hispanics
pct_20_30	24.01	11.30	53.06	Percentage in the 20–39 year age group
pm25	8.95	4.2	15.4	PM2.5 (in micrograms per cubic meter)
medinc	49.39	22.05	134.61	Median Income ('000 US dollars)
unins	11.95	2.13	33.27	Percentage of Uninsured Individuals
index_zscore	0.08	-3.65	3.56	z–score of Health Risk Index value

Notes: The sample size *N* denotes the number of U.S. counties included.

To provide some perspective on the data, the following table presents statistics for four selected counties:

1. Los Angeles, CA
2. Cook, IL, contiguous with Chicago
3. Suffolk, MA, largely contiguous with Boston
4. Brazos, Texas, the home of StataCorp

To provide some perspective on the data, the following table presents statistics for four selected counties:

1. Los Angeles, CA
2. Cook, IL, contiguous with Chicago
3. Suffolk, MA, largely contiguous with Boston
4. Brazos, Texas, the home of StataCorp

To provide some perspective on the data, the following table presents statistics for four selected counties:

1. Los Angeles, CA
2. Cook, IL, contiguous with Chicago
3. Suffolk, MA, largely contiguous with Boston
4. Brazos, Texas, the home of StataCorp

To provide some perspective on the data, the following table presents statistics for four selected counties:

1. Los Angeles, CA
2. Cook, IL, contiguous with Chicago
3. Suffolk, MA, largely contiguous with Boston
4. Brazos, Texas, the home of StataCorp

Table: COVID-19 confirmed cases as of 23 May 2020

	Los Angeles CA	Cook IL	Suffolk MA	Brazos TX
Cases	44121.00	70417.00	17291.00	362.00
Cases100K	436.60	1359.27	2141.96	159.64
PopK	10105.52	5180.49	807.25	226.76
%Male	49.30	48.57	48.32	50.55
%Black	9.97	24.81	27.18	11.82
%Hisp	48.64	25.54	23.38	26.03
%20-30	30.27	29.98	40.28	41.69
MedIncomeK	61.31	60.02	61.37	42.28
PM2.5	14.40	14.00	10.00	9.20
%Uninsured	12.49	10.63	4.64	17.38
Z HealthSc	-2.14	-1.05	-2.30	-1.92

- In the late May period, Boston was hardest hit in terms of cases per resident. At that point Texas had experienced relatively few cases.
- Boston has the largest Black population, while Los Angeles has the largest Hispanic population and the largest Black+Hispanic population fraction.
- Boston and Brazos have 10% more residents in the 20–39 age group.
- Brazos has a median income about 2/3 that of the other three counties.
- Boston and Brazos have lower levels of air pollution than LA and Chicago.
- Romneycare has driven the percentage of those lacking health insurance in Boston below 5%, while Texas has not expanded Medicaid, health coverage for the poor.
- LA and Boston have lower levels of the health score index than Chicago and Brazos, whose residents are less healthy

- In the late May period, Boston was hardest hit in terms of cases per resident. At that point Texas had experienced relatively few cases.
- Boston has the largest Black population, while Los Angeles has the largest Hispanic population and the largest Black+Hispanic population fraction.
- Boston and Brazos have 10% more residents in the 20–39 age group.
- Brazos has a median income about 2/3 that of the other three counties.
- Boston and Brazos have lower levels of air pollution than LA and Chicago.
- Romneycare has driven the percentage of those lacking health insurance in Boston below 5%, while Texas has not expanded Medicaid, health coverage for the poor.
- LA and Boston have lower levels of the health score index than Chicago and Brazos, whose residents are less healthy

- In the late May period, Boston was hardest hit in terms of cases per resident. At that point Texas had experienced relatively few cases.
- Boston has the largest Black population, while Los Angeles has the largest Hispanic population and the largest Black+Hispanic population fraction.
- Boston and Brazos have 10% more residents in the 20–39 age group.
- Brazos has a median income about 2/3 that of the other three counties.
- Boston and Brazos have lower levels of air pollution than LA and Chicago.
- Romneycare has driven the percentage of those lacking health insurance in Boston below 5%, while Texas has not expanded Medicaid, health coverage for the poor.
- LA and Boston have lower levels of the health score index than Chicago and Brazos, whose residents are less healthy

- In the late May period, Boston was hardest hit in terms of cases per resident. At that point Texas had experienced relatively few cases.
- Boston has the largest Black population, while Los Angeles has the largest Hispanic population and the largest Black+Hispanic population fraction.
- Boston and Brazos have 10% more residents in the 20–39 age group.
- Brazos has a median income about 2/3 that of the other three counties.
- Boston and Brazos have lower levels of air pollution than LA and Chicago.
- Romneycare has driven the percentage of those lacking health insurance in Boston below 5%, while Texas has not expanded Medicaid, health coverage for the poor.
- LA and Boston have lower levels of the health score index than Chicago and Brazos, whose residents are less healthy

- In the late May period, Boston was hardest hit in terms of cases per resident. At that point Texas had experienced relatively few cases.
- Boston has the largest Black population, while Los Angeles has the largest Hispanic population and the largest Black+Hispanic population fraction.
- Boston and Brazos have 10% more residents in the 20–39 age group.
- Brazos has a median income about 2/3 that of the other three counties.
- Boston and Brazos have lower levels of air pollution than LA and Chicago.
- Romneycare has driven the percentage of those lacking health insurance in Boston below 5%, while Texas has not expanded Medicaid, health coverage for the poor.
- LA and Boston have lower levels of the health score index than Chicago and Brazos, whose residents are less healthy

- In the late May period, Boston was hardest hit in terms of cases per resident. At that point Texas had experienced relatively few cases.
- Boston has the largest Black population, while Los Angeles has the largest Hispanic population and the largest Black+Hispanic population fraction.
- Boston and Brazos have 10% more residents in the 20–39 age group.
- Brazos has a median income about 2/3 that of the other three counties.
- Boston and Brazos have lower levels of air pollution than LA and Chicago.
- Romneycare has driven the percentage of those lacking health insurance in Boston below 5%, while Texas has not expanded Medicaid, health coverage for the poor.
- LA and Boston have lower levels of the health score index than Chicago and Brazos, whose residents are less healthy

- In the late May period, Boston was hardest hit in terms of cases per resident. At that point Texas had experienced relatively few cases.
- Boston has the largest Black population, while Los Angeles has the largest Hispanic population and the largest Black+Hispanic population fraction.
- Boston and Brazos have 10% more residents in the 20–39 age group.
- Brazos has a median income about 2/3 that of the other three counties.
- Boston and Brazos have lower levels of air pollution than LA and Chicago.
- Romneycare has driven the percentage of those lacking health insurance in Boston below 5%, while Texas has not expanded Medicaid, health coverage for the poor.
- LA and Boston have lower levels of the health score index than Chicago and Brazos, whose residents are less healthy

Empirical results

- We first consider a pure autoregressive model, given in the first column of the following table, where confirmed cases are explained by their 14- and 28-day lagged values and spatial factors, capturing proximity effects. The estimated dynamics indicate an explosive process.
- We then consider five specifications in which socioeconomic factors are added, retaining only the 14-day lag as an autoregressive term
- In these models, the estimated autoregressive term always exceeds 1.0, reflecting exponential growth in confirmed cases in counties where the epidemologists' R measure exceeds 1: that is, one infected person infects more than one additional person.

Empirical results

- We first consider a pure autoregressive model, given in the first column of the following table, where confirmed cases are explained by their 14- and 28-day lagged values and spatial factors, capturing proximity effects. The estimated dynamics indicate an explosive process.
- We then consider five specifications in which socioeconomic factors are added, retaining only the 14-day lag as an autoregressive term
- In these models, the estimated autoregressive term always exceeds 1.0, reflecting exponential growth in confirmed cases in counties where the epidemologists' R measure exceeds 1: that is, one infected person infects more than one additional person.

Empirical results

- We first consider a pure autoregressive model, given in the first column of the following table, where confirmed cases are explained by their 14- and 28-day lagged values and spatial factors, capturing proximity effects. The estimated dynamics indicate an explosive process.
- We then consider five specifications in which socioeconomic factors are added, retaining only the 14-day lag as an autoregressive term
- In these models, the estimated autoregressive term always exceeds 1.0, reflecting exponential growth in confirmed cases in counties where the epidemologists' R measure exceeds 1: that is, one infected person infects more than one additional person.

Table: Spatial models of COVID-19 Cases: I

	(1)	(2)	(3)	(4)	(5)	(6)
Ct14	1.187***	1.126***	1.118***	1.124***	1.116***	1.129***
	(125.12)	(203.90)	(202.34)	(201.79)	(201.90)	(204.79)
Ct28	-0.114***					
	(-7.57)					
pct_male		6.028***			6.214***	
		(4.94)			(5.10)	
pct_black			1.903***		2.012***	
			(8.04)		(8.48)	
pct_hisp			1.455***		1.326***	
			(6.15)		(5.57)	
pct_20_30				3.216***		
				(4.84)		
medinc						-0.963***
						(-3.99)
CDIST						
Ct	0.0463***	0.0354***	0.0100	0.0289***	0.0139	0.0368***
	(4.34)	(3.26)	(0.91)	(2.68)	(1.26)	(3.39)
e.Ct	0.248***	0.304***	0.321***	0.298***	0.328***	0.292***
	(7.51)	(9.35)	(9.85)	(9.15)	(10.10)	(8.95)
PseudoR2	0.937	0.936	0.937	0.936	0.938	0.936

z statistics in parentheses

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

- In all models, the test for joint significance of the spatial lag terms rejects its null hypothesis, as does the test for the overall model versus the null model
- The fraction of male residents, by itself or in conjunction with racial effects, predicts higher confirmed cases
- The two racial minority terms also predict higher confirmed cases, cet.par.
- The fraction of the county population in their 20s and 30s predicts higher confirmed cases
- Higher median income at the county level is associated with lower confirmed cases, cet.par.
- These findings all support the hypotheses that socioeconomic factors, combined with geography, are predictive of the spread of COVID-19

- In all models, the test for joint significance of the spatial lag terms rejects its null hypothesis, as does the test for the overall model versus the null model
- The fraction of male residents, by itself or in conjunction with racial effects, predicts higher confirmed cases
- The two racial minority terms also predict higher confirmed cases, cet.par.
- The fraction of the county population in their 20s and 30s predicts higher confirmed cases
- Higher median income at the county level is associated with lower confirmed cases, cet.par.
- These findings all support the hypotheses that socioeconomic factors, combined with geography, are predictive of the spread of COVID-19

- In all models, the test for joint significance of the spatial lag terms rejects its null hypothesis, as does the test for the overall model versus the null model
- The fraction of male residents, by itself or in conjunction with racial effects, predicts higher confirmed cases
- The two racial minority terms also predict higher confirmed cases, cet.par.
- The fraction of the county population in their 20s and 30s predicts higher confirmed cases
- Higher median income at the county level is associated with lower confirmed cases, cet.par.
- These findings all support the hypotheses that socioeconomic factors, combined with geography, are predictive of the spread of COVID-19

- In all models, the test for joint significance of the spatial lag terms rejects its null hypothesis, as does the test for the overall model versus the null model
- The fraction of male residents, by itself or in conjunction with racial effects, predicts higher confirmed cases
- The two racial minority terms also predict higher confirmed cases, cet.par.
- The fraction of the county population in their 20s and 30s predicts higher confirmed cases
- Higher median income at the county level is associated with lower confirmed cases, cet.par.
- These findings all support the hypotheses that socioeconomic factors, combined with geography, are predictive of the spread of COVID-19

- In all models, the test for joint significance of the spatial lag terms rejects its null hypothesis, as does the test for the overall model versus the null model
- The fraction of male residents, by itself or in conjunction with racial effects, predicts higher confirmed cases
- The two racial minority terms also predict higher confirmed cases, cet.par.
- The fraction of the county population in their 20s and 30s predicts higher confirmed cases
- Higher median income at the county level is associated with lower confirmed cases, cet.par.
- These findings all support the hypotheses that socioeconomic factors, combined with geography, are predictive of the spread of COVID-19

- In all models, the test for joint significance of the spatial lag terms rejects its null hypothesis, as does the test for the overall model versus the null model
- The fraction of male residents, by itself or in conjunction with racial effects, predicts higher confirmed cases
- The two racial minority terms also predict higher confirmed cases, cet.par.
- The fraction of the county population in their 20s and 30s predicts higher confirmed cases
- Higher median income at the county level is associated with lower confirmed cases, cet.par.
- These findings all support the hypotheses that socioeconomic factors, combined with geography, are predictive of the spread of COVID-19

We then considered the additional health-related factors in a second set of six models.

Table: Spatial models of COVID-19 Cases: II

	(1)	(2)	(3)	(4)	(5)	(6)
Ct14	1.127***	1.129***	1.127***	1.129***	1.125***	1.117***
	(203.74)	(204.79)	(204.77)	(204.17)	(204.15)	(201.77)
pm25	5.252**				8.496***	5.414**
	(2.38)				(3.69)	(2.34)
medinc		-0.963***				
		(-3.99)				
unins			2.999***		3.660***	
			(4.59)		(5.40)	
index_zscore				8.486***		
				(2.74)		
pct_black						1.809***
						(7.54)
pct_hisp						1.573***
						(6.51)
CDIST						
Ct	0.0267**	0.0368***	0.0318***	0.0311***	0.0252**	0.00687
	(2.43)	(3.39)	(2.94)	(2.90)	(2.30)	(0.62)
e.Ct	0.297***	0.292***	0.303***	0.286***	0.309***	0.324***
	(9.07)	(8.95)	(9.29)	(8.74)	(9.44)	(9.92)
PseudoR2	0.936	0.936	0.936	0.936	0.936	0.937

z statistics in parentheses
* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

- In all models, the test for joint significance of the spatial lag terms rejects its null hypothesis, as does the test for the overall model versus the null model
- Higher levels of air pollution (pm25) are associated with higher confirmed cases
- A higher fraction of county residents lacking health insurance increases confirmed cases, cet.par.
- The health score index of complicating health conditions has a very strong effect in predicting higher confirmed cases
- Combinations of these factors are also predictive, with individual factors retaining statistical significance
- These findings all support the hypotheses that health-related factors, combined with geography, are predictive of the spread of COVID-19

- In all models, the test for joint significance of the spatial lag terms rejects its null hypothesis, as does the test for the overall model versus the null model
- Higher levels of air pollution (pm25) are associated with higher confirmed cases
- A higher fraction of county residents lacking health insurance increases confirmed cases, cet.par.
- The health score index of complicating health conditions has a very strong effect in predicting higher confirmed cases
- Combinations of these factors are also predictive, with individual factors retaining statistical significance
- These findings all support the hypotheses that health-related factors, combined with geography, are predictive of the spread of COVID-19

- In all models, the test for joint significance of the spatial lag terms rejects its null hypothesis, as does the test for the overall model versus the null model
- Higher levels of air pollution (pm25) are associated with higher confirmed cases
- A higher fraction of county residents lacking health insurance increases confirmed cases, cet.par.
- The health score index of complicating health conditions has a very strong effect in predicting higher confirmed cases
- Combinations of these factors are also predictive, with individual factors retaining statistical significance
- These findings all support the hypotheses that health-related factors, combined with geography, are predictive of the spread of COVID-19

- In all models, the test for joint significance of the spatial lag terms rejects its null hypothesis, as does the test for the overall model versus the null model
- Higher levels of air pollution (pm25) are associated with higher confirmed cases
- A higher fraction of county residents lacking health insurance increases confirmed cases, cet.par.
- The health score index of complicating health conditions has a very strong effect in predicting higher confirmed cases
- Combinations of these factors are also predictive, with individual factors retaining statistical significance
- These findings all support the hypotheses that health-related factors, combined with geography, are predictive of the spread of COVID-19

- In all models, the test for joint significance of the spatial lag terms rejects its null hypothesis, as does the test for the overall model versus the null model
- Higher levels of air pollution (pm25) are associated with higher confirmed cases
- A higher fraction of county residents lacking health insurance increases confirmed cases, cet.par.
- The health score index of complicating health conditions has a very strong effect in predicting higher confirmed cases
- Combinations of these factors are also predictive, with individual factors retaining statistical significance
- These findings all support the hypotheses that health-related factors, combined with geography, are predictive of the spread of COVID-19

- In all models, the test for joint significance of the spatial lag terms rejects its null hypothesis, as does the test for the overall model versus the null model
- Higher levels of air pollution (pm25) are associated with higher confirmed cases
- A higher fraction of county residents lacking health insurance increases confirmed cases, cet.par.
- The health score index of complicating health conditions has a very strong effect in predicting higher confirmed cases
- Combinations of these factors are also predictive, with individual factors retaining statistical significance
- These findings all support the hypotheses that health-related factors, combined with geography, are predictive of the spread of COVID-19

Conclusions and extensions

- This paper is only the first step in considering how spatial econometric techniques can be combined with standard regression modeling in the context of daily-frequency panel data.
- These models illustrate the potential for spatial modeling to enhance our understanding of a fast-moving dynamic process, moderated by a number of quasi-fixed socioeconomic and demographic measures.
- We will extend the analysis from late May through the summer months, which are revealing strong shifts in the geographic patterns of confirmed cases with a resurgence of the pandemic.
- We are also exploring additional county-level data to better capture the interactions between health risk factors, medical resources and Medicaid eligibility.

Conclusions and extensions

- This paper is only the first step in considering how spatial econometric techniques can be combined with standard regression modeling in the context of daily-frequency panel data.
- These models illustrate the potential for spatial modeling to enhance our understanding of a fast-moving dynamic process, moderated by a number of quasi-fixed socioeconomic and demographic measures.
- We will extend the analysis from late May through the summer months, which are revealing strong shifts in the geographic patterns of confirmed cases with a resurgence of the pandemic.
- We are also exploring additional county-level data to better capture the interactions between health risk factors, medical resources and Medicaid eligibility.

Conclusions and extensions

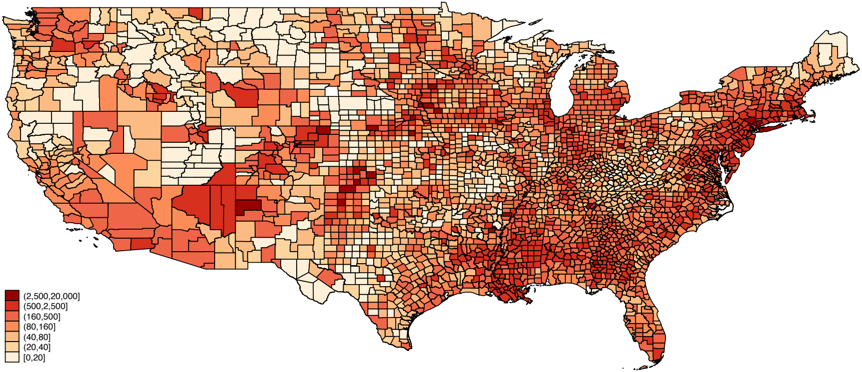
- This paper is only the first step in considering how spatial econometric techniques can be combined with standard regression modeling in the context of daily-frequency panel data.
- These models illustrate the potential for spatial modeling to enhance our understanding of a fast-moving dynamic process, moderated by a number of quasi-fixed socioeconomic and demographic measures.
- We will extend the analysis from late May through the summer months, which are revealing strong shifts in the geographic patterns of confirmed cases with a resurgence of the pandemic.
- We are also exploring additional county-level data to better capture the interactions between health risk factors, medical resources and Medicaid eligibility.

Conclusions and extensions

- This paper is only the first step in considering how spatial econometric techniques can be combined with standard regression modeling in the context of daily-frequency panel data.
- These models illustrate the potential for spatial modeling to enhance our understanding of a fast-moving dynamic process, moderated by a number of quasi-fixed socioeconomic and demographic measures.
- We will extend the analysis from late May through the summer months, which are revealing strong shifts in the geographic patterns of confirmed cases with a resurgence of the pandemic.
- We are also exploring additional county-level data to better capture the interactions between health risk factors, medical resources and Medicaid eligibility.

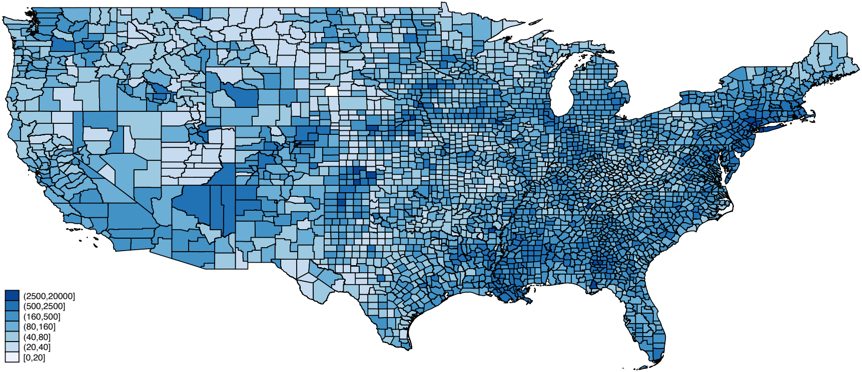
Confirmed Cases of COVID-19 in the United States

cases per 100,000 population as of 5/23/2020



Predicted Confirmed Cases of COVID-19 from Spatial Lag Model

cases per 100,000 population as of 5/24/2020



September 2020 update

To provide a more recent perspective on COVID-19 spread in the US, the state of Texas and that state's Brazos County, in which StataCorp's offices are located, we present a graph that StataCorp's Chuck Huber recently posted to Facebook,[1] based on smoothed Johns Hopkins data.

You might want to consider two hypotheses:

- Masks reduce the spread of COVID-19.
- Without strong restrictions on behavior, reopened college campuses may spread the virus.

[1] https://www.facebook.com/chuck.huber1/posts/10222655585731442

September 2020 update

To provide a more recent perspective on COVID-19 spread in the US, the state of Texas and that state's Brazos County, in which StataCorp's offices are located, we present a graph that StataCorp's Chuck Huber recently posted to Facebook,[1] based on smoothed Johns Hopkins data.

You might want to consider two hypotheses:

- Masks reduce the spread of COVID-19.
- Without strong restrictions on behavior, reopened college campuses may spread the virus.

[1] https://www.facebook.com/chuck.huber1/posts/10222655585731442

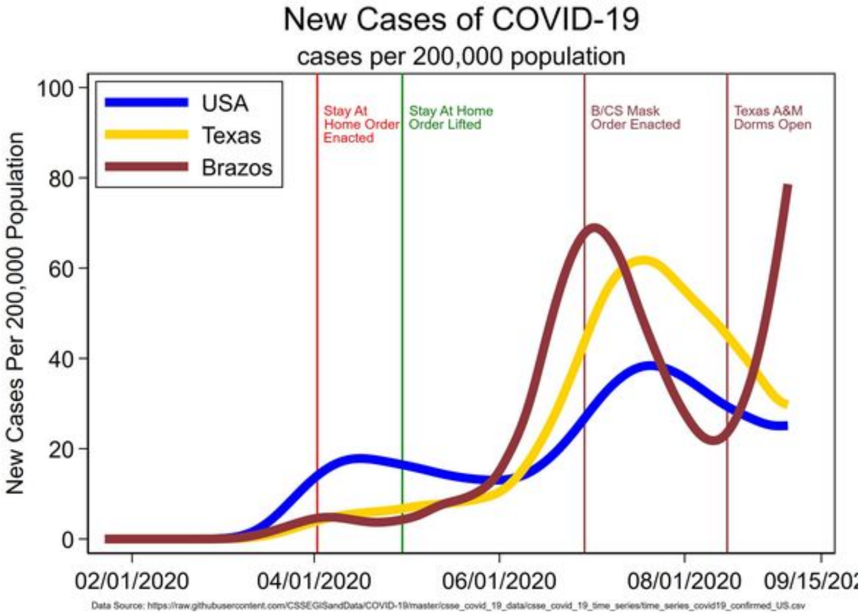
September 2020 update

To provide a more recent perspective on COVID-19 spread in the US, the state of Texas and that state's Brazos County, in which StataCorp's offices are located, we present a graph that StataCorp's Chuck Huber recently posted to Facebook,[1] based on smoothed Johns Hopkins data.

You might want to consider two hypotheses:

- Masks reduce the spread of COVID-19.
- Without strong restrictions on behavior, reopened college campuses may spread the virus.

[1] https://www.facebook.com/chuck.huber1/posts/10222655585731442



New Cases of COVID-19
cases per 200,000 population
New Cases Per 200,000 Population
100
80
60
40
20
0
USA
Texas
Brazos
Stay At Home Order Enacted
Stay At Home Order Lifted
B/CS Mask Order Enacted
Texas A&M Dorms Open
02/01/2020
04/01/2020
06/01/2020
08/01/2020
09/15/202
Data Source: https://raw.githubusercontent.com/CSSEGISandData/COVID-19/master/csse_covid_19_data/csse_covid_19_time_series/time_series_covid19_confirmed_US.csv