

Segregation and the onset of COVID-19 in American cities

Francesco Andreoli

*Department of Economics, University of Verona, Verona, Italy and
Luxembourg Institute of Socio-Economic Research, Esch-sur-Alzette, Luxembourg*

Vincenzo Prete

Department of Law, University of Palermo, Palermo, Italy, and

Claudio Zoli

Department of Economics, University of Verona, Verona, Italy

Abstract

Purpose – This paper investigates one of the potential costs of rising segregation in American cities by evaluating empirically the extent at which ethnic-based segregation contributes to the onset and the speed of propagation of the COVID-19 pandemic.

Design/methodology/approach – Regression analysis based on matched data on early incidence of COVID-19 cases, segregation and covariates. Identification resorts on variations in segregation across MSAs and heterogeneity in the geography and timing of stay-at-home orders.

Findings – One cross-MSA standard deviation increase in segregation leads to a significant and robust rise of COVID-19 cases of 8.7 per 100,000 residents across urban counties.

Originality/value – Combines spatial data on COVID-19 cases and segregation; use of a new segregation measure; focus on early incidence of the pandemic and its drivers.

Keywords Segregation, COVID-19, American community survey

Paper type Research paper

1. Introduction

There has been substantial research about the biological causes and consequences of the Coronavirus pandemic, the process of its transmission, the underlying mechanisms, alongside the economic and social consequences of its spreading. In social sciences, the COVID-19 epidemics resulting from the global spreading of the virus has been treated as a health shock, and its implications at individual and societal levels investigated (for a review, see [Rathnayaka et al., 2023](#)).

Understanding the role of mediating factors that may have contributed to the onset of COVID-19 is equally important from the perspective of public health and the economy as a whole. This paper contributes with evidence that identifies ethnic-based segregation in American cities as a driver for the onset and early spreading of the COVID-19 infection. [Wu and McGoogan \(2020\)](#) highlight the relevance of demographics and income as potential drivers. Housing characteristics, population density, and the extent of urbanization are found to be relevant for explaining the likelihood of exposure to the disease. This happens because

JEL Classification — R23, D63, D85

© Francesco Andreoli, Vincenzo Prete and Claudio Zoli. Published by Emerald Publishing Limited. This article is published under the Creative Commons Attribution (CC BY 4.0) licence. Anyone may reproduce, distribute, translate and create derivative works of this article (for both commercial and non-commercial purposes), subject to full attribution to the original publication and authors. The full terms of this licence may be seen at <http://creativecommons.org/licences/by/4.0/legalcode>

Declarations: The authors have no declarations of interest. Replication code for this paper is made available as supplementary material.



poorer households are typically larger and multi-generational, live in smaller dwellings, and are more likely to use public transport, thereby being exposed to more occasions for human interactions and facing lower chances of self-isolation. [Borjas \(2020\)](#) makes use of zip-code data on testing coverage and COVID-19 onset in New York City to highlight a correlation between the ethnic composition of the neighborhoods and the rate of infection, which is also related to non-randomness in testing.

We conjecture that if the quantity of social interactions is stronger within communities with similar traits, then societies that are more segregated along the same traits also display larger chances of interactions, which become more specialized within the same community. This mechanism has been explored in [Jackson *et al.* \(2023\)](#) and [Kim *et al.* \(2023\)](#), suggesting the possibility of a segregation gradient on COVID-19 spreading.

In this paper, we investigate empirically this hypothesis. Our focus is, in particular, on the dimension of segregation in the space of social interactions. The measure of segregation that we consider aims at capturing heterogeneity in the way people of various ethnic origins who live in different places of the city display different patterns of social interactions with the rest of the population. Segregation arises from the fact that the chances of interacting with people of the same origin or from a different origin are unevenly distributed across the population. We adopt an index of segregation that embodies these features to measure segregation in American cities, for which fine-grained information about the ethnic distribution of urban residents across neighborhoods is available from censuses and surveys. More segregated cities are cities where people tend to interact with larger frequency with other individuals sharing a particular ethnic origin background.

The analysis that follows uncovers evidence supporting our stating hypothesis: segregation, pre-determined with respect to the pandemic, is positively correlated with the onset and speed of increase of cases of COVID-19 at county level that were reported from official sources at the very onset of the pandemic, about March 2020.

2. Relevant literature

This paper adds to a growing body of research that emphasizes the role of social connections as a driver for the COVID-19 pandemic. [Kuchler *et al.* \(2022\)](#) provides empirical support to this conjecture: making use of data from social media as a proxy of social connections between regions, they find that social ties correlate with the spread of the pandemic. The extent at which exposure to social connections can be avoided or contained during the onset of the pandemic crucially depends on the means available, on the policy context and on the strength of the connections.

Concerning the means of avoidance, [Brown and Ravallion \(2020\)](#) find that social distancing remains a relevant driver to reduce the infection spread, albeit the effect being weaker in poorer and more unequal US counties, where the cost of reducing social distancing is higher. Likewise, [Jung *et al.* \(2021\)](#) show that in low-density counties, the incidence of COVID-19 infections rises with the poverty level. Conversely, in densely populated areas with limited opportunities for self-isolation, a U-shaped relationship between county poverty level and infection spread predominates. Moreover, a positive association between mobility and the spread of COVID-19 have been thoroughly documented ([Ghirelli *et al.*, 2023](#)), whereas the relation is accentuated in areas with higher poverty rates and per capita income ([Yilmazkuday, 2023](#)).

Policy actions, taking the forms of stay-at-home orders, have been put in place since March 2020 in order to restrain mobility. Evidence from the 1918 influenza pandemic suggests that strict and timely quarantine measures are among the most effective short-term measures to prevent the spread of epidemic ([Bootsma and Ferguson, 2007](#)). Differences in the timing of implementation of social distancing in American cities around 1918 have been used to

analyze such effects. Only strict and timely quarantine measures have a significant impact on the epidemic incidence and mortality (Bootsma and Ferguson, 2007), while less effective measures may also have a strong effect in reducing the virus attack rate, albeit the effect is channeled by a delaying of the pandemic duration (Markel *et al.*, 2007). The impact of mobility restrictions appears to hinge on the socioeconomic context of the places where these measures are implemented (Berry *et al.*, 2021; Chernozhukov *et al.*, 2021). Specifically, Coven and Gupta (2020) show that residents of low-income neighborhoods in New York City face greater constraints in complying with stay-at-home orders, while Chiou and Tucker (2020) report evidence of higher compliance with social distance directives among residents of higher-income regions. See Brodeur *et al.* (2021) for a survey of determinants of compliance with and effectiveness of social distancing.

Less evidence is available about the impact of the strength of social connections. Social bonds can be measured in many dimensions. Recent contributions have focused on the ethnic dimensions and the way social bond differ across ethnic groups and locations. The degree of ethnic diversity represents a relevant driver of the compliance with, and hence the effectiveness of, social distancing orders. To this regard, Zhai *et al.* (2023) find evidence that people in high ethnically diverse neighborhood show less compliance with mobility restrictions and are less likely to reduce social interactions. More generally, residential segregation has been a significant factor in rising mortality and infection rates within US counties, with Black and Latino residents exposed to higher mortality rate than white residents (Torrats-Espinosa, 2021; Trounstein and Goldman-Mellor, 2023). Moreover, the interplay of income inequality and segregation exacerbates the toll, amplifying the adverse health outcomes and resulting in a greater loss of life (Yu *et al.*, 2021).

Differently from the aforementioned contributions, this paper examines the relation between COVID-19 onset and city-level segregation in the domain of social interactions.

3. Methodology

3.1 Measuring exposure segregation

Exposure segregation is conceived as *inequality* in the distribution of interaction profiles with the relevant social groups. In this application, the focus is on groups defined on the basis of their ethnic origin. An interaction profile measures the probabilities that a given *individual* i has to interact with each of the G ethnic groups that we consider. Denote such probability $\pi_{gi} \in [0, 1]$, with $\sum_g \pi_{gi} = 1$ and $g \in G$. An interaction profile is depicted by column vector π_i . There are multiple dimensions that may affect interaction profiles. It is common in applied analysis to assume that all individuals observed in the same *location* have the same interaction probabilities and that spatial proximity can be used as a proxy for social proximity. If i are units, such as neighborhoods or ZIP codes, one can compute interaction profiles for each unit by looking at the ethnic composition of the unit and neighboring units.

Our goal is to assess exposure segregation in a specific city. In our application, i stands for a census tract of the city, and π_{gi} is the probability that individuals of census tract i interact with individuals from group g in that tract or in neighboring tracts. We adopt a spatial proximity interaction model as in Andreoli (2014) to estimate the interaction probabilities π_{gi} based on the American Community Survey data. Within a city, tracts differ in demographic size. The weights $w_i \in [0, 1]$ measure the population size of tract i as a share of the city population. Each city displays a specific social composition, represented by the group-specific probability $\bar{\pi}_g = \sum_i w_i \pi_{gi}$. It measures the probability that any random person in the city has to interact with an individual of the group g , given the distribution of groups across the city's traits. Cities differ in population size, density, and group composition. To account for such sources of within and across municipality heterogeneity, it is proposed to weight interaction probabilities by the weight of each unit—measuring the probability that a person

from that unit i is observed in the city population—and scaling that value by the average probability of interaction. This leads to the *interaction likelihood* $\ell_{gi} := w_i \frac{\pi_{gi}}{\pi_g}$. From the Bayes rule, $\ell_{ig} \in [0, 1]$ and $\sum_i \ell_{gi} = 1$.

Interaction likelihoods can be understood as the chances that the population from group g has to interact with a random inhabitant of unit i . For a given unit, we can represent these probabilities by the column vector ℓ_i . Segregation accrues from the extent to which these probabilities differ across groups.

A segregation *index* is a function that maps information provided by interaction likelihoods into a number, which we regard as the level of exposure segregation displayed by the city. Our reference measure is the multi-group *Gini Exposure Segregation* index $GS : \mathcal{M}_{GN} \rightarrow [0, 1]$. The index has been introduced by Andreoli (2014) and is inspired by the multidimensional extension of the Gini index by Koshevoy and Mosler (1997) (see also Koshevoy, 1995; Koshevoy and Mosler, 1996). For a given city, the index measures inequality in the distribution of normalized profiles ℓ_i across the population of units. Formally:

$$GS := \frac{1}{G!} \sum_{\forall \{i_1, \dots, i_G\} \subseteq \mathcal{N}} |\det(\ell_{i_1} \ \dots \ \ell_{i_G})|, \quad (1)$$

where $\{i_1, \dots, i_G\}$ is any subset of G units drawn at random without replacement among N units available in the city (we limit our analysis to cities where $N > G$, the number of tracts is larger than the number of social groups).

The index displays some important features. When all units display the same interaction profiles $\pi_i = \pi$, then all vectors ℓ_i can be expressed as linear combinations of one of the others, implying $GS = 0$. This is arguably the case in which there is a complete lack of segregation, and the residents of any unit i have the same chances as any other unit to interact with each of the G groups. Conversely, if for every i there exists a group γ such that $\pi_{i\gamma} = 1$ and $\pi_{ig} = 0$ for all other groups except γ , then $GS = 1$. Arguably, this is the case of maximal segregation in which residents of unit i interact with one and only one group, γ . Lastly, the index has been shown (see Andreoli, 2014 and Andreoli and Zoli, 2015) to be the unique index satisfying some desirable decomposition properties and behaving consistently with basic transformations of the data that are unambiguously understood as being segregation-reducing.

In the application, we compute the GS index for each American city considered in this study, and we interpret variations of the index as a measure of segregation in the city, with larger values of GS corresponding to more segregated cities.

3.2 Estimation strategy

In the absence of individual-level information about the incidence of COVID-19 infections, we can only rely on average levels of COVID-19 cases and observable covariates aggregated at the county level, the finer spatial partition for which all data are available. Using counties as units allows us to account for larger heterogeneity in healthcare availability across finer-scale partition of the territory, thereby reducing the risk that variability in COVID-19 cases reporting is driven by access to care. Second, county estimates provide enough within-MSA variation in COVID-19 cases.

Our analysis considers two periods. The earliest period encompasses the onset of COVID-19 outbreaks up to March 29, 2020, which is ten days (roughly corresponding to the average incubation period of the virus) after the first stay-at-home order was implemented in California. Therefore, COVID-19 incidence data on this date are likely unaffected by the introduction of early lock-down policies, providing insights into the initial patterns of the pandemic's evolution in response to local characteristics. The second period concludes on April 13, 2020, approximately ten days after the enforcement of lock-down restrictions across

all US states. We examine the number of new cases as of April 13 and their speed of variation over one- and two-week intervals to assess the persistence of segregation effects on COVID-19 incidence subsequent to the implementation of lock-down policies.

We employ regression analysis to study the partial correlation between exposure segregation and the evolution of COVID-19 across American counties. While COVID-19 measures change over time, the main treatment, exposure segregation, and the relevant set of covariates are pre-determined and static with respect to the evolution of COVID-19 cases. For this reason, our estimates only rely on cross-sectional variability of segregation and COVID-19 across MSAs and counties, whereas we explore the longitudinal features of the data to assess and update our estimates at different timings of early onset of the pandemic.

The estimating equation is:

$$Y_{cms}(t) = \alpha + \beta_1 GS_m + \beta_2 GS_m \cdot D_c(t - 7) + \beta_3 GS_m \cdot D_c(t - 15) + \delta X_{cm} + \theta_s + \varepsilon_{cmst}, \quad (2)$$

where $Y_{cms}(t)$ is a measure of COVID-19 cases in county c located in MSA m in State s observed at date t , whereas GS_m is the Gini Exposure Segregation index observed in MSA m and normalized by its standard deviation in the sample. $D_c(t - 7)$ and $D_c(t - 15)$ are dummies taking values 1 if county c had experienced respectively a stay-at-home order one week and two weeks before the reference date t (March 29 or April 13). When $t = \text{March } 29$, we set $D_c(t - 15) = 0$ for all counties, switching off any interaction term with the Gini exposure index ($GS_m \cdot D_c(t - 15) = 0$) for all counties. Covariates are collected on the vector X_{cm} . The model features State fixed effects, which capture differences in monitoring practices across healthcare structures.

Model (2) is estimated with OLS when the dependent variable is continuous, such as for cases normalized by the resident population. We use Poisson regression when the dependent variable corresponds to counts. We further utilize logistic regressions to analyze the impact of changes in segregation on the probability of COVID-19 infection outbreaks across counties.

Independently on the way models are estimated, the effects of interest relate to coefficients β_1 , β_2 and β_3 . More specifically, β_1 captures the effect of increasing segregation (i.e. by one standard deviation of the GS index) on new and cumulative COVID-19 cases. Coefficients β_2 and β_3 capture the additional contribution of variations in segregation originating within MSAs that have adopted lock-down measures earlier during the pandemic.

Identification hinges on a strong exogeneity assumption, positing that variations in segregation within a State are unrelated to unobservable factors influencing the pandemic and potentially associated with other variables, such as mobility restrictions imposed in a specific county. This assumption could be compromised if factors related, for instance, to the quality of healthcare services in a State (which we do not observe and that are related to the effectiveness and extent of COVID-19 monitoring) also play a significant role in driving early lock-down policies, often enacted to prevent over-crowding of ICU units. We address such issues by controlling for stay-at-home adoptions on different days by county and separating the analysis by period of analysis (the issue being more relevant for later dates in which COVID-19 is measured).

4. Data

Data on ethnic composition and covariates are from the American Community Survey (ACS) 5-year modules, covering urban counties in 366 Metropolitan Statistical Areas (MSA). Estimates of segregation are based on the 2014–2018 modules. We consider these estimates as pre-determined to the onset of the COVID-19 epidemic and, as such, they are not impacted by the restrictions put in place during and after the COVID-19 pandemic.

In 2018, the White population was 60% of urban residents on averages, and more than 25% of American MSAs now display a clear non-White majority. The Hispanic population

counts for almost 20% of the population, while the average proportion of urban Black population at about 10% and the Asian population experienced to 7% of the total urban population. Table 1 describes the Gini Exposure Segregation index in 2018 for the median, bottom and top quartile cities as ranked by the level of the index in that year.

Table 2 reports descriptive statistics of our sample based on 1,087 counties. As additional controls, we consider: (A) demographics (which includes population size, density, indicators for 10 and 5 largest cities, percentages of Black, Hispanics, Asians and Whites as well as share of individuals aged 65 or more); (B) housing market characteristics (such as the share of owner-occupied houses, the share of old houses, the median value and rent of houses and their variability within a census tract); (C) human capital (indicators for college towns and for the presence of students, average education and mean, median and variance of income); (D) access to health insurance; (E) ethnic segregation and poverty at county level as measured by dissimilarity indices for Black, Asian, Hispanic and Native origin groups. In panel (F) we consider further multi-group spatial segregation measures which are largely adopted in the literature (see Reardon and O’Sullivan, 2004). Figure A1 in the appendix shows that all indices are rank-correlated to the Gini Exposure index, albeit such correlation is very blurry and mostly driven by few large cities. The Figure highlights the fact that the Gini Exposure index captures features of segregation in American MSA that are, to a large extent, independent from what is measured by alternative indices.

The data on confirmed COVID-19 cases are from the Centers for Disease Control and Prevention and from the state and local level public agencies. These data have been recorded at county level on the New York Times dedicated data repository [1]. Data for geographic exceptions have been collected from USAFact online repository [2]. Additional data about legislation on stay-at-home and shelter-in-place orders have been collected from the New York Times interface [3].

As of March 29, 5% of the 3,220 American counties (for which data are available) record at least 100 cases tested positive to COVID-19, with a total of 136,820 cumulative positive cases. The average incidence of COVID-19 is 13.3 cases (s.d. 38.2) per 100,000 residents across counties, and 495 counties display an incidence larger than 20 cases per 100 k residents. Incidence of positive cases of COVID-19 is larger in urban counties, 1,087 in our matched MSA-county database (each MSA may gather more than one county). Among all urban counties, the average incidence is 21.7 cases per 100,000 residents, and in more than 25% of counties incidence is larger than 20 cases per 100,000 residents.

5. Results

Table 3 reports estimates of different models for the incidence of COVID-19 in American counties. Models (1)–(4) analyze the correlation between the total number of positive COVID-19 cases and the MSA-level Gini Exposure index, normalized by its standard

Measure	All MSAs			NY city (4)	LA, CA (5)	Largest MSAs		Houston, TX (8)
	Bottom quartile (1)	Median (2)	Top quartile (3)			Chicago, IL (6)	Dallas, TX (7)	
GS_m	0.083	0.116	0.157	0.255	0.185	0.256	0.228	0.159

Note(s): Bottom, median and top quartile refer to the distribution of the Gini Exposure index across the 325 MSA used in this study

Source(s): Authors own work based on ACS data

Table 1.
Gini exposure segregation for American MSAs in 2018

	Mean	S.d.	Bottom quartile	Median	Top quartile
Gini exposure (SD units)	2.66	1.08	1.80	2.56	3.33
<i>(A) Demographics</i>					
Population size (ln)	11.42	1.36	10.39	11.47	12.30
Pop. density (ln)	5.14	1.48	4.10	5.06	6.02
10 larger MSA	0.13	0.33	0.00	0.00	0.00
5 larger MSA	0.06	0.23	0.00	0.00	0.00
Black (%)	0.11	0.14	0.02	0.06	0.16
Hispanic (%)	0.10	0.13	0.03	0.05	0.11
Asian (%)	0.02	0.04	0.01	0.01	0.03
White (%)	0.75	0.19	0.64	0.80	0.91
Aged 65+ (%)	0.16	0.04	0.13	0.16	0.18
<i>(B) Housing market</i>					
Old houses, 20+ year (%)	0.64	0.14	0.56	0.65	0.74
Owner occupied (%)	0.70	0.09	0.65	0.72	0.77
Median value house by CT (ln)	10.69	1.64	9.55	10.65	12.12
S.d. value house by CT (ln)	10.70	0.82	10.27	10.74	11.19
Median rent by CT (ln)	4.53	1.62	3.37	4.49	6.08
S.d. rent by CT (ln)	4.91	0.76	4.47	4.99	5.42
<i>(C) Human capital</i>					
College-town	0.28	0.45	0.00	0.00	1.00
Student-town	0.14	0.34	0.00	0.00	0.00
Less than high school (%)	3.23	1.93	1.91	2.70	4.09
With college (%)	38.79	7.18	33.85	38.86	43.64
Employed (%)	0.92	0.03	0.91	0.93	0.94
Commute < 1/2 hour (%)	28.41	7.05	23.18	27.81	33.26
Avg. hh income (ln)	11.05	0.27	10.87	11.03	11.20
Median hh income by CT (ln)	8.92	1.53	7.85	8.95	10.37
S.d. hh income by CT (ln)	9.53	0.63	9.22	9.62	9.96
<i>(D) Health access</i>					
Without health insurance (%)	11.20	4.45	7.93	10.88	13.96
Health insurance among 65+ (%)	15.44	3.96	12.78	15.24	17.43
<i>(E) Ethnic segregation and poverty</i>					
Incidence of poor	0.15	0.06	0.11	0.14	0.18
Dissimilarity black	0.44	0.10	0.36	0.45	0.52
Dissimilarity hispanic	0.35	0.08	0.31	0.36	0.40
Dissimilarity asian	0.47	0.08	0.42	0.47	0.52
Dissimilarity white	0.14	0.08	0.07	0.12	0.19
<i>(F) Multi-group segregation measures</i>					
Dissimilarity	0.45	0.08	0.40	0.45	0.51
Entropy	0.22	0.08	0.16	0.21	0.27
Gini segregation	0.60	0.10	0.54	0.61	0.67
Squared coefficient of variation	0.10	0.05	0.06	0.10	0.13
Relative diversity	0.20	0.10	0.11	0.21	0.27
Normalized exposure	0.22	0.11	0.13	0.22	0.30
N	1,087				

Note(s): Bottom, median and top quartile refer to the distribution of each characteristic across the 1,087 counties used in this study

Source(s): Authors own work based on ACS data

Table 2.
Descriptive statistics

Dependent variable: Measure: Time frame: Method:	COVID-19 cumulative cases			COVID-19 new cases			Onset
	#Positives/100 k	#Positives	1 day	#Positives/100 k	7 days	15 days	#Positives > 100
	OLS (2)	Poisson (4)	(5)	OLS (6)	(7)	(8)	Logit (8)
Gini exposure (SD units)	13.01* (6.65)	8.77** (4.34)	19.78** (5.80)	6.70** (3.30)	5.87* (3.08)		-0.02** (0.01)
Stay-at-home order in t		-18.87** (8.96)	-3.07** (1.14)	-14.88** (6.91)	-14.55** (6.27)		0.30** (0.05)
Stay-at-home order in $t-7$		-57.15** (23.26)	-4.31* (2.43)	-40.21** (17.32)	-31.75** (11.56)		-0.16* (0.10)
Controls for (A), (B), (C), (D), (E)	✓	✓	✓	✓	✓	✓	✓
State fixed effect	1,136	1,087	1,087	1,087	1,087	1,087	944
N	0.327	0.540	0.498	0.532	0.340		
R -squared							

Note(s): Using sample limited to urban counties. Robust standard errors clustered at MSA level. Estimated effects have to be interpreted as variations in cases per 100,000 residents in the county. Models (4) and (8) report marginal effects at the average. SD stands for standard deviation units. Significance levels: * = 10% and ** = 5%.

Source(s): Authors own work based on ACS data, the New York Times online repository (all counties) and USAFact data collection (New York). Data extracted on April 14, 2020.

Table 3.
Exposure segregation
and COVID-19 – early
incidence, March
29, 2020

deviations (SD) across the using sample. Model (1) controls only for the level of Gini Exposure Segregation, all the other models control for relevant county-level covariates. According to Model (3), our preferred specification, one SD increment in the Gini Exposure corresponds to a significant increase of 8.77 cases per 100,000 residents. The effect is robust to a large number of controls and to State fixed effects, capturing differences in testing policies across States jurisdictions. Additionally, the effect is robust with respect to controls for whether stay-at-home orders were in force on March 29 or earlier (Model (2)) and closely matches the unconditional estimate reported in Model (1).

The significance of the effects of the Gini Exposure index hold when regressions are augmented with controls for alternative measures of segregation widely adopted in the literature. This result suggests that variations in the Gini Exposure index are informative of features of the distribution of ethnic groups across American cities that is relevant for explaining COVID-19 onset and early evolution and that cannot be captured by alternative measures of segregation.

In Model (4), we use a zero-inflated count data model to assess the impact of the Gini Exposure index on county-level COVID-19 incidence. The estimated effect is significant, indicating that a one SD increment in the Gini Exposure index is associated with an increase of 19.7 cases in COVID-19 incidence.

Models (5)–(7) show that one SD increase in the Gini Exposure index is positively associated with the *variation* in COVID-19 cases, the effect ranging from 1 to about 6 cases per 100,000 residents, depending on the period over which the variation is measured (daily, weekly or 15 days). Finally, Model (8) provides evidence that segregation is associated with a lower probability (about –2%) of COVID-19 onset in the county.

One explanation for the effects displayed in [Table 3](#) is that rising heterogeneity in interaction profiles (hence more segregation) may impact the quantity and quality of social interactions. In more segregated MSAs, residents are more likely to interact with similar individuals (here, from the same ethnic group), increasing overall social interactions. Moreover, limited exposure to other groups reduces access to information about epidemic spreading and relevant habits mitigating risk factors. Both occurrences raise the probability of infection and contribute to the virus spread.

[Table 4](#) investigates whether exposure segregation has implications for the effectiveness of lock-down policies. We use indicators for whether such policies were in place on April 13 (t), on April 6 (t-7) and on March 28 (t-15) in each county as additional controls and interact these indicators with the Gini Exposure level in order to evaluate the partial association with COVID-19 incidence across American urban counties. We use the same set of controls as in Model (3) in [Table 3](#) and additionally control for fixed effects attributable to New York MSA.

As expected, counties that have implemented earlier restrictions display lower COVID-19 incidence, the effect being significant across specifications only for stay-at-home orders made 15 or more days earlier than April 13. The extent to which such policy reduces the COVID-19 incidence is mitigated by the degree of exposure segregation in the city, with rising Gini Exposure that offsets (by about one-fourth in magnitude) the effects of early lock-down restrictions on the epidemic spreading. This effect suggests that the extent of exposure segregation may act as a mitigating factor which delays the effects of intervention against the COVID-19 spreading.

Our results hold even after controlling for alternative dimensions of ethnic segregation that are not captured by the Gini Exposure index. In [Tables A1 and A2](#) in the Appendix we report estimates from the same regressions in [Tables 3 and 4](#), where we additionally control for the multi-group segregation indices described in [Figure A1](#). Estimated regression coefficients for the Gini Exposure index are unaffected in sign, magnitude and significance after controlling for multi-group segregation, thereby validating our assumption that rising

Dependent variable: Measure: Time frame: Method:	COVID-19 cumulative cases			COVID-19 new cases			Onset #Positives > 100 Logit (8)
	#Positives/100 k OLS (2)	#Positives Poisson (4)	1 day (5)	#Positives/100 k 7 days OLS (6)	15 days (7)		
Gini exposure (SD units)	6.01 (8.68)	6.23 (9.26)	-0.34 (0.69)	-1.17 (4.31)	-1.08 (8.56)	0.01 (0.03)	
Gini exp. × stay-at-home t	-0.60 (8.21)	3.01 (11.81)	0.77 (0.90)	6.44 (5.08)	11.10 (10.34)	0.03 (0.08)	
Gini exp. × stay-at-home $t-7$	-18.35 (12.21)	-74.82 (84.15)	-1.48* (0.79)	-10.71** (4.89)	-18.25* (9.46)	-0.05 (0.08)	
Gini exp. × stay-at-home $t-15$	26.68** (12.99)	98.85** (29.42)	1.39** (0.61)	9.78** (4.45)	15.89* (8.58)	-0.01 (0.02)	
Stay-at-home policy in t	-15.99 (30.26)	-29.21 (44.61)	-6.47* (3.46)	-35.05* (18.84)	-27.16 (35.64)	0.51* (0.31)	
Stay-at-home policy in $t-7$	-0.33 (19.20)	38.21 (45.47)	6.45** (3.06)	31.46* (18.41)	32.06 (37.45)	0.24 (0.27)	
Stay-at-home policy in $t-15$	-21.39* (12.95)	-91.28** (38.19)	-6.93** (2.44)	-38.42** (14.00)	-52.14** (21.27)	-0.08 (0.08)	
Controls for (A), (B), (C), (D), (E)	✓	✓	✓	✓	✓	✓	
State fixed effect	✓	✓	✓	✓	✓	✓	
N	1,087	1,087	1,087	1,087	1,087	1,019	
R-squared	0.694	0.696	0.551	0.689	0.574		

Note(s): Estimates are based on urban counties. Robust standard errors clustered at MSA level. Estimated effects have to be interpreted as variations in cases per 100,000 residents in the county. Models (4) and (8) report marginal effects at the average. SD stands for standard deviation units. Significance levels: * = 10% and ** = 5%.

Source(s): Authors own work based on ACS data, the New York Times online repository (all counties) and USAFact data collection (New York). Data extracted on April 14, 2020

Table 4.
Exposure segregation
and COVID-19 – stay-
at-home period, April
13, 2020

within group exposure, and not aspects of segregation related to groups size, diversity or dissimilarity, explains the onset of the COVID-19 pandemic across American MSAs.

6. Concluding remarks

Using ACS data, we document the effect of ethnic-based segregation in American cities on the onset of the COVID-19 pandemic. While factors like human capital, location decisions, labor market attachments, or demographics may not directly impact the probability of virus transmission, the extent of location and the quantity and quality of interactions may substantially influence the uneven spread of the virus across different places. It is well-known that there are important differences between ethnic groups in terms of quantity and quality of interactions. The Gini Exposure index is capable of quantifying such differences and measuring the degree of segregation of groups along the interaction dimensions.

We find that a one-standard-deviation increase in the Gini Exposure index is associated with an increment of about 8.77 cases per 100,000 urban county residents as of March 29, 2020. The effect, which survives after controlling for relevant mechanisms related to demographics, human capital, working status, housing market conditions, social and ethnic composition, State fixed effects, and the extent of stay-at-home orders in force, is likely a lower bound of the true effect. In fact, many positive cases are not tested but contribute to the spreading, and are more likely observed in areas where more positive cases are recorded. Additionally, a separate set of regressions reveals that higher Gini Exposure limits the effectiveness of quarantine measures in reducing the COVID-19 incidence.

This paper answers the call of [Avery et al. \(2020\)](#), among others, who highlight an “urgent need” for improving our knowledge about factors that have contributed to the evolution of the pandemic. The Gini Exposure index is shown to play a role in this respect. A back of the envelop calculation determines the health and economic costs exposure segregation. A SD reduction in exposure segregation reduces by 0.174 (8.77 times 2%, the official rate of COVID-19 cases requesting hospitalization according to WHO) the demand for intensive care units, thus reducing the turnaround of IC units by 6.96% (the average number of ICU in American hospitals is 24 per 100,000 residents according to [Weiner-Lastinger et al., 2022](#)). Despite the obvious life-saving positive impact of reducing segregation, the effect leads to a reduction of financial pressure on hospitals of at least \$723 for the average COVID-19 patient treated in IC units (given that the direct cost of COVID-19 inpatient was of \$10,394 in 2020, see [Kapinos et al., 2024](#)).

The Gini Exposure index provides a simple tool for monitoring segregation with widely accessible data, for charting segregation differences across cities, and for inferring potential costs of high segregation. While this paper infers costs of segregation based on figures related to the COVID-19 pandemic, a future avenue of research could investigate other dimensions of public health or economic activities as outcomes, and employ the Gini Exposure index to quantify the hidden costs on an uneven social mix in American cities on such dimensions.

Notes

1. Data have been accesses on April 14, 2020 from the online repository <https://github.com/nytimes/covid-19-data>.
2. Geographic exceptions to the New York Times county series are New York MSA, Georgia, Kansas City (MO), Alameda County (CA), Douglas County (NB) and Chicago MSA. We refer to USAFact data provider for these exceptions. Data have been downloaded on April 14, 2020 from: <https://usafacts.org/visualizations/coronavirus-covid-19-spread-map/>.
3. Data web-scraped on April 14, 2020 from <https://www.nytimes.com/interactive/2020/us/coronavirus-stay-at-home-order.html>.

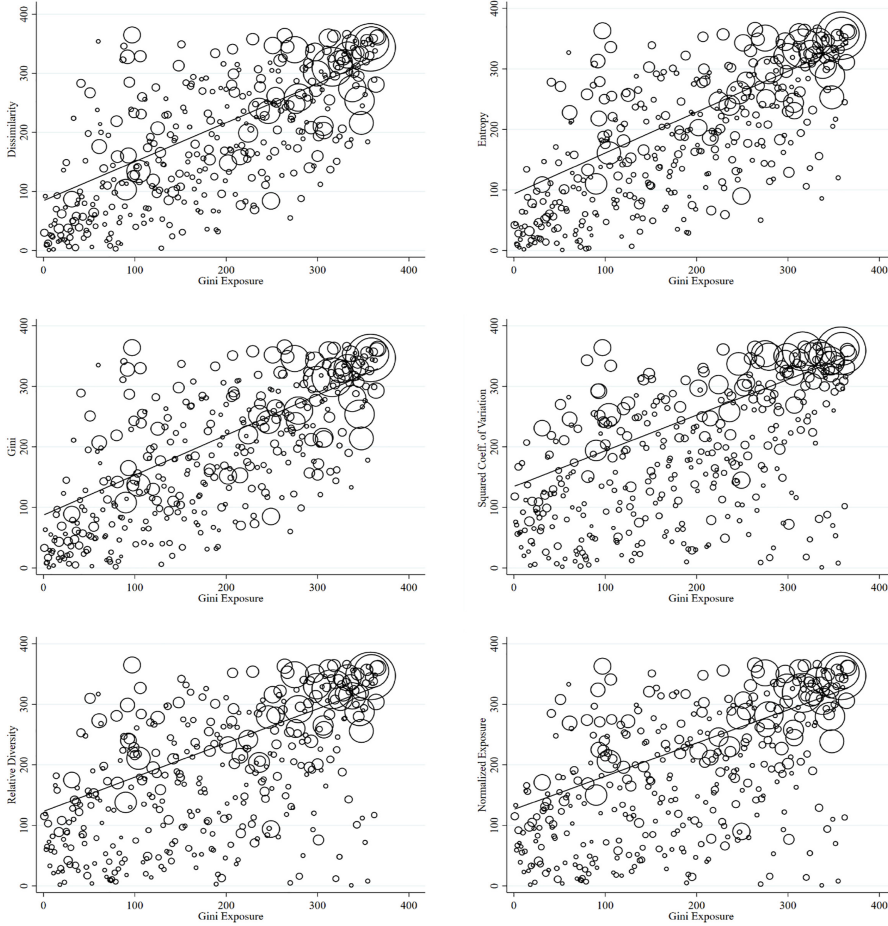
References

- Andreoli, F. (2014), "The Gini-Exposure index for measuring segregation in networks", *Rivista Italiana degli Economisti/The Journal of the Italian Economic Association*, Vol. XIX No. 1, pp. 129-168.
- Andreoli, F. and Zoli, C. (2015), "Measuring the interaction dimension of segregation: the Gini-Exposure index", University of Verona, Department of Economics Working Papers.
- Avery, C., Bossert, W., Clark, A., Ellison, G. and Ellison, S.F. (2020), "Policy implications of models of the spread of coronavirus: perspectives and opportunities for economists".
- Berry, C.R., Fowler, A., Glazer, T., Handel-Meyer, S. and MacMillen, A. (2021), "Evaluating the effects of shelter-in-place policies during the Covid-19 pandemic", *Proceedings of the National Academy of Sciences*, Vol. 118 No. 15, e2019706118, doi: [10.1073/pnas.2019706118](https://doi.org/10.1073/pnas.2019706118).
- Bootsma, M.C. and Ferguson, N.M. (2007), "The effect of public health measures on the 1918 influenza pandemic in US cities", *Proceedings of the National Academy of Sciences*, Vol. 104 No. 18, pp. 7588-7593, doi: [10.1073/pnas.0611071104](https://doi.org/10.1073/pnas.0611071104).
- Borjas, G.J. (2020), "Demographic determinants of testing incidence and Covid-19 infections in New York city Neighborhoods", Technical report, National Bureau of Economic Research.
- Brodeur, A., Gray, D., Islam, A. and Bhuiyan, S. (2021), "A literature review of the economics of COVID-19", *Journal of Economic Surveys*, Vol. 35 No. 4, pp. 1007-1044, doi: [10.1111/joes.12423](https://doi.org/10.1111/joes.12423).
- Brown, C.S. and Ravallion, M. (2020), "Inequality and the coronavirus: Socioeconomic covariates of behavioral responses and viral outcomes across US counties", Technical Report, National Bureau Of Economic Research.
- Chernozhukov, V., Kasahara, H. and Schrimpf, P. (2021), "Causal impact of masks, policies, behavior on early Covid-19 pandemic in the U.S.", *Journal of Econometrics*, Vol. 220 No. 1, pp. 23-62, *Pandemic Econometrics*, doi: [10.1016/j.jeconom.2020.09.003](https://doi.org/10.1016/j.jeconom.2020.09.003).
- Chiou, L. and Tucker, C. (2020), "Social distancing, internet access and inequality", Technical Report, National Bureau of Economic Research.
- Coven, J. and Gupta, A. (2020), "Disparities in mobility responses to COVID-19", Technical Report, NYU Stern Working Paper.
- Ghirelli, C., Gonzalez, A., Herrera, J.L. and Hurtado, S. (2023), "Weather, mobility and the evolution of the COVID-19 pandemic", *Journal of Economic Studies*, Vol. 50 No. 3, pp. 625-641, doi: [10.1108/jes-01-2022-0032](https://doi.org/10.1108/jes-01-2022-0032).
- Jackson, M.O., Nei, S.M., Snowberg, E. and Yariv, L. (2023), "The dynamics of networks and homophily", Technical Report, National Bureau of Economic Research.
- Jung, J., Manley, J. and Shrestha, V. (2021), "Coronavirus infections and deaths by poverty status: the effects of social distancing", *Journal of Economic Behavior and Organization*, Vol. 182, pp. 311-330, doi: [10.1016/j.jebo.2020.12.019](https://doi.org/10.1016/j.jebo.2020.12.019).
- Kapinos, K.A., Peters, R.M., Murphy, R.E., Hohmann, S.F., Podichetty, A. and Greenberg, R.S. (2024), "Inpatient costs of treating patients with COVID-19", *JAMA Network Open*, Vol. 7 No. 1, p. e2350145, doi: [10.1001/jamanetworkopen.2023.50145](https://doi.org/10.1001/jamanetworkopen.2023.50145).
- Kim, J.S., Patacchini, E., Picard, P.M. and Zenou, Y. (2023), "Spatial interactions", *Quantitative Economics*, Vol. 14 No. 4, pp. 1295-1335, doi: [10.3982/qe1720](https://doi.org/10.3982/qe1720).
- Koshevoy, G. (1995), "Multivariate lorenz majorization", *Social Choice and Welfare*, Vol. 12 No. 1, pp. 93-102, doi: [10.1007/BF00182196](https://doi.org/10.1007/BF00182196).
- Koshevoy, G. and Mosler, K. (1996), "The Lorenz zonoid of a multivariate distribution", *Journal of the American Statistical Association*, Vol. 91 No. 434, pp. 873-882, doi: [10.2307/2291682](https://doi.org/10.2307/2291682), available at: <http://www.jstor.org/stable/2291682>

- Koshevoy, G.A. and Mosler, K. (1997), "Multivariate Gini indices", *Journal of Multivariate Analysis*, Vol. 60 No. 2, pp. 252-276, doi: [10.1006/jmva.1996.1655](https://doi.org/10.1006/jmva.1996.1655), available at: <http://ideas.repec.org/a/eee/jmvana/v60y1997i2p252-276.html>
- Kuchler, T., Russel, D. and Stroebel, J. (2022), "JUE insight: the geographic spread of COVID-19 correlates with the structure of social networks as measured by Facebook", *Journal of Urban Economics*, Vol. 127, 103314, doi: [10.1016/j.jue.2020.103314](https://doi.org/10.1016/j.jue.2020.103314).
- Markel, H., Lipman, H.B., Navarro, J.A., Sloan, A., Michalsen, J.R., Stern, A.M. and Cetron, M.S. (2007), "Nonpharmaceutical interventions implemented by us cities during the 1918-1919 influenza pandemic", *Jama*, Vol. 298 No. 6, pp. 644-654, doi: [10.1001/jama.298.6.644](https://doi.org/10.1001/jama.298.6.644).
- Rathnayaka, I.W., Khanam, R. and Rahman, M.M. (2023), "The economics of Covid-19: a systematic literature review", *Journal of Economic Studies*, Vol. 50 No. 1, pp. 49-72, doi: [10.1108/jes-05-2022-0257](https://doi.org/10.1108/jes-05-2022-0257).
- Reardon, S.F. and O'Sullivan, D. (2004), "Measures of spatial segregation", *Sociological Methodology*, Vol. 34 No. 1, pp. 121-162, doi: [10.1111/j.0081-1750.2004.00150.x](https://doi.org/10.1111/j.0081-1750.2004.00150.x).
- Torrats-Espinosa, G. (2021), "Using machine learning to estimate the effect of racial segregation on COVID-19 mortality in the United States", *Proceedings of the National Academy of Sciences*, Vol. 118 No. 7, e2015577118, doi: [10.1073/pnas.2015577118](https://doi.org/10.1073/pnas.2015577118).
- Trounstine, J. and Goldman-Mellor, S. (2023), "County-level segregation and racial disparities in COVID-19 outcomes", *Journal of Health Politics, Policy and Law*, Vol. 48 No. 2, pp. 187-214, doi: [10.1215/03616878-10234170](https://doi.org/10.1215/03616878-10234170).
- Weiner-Lastinger, L.M., Dudeck, M.A., Allen-Bridson, K., Dantes, R., Gross, C., Nkwata, A., Tejedor, S.C., Pollock, D. and Benin, A. (2022), "Changes in the number of intensive care unit beds in US hospitals during the early months of the coronavirus disease 2019 (COVID-19) pandemic", *Infection Control and Hospital Epidemiology*, Vol. 43 No. 10, pp. 1477-1481, doi: [10.1017/ice.2021.266](https://doi.org/10.1017/ice.2021.266).
- Wu, Z. and McGoogan, J.M. (2020), "Characteristics of and important lessons from the coronavirus disease 2019 (COVID-19) outbreak in China: summary of a report of 72 314 cases from the Chinese center for disease control and prevention", *Jama*, Vol. 323 No. 13, pp. 1239-1242, doi: [10.1001/jama.2020.2648](https://doi.org/10.1001/jama.2020.2648).
- Yilmazkuday, H. (2023), "Nonlinear effects of mobility on COVID-19 in the US: targeted lockdowns based on income and poverty", *Journal of Economic Studies*, Vol. 50 No. 1, pp. 18-36, doi: [10.1108/jes-11-2021-0596](https://doi.org/10.1108/jes-11-2021-0596).
- Yu, Q., Salvador, C.E., Melani, I., Berg, M.K., Neblett, E.W. and Kitayama, S. (2021), "Racial residential segregation and economic disparity jointly exacerbate COVID-19 fatality in large American cities", *Annals of the New York Academy of Sciences*, Vol. 1494 No. 1, pp. 18-30, doi: [10.1111/nyas.14567](https://doi.org/10.1111/nyas.14567).
- Zhai, W., Fu, X., Liu, M. and Peng, Z.-R. (2023), "The impact of ethnic segregation on neighbourhood-level social distancing in the United States amid the early outbreak of COVID-19", *Urban Studies*, Vol. 60 No. 8, pp. 1403-1426, doi: [10.1177/00420980211050183](https://doi.org/10.1177/00420980211050183).

Corresponding author

Francesco Andreoli can be contacted at: francesco.andreoli@univr.it



Note(s): Each graph reports the ranking of American MSA ordered according to the Gini Exposure index (horizontal axis) and one of the most-commonly used multi-group segregation indices (Reardon and O’Sullivan 2004)
Source(s): Authors own work based on ACS data

Figure A1.
Gini exposure and
multi-group
segregation measures

Table A1.
Robustness check:
exposure segregation
and COVID-19 – early
incidence, March
29, 2020

Dependent variable: Measure: Time frame: Method:	COVID-19 cumulative cases			COVID-19 new cases			Onset
	#Positives/100 k	#Positives	1 day	#Positives/100 k	7 days	15 days	#Positives > 100
	OLS	Poisson	(4)	OLS	OLS	(6)	Logit
	(1)	(3)	(4)	(5)	(5)	(6)	(7)
Gini Exposure (SD units)	7.97* (4.37)	17.86** (6.57)	0.93* (0.52)	6.36* (3.35)	5.80* (3.19)		-0.03** (0.01)
Stay-at-home policy in <i>t</i>	-17.79** (8.08)	-12.13 (16.76)	-2.93** (1.08)	-14.01** (6.17)	-13.52** (5.56)		0.28** (0.05)
Stay-at-home policy in <i>t-7</i>	-59.71** (24.01)	7.29 (47.60)	-4.35* (2.53)	-42.34** (17.96)	-34.85** (12.28)		-0.12 (0.09)
Controls for (A), (B), (C), (D), (E)	✓	✓	✓	✓	✓	✓	✓
State fixed effect	✓	✓	✓	✓	✓	✓	✓
Controls for (F)	✓	✓	✓	✓	✓	✓	✓
N	1,087	1,087	1,087	1,087	1,087	1,087	944
R-squared	0.550	0.553	0.505	0.542	0.548		

Note(s): Estimates are based on urban counties. Robust standard errors clustered at MSA level. Estimated effects have to be interpreted as variations in cases per 100,000 residents in the county. Models (4) and (8) report marginal effects at the average. SD stands for standard deviation units. Compared to Table 3, the specifications include controls for multi-group segregation indices. Significance levels: * = 10% and ** = 5%

Source(s): Authors own work based on ACS data, the New York Times online repository (all counties) and USAFact data collection (New York). Data extracted on April 14, 2020

Dependent variable: Measure: Time frame: Method:	COVID-19 cumulative cases #Positives/100 k			COVID-19 new cases #Positives/100 k			Onset #Positives > 100 Logit (8)	
	(1)	(2)	(3)	(4)	(5)	(6)		(7)
Gini exposure (SD units)	3.35 (8.97)	5.18 (10.93)	5.50 (11.04)	76.07 (64.96)	-0.92 (0.77)	-2.17 (4.87)	1.79 (10.39)	0.00 (0.03)
Gini exp. × stay-at-home t		-2.02 (9.23)	5.51 (15.60)	-63.79 (100.96)	1.12 (0.92)	7.59 (6.09)	9.85 (13.10)	0.04 (0.08)
Gini exp. × stay-at-home t-7			-21.40 (16.37)	-52.02 (84.58)	-1.51* (0.86)	-11.49* (6.14)	-18.73 (12.51)	-0.05 (0.08)
Gini exp. × stay-at-home t-15			24.45* (12.62)	90.72** (29.88)	1.24** (0.57)	8.71** (4.26)	13.70* (8.21)	-0.01 (0.02)
Stay-at-home policy in t	-24.95 (33.00)	-21.49 (35.78)	-50.34 (61.44)	203.97 (370.71)	-7.61** (3.83)	-42.40* (24.20)	-32.00 (47.38)	0.41 (0.31)
Stay-at-home policy in t-7	5.99 (23.62)	6.26 (23.92)	57.29 (63.20)	69.46 (333.81)	6.66* (3.41)	37.25 (25.19)	39.01 (51.55)	0.30 (0.27)
Stay-at-home policy in t-15	-21.07 (13.93)	-21.17 (14.02)	-86.95** (38.52)	-361.63** (118.54)	-6.03** (2.23)	-35.55** (13.91)	-48.22** (22.20)	-0.07 (0.08)
Controls for (A), (B), (C), (D), (E)	✓	✓	✓	✓	✓	✓	✓	✓
State fixed effect	✓	✓	✓	✓	✓	✓	✓	✓
Controls for (F)	✓	✓	✓	✓	✓	✓	✓	✓
N	1,087	1,087	1,087	1,087	1,087	1,087	1,087	1,019
R-squared	0.698	0.698	0.700		0.555	0.692	0.578	

Note(s): Estimates are based on urban counties. Robust standard errors clustered at MSA level. Estimated effects have to be interpreted as variations in cases per 100,000 residents in the county. Models (4) and (8) report marginal effects at the average. SD stands for standard deviation units. Compared to Table 4, the specifications include controls for multi-group segregation indices. Significance levels: * = 10% and ** = 5%.

Source(s): Authors own work based on ACS data, the New York Times online repository (all counties) and USAFact data collection (New York). Data extracted on April 14, 2020

Table A2.
Robustness check:
exposure segregation
and COVID-19 – stay-
at-home period, April
13, 2020