Title

The Impact of Air Travel on the COVID-19 Pandemic Spread in the United States

Authors
Dai, Lu
Tereshchenko, Ivan
Hansen, Mark

Publication Date
2022-10-01

The Impact of Air Travel on the COVID-19 Pandemic Spread in the United States

Lu Dai*, Ivan Tereshchenko, Mark Hansen
Department of Civil and Environmental Engineering
University of California, Berkeley
Berkeley, CA 94720, USA
dailu@berkeley.edu, terivan2006@berkeley.edu, mhansen@ce.berkeley.edu
* *Corresponding author at: dailu@berkeley.edu, 107D McLaughlin Hall, University of California, Berkeley, CA 94720-1720, USA*

Abstract

This paper develops models to quantify the dynamics of the impact of domestic air travel on the spread of the COVID-19 pandemic in the United States, using a wide range of datasets covering the period from March to December 2020 – when international travel was restricted and became a minor source of new COVID-19 cases across country borders. With the help of flight operation data, we first develop a novel approach to estimate the county-level daily air passenger traffic, which combines passenger load factor estimates and information about the air traffic distribution. Cross-sectional models using aggregated county-level variables are estimated. While this study focuses on air travel variables, we also control for potential spatial autocorrelation and other relevant covariates, including vehicle miles traveled (VMT), road network connectivity, demographic characteristics, and climate. The model results indicate that air travel has a strong and positive impact on the initial pandemic growth rate for both case-based and fatality-based aggregate models, but this impact attenuated after the first few months of the pandemic.

Keywords:

Pandemic spread; air passenger traffic; cross-sectional model; spatial autocorrelation; network effect

Prior conference version: An earlier, shorter version of this work was presented at the Fourteenth USA/Europe Air Traffic Management Research and Development Seminar (ATM2021) under the title "Quantifying the Impact of Air Travel on Growth of COVID-19 Pandemic in the United States." The present working paper is an expanded journal-length version of that study.

I. INTRODUCTION

After the first novel coronavirus cases were observed in December 2019 in Wuhan, China, the COVID-19 pandemic has rapidly swept across the world. High-speed transportation played a key role in the global spread of the pandemic. Infected individuals could travel rapidly across China and to other cities around the world [1-5]. The COVID-19 pandemic in the United States surged in March of 2020. By April of 2020, the United States has restricted international travel, thereby making it a minor source of new COVID-19 cases within the country's borders [6, 7]. After a few large urban areas such as New York City became significant COVID-19 hot spots, the virus started to spread domestically.

Air transportation is the second largest mode of intercity transportation in the United States [8], with over 2 million people using air transportation daily for domestic travel in 2019 [9]. During the epidemic outbreak, air transportation not only incurred a massive decline in air travel demand, but may also have played an essential role in spreading the disease. On the one hand, the impact of COVID-19 on air travel is evident from Figure 1. , which shows the daily number of Transportation Security Administration (TSA) screened passengers at all the U.S. airports. The blue bars represent the screened passengers in 2020, compared with passengers screened on the corresponding day one year ago in grey. The orange curve, based on the right-hand scale, shows the percentage decline from 2019 to 2020 in daily air traffic throughput. On March 1st, a Sunday, the TSA screened over 2.2 million air passengers at all U.S. airports, 99% of the total number of passengers screened on the corresponding Sunday in 2019. Thereafter the TSA checkpoint throughput followed a mainly downward trajectory in March-April. During the second quarter of 2020, the air traffic volume decreased by as much as 96%. However, starting from May, the air traffic volume began to recover. By July 21st, the air traffic had rebounded to about 25% of the passenger throughput in 2019. On the other hand, asymptomatic infected passengers with COVID-19 could be more likely to travel by air, since the duration of air trips is much shorter than personal vehicle trips and gives less time for symptoms to manifest. Additionally, aircraft passengers share the environment with hundreds of other passengers, some of whom might be infected. Some have argued that aircraft air circulation systems may make it easy for virus particles to spread and infect other passengers [10-12]. All of these factors combined to make aviation a potential vector for COVID-19 infection to spread within the United States. In this paper, we provide a quantitative assessment of the impact of air travel on COVID-19 spread in comparison to other factors, such as weather, urban mobility, population density, and highway accessibility.

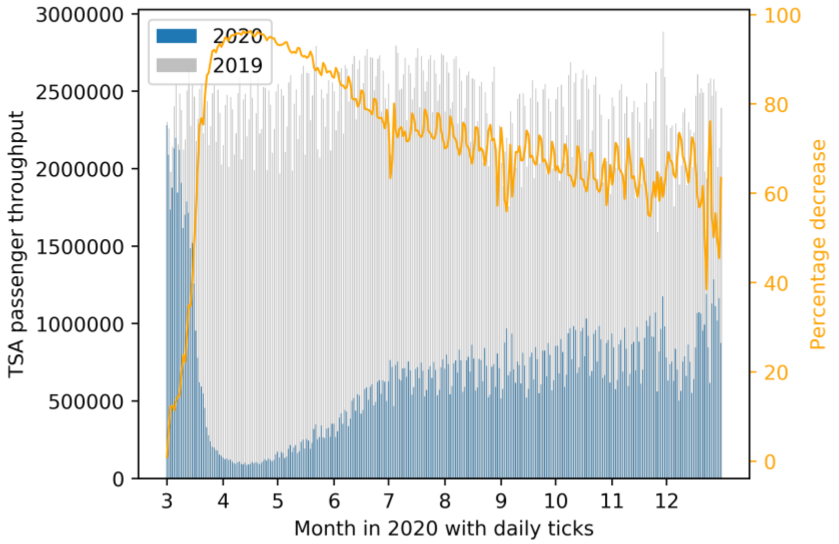

Figure 1. The daily TSA passenger throughput in 2019 and 2020

There is a considerable body of research concerning the impact of the COVID-19 pandemic on air transportation. Sun et al. [13] provided a comprehensive review of studies that looked at the impact of COVID-19 on air transportation. The studies can be classified into the following topics: impact of COVID-19 on flight suspensions [14-16], airport experience and passenger screening [17, 18], aircraft boarding patterns [19], in-flight infection risk [20, 21], passenger demand prediction [22, 23], and many others. We make two contributions to the study of COVID-19 impact on air transport. First, we propose a new metric of passenger traffic change based on the traveler throughput data from the U.S. Transportation Security Administration (TSA). This metric allows us to look at the impact of the COVID-19 pandemic at the level of individual air travel markets, while previous work has focused on aggregate traffic at the national level. Second, since the data we use is public, we propose a new open-source way to summarize intercity air travel patterns without relying on the use of proprietary data. This provides the research community with a new tool to analyze travel demand shifts caused by COVID-19. For example, this data can be used in studies aimed at developing rapid response strategies for future pandemics.

Conversely, air travel could be an essential factor in spreading diseases like COVID-19, where human-to-human contact is the primary transmission mechanism. Air transportation has been identified as a significant medium in the spread of airborne diseases like SARS, MERS [24], vector-borne diseases such as Dengue, Malaria [25, 26], infectious such as Ebola [27], as well as seasonal influenza like H1N1 [28]. However, the impact of air transportation on the spread of COVID-19 received less attention than the impact of the pandemic on global air travel patterns. The systemic review by Sun et al. [13] identified almost 100 studies by the summer of 2021 that analyzed the impact of the COVID pandemic on aviation. They were able to locate only 17 papers that quantitatively examined the role of rapid transportation in the spread of the pandemic. The main contribution of this paper is to add to this limited literature by analyzing the effect of air transportation on the spread of COVID-19 in the United States, at the county level.

Researchers have employed various approaches to investigate the role of air travel in affecting the spread of COVID-19. Compartmental models are widely used among epidemiologists to model the COVID-19 pandemic, with some focusing on the roles of airports/airlines in the disease spread using airline data [29, 30]. Some studies adopt a network science approach to investigate the disease transmissibility in response to the air transportation network dynamics under the effect of mitigation strategies, at both the global level [31, 32] and regional level [33, 34]. Li et al. [32] proposed a weighted international connectivity (IC) index to investigate the hierarchical structure of the global air transport network, which they used to investigate the impact of two entry restriction policies – direct flight suspension and complete entry suspension – imposed by different counties are reflected as node/linkage removal in the global air network. They identified countries that affect the IC index most under different entry policies. While this study deepens our understanding of international air travel from the network perspective, the results only relied on one week of pre-pandemic flight data in 2019, which might not be valid for the link weights calculation. Nikolaou and Dimitriou [34] integrated the epidemiological model with the airline and land transport network to simulate and reproduce the outbreaks of infectious diseases in the European continent. This simulation framework is able to identify critical nodes in the network, that is, airports with a high possibility to impact the disease spread. In the application of China and the United States, Peirlinck et al. [33] integrated the Susceptible-Exposed-Infectious-Recovered model with the air network constructed at the province/state level to estimate the COVID-19 outbreak dynamics.

Taking a statistical modeling approach, Lau et al. [35] investigated the relationship between international/domestic air traffic and coronavirus outbreak in China by comparing the air passenger volume and flight routes to the distribution of COVID-19 cases. The results indicate a strong linear correlation between the cases and air passenger volume. However, they used the annual passenger throughput data from 2013 to 2018, which might be inadequate since air travel patterns could have dramatically changed in 2020, especially after many flights had been canceled during the pandemic. Zhang et al. [4] estimated a gravity model to investigate the factors influencing pandemic spread from Wuhan to the rest of China, and found that the frequencies of air flights and high-speed train services are significantly associated with the number of COVID-19 cases in the destination cities. By correlating an air network mobility model to air passenger traffic data, Linka et al. [36] found that the air passenger travel data can be used to predict the emerging global diffusion pattern of a pandemic at the early stages of the outbreak. Focusing on international air travel, Zhang et al. [31] proposed the imported case risk index, which accounts for the foreign country's pandemic condition and the air connectivity with China, using the air ticket reservation data provided by a Chinese aviation service company. This is one of the few studies that utilized real flight data to measure air travel's impact on the COVID-19 pandemic.

While these studies establish a dependence of pandemic spread on air traffic mobility dynamics, none has developed a reliable estimate for measuring air passenger traffic and quantified the impact of such traffic on the pandemic spread. Studies relying on pre-pandemic aviation data, or simply considering aircraft flow, could be unreliable as they ignore the actual number of passengers on board at the time of the pandemic. Our paper aims to fill the gap by integrating data analytics and statistical modeling to provide insights about how and to what extent domestic air travel may contribute to the COVID-19 pandemic spread in different periods. In summary, our contribution is three-fold: first, a wide range of open-source datasets are collected and fused together to derive factors that may potentially influence pandemic spread at the county level in the United States. While we will focus on the domestic air travel variables in this study, other relevant covariates related to ground transportation, demography, geography, and climate are also derived and accounted for in the model. Second, we develop a novel approach for estimating the county-level air passenger traffic, which combines passenger load factor estimates and a model of traffic distribution across counties. This metric, which utilizes realized flight operation data, can serve as a proxy variable for air travel and support other pandemic-related research endeavors. Third, we estimate cross-sectional models at the county level for different periods to investigate how air travel affected pandemic spread over time, when international travel was restricted and become a minor source of new COVID-19 cases across country borders. This is the first work to utilize realized flight operation data in quantifying the air travel's impact on the pandemic spread over time (from March to December 2020) while controlling for other potential covariates and spatial autocorrelation.

The rest of this paper is organized as follows. In Section II, we introduce the data sources. Section III provides details about how we derive the air passenger traffic variable from realized flight operation data. Section IV introduces the model form and specification, including the covariates that are included. The estimation results and discussion are presented in Section V. Finally, conclusions and future work are presented in Section VI.

II. DATA SOURCES

We limit the scope of the study to the contiguous United States, which consists of the 48 adjoining U.S. states and the District of Columbia on the continent of North America. We combine a wide range of data sources covering the period from January 1st to December 31st in 2020, when international travel was restricted and become a minor source of new COVID-19 cases across country borders.

To measure COVID-19 disease incidence, we obtain county-level SARS-CoV-2 morbidity and mortality records. These are publicly available on the USAFACTS website, based on reports from state and local health agencies. This data includes the daily confirmed coronavirus infection case counts nationwide, and the number of deaths in different geographies. The data begins with the first reported coronavirus case in Washington State on Jan 21st, 2020 and is regularly updated.

We employ several sources to estimate airport passenger traffic. The FAA aviation system performance metrics (ASPM) database provides operated individual flight information at the top 77 U.S. airports. Canceled flights are not in the ASPM database. Fields of interest for each flight operation include origin airport, destination airport, departure time, arrival time, air carrier, aircraft type, user class (commercial, air taxi, freight, general aviation), and tail number. We remove cargo aircraft and flights that do not arrive in the continental United States. There are over 20,000 flights without aircraft types specified. We impute the missing aircraft type by searching the database for the flight operation with the same tail number, which is physically unique to an airplane. For those that are not matched, we check the historical flight activity log on FlightAware. Transportation Security Administration (TSA) checkpoint traveler throughput for each airport in the U.S. was obtained from the Freedom of Information Act (FOIA) Electronic Reading Room. The data records the total number of screening passengers, not including the airport crew members, passing through TSA security checkpoints at each departing airport by the hour. Airport data were obtained from multiple sources – the Bureau of Transportation Statistics (BTS) Master Coordinate table, the Global Airport Database (GADB), and the OpenFlights Airports Database. The latter database provides a list of domestic and foreign airport codes and their associated airport name, country, latitude, longitude, and altitude information. Finally, estimates of aircraft seat counts were obtained from Aviation Encyclopedia and Pilot Book, which lists information on various aircraft types, manufacturers, standard seating capacity, and model names.

County geography and road network data were obtained from the U.S. Census Bureau's Master Address File / Topologically Integrated Geographic Encoding and Referencing (MAF/TIGER) Database (MTDB). This dataset provides us with geospatial information of road networks, county cartographic boundary, and county population centers in shapefile format. The roads line shapefile provides highways, major roads, and secondary roads for national, state, and regional display. The 2019 county cartographic boundaries polygon shapefile is a simplified representation of counties with a resolution level of 1:500,000. The mean centers of population for each county are coordinates (latitude, longitude) based on the 2010 Census.

We also obtained daily county-by-county vehicle miles traveled (VMT). It measures the sum of the number of miles traveled by each vehicle in a county over a one-day period, which we think is the best available proxy for within-county traffic. The data is provided by StreetLight Data Inc[37], which collects and analyzes anonymous records from location-based service (LBS) installed on mobile devices and other GPS-enabled devices.

Hourly weather and climate data are retrieved from Google Earth Engine. We download hourly records of convective fraction, potential evaporation, shortwave radiation, specific humidity, surface temperature, total precipitation for each county.

Finally, we obtain cross-sectional socio-demographic characteristics of counties that may affect the rate of epidemic spread. These data include population density and urban population from the 2010 Census, and demography, educational attainment from the Bureau of Economic Analysis county-level economic data.

III. Air Passenger Traffic

In this section, we present our approach to estimating air passenger traffic for U.S. counties during the pandemic, which serves as an explanatory variable to represent the impact of air travel when we statistically model the pandemic spread using cross-sectional models presented in the next section. Due to the coronavirus outbreak, U.S. airlines' passenger load factor declined to 10% from 81% in January 2020 [38]. Therefore, aircraft flow is an unreliable estimate of passenger traffic and may not accurately capture the impact of air travel on the pandemic spread. However, air passenger traffic flow data availability from conventional sources is sparse and incomplete due to the sudden emergence of COVID-19.

A. Passenger Load Factor

Given the realized flight activity provided by the ASPM dataset, it is crucial for us to estimate the passenger load factor to convert the aircraft flows to passenger flows. One naïve approach would be to treat the daily ratio of 2020 TSA throughput to the 2019 TSA throughput as the passenger load factor, shown as the grey line in Figure 2. However, this is problematic and runs the danger of underestimation. As the traffic demand went down, the airlines also reduced the number of flights in order to reduce economic loss. Therefore, the average load factor should not be simply the traffic ratio, but should also capture reductions in flight schedules during the pandemic.

Toward this end, we present an approach that can approximate the passenger load factor, considering both traffic volume and the flight schedule. To do so, we must assume that flights departing from the same airport on a particular day would have the same passenger load factor. We then calculate the passenger load factor of all flights that depart from airport o on day t as PLF_{ot} using TSA checkpoint throughput and aircraft seating capacity information.

$$PLF_{ot} = \frac{S_{ot}}{\sum_{m}^{M_{ot}} C_m} \tag{1}$$

S_{ot} is the total number of screened passengers, not including crew members, passing through TSA security checkpoints at the airport o on day t, which is available from the FOIA dataset. M_{ot} is the total number of flights departing from airport o on day t, while m indexes these flights. The count M_{ot} is based on the ASPM individual flight dataset, in which canceled flights, cargo flights, and flights that do not arrive in the contiguous U.S. are removed. It varies by airport and by day. C_m is the standard seating capacity for each flight m according to its aircraft type. Thus, the denominator in (1) is an estimate of the total seats departing from airport o on day t, which when divided into the number of screened passengers yields to load factor.

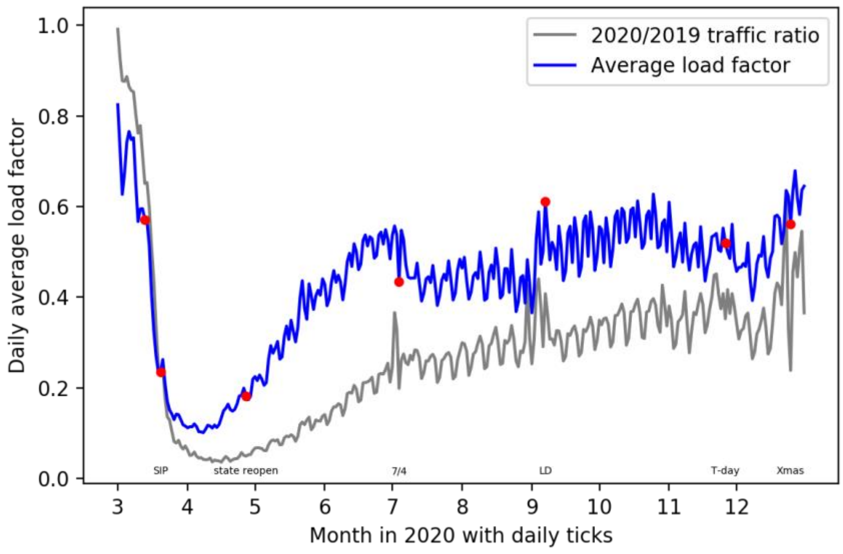

Figure 2. The passenger load factor results

The blue curve in Figure 2. represents the daily passenger load factor calculated using Equation (1), averaging over all the airports in the U.S. Our estimated passenger load factor is, in general, higher than the "load factor" approximated by the traffic ratio in grey. This reflects the fact that although traffic is reduced a lot during the pandemic, the scheduled flights are also reduced, and thus the passenger load factor would not be as low as the traffic ratio we observe. The red dots labeled on the curve are the dates with special events, such as the first Shelter In Place (SIP) restriction approved date, the first state-reopening date, Fourth of July, Labor Day, Thanksgiving Day, and Christmas. Figure 2 shows that in early March 2020, after the Center for Disease Control (CDC) issued the domestic Level 2 travel advisory, the load factor fell most steeply. Starting from March 19th, state-wide stay-at-home orders became the norm across the U.S., and soon thereafter, airlines began to adapt their flight schedules. Starting in late April 2020, states gradually eased COVID-19 restrictions, and the summer travel season caused some rebound in load factor.

B. Airport-Level Passenger Traffic

We construct the air transportation network based on the realized flight operations collected from ASPM individual flight dataset. The nodes of the network are the operating airports, and the links between them are inferred from daily flight activity. From the previous section, we obtain the estimated passenger load factor for all flights that depart from a given airport on a given day. The estimated origin-specific daily passenger load factor is then applied to the realized air transportation network to approximate the volume of passengers on each link. For a given day t, we are able to calculate the total number of passengers traveling from airport o to the airport d as N_{odt}, by adding up the number of passengers on each aircraft:

$$N_{odt} = \sum_{m}^{M_{ot}} PLF_{ot} \cdot C_m \cdot \mathbb{I}(m: o \longrightarrow d) \tag{2}$$

where PLF_{ot} is the estimated passenger load factor applied to all flights departing from airport o on day t; C_m is the standard seating capacity for each flight m according to its aircraft type; $\mathbb{I}(m: o \longrightarrow d)$ is the indicator variable which equals one if the flight m flies from airport o to airport d, zero otherwise; the M_{ot} is the total number of flights departing from airport o on day t, of which the destination could be any airports in the U.S.

In Figure 3. , we visualize the estimated number of air passengers between airport pairs on March 1st, April 1st, and July 1st from top to bottom, with warmer links indicating higher air passenger traffic volume. As expected, the passenger traffic dropped significantly during late March and early April, due to the COVID-19 shocks of lockdowns and travel restrictions. There was a moderate rebound during the summer travel period, but it did not return to the pre-pandemic levels.

After we obtain the air passenger traffic on each airport OD pair, we compute the total number of air passengers arriving at airport d on day t by summing up the traffic flows from all other airports in the U.S.:

$$N_{dt} = \sum_{o \neq d} N_{odt} \tag{3}$$

Note that the incoming traffic flow N_{dt} includes both connecting and terminating passengers, as we lack the data to distinguish between these groups. We propose the hypothesis that connecting traffic should have a more negligible effect on the estimation of passenger traffic than usual, since travelers may be more inclined to fly direct routes over connecting routes during the pandemic. Nonetheless, our passenger traffic estimates will overcount passengers into airports that serve as major connecting hubs.

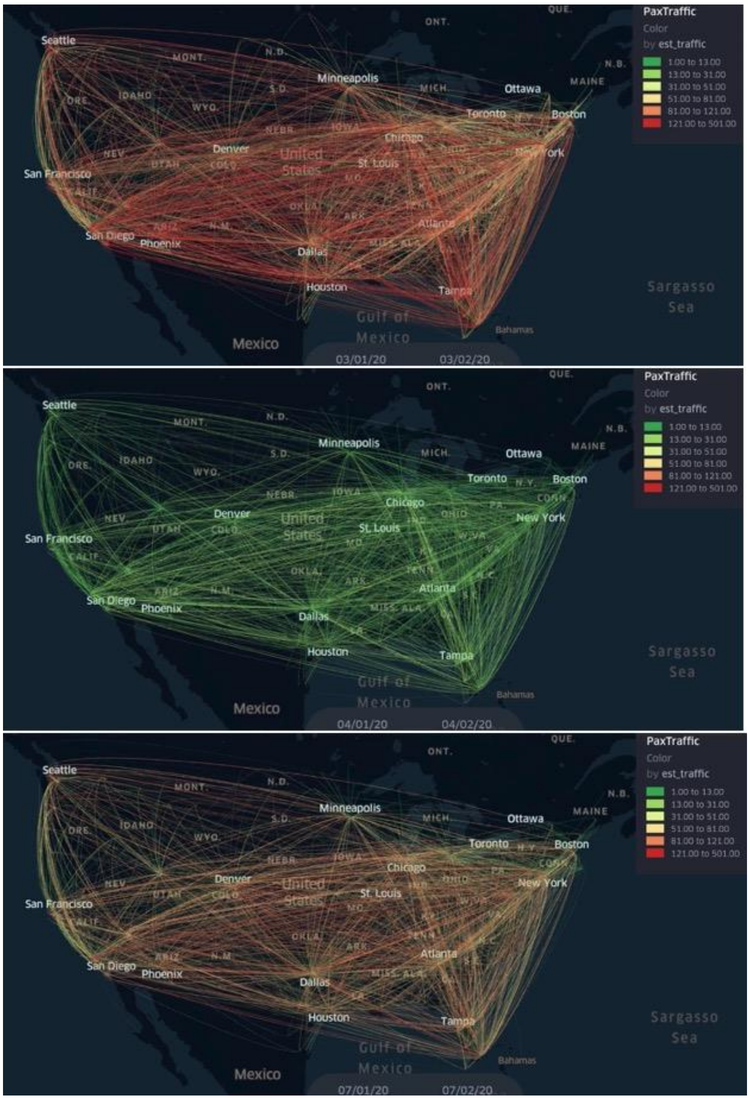

Figure 3. The air passenger flow between airports on March 1st, April 1st, and July 1st

C. County-Level Passenger Traffic

To model pandemic spread at the county scale, we need to transform airport-level air passenger traffic to be county-specific. We wish to answer the following question: of all the air passengers arriving at airport d on day t, N_{dt}, how many of them go to county i? In the transportation planning literature, this is known as a trip distribution problem.

Guided by the literature on airport passenger leakage studies, we decide to consider counties within 300 miles of each airport as the targeted counties to which the air passenger flow from that airport will be allocated. Studies have found that the airport market leakage is typically in the range of 200 miles [39], up to 230 miles away [40], 250-260 miles [41], and 300 miles [42]. We choose the most recent estimate of 300 miles. Thus, all the air passengers arriving at a given airport would only be distributed to the neighboring counties within 300 miles.

Next, we need to decide how to distribute airport arrival passengers to all these neighboring counties. According to the data availability, we assume the traffic distribution is based on county population and distance to the airport. The shorter the distance between an airport and a given county, the more convenient the ground access between the county and the airport, and thus more arriving traffic to this county. Counties with larger populations have more potential airport users increasing their shares of arriving passengers. Based on these assumptions, we define the traffic distribution probability in the form of an exponential function and calculate the number of air passengers going from airport d to county i on day t as Q_{kjt}:

$$Q_{idt} = N_{dt} \cdot \frac{P_i D_{id}^{-\alpha}}{\sum_i^I P_i D_{id}^{-\alpha}} \tag{4}$$

where P_i is the population of the county i, which is assumed to be constant over the analysis period; D_{id} is the Euclidean distance between airport d and the population centroid of county i, which is also constant over the analysis period; N_{dt} is the total number of air passengers arriving at airport d on day t; α is the unknown parameter that needs to be estimated. Note that I is the total number of counties considered in the traffic distribution, which means there are I counties within 300 miles of the airport d.

We estimate α based on passenger survey reports available for the following nine airports in the U.S.: San Francisco International Airport (SFO), Oakland International Airport (OAK), Norman Y. Mineta San Jose International Airport (SJC), John F. Kennedy International Airport (JFK), LaGuardia Airport (LGA), Newark Liberty International Airport (EWR), Los Angeles International Airport (LAX), Seattle-Tacoma International Airport (SEA), and Dallas Love Field Airport (DAL). These surveys are conducted by airports to obtain information about air travelers, including their ground access to the airport. It usually includes questions on journey purpose, origins/destinations, means of transport to and from airports, route flown, residence and income, etc. The survey question, or a similar question, we are particularly interested in is “from which county did you leave today?”. Though the survey results are based on people leaving the county on outbound trips, we assume this is applicable to inbound passengers as well. We fit Equation (4) with 1,457 samples collected from the survey responses. The optimal value of α is estimated to be 1.96 by minimizing the root mean squared error of Q_{idt}.

Lastly, our explanatory variable of the cross-sectional model – the number of air passengers traveling to county i on day t – by summing across all airports:

$$Q_{it} = \sum_d Q_{idt} \tag{5}$$

As an example from the San Francisco Bay Area, we first distribute the arriving air passenger traffic at the airports (e.g., SFO, OAK, SJC) within 300 miles of the Alameda county according to the county population and its distance to these airports. Then we aggregate the allocated traffic from these two airports. The sum of traffic is the number of air passengers traveling to Alameda County on a given day.

IV. Model Form and Variables

In this section, we present a county-level cross-sectional model to provide insights about how and to what extent the different structural determinants may contribute to the COVID-19 pandemic spread.

A. Model Specification

In this paper, we focus on an Ordinary Least Squares (OLS) regression-based cross-sectional model:

$$y_i = \boldsymbol{\beta} \cdot \boldsymbol{X_i} + \varepsilon_i \tag{6}$$

where i is the county index and $i = 1, 2, \ldots,$ N; $\boldsymbol{\beta}$is the vector of regression coefficients that reflect the sensitivity of the dependent variable to the predictors; $\boldsymbol{y}, \boldsymbol{X}$ are log-scaled dependent and independent variables respectively, and ε_i is the regression residual. The variables of this model are discussed in the following sections.

Cross-sectional analysis using measures aggregated over counties at the same period in time provides us with insights into the effects of structural determinants that may cause some regions to have higher or lower rates of pandemic spread. In this study, we estimate an aggregated model for the whole analysis period from March 1st to December 31st, 2020, and three cross-sectional models for three sub-periods: (1) 03/01/2020 – 06/30/2020, (2) 07/01/2020 – 09/30/2020, and (3) 10/01/2020 – 12/31/2020. Note that all these cross-sectional models are estimated at the county level but aggregated over different periods. The three sub-models are analyzed to investigate how the effects of the predictors vary across different epochs of the pandemic, when international travel was restricted and become a minor source of new COVID-19 cases across country borders.

B. Logarithm Growth Rate

For the dependent variable, we compute the intrinsic growth rate of the cumulative number of confirmed cases/fatalities for each county i within a given interval of time T.

$$y_i = \ln(N_{it}) - \ln(N_{i0}) \tag{7}$$

where N_{it} is the number of confirmed cases/fatalities on the last day of the analysis period; N_{i0} is the number of confirmed cases/fatalities on the first day of the analysis period for aggregated and sub-period (1) models, or the last day of the previous period for the subperiod (2) and (3) models.

We consider both the county case growth and the county fatality growth, and run the models independently for these two response variables. The fatality growth is expected to reflect the epidemic situations more accurately since the infected case data in the early phase of the coronavirus outbreak may be unreliable due to limited testing. Fatalities are also influenced by access to quality health care and individual characteristics of COVID patients, which may introduce some noise into the fatality growth model.

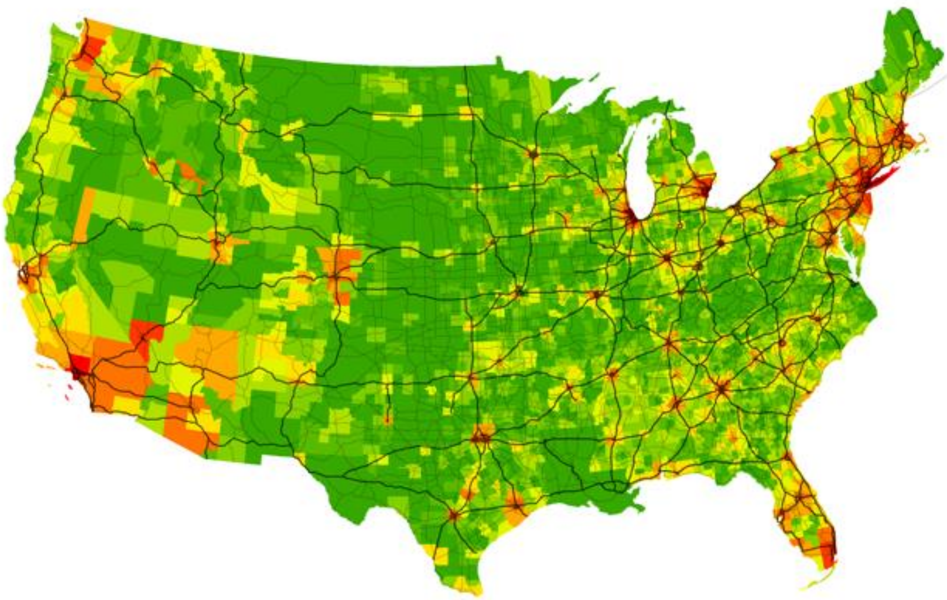

Figure 4. The county choropleth map colored by logarithm case growth rate, with the road network shown in black

C. *Ground Transportation*

To control the effects of ground transportation, we derive two variables that characterize road network connectivity and human movements. Employing complex network theory, we explore how the location of a county with respect to the road network may influence disease spread. In Figure 4. , we visualize the logarithm case growth rate on the county choropleth map from March 1st to April 15th, with warmer colors indicating faster growth. By overlaying the national highway network on the map, we observe that the logarithm case growth rate (dependent variable) is much higher in counties near geographical conjunctive points of freeways. This suggests that ground transportation network connectivity may be a significant causal factor in disease growth.

To quantify road network connectivity, we calculate the closeness centrality for each county. The closeness centrality measures how far a given county is from all other counties based on the shortest path through the road network. The closeness centrality metric reflects the position of the county in the road network. In addition, we consider the effect of the population size of the reachable counties by introducing the population multiplier. The resulting population-weighted closeness centrality for county i is denoted as PC_i:

$$PC_i = \sum_{j \neq i}^{J_i} \frac{P_j}{D_{ij}} \tag{8}$$

where J_i represents all the counties that are reachable by county i via the road network by car; P_j is the population of the county j, which is assumed to be constant over the analysis period; D_{ij} is the shortest-path distance between county i and county j via the constructed road network, which is also considered to be unchanged over the analysis period.

We choose the national major roads system to construct the network in ArcGIS. As shown in Figure 5. , the nodes are the population centers for each county based on the 2010 Census. If there is a direct road connection between two counties, a link is drawn and is weighted by the distance. With this weighted undirected graph, the population-weighted closeness centrality for each county in the U.S. can be calculated using Equation (8) and is colored on the nodes in Figure 5. Counties in the New York metropolitan area have a higher population-weighted closeness centrality score, which means, in general, they have the shortest distances to all other populous counties.

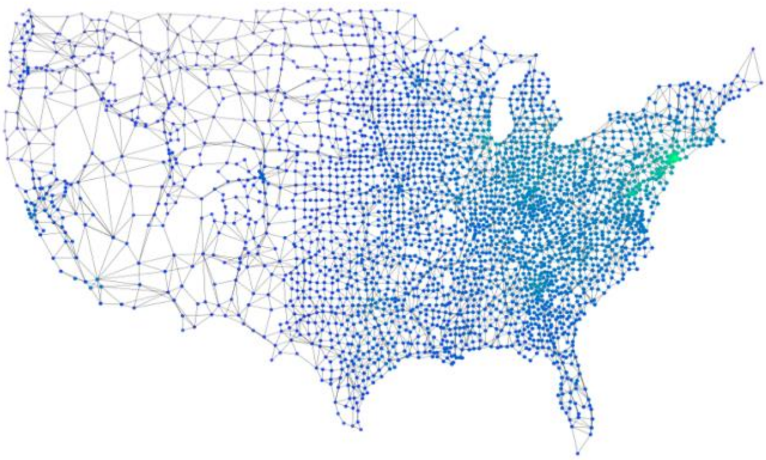

Figure 5. Population-weighted closeness centrality (in logarithm scale)

In order to control for the effect of intra-county trips on the pandemic spread (i.e., case/fatality growth rate), we also aggregate the daily VMT for each county and take the average over different analysis periods. The county-level average daily VMT per capita is used as the explanatory variable in the cross-sectional models.

TABLE I. SUMMARY STATISTICS FOR REGRESSION VARIABLES.

County-level Variable	**Description**	**Mean (standard deviation)**			
		Mar. – Dec.	***Mar. – Jun.***	***Jul. – Sep.***	***Oct. – Dec.***
Case growth rate	Logarithm case growth rate over the analysis period	7.47 (1.51)	4.48 (2.11)	1.71 (0.89)	1.38 (0.62)
Mortality growth rate	Logarithm mortality growth rate over the analysis period	3.27 (1.59)	1.30 (1.71)	0.91 (0.90)	1.15 (0.86)
Air passengers	Average number of air passengers traveling to county i, Q_i	250.26 (1643.57)	192.14 (1239.23)	273.30 (1769.04)	333.82 (2160.19)
VMT per capita	Average daily VMT per capita for county i	47.04 (15.60)	43.52 (11.19)	47.97 (11.31)	45.27 (11.89)
Closeness centrality	Normalized population-weighted closeness centrality for county i, PC_i	0.26 (0.09)			
Urban population	Percentage of urban population in county i	42.34 (30.69)			
Population density	Population density of county i	262.41 (1850.13)			
LTHSG	The number of less than high school graduates per capita	0.09 (0.04)			
Shortwave radiation	Average of daily average shortwave radiation	208.04 (17.33)	237.98 (19.23)	256.26 (14.75)	120.28 (24.23)
Precipitation	Average of daily cumulative precipitation	2.76 (1.29)	3.13 (1.44)	2.96 (1.58)	2.25 (1.41)
Temperature	Average of daily average temperature ℃	15.93 (4.31)	15.41 (4.76)	23.87 (3.21)	9.09 (5.08)

D. Climate

Studies come to conflicting results about the associations between climatic factors and the pandemic spread. Most researchers found that the transmission of COVID-19 occurs in regions exposed to cool and dry conditions, and near the lower end of the radiation gradient [43]. Most studies also reported a positive association between temperature and the COVID-19 spread due to human movements, whereas some found a negative relationship since a cold environment may provide a more favorable condition for the virus survival [44]. To investigate the climatic effects on the pandemic spread, we incorporate the average values of daily average shortwave radiation, daily cumulative precipitation, and the daily average temperature in the cross-sectional model.

E. Socio-demographic Characteristics

For each county, we include the population density, the percentage of the urban population, the number of less than high school graduates per capita, and dummy variables for different ethnic groups in the cross-sectional model. We also add dummy variables for each state to control the fixed effects at the state level, such as testing ability and lockdown policy.

After data preprocessing and filtering out counties in Alaska, Hawaii, and Puerto Rico, there are 3,056 observations (counties) remaining to estimate the model. The summary statistics of the variables are presented in TABLE I. The average case growth rate decreases over time, while the average mortality growth rate is relatively stable. Air travel, on average, grows from its nadir in the March-June period, as does its standard deviation. The VMT seems to spike in the summer. All the climatic variables match our expectations. All the explanatory variables are log-transformed except for the temperature and state dummies.

V. RESULTS AND DISCUSSION

A. Estimation Results

In this section, we present the estimation results for the model described in the previous section. We estimate two types of models – one type, where the dependent variable is the growth of the number of cases, and another type, where the dependent variable is the growth rate of the number of deaths. For each model type, we estimate four models. First, we use the average growth rate from March to December as the dependent variable ("aggregate model"), second – growth rate between March and June, third – growth rate between July and September, fourth – growth rate between October and December. The same set of predictors is used in all eight estimated models.

We estimate both case-based and fatality-based models because the two available variables have their advantages and disadvantages. On the one hand, the case-specific data is available for the majority of counties, while for many smaller counties, the number of deaths is very small, which results in an apparently small growth rate of fatalities. On the other hand, the case-specific data depends on the amount of COVID testing. The fatality growth rate more accurately reflects the spread of the epidemic for many counties, but is also influenced by omitted factors that influence mortality risk from contracting the disease. The purpose of estimating separate models for every period is to capture changes in the dynamics of virus spread over time.

All coefficient estimates are for standardized predictors – centered around the mean and divided by the standard deviation of each variable. This was done in order to compare the relative importance of each variable. We do not report the coefficient estimates for the dummy variables used in the regressions as they are not the primary focus of the paper. The dummy fixed effects were included in order to reduce the potential omitted variable bias.

Let us start by examining the R^2 values. For both case-based and fatality-based aggregate models, R^2 is very high – .90 and .80 of the variation is explained by the predictors. However, when estimated for individual periods, the R^2 decreases, particularly after the initial March-June period. This suggests that omitted variables, for example, local government policies and individual choices related to social distancing and mask wearing, played a larger role in determining COVID spread in the summer and fall of 2020. We also observe that the case models have higher R^2 than the fatality models. Several factors may contribute to this difference. Firstly, the models do not include factors related to the quality of medical care, which affects fatalities but not cases. Second, the models do not account for individual factors, such as age, that affect the risk of mortality from COVID. Third, in addition to capturing the pandemic dynamic, the case growth rate reflects the scale of testing efforts undertaken in individual counties. The amount of testing is influenced by the economic prosperity and cultural characteristics of locations. Our models capture this economic and cultural influence, which leads to better models.

TABLE II. REGRESSION ESTIMATES FOR 4 CROSS-SECTIONAL MODELS (DEPENDENT VARIABLE – AVERAGE GROWTH RATE OF COVID CASES)

Period	Mar-Dec	Mar–Jun	Jul–Sep	Oct–Dec
Air Passengers	0.311*** (0.011)	0.310*** (0.021)	-0.015 (0.015)	0.006 (0.009)
VMT Per Capita	-0.433*** (0.054)	0.678*** (0.113)	-0.646*** (0.087)	-0.135*** (0.047)
Closeness	0.038*** (0.007)	0.034** (0.014)	-0.011 (0.011)	0.014* (0.006)
Urban Population %	0.141*** (0.007)	0.138*** (0.014)	0.005 (0.010)	-0.002 (0.006)
Population Density	0.390*** (0.017)	0.578*** (0.032)	-0.111*** (0.023)	-0.068*** (0.014)
LTHSG	0.278*** (0.029)	0.419*** (0.057)	-0.061 (0.043)	-0.068** (0.024)
Shortwave Radiation	-0.819* (0.415)	0.103 (0.640)	0.291 (0.553)	-0.891*** (0.159)
Precipitation	-0.114*** (0.034)	0.105 (0.063)	-0.007 (0.030)	-0.120*** (0.016)
Temperature	0.029*** (0.007)	0.003 (0.011)	0.035*** (0.010)	-0.015*** (0.005)
Adjusted R^2	0.902	0.828	0.435	0.626

Variables are significant at the 0.1% level***, 1% level**, 5% level*.

In interpreting the regression results, let us first focus on the primary variable of interest – the number of air passengers traveling to a given county. Note that all predictor variables, except for the dummy variables, have been log-transformed. The dependent variable is also log-transformed. This means that we can interpret the regression coefficients as elasticities of case/fatality growth rates with respect to the predictor. For example, the air passenger coefficient for the aggregate case-based model is equal to 0.331. We can interpret it the following way: for every 10% increase in the number of incoming air passengers, the ratio of COVID cases at the end of the period to the beginning of the period N_{it}/N_{i0} increases by 3.11%.

For both case-based and fatality-based aggregate models, the influence of the number of air passengers on the pandemic spread rate is strong and positive. However, if we look at the period-specific models, we see a more nuanced picture. The regression coefficient is positive for the March-June models, suggesting a positive effect of air traffic on the pandemic spread rate. After that (June-September and October-December models), the effect is weak and generally not statistically significant.

There are several possible explanations for this result. First, the spread of the pandemic is determined to a significant extent at its beginning. For example, the Chinese authorities were able to curb the pandemic by introducing strict lockdowns at the very beginning, not letting the virus spread around the country. However, air transportation played a role in the early seeding of the pandemic in countries around the world. Aviation might have played a similar role in the U.S. Counties with more incoming air traffic were more likely to be exposed to the virus early on. At the early phase of the pandemic (March-June), the virus spread uncontrollably throughout the country, even as the air traffic collapsed. After the lockdowns started, the pandemic spread locally, determined mainly by local factors, such as population density, local health policies, and individual behaviors. However, all else equal, counties with stronger air connections at the beginning of the pandemic had a higher early virus "load", which determined the remaining course of the pandemic.

The second explanation for the weak effect of air transportation in the later phases of the COVID pandemic is the behavioral adjustment. As the pandemic progressed, more information, such as location-specific case numbers, became available. This may have deterred travel to places with high caseloads. Moreover, travelers with even mild symptoms—or no symptoms but recent contact to an infected person--were urged not to travel. Finally, as the pandemic progressed, air travelers were urged or required to take precautions such as self-quarantining when they arrived at their destinations. This appears to have worked as intended in cutting the link between air travel and COVID spread.

TABLE III. REGRESSION ESTIMATES FOR 4 CROSS-SECTIONAL MODELS (DEPENDENT VARIABLE – AVERAGE GROWTH RATE OF COVID-19 MORTALITY)

Period	**Mar-Dec**	**Mar–Jun**	**Jul–Sep**	**Oct–Dec**
Air Passengers	0.327*** (0.016)	0.409*** (0.021)	-0.017 (0.017)	-0.056*** (0.016)
VMT Per Capita	-0.368*** (0.081)	0.249* (0.117)	-0.128 (0.096)	-0.320*** (0.087)
Closeness	0.069*** (0.011)	0.030* (0.014)	0.017 (0.012)	0.027** (0.011)
Urban Population %	0.125*** (0.011)	-0.046*** (0.014)	0.106*** (0.012)	0.067*** (0.011)
Population Density	0.335*** (0.026)	0.260*** (0.033)	0.114*** (0.026)	0.017 (0.026)
LTHSG	0.477*** (0.043)	0.343*** (0.059)	0.151*** (0.048)	0.071 (0.045)
Shortwave Radiation	0.193 (0.619)	2.483*** (0.667)	0.104 (0.615)	-0.871*** (0.291)
Precipitation	-0.125** (0.050)	-0.023 (0.066)	0.002 (0.034)	-0.131*** (0.030)
Temperature	0.045*** (0.010)	0.009 (0.011)	0.037*** (0.011)	-0.038*** (0.009)
Adjusted R^2	0.801	0.717	0.325	0.356

Variables are significant at the 0.1% level***, 1% level**, 5% level*.

The interpretation of the remaining variables was not the focus of this paper. We included them in the regression analysis in order to minimize the possible omitted variable bias. For example, population density has a positive effect on both the pandemic growth rate and the number of air passengers. If we do not include the population density variable, the effect of air transportation on the pandemic spread will be overestimated. To avoid such occurrences, we controlled for as many variables as possible.

The estimation results underscore the importance of estimating different models for different time periods. Almost none of the predictors have the same effect across all time periods. Since the biology of the virus remained approximately the same between March and December of 2020, this result is likely the reflection of the social and behavioral response to the pandemic, population-level dynamics of the virus, such as herd immunity, and asynchronicity of the dynamics of the virus – various counties were at different stages (waves) of the pandemic at a given time. In some cases, coefficients change not only in magnitude but also in the sign. For example, in the fatality-based model for March-June, the VMT per capita coefficient was equal to 0.249, while in October-December, it was -0.320. In March-June, high VMTs contributed to the initial spread of the virus, while in subsequent time periods, when the virus was established virtually everywhere, the importance of VMT as a source of virus spread diminished. Rather, high VMT may have become more of an indicator of the degree of virus containment.

In the case of the fatality-based model, only a few variables had a consistent effect on the growth rate – closeness centrality, population density, and percentage of people with less than a high school degree. However, the effect of closeness centrality is very small compared to other variables. The population density effect was not statistically significant in October-December. Similarly, the effect of education was positive at first (a more educated population resulted in a slower growth rate) but was not significant at the end of 2020.

The effects of precipitation and temperature are small, albeit often statistically significant. The amount of shortwave radiation was statistically significant for the aggregate, March-June, and October-December models, but the sign of the effect was first positive, then negative. The surface-level shortwave radiation reflects the amount of sunlight that reaches the surface. Sunlight increases outdoor activity, but the UV radiation might destroy Coronavirus particles. Our model shows that the average effect of sunlight over nine months is positive. However, it was large and positive in March-June (especially for the fatality-based model), and large and negative in October-December.

B. Spatial Autocorrelation

Counties, as geographic units, may exhibit spatial autocorrelation in the residuals of models such as the ones estimated here. The outcome (growth rate) in a given county may depend not only on the characteristics of that county but also on the outcomes in adjacent counties. To control for this potential spatial autocorrelation in our models, Moran's I for regression residuals ε is calculated using the following equation:

$$I = \frac{n}{\boldsymbol{W}} \frac{\sum_i \sum_j w_{ij} (\varepsilon_i - \bar{\varepsilon}) (\varepsilon_j - \bar{\varepsilon})}{\sum_i (\varepsilon_i - \bar{\varepsilon})^2} \tag{9}$$

where n is the number of counties; $\boldsymbol{W} = \sum_i \sum_j w_{ij}$ is the spatial weights matrix with zeros on the diagonal ($w_{ii} = 0$), and $I \in [-1,1]$ indicating the degree to which a spatial pattern is clumped, random, or dispersed.

Assuming that areas near each other are more similar than areas that are far apart, we first employ the Queen contiguity to define neighbors of each county. The Queen adjacency neighbors are defined as counties that share a common edge or node with the subject county. In Figure 6. , the left plot shows the histogram of the number of neighbors. Most counties in the U.S. share edges or nodes with six other counties. San Juan County, Utah, which represents as the white polygon in Figure 6. on the right, has the most neighbors (i.e., colored polygons around the San Juan county in the middle).

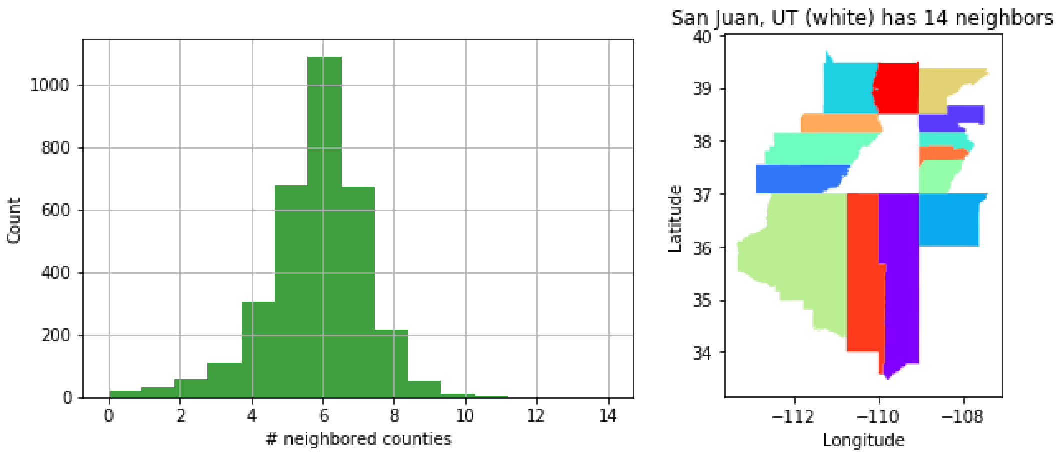

Figure 6. Histogram of Queen adjacency neighbors (left); Queen adjacency neighbors of the San Juan county (right)

Then we assign standardized weights to these neighbors by dividing the number of neighbors around the subject county. This ensures that neighboring counties exert the same collective influence irrespective of how many neighbors the subject county has. Lastly, with the spatial matrix constructed, we perform the Moran's I statistical test using the residuals for the fitted models presented in Section V.A. The null hypothesis states that the spatial distribution of the model residuals is independently distributed across the U.S. counties, including adjacent ones. TABLE IV. summarizes the Moran's I statistical test results for the case growth rate models and the mortality growth rate models.

TABLE IV. SUMMARY OF SPATIAL AUTOCORRELATION OF REGRESSION RESIDUAL ACROSS U.S. COUNTIES USING MORAN'S I METRIC

		Mar. – Dec.	**Mar. – Jun.**	**Jul. – Sep.**	**Oct. – Dec.**
Case growth rate models	***Moran's I***	0.005	0.008	0.003	-0.008
	Z-score	0.487	0.776	0.305	-0.685
	P-value	0.313	0.219	0.380	0.247
Mortality growth rate models	***Moran's I***	-0.004	-0.005	-0.019	0.004
	Z-score	-0.314	-0.458	-1.674	0.407
	P-value	0.377	0.323	0.047	0.342

For the case growth rate models, the Moran's I values indicate fairly weak positive spatial autocorrelation across all the periods, but are not significant. We fail to reject the null hypothesis and conclude that the model residuals do not exhibit spatial autocorrelation.

For the mortality growth rate models, Moran's I values indicate weak negative spatial autocorrelation for the whole analysis period, except for the period from October to December. In the second period from July to September, there is a significant negative spatial autocorrelation across the U.S. counties, suggesting a dispersed spatial pattern of the feature values. If the fatality growth rate is large in the neighboring counties, the subject county likely has a low fatality growth rate. This could be explained by the uneven distribution of nursing and medical resources—particularly nursing homes--among counties. Thus, counties in which such facilities are scarce export residents to adjacent counties in which they are more plentiful. In any case, the spatial autocorrelation, even if present, is small and its effect on our estimation results can be safely ignored.

VI. CONCLUSIONS AND FUTURE WORK

In this study, we estimate how domestic air travel contributed to the Covid-19 spread in the United States, using county-level data. Toward that end, we develop a novel approach for estimating the county-level daily air passenger traffic, which incorporates the estimation of passenger load factor and traffic distribution. This metric, which utilizes realized flight operation data, can serve as a proxy variable for air travel and support other coronavirus-related research projects. Cross-sectional models using county-level aggregate variables are employed to quantify the impact of domestic air travel on the pandemic growth rate in different time periods when international travel was restricted and become a minor source of new COVID-19 cases across country borders. An aggregated model from March to December, and three sub-period models (March – June, July – September, October – December) are estimated in order to investigate the time-varying effects of the predictors. While this study focuses on the air travel variables, other relevant covariates, including VMT, road network closeness centrality, urban population, population density, education level, shortwave radiation, precipitation, and temperature, are also derived from multiple datasets. We include them in the regression analysis to minimize the possible omitted variables bias.

The aggregate and first sub-period models have higher goodness-of-fit than the models for the later periods. This suggests that in later months of 2020, other factors not included in our models played a larger role in COVID spread, and also that the overall growth rate between March and December of 2020 was predominantly influenced by the rate in the first three months of this period. In addition, the goodness-of-fit of case-based models is higher than for the fatality-based models, probably because the models do not include a number of factors that probably influence mortality rates of COVID-infected individuals. As to the primary variable of interest – the number of air passengers traveling to a given county, we find that it has a strong, positive influence on pandemic spread in both the case-based and fatality-based aggregate models. However, the effect seems to become weak and not statistically significant for the June – September, and October – December models. This suggests air travel plays a role in the early seeding of the pandemic as counties with more incoming air traffic were more likely to be exposed to the virus early on. After the lockdowns started, the pandemic spread is mainly determined by local factors such as population density. Moreover, as the pandemic progressed, potential air passengers adjusted their behavior in ways that reduced risks of COVID spread. Finally, we perform Moran's I statistical test for our model residuals and find no significant spatial autocorrelation of residuals that we fail to control for.

We will continue working on deriving robust signatures of the air travel influence on COVID-19 transmission. First, the current cross-sectional model should be improved. For example, a deep learning model – Graphical Neural Networks – might be developed to capture the dynamics of the transportation influences on the pandemic spread at more granular temporal and spatial scales. With the graph structure of transportation networks added to the hidden layers, the deep learning model could capture the latent temporal and spatial dependencies and predict pandemic spread dynamics. Second, once a more granular model is developed, it could be used to predict how various transportation interventions might have affected the growth of coronavirus transmission, and assess their effectiveness of potential targeted transportation interventions, depending on when, where, and for how long they are deployed. For example, certain air services or even entire airports might have been closed at different times. The counterfactual analysis could be employed to investigate how specific transportation interventions would have controlled the spread of the pandemic in 2020, or might do so in the future. While the models presented here are not capable of informing such detailed policy designs, they do carry a clear policy implication: if transportation interventions are to be effective in mitigating the spread of a virus like COVID, they must be taken sooner rather than later.

References

[1] Nakamura, H. and S. Managi. *Airport risk of importation and exportation of the COVID-19 pandemic.* Transport Policy, 2020. **96**: p. 40-47.

[2] Zhao, S., Z. Zhuang, J. Ran, J. Lin, G. Yang, L. Yang, and D. He. *The association between domestic train transportation and novel coronavirus (2019-nCoV) outbreak in China from 2019 to 2020: a data-driven correlational report.* Travel medicine and infectious disease, 2020. **33**: p. 101568.

[3] Kraemer, M.U., C.-H. Yang, B. Gutierrez, C.-H. Wu, B. Klein, D.M. Pigott, O.C.-D.W. Group†, L. du Plessis, N.R. Faria, and R. Li. *The effect of human mobility and control measures on the COVID-19 epidemic in China.* Science, 2020. **368**(6490): p. 493-497.

[4] Zhang, Y., A. Zhang, and J. Wang. *Exploring the roles of high-speed train, air and coach services in the spread of COVID-19 in China.* Transport Policy, 2020. **94**: p. 34-42.

[5] Gaskin, D.J., H. Zare, and B.A. Delarmente. *Geographic disparities in COVID-19 infections and deaths: The role of transportation.* Transport policy, 2021. **102**: p. 35-46.

[6] Travel.State.Gov. *COVID-19 Traveler Information*. The U.S. Department of State 2020; Available from: https://travel.state.gov/content/travel/en/traveladvisories/ea/covid-19-information1.html.

[7] Robertson, L. *Trump's Snowballing China Travel Claim*. 2020, Factcheck.org.

[8] *Table 1: Annual Long-Distance Trips by Mode*. 2019, U.S. Department of Transportation: Bureau of Transportation Statistics.

[9] *TSA checkpoint travel numbers (current year versus prior year(s)/same weekday)*. Transportation Security Adiministration; Available from: https://www.tsa.gov/coronavirus/passenger-throughput.

[10] Colizza, V., A. Barrat, M. Barthélemy, and A. Vespignani. *The role of the airline transportation network in the prediction and predictability of global epidemics.* Proceedings of the National Academy of Sciences, 2006. **103**(7): p. 2015-2020.

[11] Mangili, A. and M.A. Gendreau. *Transmission of infectious diseases during commercial air travel.* The Lancet, 2005. **365**(9463): p. 989-996.

[12] Wan, M.P., G.N. Sze To, C.Y.H. Chao, L. Fang, and A. Melikov. *Modeling the fate of expiratory aerosols and the associated infection risk in an aircraft cabin environment.* Aerosol Science and Technology, 2009. **43**(4): p. 322-343.

[13] Sun, X., S. Wandelt, C. Zheng, and A. Zhang. *COVID-19 pandemic and air transportation: Successfully navigating the paper hurricane.* Journal of Air Transport Management, 2021: p. 102062.

[14] Albers, S. and V. Rundshagen. *European airlines' strategic responses to the COVID-19 pandemic (January-May, 2020).* Journal of air transport management, 2020. **87**: p. 101863.

[15] Bombelli, A. *Integrators' global networks: A topology analysis with insights into the effect of the COVID-19 pandemic.* Journal of transport geography, 2020. **87**: p. 102815.

[16] Sun, X., S. Wandelt, and A. Zhang. *How did COVID-19 impact air transportation? A first peek through the lens of complex networks.* Journal of Air Transport Management, 2020. **89**: p. 101928.

[17] Serrano, F. and A. Kazda. *The future of airport post COVID-19.* Journal of Air Transport Management, 2020. **89**: p. 101900.

[18] Dollard, P., I. Griffin, A. Berro, N.J. Cohen, K. Singler, Y. Haber, C. de la Motte Hurst, A. Stolp, S. Atti, and L. Hausman. *Risk assessment and management of COVID-19 among travelers arriving at designated US Airports, January 17–September 13, 2020.* Morbidity and Mortality Weekly Report, 2020. **69**(45): p. 1681.

[19] Cotfas, L.-A., C. Delcea, R.J. Milne, and M. Salari. *Evaluating classical airplane boarding methods considering COVID-19 flying restrictions.* Symmetry, 2020. **12**(7): p. 1087.

[20] Pavli, A., P. Smeti, S. Hadjianastasiou, K. Theodoridou, A. Spilioti, K. Papadima, A. Andreopoulou, K. Gkolfinopoulou, S. Sapounas, and N. Spanakis. *In-flight transmission of COVID-19 on flights to Greece: An epidemiological analysis.* Travel Medicine and Infectious Disease, 2020. **38**: p. 101882.

[21] Bielecki, M., D. Patel, J. Hinkelbein, M. Komorowski, J. Kester, S. Ebrahim, A.J. Rodriguez-Morales, Z.A. Memish, and P. Schlagenhauf. *Reprint of: Air travel and COVID-19 prevention in the pandemic and peri-pandemic period: A narrative review.* Travel medicine and infectious disease, 2020. **38**: p. 101939.

[22] Abate, M., P. Christidis, and A.J. Purwanto. *Government support to airlines in the aftermath of the COVID-19 pandemic.* Journal of air transport management, 2020. **89**: p. 101931.

[23] Warnock-Smith, D. *Air traffic will take years to return to 2019 levels.* Oxford Analytica Daily Brief, 2020.

[24] Wong, G., W. Liu, Y. Liu, B. Zhou, Y. Bi, and G.F. Gao. *MERS, SARS, and Ebola: the role of super-spreaders in infectious disease.* Cell host & microbe, 2015. **18**(4): p. 398-401.

[25] Semenza, J.C., B. Sudre, J. Miniota, M. Rossi, W. Hu, D. Kossowsky, J.E. Suk, W. Van Bortel, and K. Khan. *International dispersal of dengue through air travel: importation risk for Europe.* PLoS neglected tropical diseases, 2014. **8**(12): p. e3278.

[26] Tatem, A.J., S.I. Hay, and D.J. Rogers. *Global traffic and disease vector dispersal.* Proceedings of the National Academy of Sciences, 2006. **103**(16): p. 6242-6247.

[27] Pigott, D.M., N. Golding, A. Mylne, Z. Huang, A.J. Henry, D.J. Weiss, O.J. Brady, M.U. Kraemer, D.L. Smith, and C.L. Moyes. *Mapping the zoonotic niche of Ebola virus disease in Africa.* Elife, 2014. **3**: p. e04395.

[28] Khan, K., J. Arino, W. Hu, P. Raposo, J. Sears, F. Calderon, C. Heidebrecht, M. Macdonald, J. Liauw, and A. Chan. *Spread of a novel influenza A (H1N1) virus via global airline transportation.* New England journal of medicine, 2009. **361**(2): p. 212-214.

[29] Chinazzi, M., J.T. Davis, M. Ajelli, C. Gioannini, M. Litvinova, S. Merler, A.P. y Piontti, K. Mu, L. Rossi, and K. Sun. *The effect of travel restrictions on the spread of the 2019 novel coronavirus (COVID-19) outbreak.* Science, 2020. **368**(6489): p. 395-400.

[30] Hwang, G.M., P.J. Mahoney, J.H. James, G.C. Lin, A.D. Berro, M.A. Keybl, D.M. Goedecke, J.J. Mathieu, and T. Wilson. *A model-based tool to predict the propagation of infectious disease via airports.* Travel medicine and infectious disease, 2012. **10**(1): p. 32-42.

[31] Zhang, L., H. Yang, K. Wang, Y. Zhan, and L. Bian. *Measuring imported case risk of COVID-19 from inbound international flights---A case study on China.* Journal of Air Transport Management, 2020. **89**: p. 101918.

[32] Li, S., Y. Zhou, T. Kundu, and F. Zhang. *Impact of entry restriction policies on international air transport connectivity during COVID-19 pandemic.* Transportation Research Part E: Logistics and Transportation Review, 2021: p. 102411.

[33] Peirlinck, M., K. Linka, F.S. Costabal, and E. Kuhl. *Outbreak dynamics of COVID-19 in China and the United States.* Biomechanics and modeling in mechanobiology, 2020. **19**(6): p. 2179-2193.

[34] Nikolaou, P. and L. Dimitriou. *Identification of critical airports for controlling global infectious disease outbreaks: Stress-tests focusing in Europe.* Journal of Air Transport Management, 2020. **85**: p. 101819.

[35] Lau, H., V. Khosrawipour, P. Kocbach, A. Mikolajczyk, H. Ichii, M. Zacharski, J. Bania, and T. Khosrawipour. *The association between international and domestic air traffic and the coronavirus (COVID-19) outbreak.* Journal of Microbiology, Immunology and Infection, 2020. **53**(3): p. 467-472.

[36] Linka, K., M. Peirlinck, F. Sahli Costabal, and E. Kuhl. *Outbreak dynamics of COVID-19 in Europe and the effect of travel restrictions.* Computer Methods in Biomechanics and Biomedical Engineering, 2020. **23**(11): p. 710-717.

[37] StreetLight Data, I. *Vehicle Miles Traveled (VMT) data for U.S. counties*, I. StreetLight Data, Editor. 2021.

[38] Mazareanu, E. *Monthly passenger load factor (PLF) on international flights by region 2020*, in *Transportation & Logistics*, Statista, Editor. 2021.

[39] KIMLEY-HORNANDASSOCIATES, I. *Airport Air Service Profile: Orlando-Sanford International Airport.* 2012.

[40] Suzuki, Y., M.R. Crum, and M.J. Audino. *Airport choice, leakage, and experience in single-airport regions.* Journal of transportation engineering, 2003. **129**(2): p. 212-218.

[41] Phillips, O.R., L.R. Weatherford, C.F. Mason, and M. Kunce. *Passenger leaks and the fate of small community air service.* Economic Inquiry, 2005. **43**(4): p. 785-794.

[42] Ryerson, M.S. and A.M. Kim. *A drive for better air service: How air service imbalances across neighboring regions integrate air and highway demands.* Transportation Research Part A: Policy and Practice, 2018. **114**: p. 237-255.

[43] Araujo, M.B. and B. Naimi. *Spread of SARS-CoV-2 Coronavirus likely to be constrained by climate.* MedRxiv, 2020.
[44] Xie, J. and Y. Zhu. *Association between ambient temperature and COVID-19 infection in 122 cities from China.* Science of the Total Environment, 2020. **724**: p. 138201.

Declaration of Interest

None.