JUE Insight: Urban flight seeded the COVID-19 pandemic across the United States

Joshua Coven, Arpit Gupta*, Iris Yao

NYU Stern School of Business, United States

ARTICLE INFO

ABSTRACT

We document large-scale urban flight in the United States during the COVID-19 pandemic. Regions that saw migrant influx experienced greater subsequent new COVID-19 cases, linking urban flight (as a disease vector) and coronavirus spread in destination areas. Urban residents fled to socially connected areas, consistent with the theory that individuals sheltered with friends and family, or in second homes. Populations that fled were disproportionately younger, whiter, and wealthier. The association between migration and subsequent new cases persists when instrumenting for migration with social networks.

1. Introduction

"Rumors that cholera was moving west and not south from Canada could not stem the growing panic; mass exodus from the city had already begun. A hyperbolic and sarcastic observer remarked later that Sunday had seen 'fifty thousand stout hearted' New Yorkers scampering 'away in steamboats, stages, carts, and wheelbarrows.' " — The Cholera Years: The United States in 1832, 1849, and 1866 Rosenberg, 1968.

Cities are vulnerable to contagious disease due to population density and international connections, the same qualities that foster human interaction and economic activity. The role of cities in pandemics has seen renewed attention in the context of the novel coronavirus disease 2019 (COVID-19), especially in the United States where New York City was an early epicenter, had a high death rate, and experienced massive urban flight. The behavioral responses of those in cities to mitigate personal disease risk have been studied on the intensive margin, including sheltering in place. However the migratory response of people in cities, who employ their resources to mitigate personal disease risk, and the spillovers of these actions on broader community transmission, remain unclear. A better understanding of the relationship between migration and disease transmission has implications for disease mitigation policy, including travel restrictions. The possible persistence of moves and the demographics of movers also has large public finance implications for cities.

This paper documents how urban flight seeded the pandemic across the United States and quantifies the extent of urban flight in response

to COVID-19 in its initial phase. We use mobile phone geolocation data, which allows for a higher frequency analysis than has been possible in prior studies, to analyze migration in the United States. We find large outflows from cities early in the pandemic. We use Facebook friendship data to establish that migration was high between socially connected regions, consistent with the idea that urban flight led to sheltering with friends and family or in second homes. In New York City, we find that a change from the bottom decile of tracts by income to the top increases the likelihood of having left the city from 0.8% to 12%. Regions with greater flight were generally richer, whiter, and younger, pointing to important disparities in the availability of migration as a risk-mitigating strategy.

We identify the association of urban flight and migration on increases in COVID-19 cases in the destination counties with an instrumental variables strategy that leverages social connections between counties. We use the Social Connectivity Index (SCI) measure from Facebook, as discussed in Bailey et al. (2018), which measures the normalized count of friendships between geographies, as of April 2016, to instrument for migration. We find that a one standard deviation increase in SCI-instrumented per capita inflow is associated with a one standard deviation increase in new cases per capita. Equivalently, an additional 7 travelers per 1000 individuals is associated with 1 additional case per 1000 individuals. Our estimates are substantial and point to urban migration as an important link for COVID-19 spread across the United States.

Our results are also relevant to questions surrounding public finance shocks to local governments, and to questions about the long-term future

of cities. Because the populations that left urban areas were disproportionately wealthy, their flight deprives cities of valuable tax revenue in the short-run. To the extent that urban migration remains persistent, cities may also face long-term challenges around budget shortfalls, real estate prices, and population size. While our geolocation data end in December 2020, other data from USPS change of address information indicate persistence in urban migration, suggesting that urban flight-related disruptions have important repercussions in the wake of COVID-19.

Additionally, our results have implications for the potential value of COVID-19 travel restrictions. Virtually every country has put in place stringent restrictions on international travel, and many countries have put further restrictions on regional travel within their own borders. However, the potential value of these travel restrictions remains unclear. Our results suggest that intra-US travel was associated with COVID-19 spread, and that the existing level of travel restrictions and self-suggested quarantine orders were insufficient to prevent out-of-state migration from impacting local spread.[1]

In Florida in March 2020, for which we have data distinguishing between travel cases and non-travel cases, as well as the origin of the travel, we find that as many as 40% of all cases were directly attributable to travel, and 10% could be attributed to New York City directly. Inflows are correlated with cases even after removing travel related cases, suggesting the importance of spillovers in cases from infected individuals due to travel.

Our work is closely linked to a rapidly growing literature that uses mobile phone geolocation data to assess the spread of COVID-19. Our demographic results are related to work by Chiou and Tucker (2020), which finds that shelter-in-place effects vary by income. This paper differs by considering the role of leaving the city, connecting mobility with actual COVID-19 exposure, and incorporating analysis of other demographic groups. Couture et al. (2022) uses mobile phone data to examine mobility during the pandemic. Our work differs by focusing on the role of urban flight in spreading COVID-19. Glaeser et al. (2022) examines mobility changes within regions, while our analysis examines mobility across regions. Other work (Allcott et al., 2020; Barrios and Hochberg, 2021; Barrios et al., 2021; Engle et al., 2020; Painter and Qiu, 2021; Andersen, 2020) has looked at political partisanship and COVID-19 responses. Our work is closely related to Brinkman and Mangum (2022), which examines the effect of changes in quantity and distribution of travel on COVID-19 cases, and Mangrum and Niekamp (2022), which studies the role of college spring break travel on COVID-19 cases. Our work differs in its focus on urban flight, as well as the demographics of the flight.

Prior work by Athey et al. (2021); Chen et al. (2020), and Chen and Rohla (2018), has used mobile phone geolocation data to examine segregation, racial disparities in voting waiting times, and partisanship. Another use of individual ping-level geolocation data includes research by Chen et al. (2021), who examine nursing home networks in the wake of COVID-19. Holtz et al. (2020a,b) also use Facebook data to show that socially connected areas have comparable social distancing responses. These may arise through a variety of mechanisms including both direct migration as well as communication. We argue for a specific mechanism of direct travel, which is complementary to other possible explanations.

2. Data and specification

2.1. Data

Mobile location data were sourced from VenPath–a holistic global provider of compliant smartphone data. We obtained unique data on smartphone Global Positioning Systems (GPS) signals. Our data provider aggregates information from approximately 120 million smartphone users across the United States. GPS data were combined across applications for a given user to produce pings corresponding to timestamp-location pairs. Ping data includes both background pings (location data provided while the application is running in the background) and foreground pings (activated while users are actively using the application). Our sample period covers the period from February 1, 2020 to July 13, 2020, and from October 1, 2020 to December 31, 2020.

We supplement our mobility data with county-level coronavirus case counts from the COVID-19 Data Repository by the Center for Systems Science and Engineering at Johns Hopkins University.[2] We also incorporate nursing home data from the Centers for Medicare and Medicaid Services[3] and covid case data from the Florida Department of Health, which records whether cases are associated with out-of-study travel.[4] We join this with a county-to-county Social Connectivity Index (SCI) measure from Facebook as discussed in Bailey et al. (2018) and applied in reference to COVID-19 in Kuchler et al. (2022). We also include demographic data from the Census American Community Survey, and urban-rural county classifications from the National Center for Health Statistics (NCHS).

We isolate the migration behavior of users in the US by identifying each user's modal nightly census tract (6pm–8am), provided they ping in that tract three or more times that night. We do this each night for a given month. If a user's most frequent modal night tract appears as their modal night tract on at least five nights in a month, we define it as their "home tract." We repeat this process each month to analyze mobility. We use only one month of data at a time to identify residents' home tracts. We analyze user data in the month that immediately follows the month used to identify home locations. The resulting sample includes a population of 9 to 11 million unique users per month for our base analysis across the United States. In NYC, we observe a 0.89 correlation between the population of each zip code and our observed mobile phone population in that area.

To calculate county-to-county or ZIP code-to-ZIP code flow, we observe the count of users spending the night in a given census tract and aggregate up to the county or ZIP level for each date. We aggregate tracts to counties, and link tracts to ZIP codes using a crosswalk provided by the Department of Housing and Urban Development.[5] When tracts map to multiple ZIP codes, we select the ZIP code with the highest number of residents. We aggregate resident counts to the home geography, current geography, and date to see where people from a given geography are spending the night on each date. After aggregating mobile phone migration data into county-day information, we sum, for a given day t, all net inflows into county i.

2.2. Empirical specification

Our core empirical specifications examine the determinants and consequences of urban flight in the context of the COVID-19 pandemic in the United States. We use an OLS specification which measures new cases in a destination county as a function of gross inflow into that county:

$$
\begin{aligned}
\text{New Cases}_{i,t} = {} & \beta_0 \cdot \text{Inflow}_{i,t} + \beta_1 \cdot \mathbb{1}(\text{High Cases in Originating Counties})_{i,t} \\
& + \beta_2 \cdot \mathbb{1}(\text{Far})_i + \beta_3 \cdot \text{Inflow}_{i,t} \times \mathbb{1}(\text{Far})_i + \gamma_1 \cdot X_i + \delta_{s,m,p} + \varepsilon_{i,t}. \quad (1)
\end{aligned}
$$

We are primarily interested in the β_0 coefficient, which measures the effect of inflows on new COVID-19 cases. We measure new case activity both in levels of inflow and per capita. In addition, we test whether inflow from counties which experience high case counts and inflow from

---

more distant counties have a differential impact on new cases. The indicator for high cases in originating counties is equal to 1 for counties where the inflow-weighted cases from incoming counties fall within the top quartile in the current month. We construct $\mathbb{1}(\text{Far})$ by assigning 1 to counties where the inflow-weighted distance is at least 500km (roughly equal to the distance between NYC and Pittsburgh).

We further include a number of county-level control variables in our regression specification to account for sources of new coronavirus cases orthogonal to inflows from outside counties. Controls include the distance between the home and destination counties, mean household income, population density, NCHS urban-rural classification, share of the population above 60 years of age, the share of essential workers, and the number of nursing homes. Finally, we sort counties into deciles by population and include state by month by population decile fixed effects to isolate the effect of migration from unobservable heterogeneities across city size.

While our specification includes a number of plausible control variables, an important potential identification concern with Eq. (1) is the endogenous nature of migration. If counties that receive higher domestic migration are also more likely to be susceptible to new cases for other reasons, a positive β_0 may reflect spurious correlation, rather than measure the association of migration on new COVID-19 cases. Therefore, we develop an identification strategy to address endogenous migration decisions based on Facebook connectivity in order to establish the link between inflow and new cases. We draw on prior research, as mentioned in Bailey et al. (2018), that suggests social connectivity is a driver of migration decisions when measured at annual frequencies. Our analysis establishes that social connections explain the high-frequency migration observed during the COVID-19 epidemic. To examine the relationship between migration and social connectivity, we first run a first stage regression of migration against social connectivity between regions:

$$\text{Inflow}_{i,t} = \gamma_0 SCI_{i,t} + \sum_{d=2}^{10} \gamma_d \cdot \text{Distance Decile}_{i,t} + \gamma_1 \cdot X_i + \delta_{s,m} + v_{i,t}. \quad (2)$$

The SCI between two counties or ZIP codes i and j measures the strength of social connections between them, and is defined (as in Kuchler et al., 2022) based on the friendship links between two regions i and j and normalized by the number of users in each region:

$$SCI_{ij} = \frac{FB\_Connections_{ij}}{FB\_Users_i \times FB\_Users_j}. \quad (3)$$

Our primary specification examines social connectivity at the county level. We also examine connectivity between ZIP pairs. The coefficient γ_0 measures the strength of our first stage—the predictiveness of social connectivity in a gravity regression on migration, controlling for the decile of physical distance and other factors. In order to isolate the effect of county inflows due to migration, we instrument county-county flows with the county SCI measure. Our instrumental variables specification first instruments for inflow using Eq. (2) and uses predicted inflow instead of realized inflow as a covariate in Eq. (1). We conduct our main analysis at the county level, where we have case data nation-wide, but are also able to establish the relationship between migration flows and SCI at the ZIP code level. Our main focus is on the domestic determinants of transmission. We also explore the initial seeding of cases in the United States in Fig. A4, which finds that areas with high international social connectivity tended to have higher cases earlier on (but not later).

The identifying assumption is that Facebook connections between county i and any other county do not correlate with the trajectory of new cases, *except* through the inflow of people into the county. This is a plausible assumption during the early stage of the pandemic, during which most regions were unlikely to see pandemic spread except through the inflow of individuals exposed to COVID-19 from other regions. The remaining threat to identification relates to the precise mechanisms of inflow. For instance, regions socially connected to New York

may indeed see rises in cases arising from an influx of residents; but the links to New York may simply proxy for links to other destinations which were the real source of spread. We further isolate the effect of spread from the source of travel rather than alternative sources in two ways: first, by looking at the timing of when inflows are followed by case increases, and second by examining travel-related cases in Florida, where the origins of travel-associated cases can be determined.

3. Results

3.1. Associations of national migration

We begin by descriptively analyzing the nationwide migration in the context of COVID-19. Fig. 1 documents the net flow and outflow of residents across counties in the United States as of the end of each month.[6] Map colors indicate the fraction of residents who left or entered the county, while the size of the circles indicate the size of the flow.

By the end of March, we document substantial flight out of NYC as well as several other metropolitan areas (including Boston, Los Angeles, San Francisco, and Phoenix). Travelers went to a mix of interior locations, including rural areas across the country and urban areas in the South. Several cities in the Sunbelt, in particular Atlanta, Houston, Charlotte, and Austin saw substantial net in-migration during this period. Some other cities in the North, such as Des Moines, Chicago, Detroit, Kansas City, and St. Louis, also saw substantial inflow. We also observe substantial inflow to numerous smaller counties in the vicinity of NYC, in the Hamptons and Hudson Valley. Broadly, the pattern of migration reflects flight away from the initial epicenters of the pandemic, coastal cities, towards the national interior.

We observe continued urban flight from NYC, as well as additional flight away from Phoenix, Florida, and some Californian and Texan cities in April. By May, we observe substantial inflow into coastal regions for vacation purposes. Our focus largely centers around the substantial migratory response in the wake of the first migration event, around mid-March of 2020. We have additional data which go through the end of 2020.

### 3.2. Spatial, temporal, and demographic patterns of urban flight

3.2.1. Demographic attributes of flight

We focus on specific cities to further characterize the urban flight in Fig. 2, which focuses on the propensity to remain in each of the six cities sampled (New York City, San Francisco, Los Angeles, Washington DC, Seattle, and Boston; which are chosen to be broadly representative of major urban areas). We plot background demographic associations at the ZIP code level against the fraction of the population that stays in the ZIP code. Background dots show all data points, while binscatter dots plot the average population within 25 quantiles. We find that the fraction of residents who remained in cities strongly decreases with higher tract income, the fraction of the tract that is white, and the proportion of residents aged 18–45.

3.2.2. Flight from New York City

We highlight the flight responses of individuals to the COVID-19 crisis in New York City specifically in Panel A of Fig. A1. We observe stark differences in the flight response of individuals along the following extensive margin: Manhattan residents are substantially more likely to leave the city after the crisis, as are individuals in wealthy parts of Brooklyn. We find that as much as 10–15% of the population of Manhattan, formerly residing in the city in February, leaves NYC by April 15. By contrast, residents in Queens—the epicenter of the COVID-19 pandemic in New York City—Brooklyn, and the Bronx are overwhelmingly

---

[6] Appendix Fig. A7 examines inflow.
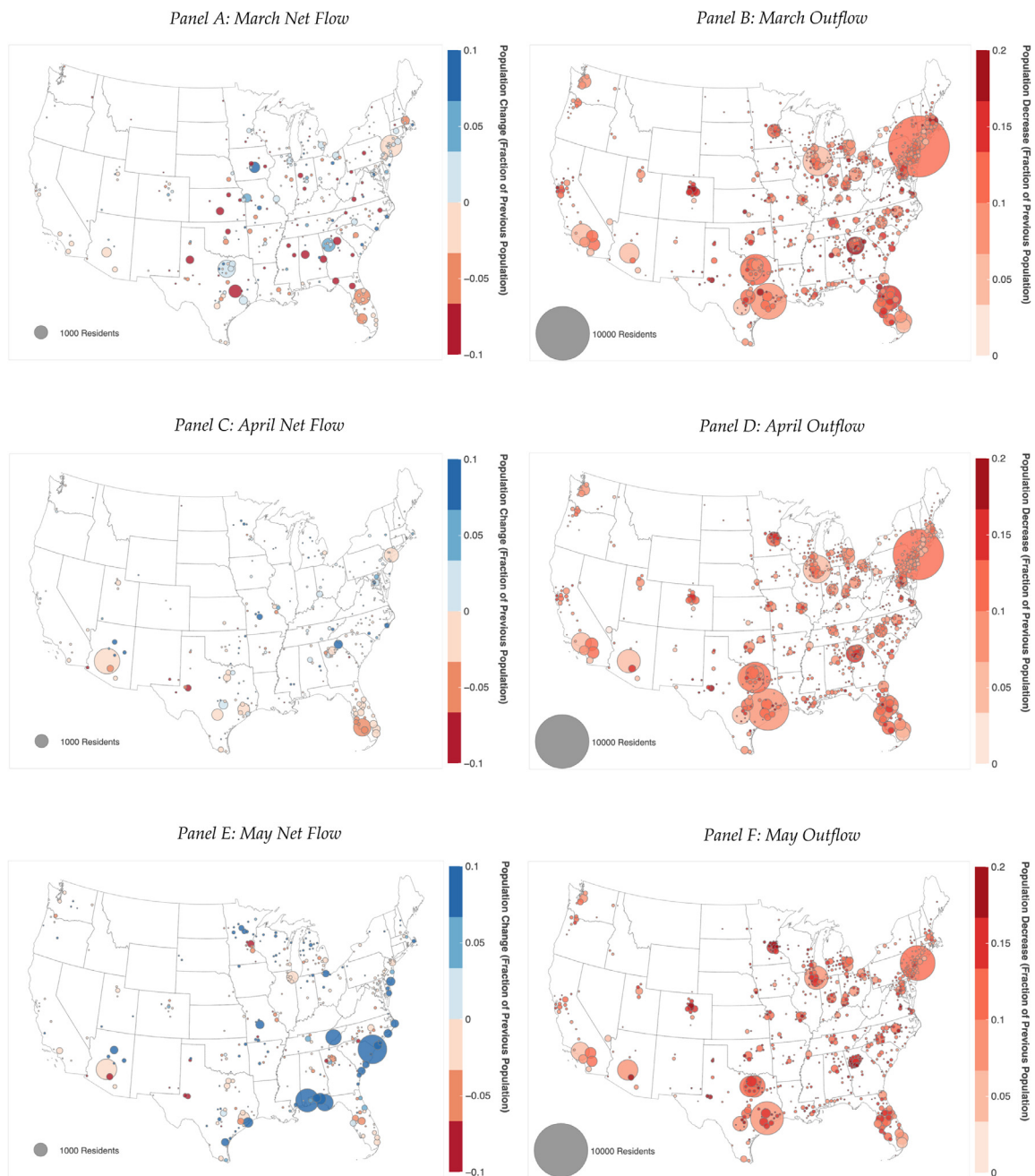
Fig. 1. Nationwide migration at the county level. *Notes:* Outflow is the number of people who were residents of a county in the previous month, who on a given date in the current month have a different modal night county. The population change fraction is this number divided by the total number of residents of that county who were in the data on that date. Inflow counts the number of people who were not residents in a county in the previous month, who have the county as their modal night county on a given date. Net flow is inflow minus outflow. All measures are averaged across the last week of the month.

more likely to stay in the city. This urban exodus continues through the end of our sample. By 10 July, we observe about 17% of the population of Manhattan remains absent in the city. These shifts in leaving the city are concentrated in the higher-income census tracts, suggesting that richer NYC residents were disproportionately able to take advantage of the option to flee the city and escape physical COVID-19 exposure.

We confirm the role of income as a factor in explaining moves away from the city in Panel B of Appendix Fig. A1, which shows a heat map of responses by tract and date. We find a large break-point in our sample on 14 March, as reflected in the stark coloration changes on that date in a number of tracts which correspond to a sharp rise in NYC inhabitants departing the city. This break comes just before Mayor Bill de Blasio or-

dered schools, restaurants, bars, cafes, entertainment venues, and gyms in the city to be closed on 16 March.[7] Following 14 March, we observe increased flight behavior in the highest-income NYC census tracts, observed as an increase in red colors in the bottom-right corner of the heat map.

3.2.3. Urban flight across metropolitan areas

We examine the spatial distribution of flight patterns in our six city sample in Fig. 3, which shows the fraction of residents who leave.

---

[7] See https://www1.nyc.gov/assets/home/downloads/pdf/executive-orders/2020/eeo-100.pdf.

Panel A: Change in Out of Town Visits by ZIP Income

Panel B: Change in Out of Town Visits by Fraction White

Panel C: Change in Out of Town Visits by Fraction Aged 18-45

Panel D: ZIP Inflows against SCI

Panel E: Inflows against New Cases
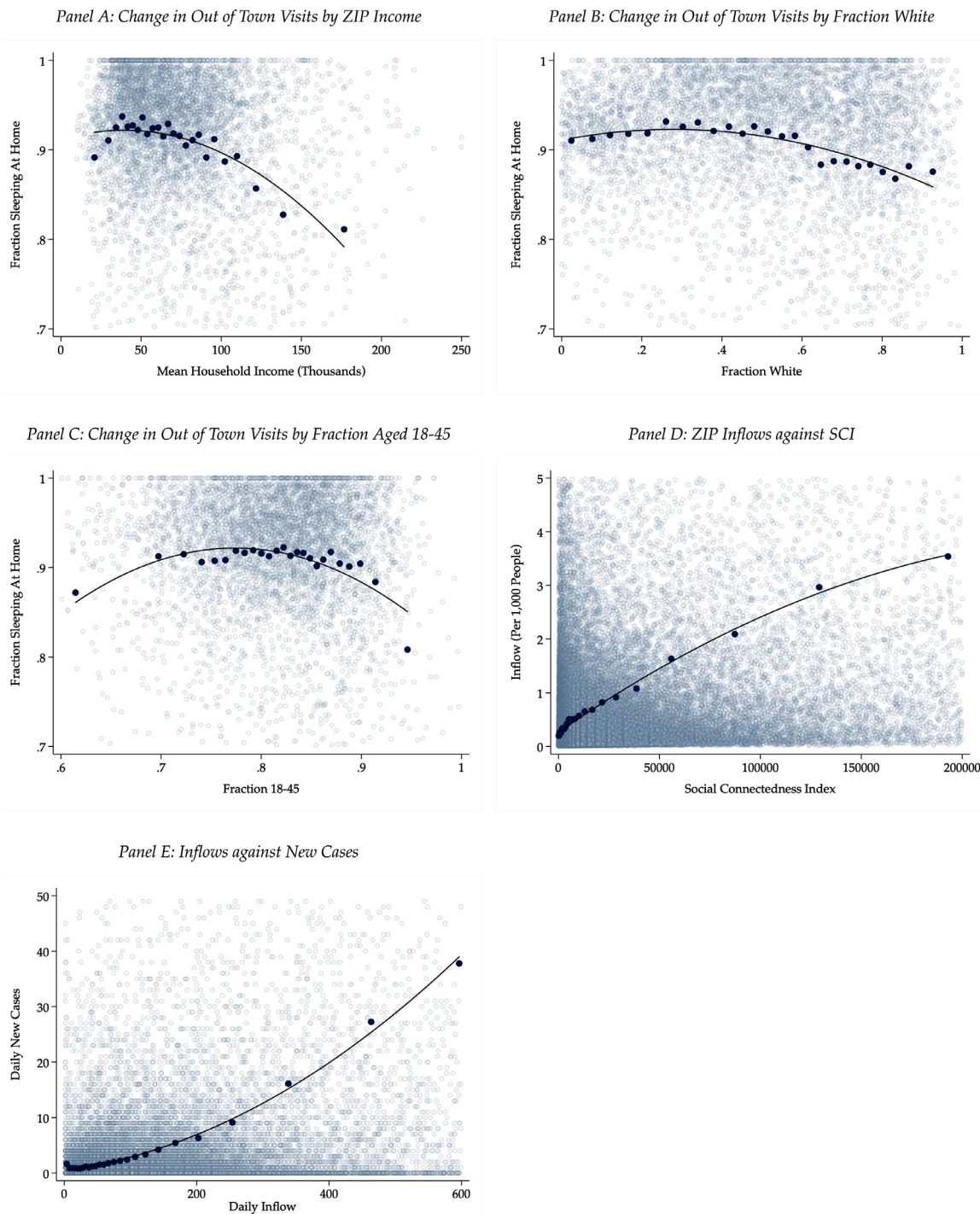
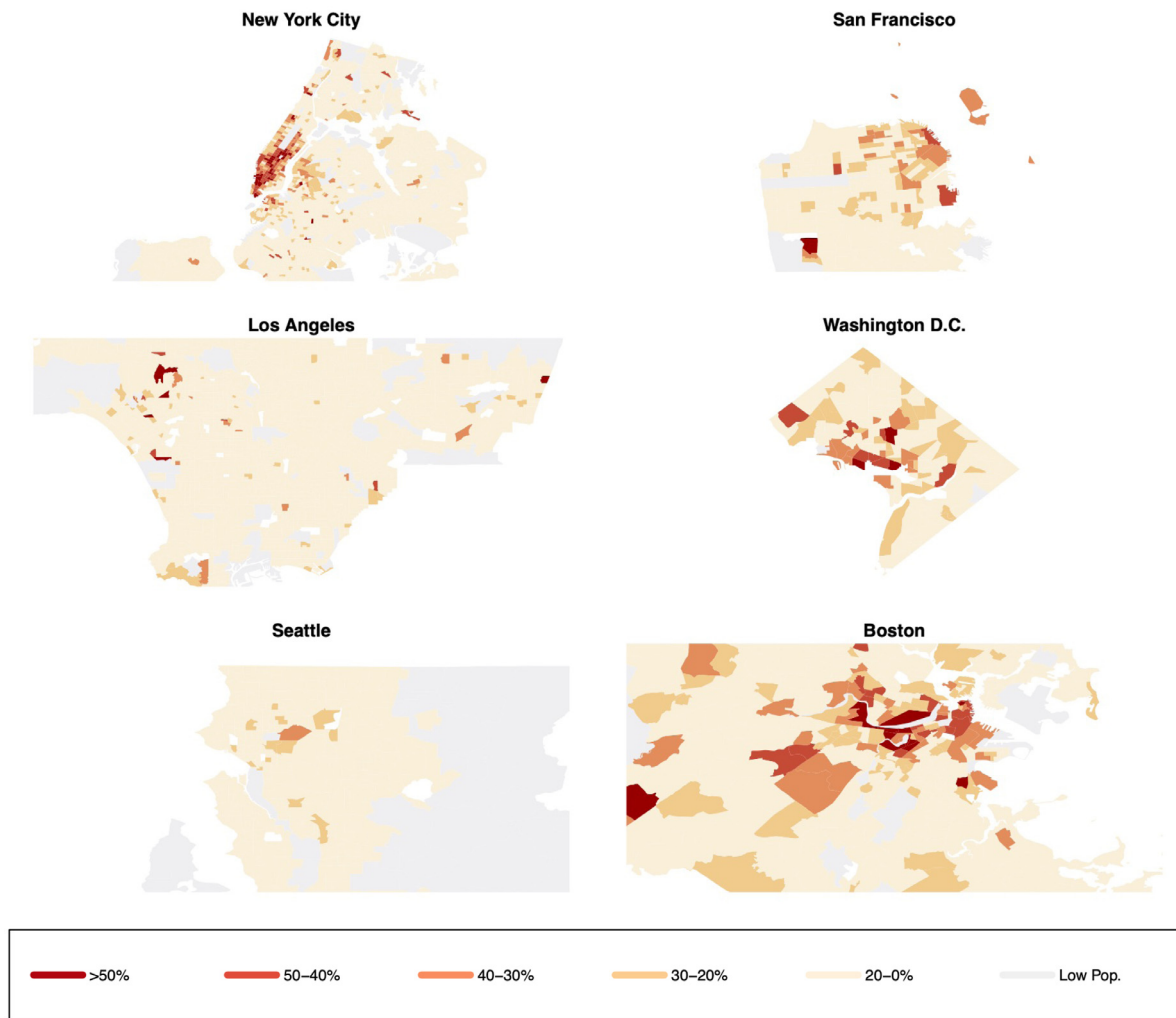**Fig. 2.** Demographic association of urban flight. *Notes:* Panel A shows average household income against the change in the fraction of out of town visits. Panel B repeats the exercise for the fraction who are white, and Panel C for the fraction between the ages of 18–45. Data is tract-level and shown only for New York City, San Francisco, Los Angeles, Washington DC, Seattle, and Boston. Out of town visits are winsorized at the 5% level. All demographic variables are left un-winsorized, with the exception of the fraction between 18 and 45 in Panel C. Panel D show inflows per 1000 people between pairs of ZIP codes against SCI between the two ZIP codes. Panel E shows daily inflows against new cases at the county level. Light gray points show 1/100th of the entire sample; for each demographic variable, the data is divided into 25 quantiles, and each dark blue dot represents the average fraction of the population sleeping at home and average demographic variable within each quantile. Income data are drawn from the IRS SOI Tax Statistics at https://www.irs.gov/statistics/soi-tax-stats-individual-income-tax-statistics-zip-code-data-soi, and demographic data are drawn from the ACS. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

We measure the pre-existing urban population in each tract and plot the fraction that has left by 29 March 2020, corresponding to heightened COVID-19 lockdown restrictions nationwide. New York City experienced extremely high flight concentrated in Manhattan, with several census tracts seeing over 50% of the resident population leaving by 29 March. Flight was concentrated in the downtown and midtown regions, though we also observe extensive urban flight in the Upper West Side, Upper East Side, and the wealthier regions of Brooklyn. We also observe distinctive patterns of urban flight in San Francisco (concentrated in the downtown regions) as well as Boston (high levels of exodus in

New York City

San Francisco

Los Angeles

Washington D.C.

Seattle

Boston

>50%	50–40%	40–30%	30–20%	20–0%	Low Pop.

Fig. 3. Propensity to leave cities by tract. *Notes:* Flight is at the census tract level and represents the average of the fraction of residents identified in February 2020 who were not present in their home county in the last week of March 2020. For example, we can think about a person identified in February as having a modal night tract in Greenwich Village, Manhattan. If this person on March 28th has more than half of their night pings in a different county, outside of Manhattan, we count them as having left. For this tract in Greenwich Village on March 28th, we count the number of February residents present in our data on March 28th who are outside of Manhattan on the night of March 28th. We divide this by the total number of February residents who are in the data on March 28th. We average this across the last week of March to get the propensity number above.

Cambridge and downtown Boston). These maps suggest that large-scale urban flight was a major reaction to the COVID-19 pandemic in its early stages, with responses concentrated in the richer parts of several major metropolitan areas.

We examine the persistence of these moves on the dimension of out-migration in Fig. A2. In each period, we examine the fraction residents still present in the city, across the six city sample. This plot measures for those identified as residents of a city in February, the fraction who are still present in the city relative to those that have left (and can be established to be present elsewhere in the country). We find that as many as 15% of the residents in Manhattan were located elsewhere in the country before the Thanksgiving and Christmas holidays, while up to 30% are elsewhere in the country at the end of our sample during the Christmas holiday. These suggest that a substantial component of urban flight appears persistent, as least through the year of 2020. We observe similar declines in the propensity to stay across residents of other cities. Washington DC, for instance, saw relatively little flight early on—but substantial urban exits over time.

We also measure the persistence of urban flight in a complementary dataset, the USPS change of address data, in Fig. A3. This dataset has been used to measure migration by multiple publications including

the New York Times,[8] and The City.[9] In contrast to the high-frequency moves corresponding to physical location in the VenPath dataset, the USPS moves are designated as permanent changes of address, and are observed more often in summer months (when residential leases often end). The longer time frame of this dataset, July 2018 to June 2022, allows us to compare migration during COVID to baseline migration and then assess the persistence of the moves. We observe an elevated rate of urban moves for several cities in our sample during the start of the pandemic. There is no net inflow at all in 2021 or 2022, let alone an inflow close to the magnitude of the net outflow during 2020. The more permanent nature of these moves and the lack of net inflow suggest that many moves during 2020 were persistent.

3.2.4. Social connectedness and urban migration

To further understand the determinants of this pattern of domestic migration, we examine role of social connections in determining where

---

[8] https://www.nytimes.com/interactive/2021/04/19/upshot/how-the-pandemic-did-and-didnt-change-moves.html
[9] https://www.thecity.nyc/2022/5/31/23145072/nycs-population-plummeted-during-peak-covid-and-its-still-likely-shrinking

individuals flee to in Fig. 2. In Panel D of this figure, we plot pre-existing social connections measured using the Facebook SCI variable and migration between March 1st and July 13th at the ZIP code level across the entire United States. Background points show a 1/100th random sample, while dark points show a binscatter of the 25 quantiles. We find a very strong positive association between higher social connectivity between ZIP codes and migration over this period. The strong relationship between inflow and SCI suggests that individuals with the ability to leave disproportionately went to areas where they had pre-existing social networks, and could take refuge with friends and family.

We then examine the relationship between migratory inflow and subsequent new cases in Panel E of Fig. 2. In these plots, we move to the county-level, in which we have COVID-19 case information. We plot the daily new cases against daily inflow for all counties over our entire sample period. Binscatter dots show the 25 quantiles of the distribution, and suggest a strong relationship between migratory inflows and new cases. Our graphical evidence suggests that urban migration, directed towards socially connected regions, had spillover effects on destination regions by increasing COVID-19 case counts for destination counties.

3.3. Impact of urban flight on nationwide COVID-19 cases

Having established the nature of urban flight over the course of the COVID-19 pandemic, we turn next to an analysis on the implications of this flight on destination regions. Because coronavirus is a predominantly respiratory disease spread in close contact, direct exposure with individuals formerly living in high-risk areas is a plausible vector for disease spread. While urban areas with international connections (particularly Seattle and NYC) appear to have been the initial hotspots for COVID-19, the disease appears to have quickly spread from those areas to outlying regions through the travel patterns of affected individuals. We explore the idea that urban flight by individuals avoiding the risk of contagion in urban areas may have seeded the pandemic in the rest of the country.

We highlight the impact of increased migration on increased COVID-19 cases in destination counties and focus on migration from NYC to illustrate our key mechanism and hypothesis (Fig. 4). The city is central to our analysis, both due to the size of its urban flight as well the its early presence of COVID-19. We first separate our analysis into different categories of urban areas, based on attributes of the destination regions. Regions differ in their exposure to infectious disease on the basis on urban features, so we analyze separately the impact of inflows of New York City residents to large and medium sized metropolitan areas (NCHS category 1, 2 & 3 in Panel A) as well as micropolitan and non-core areas (NCHS categories 4, 5 & 6 in Panel B). Within each category, we compare total cases among counties that received the highest quartile of inflow of New Yorkers, compared with counties that saw the lowest quartile of inflow. Left panels show per capita cases in logs, while right panels show total cases.

We find sizable impacts of urban inflow from NYC on increasing COVID-19 cases across regions. In the largest urban areas, we find that cases start to increase for counties that receive high inflow from NYC beginning in March. We plot the log of the seven day average in total cases for the counties receiving highest and lowest quartile of NYC inbound residents, and plot the difference of log(total cases) between these regions in gray bars in the background. Regions that saw high inbound migration see the greatest relative difference in cases in the beginning of April, a difference that declines over time. The timing of case growth matches the period of influx of NYC residents with a lag, consistent with a channel of direct infection.

The difference between areas with high inflow of NYC residents and areas with low inflow starts to decline in July. We show similar plots that depict the rate of change of new cases in Fig. A10, which confirm that areas that initially saw higher NYC resident influx later experienced negative relative growth in new cases into the summer. This result suggests that NYC inflow brought forward some cases which may

have been counterfactually experienced later in the course of the pandemic. Urban flight would still be quite important, even if it only accelerated case growth, because of steady improvements in treatment, the expanded supply of personal protective equipment such as masks which lowered mortality rates among those infected, and the avoidance of hospital overcrowding during the early period of the pandemic.[10] Another possibility is that case counts have also converged over time because the number of cases in New York declined over time, into the summer.

We observe that the impacts of NYC influx are increasing in city size. Urban areas which saw higher influx of NYC residents saw the greatest increase in new cases (a representative example would be Atlanta, which saw substantial case growth in the first wave of the pandemic). Large fringe and medium metropolitan areas also see substantial increases in early cases as a result of New York City inflow, but to a lesser degree than the largest urban areas. However, micropolitan and non-core areas see substantially weaker effects, which also turn negative around mid-May. Urban influx could be most related to subsequent case growth in the largest urban areas due to greater realized population density and possibility for individuals to interact in the crowded, indoor environments which are most conducive to COVID-19 spread.

3.4. OLS and IV

To expand our focus to migration across the entire United States, account for endogeneity in the migration decision, and control for additional factors, we turn to our core regression specification in Table 1 which follows our primary specification in Eq. (1) and covers inflows and new cases from March 1, 2020 to July 13, 2020, and from November 1, 2020 to December 31, 2020. We first show results for our OLS specifications (columns 1–6) and then instrument for migration patterns with Facebook friendship linkages (columns 6–10) to confirm a relationship between migration and subsequent new cases.[11] Our IV estimates suggest statistically significant and economically substantial estimates. A one standard deviation increase in instrumented per capita inflows is associated with a one standard deviation increase in new cases per capita. Alternatively, an additional 7 travelers per 1000 individuals is associated with 1 additional case per 1000 individuals (column 12).

Inflow per capita is a significant driver of cases per capita throughout our sample period in both the OLS and IV specifications, while the impact of the level of inflows on new cases in level terms dissipates over time as shown in Fig. 5. In the initial phase in our sample (March 2020–May 2020), for every additional 100 people who enter a county, new cases increase by 0.6, dropping to 0.2 during the summer months, and increasing to 0.9 in November and December, coincident with higher inflows during the winter holidays. We consistently observe that migration from areas with higher case loads, and influx from areas farther away, lead to higher infection rates. These results are consistent with long-distance and inter-state migration trends, especially from NYC, which also contribute to the greater spread of COVID-19 around the country. Due to a gap in our data, we are unable to estimate the association between inflow and new cases in the second half of July, August, September, and October 2020 due to a gap in our data, limiting our ability to answer questions about travelers and COVID during those months.

The difference between the impact of inflows on new cases in level terms relative to per capita terms is driven in part by geographic variation in where new cases emerge over different phases of the pandemic. Early in the pandemic, New York City and other large urban areas ac-

---

[10] See Horwitz et al. (2021) and Ciceri et al. (2020) on improved mortality and Gandhi and Rutherford (2020) who connect the improved mortality to increased mask adherence. Other explanations for greater mortality at the early stage of the pandemic including crowding at hospitals, learning-by-doing in medical care, and improved treatments over time.

[11] Coefficients in this table, other than columns 4–6, are scaled by 1×10^3, and so correspond to the case increase resulting from an additional influx of 1000 people. Columns 4–6 are scaled by 1×10^6.

Table 1

Impact of migration on new COVID-19 cases .

| | Panel A: New COVID-19 Cases Against Inflow | | | | | | | | | | | |
| | OLS | | | | | | IV | | | | | |
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Inflow	98.388***	6.142	−8.367				484.466***	154.914***	−40.224			
	(4.654)	(4.784)	(5.184)				(25.060)	(19.663)	(52.366)			
Per Capita Inflow				177.069	255.225	233.067				750.188***	76.223***	144.744***
				(186.295)	(195.587)	(200.882)				(67.740)	(19.096)	(20.617)
High Incoming Cases	11975.997***	12314.294***	12320.634***				13933.822***	13166.577***	12640.649***			
	(251.414)	(287.118)	(289.223)				(418.898)	(335.291)	(393.275)			
High Incoming Cases Per Capita				209.552***	211.982***	211.981***				0.271***	0.218***	0.223***
				(1.279)	(1.292)	(1.292)				(0.008)	(0.002)	(0.002)
Far Indicator	24729.038***	15438.162***	−12824.185***	12.105***	7.697*	6.800	27584.627***	18805.957***	−146091.950*	0.005	0.025***	0.261***
	(2296.018)	(1924.813)	(2364.570)	(3.919)	(4.050)	(4.438)	(2666.766)	(1982.508)	(87068.460)	(0.026)	(0.007)	(0.023)
Far Indicator× Inflow			103.759***						595.945*			
			(9.423)						(320.225)			
Far Indicator× Per Capita Inflow						278.969						−69.070***
						(681.191)						(6.590)
All Controls	N	Y	Y	N	Y	Y	N	Y	Y	N	Y	Y
State × Month × Population Decile FE	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
N	595,945	573,968	573,968	595,945	573,968	573,968	595,945	573,968	573,003	595,945	573,968	573,968
	Panel B: First Stage Estimates of Inflow on SCI											
Weighted SCI							42277.757***	32742.431***	34931.996***	0.009***	0.015***	0.013***
							(592.704)	(625.707)	(678.997)	(0.001)	(0.001)	(0.001)
F							5088	2738	1490	150	297	171

Columns 1–3 show our main regression specification (1). Columns 4–6 repeat the exercise using inflow per capita as the explanatory variable. Columns 7–12 repeat the exercise for columns 1–6, where inflow is instrumented with the weighted SCI, as in (2). The sample period is March 1, 2020 through December 31, 2020. Standard errors are in parentheses, and * denotes 10% significance, ** denotes 5% significance, *** denotes 1% significance. Note that all coefficients and standard errors in Panel A are scaled up by 1×10^3, with the exception of columns 4–6, where coefficients and standard errors are scaled up by 1×10^6. All coefficients and standard errors in Panel B are scaled up by 1×10^6. The indicator for high cases in originating counties is equal to 1 for counties where the inflow-weighted cases from incoming counties fall within the top quartile in the current month. We construct $\mathbb{1}$(Far) by assigning 1 to counties where the inflow-weighted distance is at least 500km (roughly equal to the distance between NYC and Pittsburgh).

Panel A: Large Central, Fringe, and Medium Metropolitan Areas
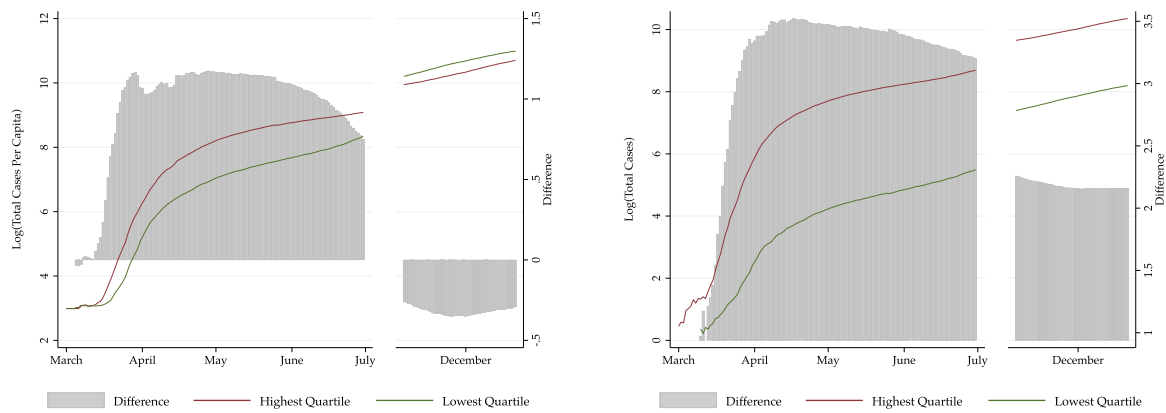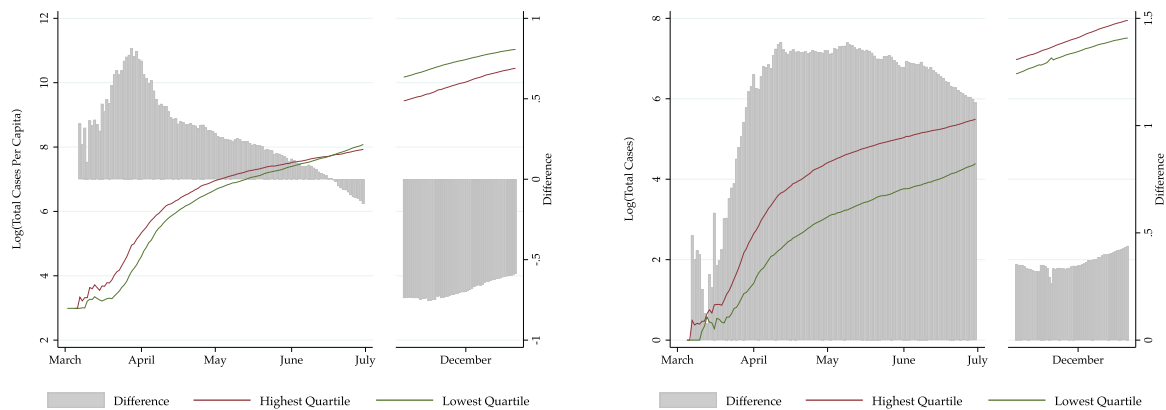

Panel B: Micropolitan and Non-Core Areas

Fig. 4. March inflows from NYC vs. log total cases. *Notes:* Counties are split into quartiles based on total inflows from New York City during the month of March. The two charts on the left show log(total cases per capita) over time for counties in the fourth quartile and the first quartile of inflows. The two charts on the right show log(total cases). Urban classification based on the NCHS urban-rural classification scheme: large central metros, large fringe and medium metropolitan areas (categories 1, 2, & 3) and micropolitan and non-core areas (categories 4, 5 & 6).

counted for the majority of new cases. In the summer and fall months, other regions of the country such as the Midwest and southern states accounted for the bulk of new cases. Consistent with our observation that counties which received more inflow from New York City experienced new cases earlier in the year, the Midwest and southern states are regions which saw lower inflows from New York City in the initial stages of the pandemic and correspondingly saw new cases arise relatively later during the year. Other possible explanations for regional variation in when the bulk of new cases emerge include increased community transmission. We note that counties which experience new cases during the summer and fall months tend to be smaller in terms of population, which is a contributing factor to higher new cases per capita, and helps explain why per capita inflow remains a significant driver of new cases per capita throughout the sample. As an illustrative example, consider Kings County, NY and Kings County, CA, which both saw an average per capita inflow of approximately 900 individuals per 100 K people. New cases per capita in Kings County, New York, the most populous county in the state, reached a high of 737 cases per 100 K individuals in our sample; in contrast, Kings County, California (located roughly halfway between San Jose and Los Angeles) saw a high of 2655 cases per 100 K individuals in late November, despite the level of new cases remaining an order of magnitude lower than that of its New York counterpart.

We also examine the relative importance of raw inflow as compared to inflows from high-case areas and inflow from distant counties. These considerations may be relevant to policymakers when exploring possible travel restrictions. Table A2 shows a standardized version of Table 1, with all non-indicator variables standardized to have zero mean and unit variance. In terms of relative magnitudes, we show that inflows from distant counties have the largest impact on the number of new cases in levels, whereas raw inflows have the largest impact on new cases per capita.

Our IV estimates are larger than our OLS estimates, since raw inflow tends to overweight areas which exhibit lower new cases relative to the SCI-instrumented inflow. This appears likely because a substantial component of urban flight was motivated by fleeing to geographically remote areas where new cases were more likely to be low. These remote regions tend to be places where travellers do not have many existing social connections (e.g. renting a temporary property in upstate New York, or visiting a second home in a region where individuals have fewer contacts). Our IV, by contrast, identifies a LATE based on migration to socially connected regions. Migration towards these areas, which is instead highlighting the flow based on the migration towards friends and families, appears to be more conducive to COVID-19 case transmission. Our IV, additionally, cleans up potential measurement error in our measurement of migration.
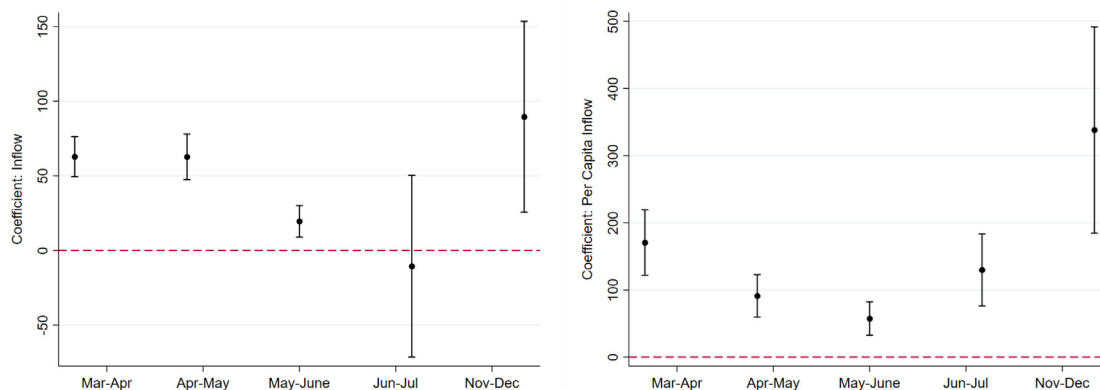
Fig. 5. Dynamic impact of inflows on new cases. *Notes:* This figure shows the impact of inflows on new cases in levels (left panel) and in per capita terms (right). We plot the coefficient on inflows and per capita inflows from our IV specification—columns (9) and (12) in our main regression specification (1) over rolling 60 day periods.

To provide a descriptive sense for the differing changes in sample and provide further support for why we see larger IV estimates, we provide some evidence in Appendix Fig. A8. We show that instrumented inflow tends to be higher in regions with fewer new cases. These regions tend to have lower population density and a higher proportion of seasonal homes. In short, our OLS estimate underestimates the impact of inflow on new cases, because inflow is positively correlated with a variable which drives lower new cases (a characteristic which naturally makes these regions more desirable destinations for those fleeing urban areas). Our focus on the SCI measure effectively picks a different set of regions across the United States based on predicted inflows due to social connections, rather than realized migration activity. Raw inflows tend to pick up coastal and rural areas, in particular, relative to more connected urban areas in the SCI measure.

Finally, we find a strong and statistically significant first stage, and report the coefficient and F-statistic from the first stage regression (2) in Panel B of Table 1 for the IV regressions.

3.5. Robustness

We perform several key robustness tests on our primary sample. We examine specifications that restrict on flight from New York City specifically. New York City's large exodus, substantial case load, and early timing in the pandemic make it an ideal focus for our analysis. In Table A3, we regress new local cases against an indicator for counties in the top quartile of those receiving inflows from NYC in March, which is most comparable with our graphical evidence. In Table A4, we repeat our full analysis but subset to include inflows from NYC specifically. Both specifications reveal large and statistically significant effects. An additional robustness table, A5 clusters standard errors at the Commuting Zone-level. While we lose significance in our per capita results, our core inflow measure remains statistically significant.

These results suggest that urban flight, and specifically the large urban exodus from NYC, was important in the spread of the COVID-19 pandemic across the United States. Our findings also potentially help explain the result in Kuchler et al. (2022), that greater social connections with Westchester (another pandemic hub) helps to predict subsequent COVID-19 deaths. A plausible transmission mechanism which we explore here, is the refugee behavior of NYC residents who migrate into these socially connected regions.

We connect the cross-sectional evidence with information on the specific timing of case growth, in Appendix Fig. A5 we examine a projection approach which looks at when new cases increase after migration flows. We find that new cases and deaths tend to increase two weeks after the initial migration event, consistent with the typical incubation times for COVID infection (Wilson, 2020). While we have assembled evidence in both the cross section and the time series on the associations between

urban flight and new cases, without more exogenous variation it is not possible to disprove that an alternative variable like low COVID restrictions causes new cases and is correlated with inflows.

A key underlying assumption in our analysis is that social connectivity between areas drives inflows between these specific places, ultimately driving case growth. A potential concern is that social connectivity simply proxies for inflows from a variety of destinations instead. In order to address this possibility, and to further descriptively analyze the role of travel on spread, we turn data from the Florida Department of Health, which notes whether a recorded case was travel-related. A case is defined as travel-related if there is a known history of exposure to COVID associated with travel outside of Florida.

We find that many cases in Florida in March came from travelers who were possibly exposed while outside of Florida. Fig. A6 shows that travel-related cases account for 40% of all new reported cases in March, and 10% of all cases are attributed to travel from New York State: highlighting again the importance of New York. By contrast, the majority of cases after March were non-travel related, and were not associated with exposure to COVID due to travel outside of Florida. To clarify, these cases could come from contact with those who were exposed outside of Florida. In Panel B, we show that social connections between Florida counties and originating states predict the number of travel-associated cases from that state specifically, highlighting how travel during COVID moved along social networks. A traveler with COVID-19 can cause new cases to increase in the destination region in two ways. Cases mechanically increase by the traveler's presence in the new county. Cases could also increase through that traveler spreading COVID-19 to those in the destination region. This spillover, not the mechanical increase, is what we are interested in. To examine the impact of inflow on COVID-19 spillovers, in Panel C we remove the mechanical travel related cases. We show that the impact of the SCI-instrumented inflows on cases without the mechanical travel related cases is not significantly different from the impact on total cases. The implication is that the role of inflows is largely felt on community transmission in affected areas in the form of spillovers, not just on the direct infections by individuals through the mechanical channel.

4. Conclusion

We document substantial urban flight in the wake of the COVID-19 pandemic and find large effects of this migration on the spread of COVID-19 elsewhere in the country. Migration responses were widespread among individuals living in major urban areas, such as New York City. As much as 15–20% of Manhattan's population had fled by the middle of the summer in 2020. These individuals came from areas which were disproportionately wealthy, white, and young. These individuals appear to have left for regions with a high degree of social connections

to NYC, suggesting that individuals took shelter with friends and family. We then use the social networks structure to develop an estimate of the impact of migration on new cases.

We find that instrumented migration patterns predict subsequent rise in cases in destination counties, suggesting that urban flight contributed to the pandemic, changing it from an initially urban disease to a more widespread, nationwide pandemic. These results demonstrate that a relationship exists between urban flight, as a massive response to the pandemic in its initial phase, and the spread of COVID-19 to city dwellers' socially-connected friends, families and host regions of the United States.

Our work has implications for public policy in the wake of the disease. First, it highlights an important feature of urban flight. Wealthy individuals, who contribute disproportionately to the local revenue and tax base of cities, are more likely to flee cities. This finding points to important challenges for municipal finances in the wake of the pandemic, and has implications for the future of cities. Second, our work highlights the role of domestic migration in spreading the pandemic. As such, our work suggests the possible value of travel restrictions—in the form of limits, quarantine periods, or testing requirements as a precondition for entry—to help curb the spread of COVID-19.

CRediT authorship contribution statement

Joshua Coven: Formal analysis, Writing – review & editing, Software. **Arpit Gupta:** Conceptualization, Methodology, Writing – original draft. **Iris Yao:** Formal analysis, Writing – review & editing, Software.

Supplementary material

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.jue.2022.103489

References

Allcott, H., Boxell, L., Conway, J., Gentzkow, M., Thaler, M., Yang, D., 2020. Polarization and public health: partisan differences in social distancing during the coronavirus pandemic. J. Public Econ. 191, 104254. doi:10.1016/j.jpubeco.2020.104254.

Andersen, M. S., 2020. Early evidence on social distancing in response to COVID-19 in the United States. Working Paper.

Athey, S., Ferguson, B., Gentzkow, M., Schmidt, T., 2021. Estimating experienced racial segregation in US cities using large-scale GPS data. Proc. Natl. Acad. Sci. 118 (46). doi:10.1073/pnas.2026160118. e2026160118.

Bailey, M., Cao, R., Kuchler, T., Stroebel, J., Wong, A., 2018. Social connectedness: measurement, determinants, and effects. J. Econ. Perspect. 32 (3), 259–280.

Barrios, J.M., Benmelech, E., Hochberg, Y.V., Sapienza, P., Zingales, L., 2021. Civic capital and social distancing during the COVID-19 pandemic. J. Public Econ. 193, 104310. doi:10.1016/j.jpubeco.2020.104310.

Barrios, J.M., Hochberg, Y.V., 2021. Risk perceptions and politics: evidence from the COVID-19 pandemic. J. Financ. Econ. 142 (2), 862–879. doi:10.1016/j.jfineco.2021.05.039.

Brinkman, J., Mangum, K., 2022. The geography of travel behavior in the early phase of the COVID-19 pandemic. J. Urban Econ. 127, 103384. doi:10.1016/j.jue.2021.103384. JUE Insights: COVID-19 and Cities.

Chandrasekhar, A.G., Goldsmith-Pinkham, P., Jackson, M.O., Thau, S., 2021. Interacting regional policies in containing a disease. Proc. Natl. Acad. Sci. 118 (19). doi:10.1073/pnas.2021520118. e2021520118.

Chen, M.K., Chevalier, J.A., Long, E.F., 2021. Nursing home staff networks and COVID-19. Proc. Natl. Acad. Sci. 118 (1). doi:10.1073/pnas.2015455118. e2015455118.

Chen, M.K., Haggag, K., Pope, D.G., Rohla, R., 2020. Racial disparities in voting wait times: evidence from smartphone data. Rev. Econ. Stat. 1–27. doi:10.1162/rest_a_01012.

Chen, M.K., Rohla, R., 2018. The effect of partisanship and political advertising on close family ties. Science 360 (6392), 1020–1024. doi:10.1126/science.aaq1433.

Chiou, L., Tucker, C., 2020. Social Distancing, Internet Access and Inequality. Working Paper. National Bureau of Economic Research doi:10.3386/w26982.

Ciceri, F., Ruggeri, A., Lembo, R., Puglisi, R., Landoni, G., Zangrillo, A.on behalf of the COVID-BioB Study Group, 2020. Decreased in-hospital mortality in patients with COVID-19 pneumonia. Pathog. Glob. Health 114 (6), 281–282. doi:10.1080/20477724.2020.1785782. PMID: 32584660.

Couture, V., Dingel, J.I., Green, A., Handbury, J., Williams, K.R., 2022. JUE insight: measuring movement and social contact with smartphone data: a real-time application to COVID-19. J. Urban Econ. 127, 103328. doi:10.1016/j.jue.2021.103328. JUE Insights: COVID-19 and Cities.

Engle, S., Stromme, J., Zhou, A., 2020. Staying at home: mobility effects of COVID-19. Working Paper.

Gandhi, M., Rutherford, G.W., 2020. Facial masking for COVID-19 potential for "variolation" as we await a vaccine. N. Engl. J. Med. doi:10.1056/NEJMp2026913.

Glaeser, E.L., Gorback, C., Redding, S.J., 2022. JUE insight: how much does COVID-19 increase with mobility? Evidence from New York and four other U.S. cities. J. Urban Econ. 127, 103292. doi:10.1016/j.jue.2020.103292. JUE Insights: COVID-19 and Cities.

Holtz, D., Zhao, M., Benzell, S.G., Cao, C.Y., Rahimian, M.A., Yang, J., Allen, J., Collis, A., Moehring, A., Sowrirajan, T., Ghosh, D., Zhang, Y., Dhillon, P.S., Nicolaides, C., Eckles, D., Aral, S., 2020. Interdependence and the cost of uncoordinated responses to COVID-19. Proc. Natl. Acad. Sci. doi:10.1073/pnas.2009522117.

Holtz, D., Zhao, M., Benzell, S.G., Cao, C.Y., Rahimian, M.A., Yang, J., Allen, J., Collis, A., Moehring, A., Sowrirajan, T., et al., 2020. Interdependence and the cost of uncoordinated responses to COVID-19. Proc. Natl. Acad. Sci. 117 (33), 19837–19843.

Horwitz, L.I., Jones, S.A., Cerfolio, R.J., Francois, F., Greco, J., Rudy, B., Petrilli, C.M., 2021. Trends in COVID-19 risk-adjusted mortality rates. J. Hosp. Med. 16 (2), 90–92. doi:10.12788/jhm.3552.

Kuchler, T., Russel, D., Stroebel, J., 2022. JUE insight: the geographic spread of COVID-19 correlates with the structure of social networks as measured by facebook. J. Urban Econ. 127, 103314. doi:10.1016/j.jue.2020.103314. JUE Insights: COVID-19 and Cities.

Lee, J.N., Mahmud, M., Morduch, J., Ravindran, S., Shonchoy, A.S., 2021. Migration, externalities, and the diffusion of COVID-19 in South Asia. J. Public Econ. 193, 104312. doi:10.1016/j.jpubeco.2020.104312.

Mangrum, D., Niekamp, P., 2022. JUE insight: college student travel contributed to local COVID-19 spread. J. Urban Econ. 127, 103311. doi:10.1016/j.jue.2020.103311. JUE Insights: COVID-19 and Cities.

Painter, M., Qiu, T., 2021. Political beliefs affect compliance with government mandates. J. Econ. Behav. Organ. 185, 688–701. doi:10.1016/j.jebo.2021.03.019.

Rosenberg, C.E., 1968. The Cholera Years: The United States in 1832, 1849, and 1866. University of Chicago Press.

Wilson, D.J., 2020. Weather, Social Distancing, and the Spread of COVID-19. Working Paper.