URBAN FLIGHT SEEDED THE COVID-19 PANDEMIC ACROSS THE UNITED STATES

Joshua Coven*
Arpit Gupta†
Iris Yao‡

October 14, 2020

Abstract

We document large-scale urban flight in the United States in the wake of the COVID-19 pandemic. Populations that flee are disproportionately younger, whiter, and wealthier. Regions that saw migrant influx experience greater subsequent COVID-19 case growth, suggesting that urban flight was a vector of disease spread. Urban residents fled to socially connected areas, consistent with the notion that individuals were sheltering with friends and family or in second homes. The association of migration and subsequent case growth persists when instrumenting for migration with social networks, pointing to a causal association.

Keywords: Migration, COVID-19, Contagion
JEL Classification: I140, J610, R410

*NYU Stern School of Business, jcoven@stern.nyu.edu
†NYU Stern School of Business, arpit.gupta@stern.nyu.edu (Corresponding Author)
‡NYU Stern School of Business, yyao2@stern.nyu.edu

I 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 by Charles E. Rosenberg

The density and international connections of cities foster human interaction and economic activity. These same associations, however, have historically made urban areas uniquely vulnerable to contagious disease. The role of urban proximity in pandemics has seen renewed attention in the context of novel coronavirus disease 2019 (COVID-19). However, the channels through which communities employ resources to mitigate their own risk, and the spillovers of these actions on broader community transmission, remain unclear.

This paper quantifies the extent of urban flight in response to COVID-19 in its initial phase in the United States, and documents how this migratory behavior seeded the pandemic in the rest of the country. We take advantage of mobile phone geolocation data, which allows for much higher frequency analysis than has been possible in prior studies of migratory behavior. We find that as much as 15–20% of the population of some high-income urban regions, such as Manhattan, fled in response to COVID-19. Regions that saw greater flight were generally richer, whiter. and younger than other areas, pointing to important disparities in the availability of migration as a risk-mitigating technique during the pandemic. We use Facebook friendship data to establish that migration was especially high between socially connected regions, consistent with the idea that urban flight led to sheltering with friends and family, or in second homes.

We then document the impact of this urban flight on increases in COVID-19 cases in the destination counties. Our instrumental variable strategy leverages the social connections between countries to causally identify the relationship between migratory flows and increased cases. We find that a one standard deviation increase in SCI-instrumented inflow is associated with a 0.5 standard deviation increase in case growth. Alternatively, an increase in inflows by an additional 100 residents raises local cases by about 20. The median county sees an average instrumented inflow of 120 residents, and over 10% of all counties see instrumented inflows of over 200 residents during our sample period. Our estimates are substantial and point to urban migration as an important vector of COVID-19 spread across the United States.

Our results are also relevant to both questions of local government public finance shocks, as well as the long-term future of cities. Because the population which left was disproportionately rich, 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 results end in July 2020, and so we are limited in addressing these long-term questions, we quantify important urban disruptions in the wake of COVID-19.

Our results additionally have implications on the possible value of travel restrictions. Virtually every country has placed severe restrictions on international travel, against the initial advice of the World Health Organization. Many countries have additionally placed restrictions on regional travel within countries. These include complete bans on interstate travel, as between Victoria and New South Wales in Australia[1], as well as requested isolation orders for out-of-state visitors, as prevailed in some U.S. states in our period, such as Rhode Island.[2] However, the potential value of these travel restrictions remains very unclear. Our results suggest that intra-US travel was a main vector spread, and therefore

[1]See https://www.npr.org/2020/07/06/887659557/australia-closes-interstate-border-because-of-coronavirus-outbreak.

[2]See Rhode Island Government official press release from April 7: https://www.ri.gov/press/view/38091.

the existing level of travel restrictions and self-suggested quarantine orders were insufficient to prevent out-of-state migration from impacting local spread. As a consequence, our work points to the value of potential policies related to travel restrictions in curbing spread.[3]

Our work is closely linked to a rapidly growing literature using mobile phone geolocation data to assess the spread of COVID-19. Most closely related is work by Chiou and Tucker (2020), which finds shelter-in-place effects vary by income. This paper differs by considering the role of leaving the city, connecting mobility with actual COVID exposure, and incorporating analysis of other demographic groups. Glaeser, Gorback, and Redding (2020) examines the change in mobility within regions, while we examine the flight response across areas. Other work (Allcott, Boxell, Conway, Gentzkow, Thaler, and Yang, 2020; Barrios and Hochberg, 2020; Engle, Stromme, and Zhou, 2020; Painter and Qiu, 2020; Andersen, 2020) has looked at political partisanship and COVID-19 responses.

Prior work, such as Athey, Ferguson, Gentzkow, and Schmidt (2019), Chen, Haggag, Pope, and Rohla (2019), 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 Chen, Chevalier, and Long (2020), who examine nursing home networks in the wake of COVID-19. Holtz, Zhao, Benzell, Cao, Rahimian, Yang, Allen, Collis, Moehring, Sowrirajan, Ghosh, Zhang, Dhillon, Nicolaides, Eckles, and Aral (2020) also uses Facebook data to shows similarity in social distancing responses among regions connected through friendship links. Our work highlights a direct migration linkage between socially connected regions.

Our work also relates to a literature that uses geonotyping of strains to establish both international chains of transmission in Europe (Pybus, Rambaut, COG-UK-Consortium,

[3]Chandrasekhar, Goldsmith-Pinkham, Jackson, and Thau (2020) also highlights the importance of regional spillovers and network interactions. We contribute to this work by quantifying the role of the migration channel in contributing to case growth at an early stage of the COVID-19 pandemic. Lee, Mahmud, Morduch, Ravindran, and Shonchoy (2020) also finds an association of migration and cases in the context of COVID-19 in South Asia.

et al., 2020) as well as domestic transmission in the United States (Fauver, Petrone, Hodcroft, Shioda, Ehrlich, Watts, Vogels, Brito, Alpert, Muyombwe, et al., 2020). While this work establishes the strain similarities across regions in the United States points to a role for domestic transmission, we provide additional evidence on the physical migration behind this regional transmission.

II Data and Specification

II.A Data

Mobile location data was sourced from VenPath—a holistic global provider of compliant smartphone data. We obtain unique data on smart phone GPS signals. Our data provider aggregates information from approximately 120 million smart phone users across the United States. GPS data were combined across applications for a given user to produce pings corresponding to time stamp-location pairs. Ping data include 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 February 1 to July 13, 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.[4] We join this with a county-to-county Social Connectivity Index (SCI) measures from Facebook as discussed in Bailey, Cao, Kuchler, Stroebel, and Wong (2018) and applied in reference to COVID-19 in Kuchler, Russel, and Stroebel (2020). We also include demographic data from the Census ACS, and urban-rural county classifications from the National Center for Health Statistics.

[4]Drawn from https://coronavirus.jhu.edu/. We also incorporate nursing home data from the Centers for Medicare and Medicaid Services and https://www.cms.gov/.

Appendix Table A1 summarizes some key statistics—in particular, we note that the distribution of caseloads across counties is highly skewed, with several key counties exhibiting extremely high growth in new cases while approximately half of the counties in our sample had fewer than 1 new case per day. Secondly, the difference between the SCI and its inflow-weighted counterpart shows that net inflows between counties are occurring most frequently between highly connected counties, giving some preliminary evidence in support of the SCI as a relevant instrument for county-county flows.

We isolate the migration behavior of users in the US by identifying each user's modal census tract each night (6pm - 8am) if they ping in it three or more times. We do this for each night in a month. If a user has the same tract as their modal night tract on at least five nights in that tract, we define it as their "home tract:" the Census tract that they spend the most time in during night hours. We repeat this process each month from February to June to analyze mobility from March to July. We use only one month of data at a time to identify residents' home tracts. We then analyze their data in the month immediately following the month that was used to identify their home locations. The resulting sample includes a population of 9–11 million unique users per month for our base analysis across the United States. In New York City, 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 flow 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 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] In cases where 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 level to see where people from a given geography are spending the night on each date. The resulting measures estimate flows from home geographies to new geographies on each date.

[5]See: https://www.huduser.gov/portal/datasets/usps_crosswalk.html for the crosswalk.

II.B 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. After aggregating mobile phone migration data into county-day information, we sum, for a given day t, all net inflows into county i. Our core OLS specification measures case growth in a destination county as a function of gross inflow into that county:

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

We are primarily interested in the β_0 coefficient, which measures the effect of inflows on cases. We measure cases both in levels as well as 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 case growth. The indicator for high cases in originating counties 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 coronavirus case growth orthogonal to inflows from outside counties. Controls include the distance between the home and destination counties, mean household income, population density, the NHS 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 equation 1 is the endogenous nature of migration. If counties that receive higher domestic migration are also more likely to be susceptible to case growth for other reasons, a positive β_0 may simply reflect spurious correlation, rather than measuring the causal effect of migration on COVID-19 case growth. We develop an identification strategy to address endogenous migration decisions based on Facebook connectivity. We draw on prior research, as mentioned in Bailey, Cao, Kuchler, Stroebel, and Wong (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 pandemic-driven migration observed during COVID-19. To examine the relationship between migration and social connectivity, we first run a first stage regression of migration against social connectivity between regions:

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

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, Russel, and Stroebel (2020)) based on the number of users in two regions i and j as well as the friendship links between them:

$$SCI_{ij} = \frac{FB\_Connections_{ij}}{FB\_Users_i \times FB\_Users_j}. \tag{3}$$

Our primary specification examines social connectivity at the county level, which measures the total degree of connectivity across the country. 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 urban flight, we instrument county-county flows with the county SCI measure. The identifying assumption is that Facebook connections between county i and other counties does not correlate with the trajectory of case growth, *except* through the inflow of people into the county. Our instrumental variables specification first instruments for inflow using equation 2, and uses predicted inflow instead of realized inflow as a covariate in equation 2. 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.

III Results

III.A Temporal and Spatial Patterns of Urban Flight

We begin by highlighting the demographic characteristics of individuals who leave urban areas in the wake of the COVID-19 pandemic. Our initial focus is New York City, which was the epicenter of the pandemic in the initial stage, and we emphasize the scale of the migration response. However, subsequent results analyze urban flight across the country.

In Panel A of Figure A1, we show responses of individuals in leaving New York City by borough. We observe stark patterns in the response of individuals along this extensive margin: residents of Manhattan are substantially more likely to leave the city after the crisis, as are other wealthy parts of the city in Brooklyn. We find that as many as 10–15% of the population of Manhattan, formerly resident in the city in February, leaves the city 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 more likely to stay in the city. This urban exodus continues through the end of our sample. By July 10, we observe about 17% of the population of Manhattan remains absent in the city.

These shifts in leaving the city, however, are concentrated in the higher-income Census tracts, suggesting that richer New York City residents were disproportionately able to

take advantage of the option to flee the city and escape physical COVID-19 exposure in the city.

We confirm the role of income as a factor in explaining moves away from the city in Panel B of Appendix Figure A1, which shows a heatmap of responses by tract and date among tracts with median household incomes in excess of $100k. We find a large breakpoint in our sample in March 14, as reflected in the sharp changes in colors beginning on that date in a number of tracts, corresponding to a sharp rise in the increase of former New York City inhabitants leaving the city. This break comes just before Mayor Bill Di Blasio ordered schools, restaurants, bars, cafes, entertainment venues, and gyms in the city closed on March 16.[6] We observe very high flight behavior in the highest-income Census tracts after that date.

The fact that we observe migration response prior to the introduction of nonpharmaceutical interventions such as lockdowns is also consistent with Goolsbee and Syverson (2020), and points to important behavioral responses separate from policy responses. Unlike other mobility changes, such as sheltering at home—migration entails possible externalities on the exposure of others which we explore.

We examine the spatial distribution of flight patterns in Figure I, which shows the fraction of residents who leave across six cities (New York, San Francisco, Los Angeles, Washington DC, and Boston). We measure the pre-existing urban population in each Zip Code, and plot the fraction that has left by March 29, 2020. New York City experiences extremely high flight concentrated in Manhattan, with several Zip Codes seeing over 50% of the resident population leaving. Flight is 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 downtown regions) as well as Boston (high in Cambridge and downtown Boston). These maps suggest that large-scale urban flight

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

in response to the COVID-19 pandemic, with responses concentrated in the richer parts of several major metropolitan areas.

III.B Demographic Associations of Urban Flight

We examine the demographic associations of urban flight in Figure III, which focuses on the propensity to remain in our six city sample. We plot background demographic associations at the ZIP Code level against the fraction of ZIP Code population that stays. Background dots show all data points, while binscatter dots plot the average population within 25 quantiles. We find that the fraction of residents who remain in cities is strongly decreasing in tract income, decreasing in the fraction of the tract that is White, and the fraction of residents aged 18–45. These results are large in magnitude and statistically significant—moving from a tract at the bottom quartile of income to the top quartile raises the fraction of New York City residents who leave, for instance, by about 3 percent, or about the same as the unconditional average of the number of New York City residents who leave. The income and demographic associations are even stronger in the panels on the right of this Figure, which winsorize the fraction of out of town visits at a 5% level.

III.C Associations of National Migration

We first begin by descriptively analyzing the nation-wide nature of migration in the context of COVID-19. Figure II documents the net flow and outflow of residents across counties in the United States as of the end of each month.[7] 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 New York City as well as several other metropolitan areas (including Boston, Los Angeles, San Francisco, and Phoenix). This inflow went to a mix of interior locations including rural areas across

[7] Appendix Figure A2 examines inflow.

the country as well as Southern urban areas. Several cities in the Sunbelt, in particular Atlanta, Houston, Charlotte, and Austin saw substantial net in-migration during this period. Some other cities in 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 New York City, in the Hamptons and Hudson Valley. Broadly, the pattern of migration reflects flight away from the initial waves of the pandemic which hit the coasts more strongly, towards the national interior.

We observe continued urban flight from New York City, as well ad 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. Because our analysis ends in mid-July, we are limited in our ability to observe further migratory events associated with subsequent case increases across the Sunbelt regions of the United States over the period beginning in June. Our focus largely remains the substantial migratory response in the wake of the first migration event around the beginning of the pandemic in March, 2020.

To further understand this unprecedented pattern of domestic migration, we begin by examining role of social connections in determining where individuals flee to in Figure III. In Panel A of this figure, we plot pre-existing social connections measured using the Facebook SCI variable and migration between March 1 and July 13 at the ZIP Code level across the entire United States. Background light points show a $1/100^{th}$ random sample, while dark points show a binscatter of the 25th quintiles. We find a very strong positive association between higher social connectivity between ZIP Codes and migration over this period. We winsorize both SCI and inflows at a 1% level in the left plots, and at a 5% level in plots on the right and find similar results across both specifications. 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 case growth in Panel B of IV. In these plots, we move to the county-level in which we have COVID-19 case information. We plot the daily growth in cases against daily inflow for all counties over our entire sample period (March 1 to July 13). Binscatter dots show the 25 quantiles of the distribution, and suggest a strong relationship between migratory inflows and case growth. Our graphical evidence suggests that urban migration, directed towards socially connected regions, had spillover effects on destination regions in increasing COVID-19 case counts for destination counties.

III.D Impact of Urban Flight on Nationwide Case Growth

Having established the nature of urban flight over the course of the COVID-19 pandemic, we turn next to our main 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 the spread of the disease. While urban areas with international connections (particularly Seattle and New York City) appear to have been the initial hotspots for spread; the disease appears to have quickly spread from there to outlying regions through the travel patterns of affected individuals. We explore the idea that this pattern of extraordinary urban flight, by individuals perhaps avoiding the risk of contagion in urban areas, may have seeded the pandemic through the rest of the country.

To illustrate our key mechanism and hypothesis in graphical form, we begin by illustrating the relationship in Figure V. This figure highlights the impact of increased migration on increased cases in the destination counties and focuses in flight form New York City. The city is central to our analysis both due to the size of its urban flight, as well the early presence of COVID-19 in the urban area. We first begin by separating our analysis into different urban categories based on attributes of the destination regions. Regions differ in their exposure to infectious disease spread based on urban form, so we

analyze separately the impact of inflows of New York City residents contrasting the large and medium sized metros (NCHS category 1, 2 & 3 Panel A), compared with micropolitan and non-core areas (NCHS categories 4, 5 & 6, Panel B). Within each category, we compare case growth 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 New York on COVID-19 cases across regions. In the largest urban areas, we find that case growth starts to increase for counties that receive high inflow from New York City beginning in March. We plot the raw seven day average in case growth for the counties receiving highest and lowest quartile of New York City inbound residents, and plot the difference between these regions in grey bars 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 New York City residents with a lag, consistent with a channel of direct infection.

The difference between areas with high inflow of New York City residents and areas that see low inflow starts to decline in July. We show similar plots that plot case growth in A5, which confirm that areas that initially saw higher New York City influx see negative relative case growth into the summer. This result suggests that New York City inflow brought forward some cases which may have counterfactually been experienced later in the course of the pandemic. Urban flight would still be a quite important, even it only accelerates case growth, because of steady improvements in treatment and expanded supply of personal protective equipment such as masks over the pandemic, which lowered mortality rates among those infected later in the course of the pandemic.[8]

[8]See Horwitz, Jones, Cerfolio, Francois, Greco, Rudy, and Petrilli (2020) and Ciceri, Ruggeri, Lembo, Puglisi, Landoni, Zangrillo, and on behalf of the COVID-BioB Study Group (2020) on improved mortality and Gandhi and Rutherford (2020) who connects 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.

We observe that the impacts of New York City influx are increasing in city size. Urban areas which saw higher influx of New York City residents saw the greatest increase in new cases (a representative example would be Atlanta, which saw large case growth in the first wave of the pandemic). Large fringe and medium metropolitan areas also see substantial increase 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.

Overall, we find evidence across the cross-section and time-series consistent with urban migration being a substantial vector of spread for COVID-19 across the country. Areas with greater influx of New York City residents see greater case growth exactly in the time period when it would be most expected. Our effects are largest for urban areas that receive New York City residents, and partially reverse over the course of the pandemic—suggesting that this migration brought forward additional cases. We next move on to regression analysis, which allows us expand our focus to migration across the entire United States, control for additional factors, and account for endogeneity in the migration decision by using the social connectivity of regions as an instrumental variable.

Our core regression specification in I follows our primary specification in equation 2 and covers inflows and case growth from March 1 to July 13. We first show that inflows lead to large increases in county caseloads, in our OLS specifications (columns 1–6). We find that for every additional 100 people who enter a county, case growth is increased by 0.8.[9] We also see that counties which experience inflows originating from counties which fall within the top quartile of case growth in a given month also see higher case growth on average. This large and statistically significant relationship parallels our graphical results

[9]Coefficients in this table other than 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.

in highlighting a role for case influx on cases. We also find evidence that migration from areas with higher case loads, and influx from areas further away lead to higher infection rates. These results are consistent with long-distance and inter-state migration, especially from New York City, contributing to greater spread around the country.

We then instrument for migration patterns with Facebook friendship linkages (columns 6–10) to confirm a causal relationship between migration and subsequent case growth. Our IV estimates suggest statistically significant and economically substantial estimates. A one standard deviation increase in inflows instrumented for by social connections (roughly 116 people) causes a 0.5 standard deviation in daily case growth (roughly 23 cases). To put this into context, an increase in instrumented inflows by one standard deviation is akin to the difference in inflows experienced by Essex County in upstate New York (home to Lake Placid) as compared to Mercer County, NJ (home to the state capital). In this particular example, we also see that our back of the envelope calculations for mobility-linked case-growth is significant, as Mercer County's average case growth was 58 per day over our sample period, as compared to Essex County's daily case growth of 0.39. Our IV results suggest that migration accounts for slightly less than half of the difference in case growth across these two counties.

Our IV estimates are larger than our OLS estimates since raw inflow tends to overweight areas which exhibit lower case growth 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 case growth is 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 do not know many others). Our IV, by contrast, identifies a LATE based on migration towards 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.

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 Figure A3. We show that instrumented inflow tends to be higher in regions with lower case growth—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 case growth, because inflow is positively correlated with a variable which drives lower case growth (a characteristic which naturally makes these regions more desirable destinations for those fleeing urban areas). Additionally, we contrast the raw inflows against the instrumented inflows from the SCI measure. 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 for the IV regressions.

We perform several key robustness tests on our primary sample. In Appendix Table A2, we include additional lags to account for past inflows. These additional lagged controls account for the incubation period between contracting the virus, exhibiting symptoms, and receiving a positive test result. These results do not suggest meaningfully different impacts from lagged inflows as compared to concurrent inflows.

We also examine specifications that restrict on flight from New York City specifically. The large exodus from New York City, and the substantial case load in the city overall, make it a key focus of our analysis. In A3, we regress local cases against an indicator for counties in the top quartile of counties 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 just inflows from New York City specifically. Both specifications find large and statistically significant effects. These results suggest that urban flight, and specifically the large urban exodus from New York City, was a central feature of the spread of

the COVID-19 pandemic across the United States. 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.

Overlal, our findings also potentially help explain the result in Kuchler, Russel, and Stroebel (2020), 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 New York City residents into these socially connected regions, which has been frequently cited by local officials as a possible transmission mechanism of the disease.[10]

IV Conclusion

We document substantial urban flight in the wake of the COVID-19 pandemic and find large effects of this migration on the growth in COVIV-19 cases elsewhere in the country. We observe large migration responses by individuals living in major urban areas. In New York City, for instance, as much as 15-20% of Manhattan 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 New York City, suggesting that individuals are taking shelter with friends and family.

We then use the social networks structure to develop a causal estimate of the impact of migration on case growth. We find that instrumented migration patterns predict subsequent rise in cases in destination counties, suggesting that urban flight helped to seed the pandemic from an initially urban disease to a more widespread nation-wide pandemic.

[10]For instance, the Governor of Rhode Island mandated a 14-day quarantine for New York City residents entering the state, punishable with a fine or arrest, see https://www.nytimes.com/2020/03/28/us/coronavirus-rhode-island-checkpoint.html

Our work has implications for public policy in the wake of the disease. First, it highlights an important phenomenon 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 banks, quarantine periods, or requiring testing as a precondition for entry—to help curb the spread of COVID-19.

REFERENCES

Allcott, H., L. Boxell, J. C. Conway, M. Gentzkow, M. Thaler, and D. Y. Yang, 2020, "Polarization and Public Health: Partisan Differences in Social Distancing during the Coronavirus Pandemic," Working Paper 26946, National Bureau of Economic Research.

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

Athey, S., B. Ferguson, M. Gentzkow, and T. Schmidt, 2019, "Experienced Segregation," Working paper.

Bailey, M., R. Cao, T. Kuchler, J. Stroebel, and A. Wong, 2018, "Social connectedness: Measurement, determinants, and effects," *Journal of Economic Perspectives*, 32(3), 259–80.

Barrios, J. M., and Y. V. Hochberg, 2020, "Risk Perception Through the Lens of Politics in the Time of the COVID-19 Pandemic," *Working Paper*.

Chandrasekhar, A. G., P. S. Goldsmith-Pinkham, M. O. Jackson, and S. Thau, 2020, "Interacting Regional Policies in Containing a Disease," .

Chen, M. K., J. A. Chevalier, and E. F. Long, 2020, "Nursing Home Staff Networks and COVID-19," Working Paper 27608, National Bureau of Economic Research.

Chen, M. K., K. Haggag, D. G. Pope, and R. Rohla, 2019, "Racial Disparities in Voting Wait Times: Evidence from Smartphone Data," Working Paper 26487, National Bureau of Economic Research.

Chen, M. K., and R. Rohla, 2018, "The effect of partisanship and political advertising on close family ties," *Science*, 360(6392), 1020–1024.

Chiou, L., and C. Tucker, 2020, "Social Distancing, Internet Access and Inequality," Working Paper 26982, National Bureau of Economic Research.

Ciceri, F., A. Ruggeri, R. Lembo, R. Puglisi, G. Landoni, A. Zangrillo, and on behalf of the COVID-BioB Study Group, 2020, "Decreased in-hospital mortality in patients with COVID-19 pneumonia," *Pathogens and Global Health*, 114(6), 281–282, PMID: 32584660.

Engle, S., J. Stromme, and A. Zhou, 2020, "Staying at Home: Mobility Effects of COVID-19," *Working Paper*.

Fauver, J. R., M. E. Petrone, E. B. Hodcroft, K. Shioda, H. Y. Ehrlich, A. G. Watts, C. B. Vogels, A. F. Brito, T. Alpert, A. Muyombwe, et al., 2020, "Coast-to-coast spread of SARS-CoV-2 during the early epidemic in the United States," *Cell*.

Gandhi, M., and G. W. Rutherford, 2020, "Facial Masking for Covid-19 — Potential for "Variolation" as We Await a Vaccine," *New England Journal of Medicine*.

Glaeser, E. L., C. S. Gorback, and S. J. Redding, 2020, "How Much does COVID-19 Increase with Mobility? Evidence from New York and Four Other U.S. Cities," Working Paper 27519, National Bureau of Economic Research.

Goolsbee, A., and C. Syverson, 2020, "Fear, Lockdown, and Diversion: Comparing Drivers of Pandemic Economic Decline 2020," Working Paper 27432, National Bureau of Economic Research.

Holtz, D., M. Zhao, S. G. Benzell, C. Y. Cao, M. A. Rahimian, J. Yang, J. Allen, A. Collis, A. Moehring, T. Sowrirajan, D. Ghosh, Y. Zhang, P. S. Dhillon, C. Nicolaides, D. Eckles, and S. Aral, 2020, "Interdependence and the cost of uncoordinated responses to COVID-19," *Proceedings of the National Academy of Sciences*.

Horwitz, L., S. A. Jones, R. J. Cerfolio, F. Francois, J. Greco, B. Rudy, and C. M. Petrilli, 2020, "Trends in Covid-19 risk-adjusted mortality rates in a single health system," *medRxiv*.

Kuchler, T., D. Russel, and J. Stroebel, 2020, "The Geographic Spread of COVID-19 Correlates with Structure of Social Networks as Measured by Facebook," Working Paper 26990, National Bureau of Economic Research.

Lee, J., M. Mahmud, J. Morduch, S. Ravindran, and A. Shonchoy, 2020, "Migration, Externalities, and the Diffusion of COVID-19 in South Asia," .

Painter, M. O., and T. Qiu, 2020, "Political Beliefs affect Compliance with COVID-19 Social Distancing Orders," *Working Paper*.

Pybus, O., A. Rambaut, COG-UK-Consortium, et al., 2020, "Preliminary analysis of SARS-CoV-2 importation and establishment of UK transmission lineages," *Preprint at https://virological. org/t/preliminary-analysis-of-sars-cov-2-importation-establishment-of-uktransmission-lineages/507*.

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

V TABLES AND FIGURES

FIGURE I: Propensity to Leave Cities

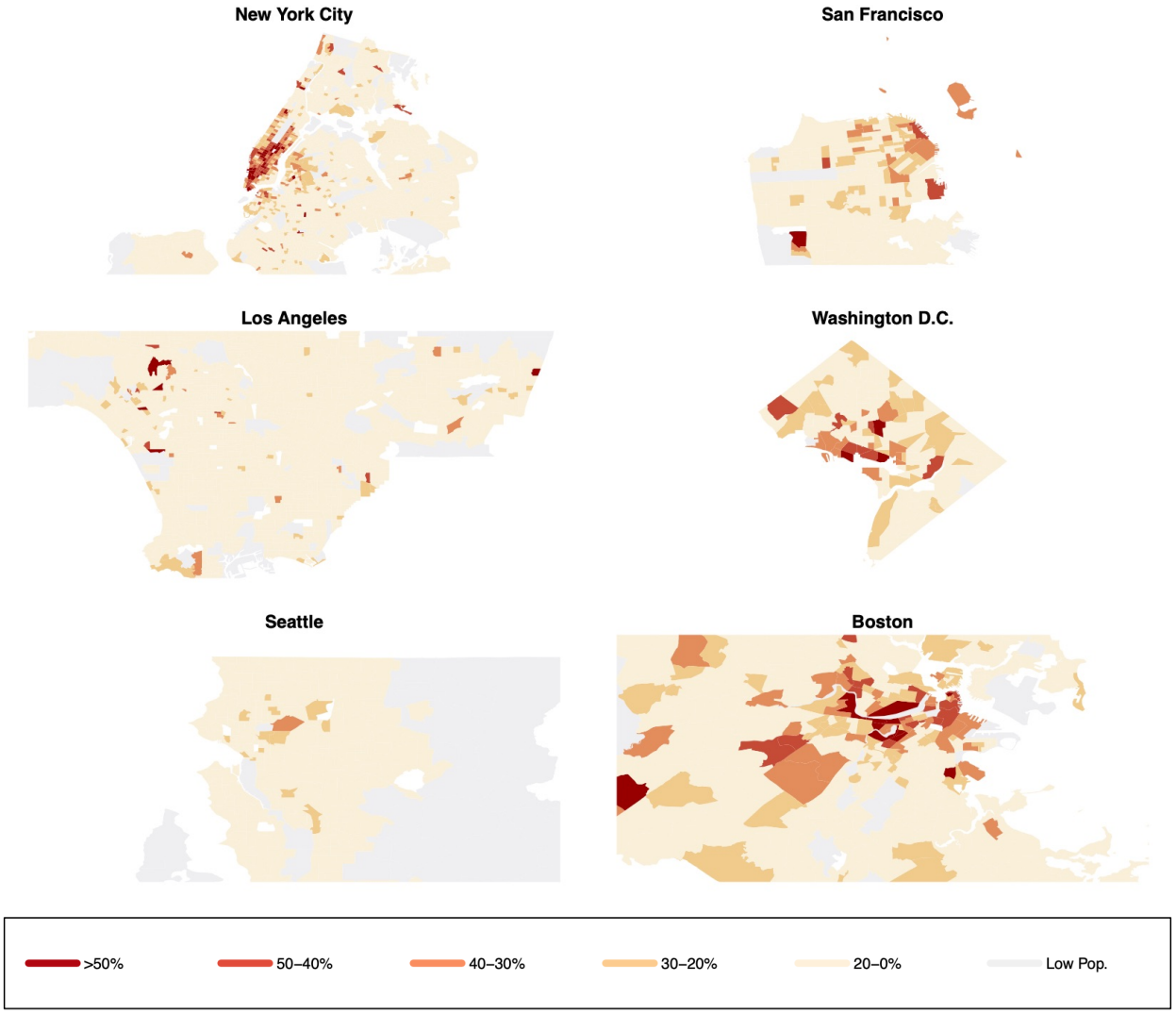
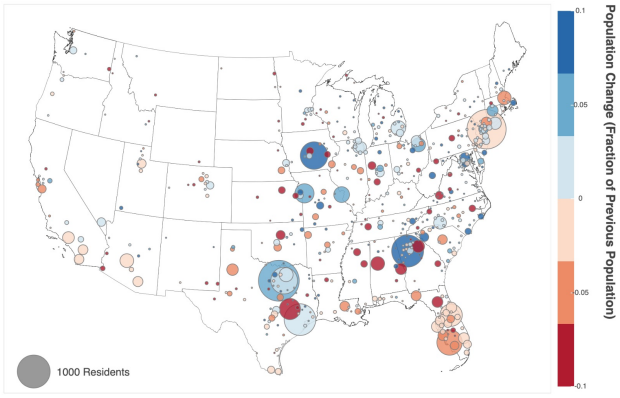
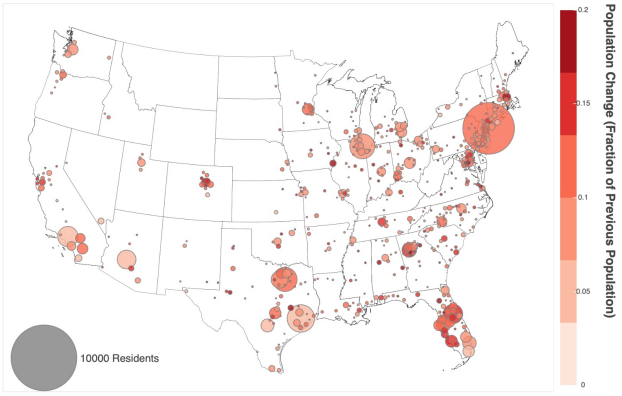


FIGURE II: Nationwide Migration

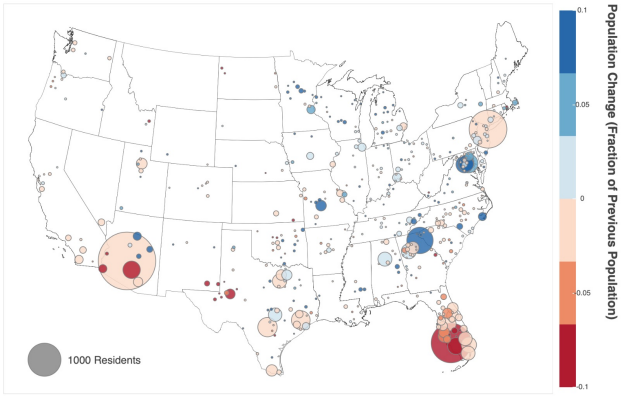
Panel A: March Net Flow



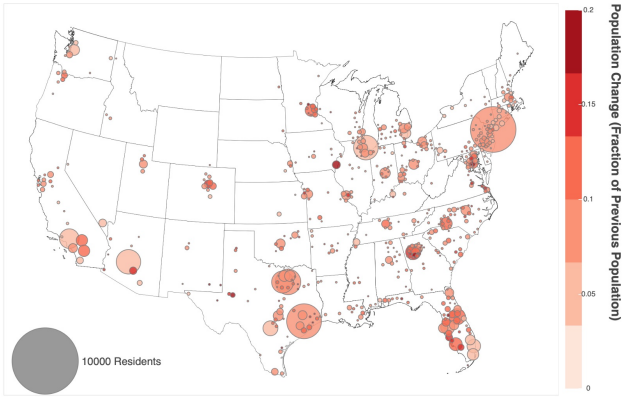
Panel B: March Outflow



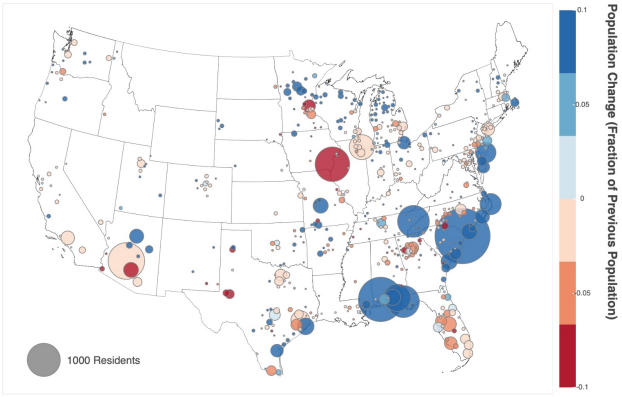
Panel C: April Net Flow



Panel D: April Outflow



Panel E: May Net Flow



Panel F: May Outflow

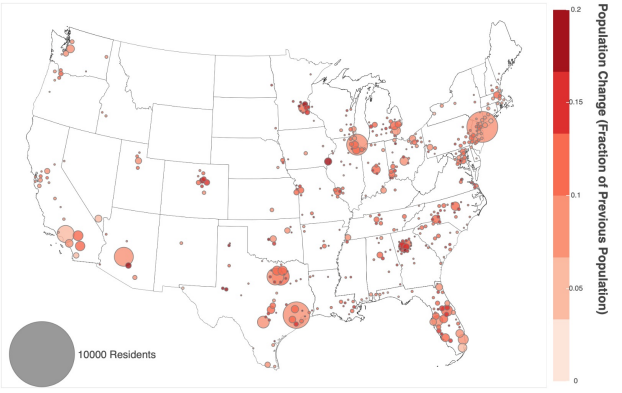
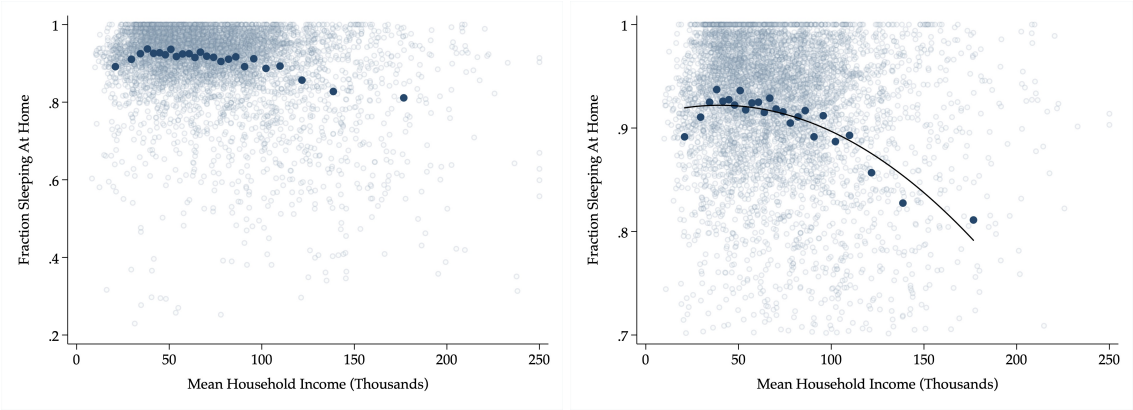
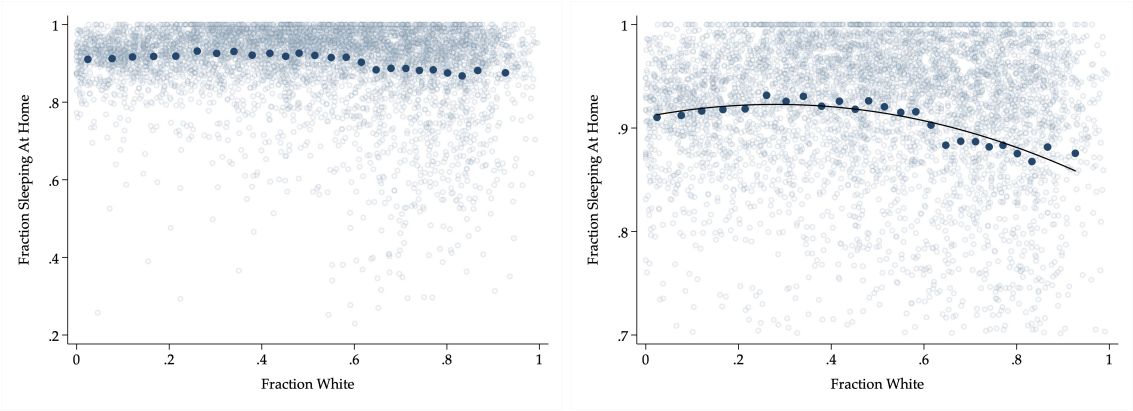


FIGURE III: Fraction Change in Out of Town Visits by Demographics

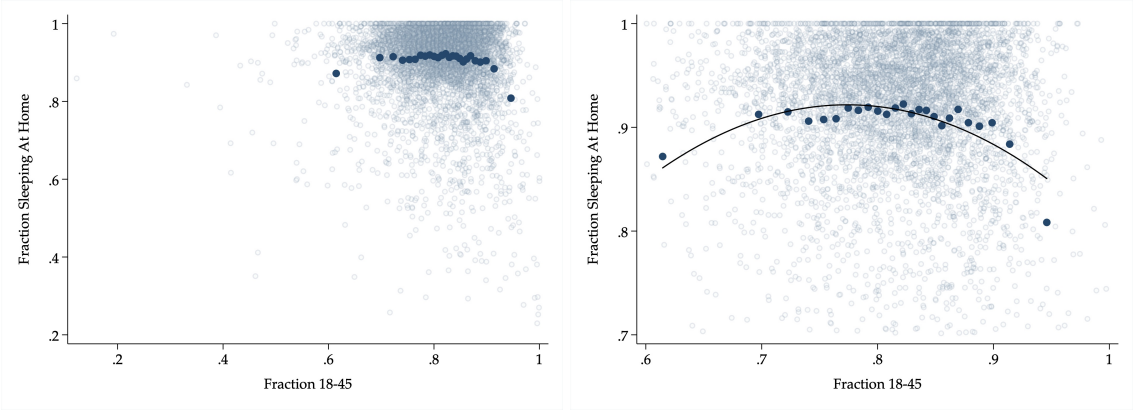
Panel A: Change in Out of Town Visits by ZIP Income



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



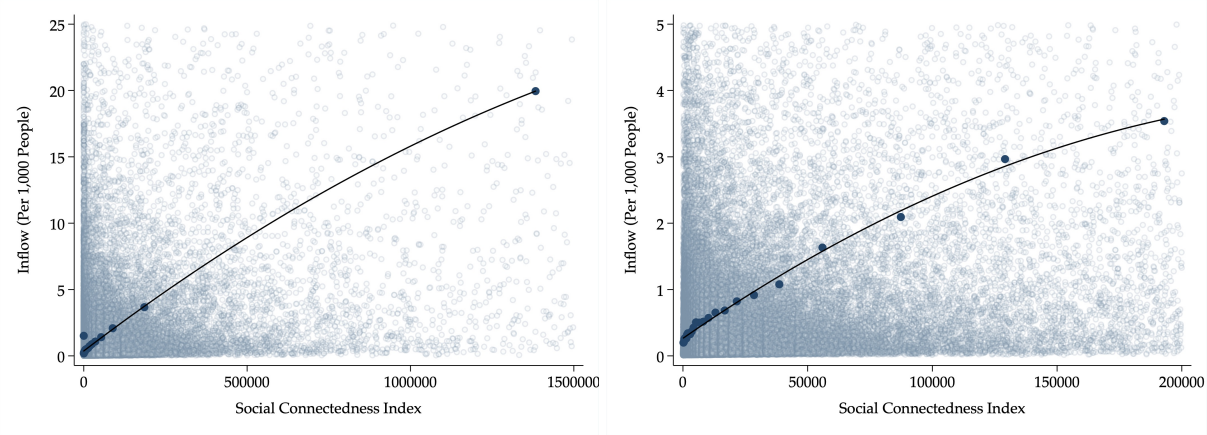
Panel C: Change in Out of Town Visits by Fraction Aged 18-45



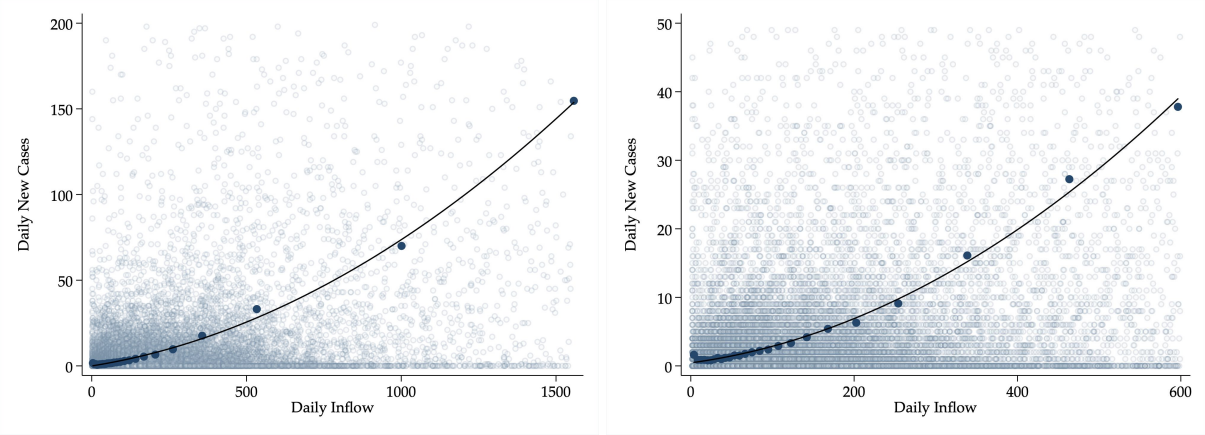
Notes: Panel A shows average household income against the change in the fraction of out of town visits. Panel B repeats the exercise for the fraction who are white, and Panel C for the fraction between the ages of 18-45. Data is tract-level and shown only for New York City, San Francisco, Los Angeles, Washington DC, Seattle, and Boston. Charts on the left represent the entire sample; charts on the right winsorize the fraction of out of town visits at the 5% level. All demographic variables are left un-winsorized, with the exception of the fraction between 18-45 in Panel C. Light gray points show all the data points; 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 is 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.

FIGURE IV: SCI-Related Mobility and Case Growth

Panel A: ZIP Inflows against SCI



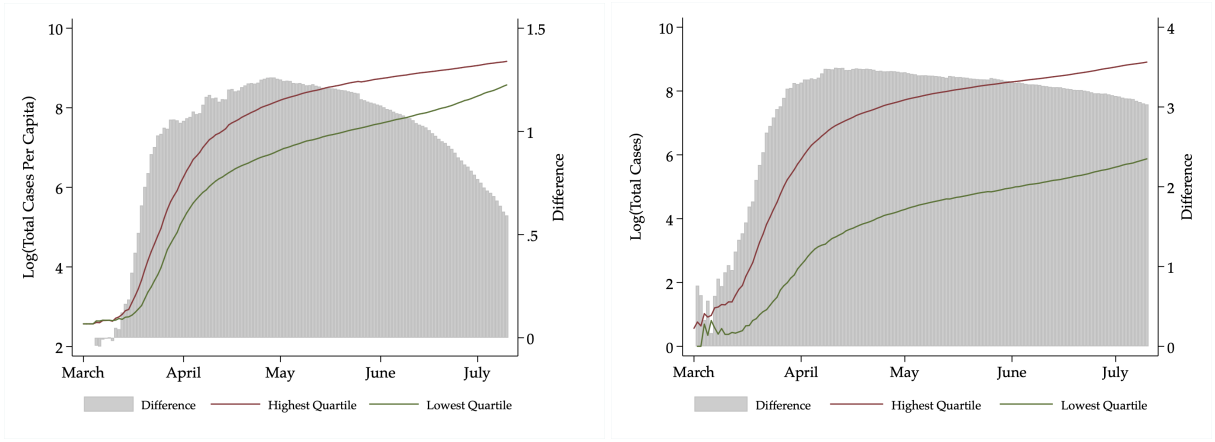
Panel B: Inflows against Case Growth



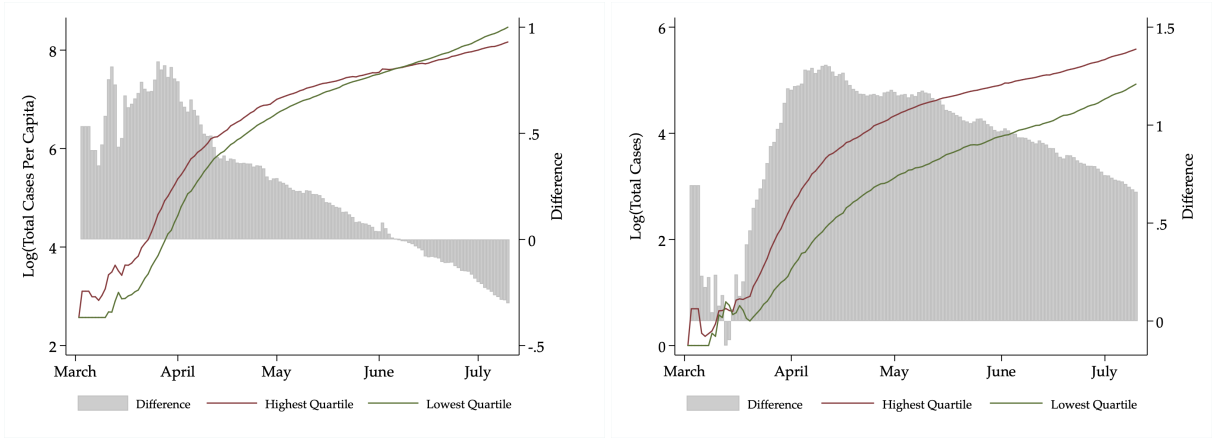
Notes: Panel A show inflows per 1000 people between pairs of zip codes against SCI between the two zip codes. Light gray points show 1/100th of the entire sample. Panel B shows daily inflows against new cases at the county level. Light gray points show one-tenth of the entire sample. In the left column, SCI and daily inflow are winsorized at the 1% and 99% levels; in the right column, SCI and daily inflow are winsorized at the 5% and 95% levels. Dark blue dots are obtained by dividing the data into 25 quantiles and calculating the mean x-axis and y-axis value in each quantile.

FIGURE V: March Inflows From NYC vs. Case Growth

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



Panel B: Micropolitan and Non-Core Areas



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).

TABLE I: Impact of Migration on COVID-19 Cases

Panel A: COVID-19 Case Growth Against Inflow

	OLS						IV					
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Inflow	92.585***	12.022***	8.288***				267.821***	197.785***	194.605***			
	(2.745)	(2.556)	(2.644)				(9.468)	(12.621)	(12.551)			
Per Capita Inflow				-771.807***	-151.848*	-199.698**				-2969.855***	261.882***	197.772***
				(82.935)	(90.625)	(94.784)				(931.814)	(31.698)	(20.711)
High Incoming Cases	4779.249***	4563.025***	4595.383***				5444.178***	6661.104***	7040.705***			
	(113.903)	(134.450)	(134.190)				(187.469)	(235.459)	(264.511)			
High Incoming Cases Per Capita				48.924***	49.372***	49.372***				-0.090*	0.063***	0.060***
				(0.438)	(0.443)	(0.443)				(0.050)	(0.002)	(0.002)
Far Indicator	3622.632***	2333.182***	-1362.127	0.427	-0.119	-1.102	3214.693***	1842.587**	-18806.066***	0.317***	0.008	-0.074***
	(471.740)	(691.956)	(948.833)	(0.541)	(1.114)	(1.279)	(491.789)	(792.740)	(2961.841)	(0.111)	(0.006)	(0.012)
Far Indicator × Inflow			20.448***						113.999***			
			(4.690)						(16.657)			
Far Indicator × Per Capita Inflow						369.540						30.254***
						(247.614)						(4.678)
Controls	N	Y	Y	N	Y	Y	N	Y	Y	N	Y	Y
State x Month x Population Decile FE	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
N	414,847	399,083	399,083	414,847	399,083	399,083	414,847	399,083	399,083	414,847	399,083	399,083

Panel B: First Stage Estimates of Inflow on SCI

Weighted SCI							42947.436***	29688.738***	32749.326***	-0.002***	0.006***	0.003***
							(443.427)	(456.984)	(526.249)	(0.001)	(0.001)	(0.001)
F							9,381	2,651	1,346	10	106	115

Columns 1-3 shows our main regression specification 1. Columns 4-6 repeat the exercise using inflow per capita as the outcome variable. Column 7-12 repeats the exercise for columns 1-6, where inflow is instrumented with the weighted SCI, as in 2. The sample period is March 1 through July 13. 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.

INTERNET APPENDIX

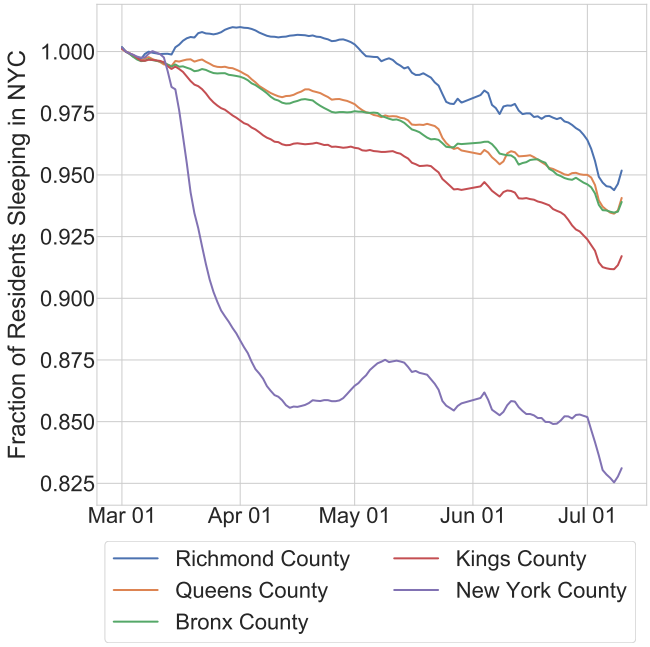
TABLE A1: Summary Statistics

	1st Quartile	Median	Mean	3rd Quartile
Daily New Cases	0.00	0.00	8.19	2.00
Daily New Cases per 1000 people	0.00	0.00	0.04	0.04
Net Inflow	31.00	67.00	152.73	153.00
SCI-Instrumented Inflow	107.18	123.70	152.67	153.05
Social Connectedness Index	31.68	63.07	111.19	121.10
Social Connectedness Index (inflow-weighted)	62.64	171.03	757.78	515.25
Distance	311.35	453.03	520.15	635.14
Distance (inflow-weighted)	182.20	250.02	301.96	349.29
Number of nursing homes	2.00	5.00	7.67	8.00

Summary statistics across all counties in our sample, with caseload and inflow data between 3/1/2020 and 7/13/2020. SCI and distance are calculated as the raw arithmetic average between the destination county and other counties from which there is a positive inflow. The inflow-weighted counterpart of these measures are calculated by using inflows from a particular county over total inflows as the weight in the weighted average.

FIGURE A1: Propensity to Leave NYC by Tract

Panel A: Propensity to Leave NYC by Borough



Panel B: Propensity to Leave NYC by ZIP Income, $100k > tracts

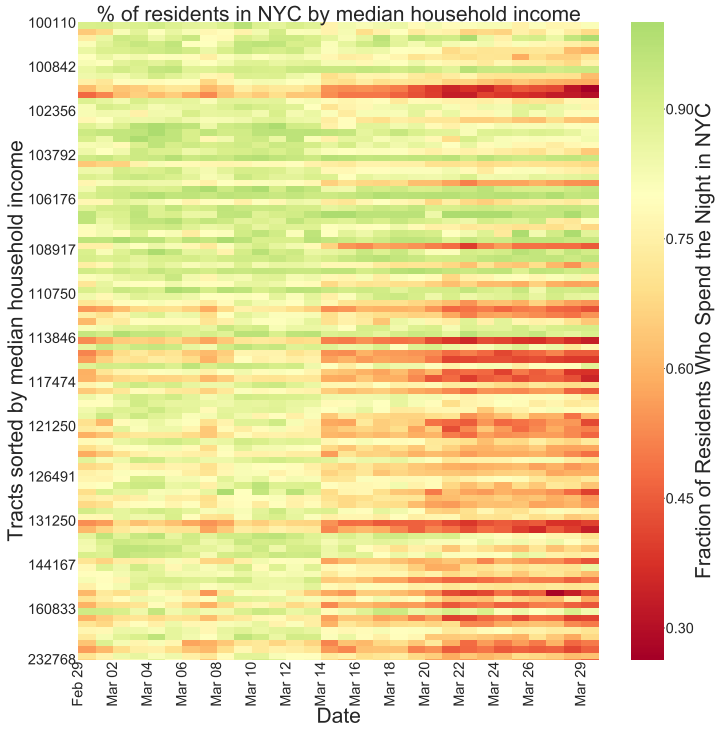
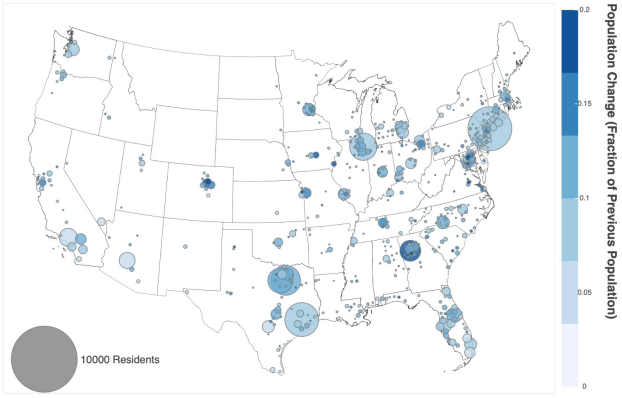
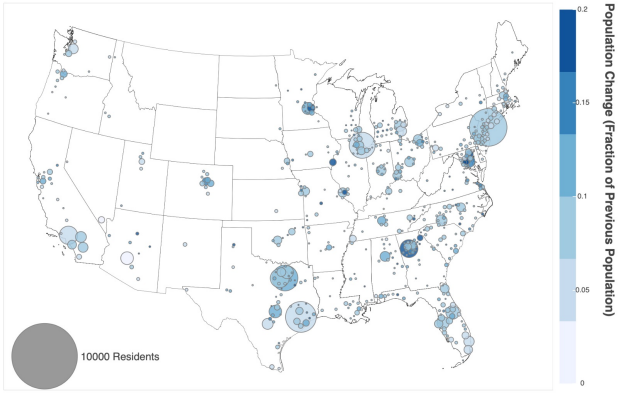


FIGURE A2: Propensity to Leave Cities: Inflow

Panel A: March



Panel B: April



Panel C: May

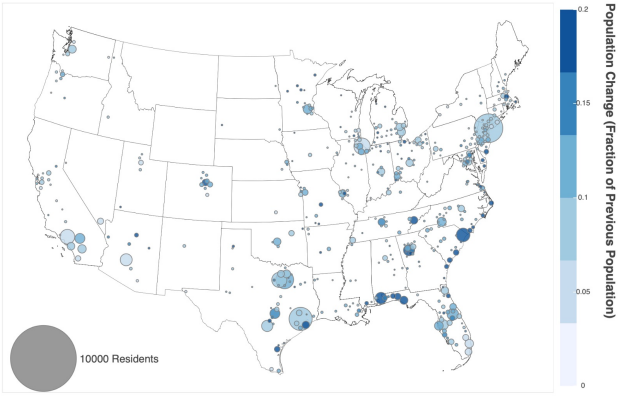
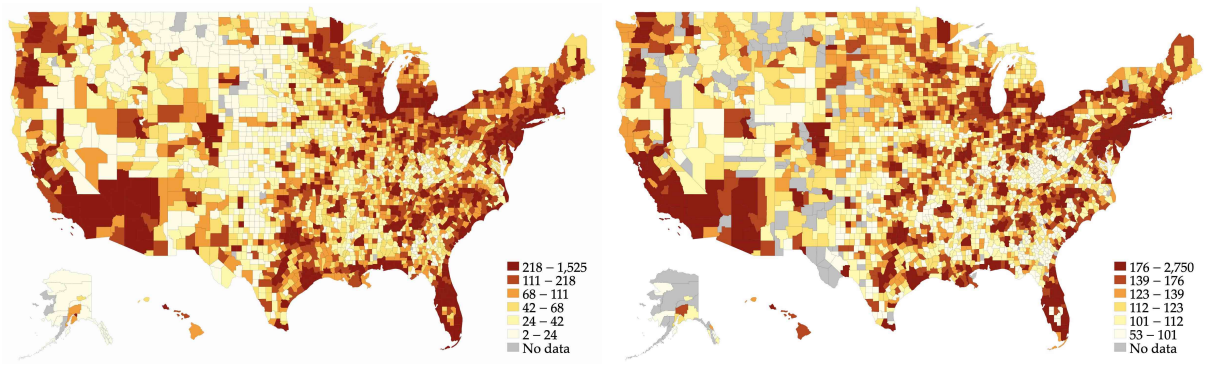
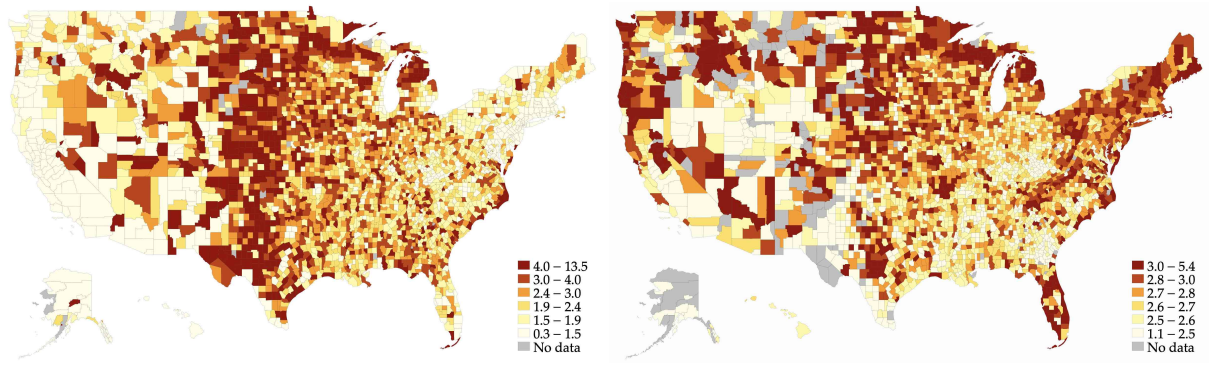


FIGURE A3: Raw vs. SCI-Fitted Inflows

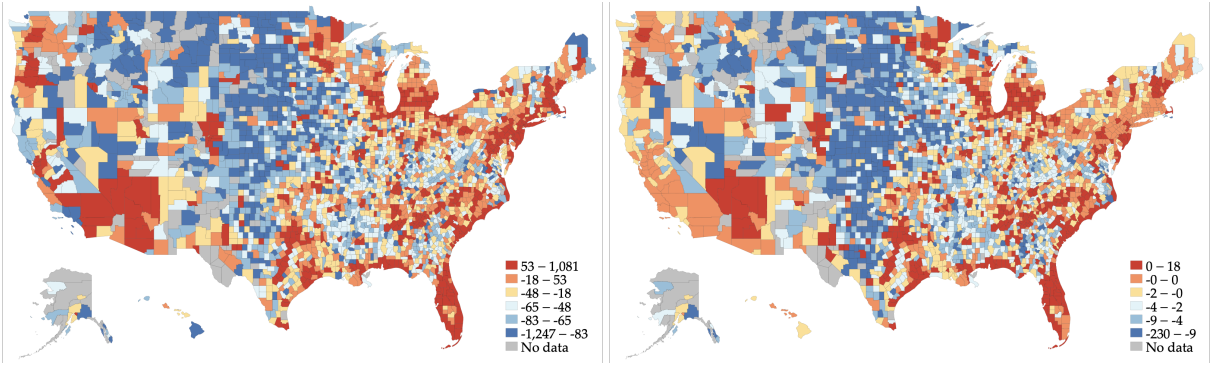
Panel A: Total Inflow



Panel B: Inflow Per Capita

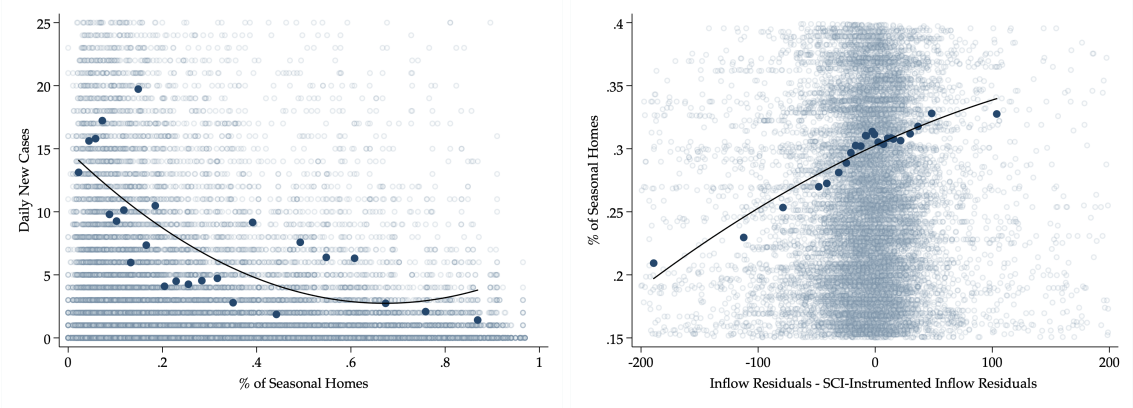


Panel C: Difference Between Raw and SCI-Instrumented Inflows



Notes: Panels A & B show raw inflows and raw per capita inflows (left column) and SCI-fitted inflows and inflows per capita (right column). Darker red represent higher values; each map is colored by splitting the data into sextiles, resulting in different cut-off values for different colors vary across maps. Panel C shows the difference between raw and SCI-instrumented inflows (inflow − SCI-instrumented inflows); red counties represents regions where raw inflow is higher than instrumented inflow, and blue counties represent regions where raw inflow is lower than instrumented inflow.

FIGURE A4: Differences in IV and OLS Estimates Due to Differences in Instrumented and Raw Inflow



Notes: The left binscatter shows case growth tends to be lower in regions with a greater proportion of seasonal homes. The right binscatter shows the unexplained portion of raw inflows (after partialling out the impact of the control variables we use in our baseline specification) tends to be higher than the unexplained portion of instrumented inflows, precisely in regions with more seasonal homes, and hence, regions with lower case growth. We point to seasonal homes as an imperfect measure of a possible omitted variable, which is positively correlated with the raw (not instrumented) inflow and negatively correlated with case growth.

TABLE A2: Impact of Migration on COVID-19 Cases: Lagged Inflows

Panel A: COVID-19 Case Growth Against Lagged Inflows

	OLS				IV			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Inflow (1 Week Lag)	23.299***				193.162***			
	(2.721)				(12.468)			
Inflow (1 Month Lag)		27.190***				197.702***		
		(3.357)				(13.490)		
Per Capita Inflow (1 Week lag)			722.257***				175.463***	
			(101.890)				(16.038)	
Per Capita Inflow (1 Month Lag)				521.679***				201.144***
				(122.954)				(20.681)
High Incoming Cases (1 Week Lag)	3747.654***				4436.659***			
	(139.241)				(196.018)			
High Incoming Cases (1 Month Lag)		2224.591***				4113.848***		
		(175.153)				(301.818)		
High Incoming Cases Per Capita (1 Week Lag)			32.141***				0.026***	
			(0.483)				(0.001)	
High Incoming Cases Per Capita (1 Month Lag)				21.342***				0.028***
				(0.517)				(0.002)
Far Indicator (1 Week Lag)	-2808.387***		-2.731***		-1.41e+04***		-0.045***	
	(589.952)		(0.918)		(2397.465)		(0.009)	
Far Indicator (1 Month Lag)		-1089.620		0.089		-1.48e+04***		-0.060***
		(662.542)		(1.016)		(2785.041)		(0.011)
Far Indicator × Inflow (1 Week Lag)	27.131***				101.291***			
	(5.066)				(16.842)			
Far Indicator × Inflow (1 Month Lag)		23.243***				110.110***		
		(5.981)				(19.830)		
Far Indicator × Per Capita Inflow (1 Month Lag)			83.988				25.628***	
			(264.241)				(3.840)	
Far Indicator × Per Capita Inflow (1 Month Lag)				12.958				33.710***
				(313.248)				(5.178)
Controls	Y	Y	Y	Y	Y	Y	Y	Y
State × Month × Population Decile FE	Y	Y	Y	Y	Y	Y	Y	Y
N	377,556	335,587	377,556	335,587	377,556	335,587	377,556	335,587

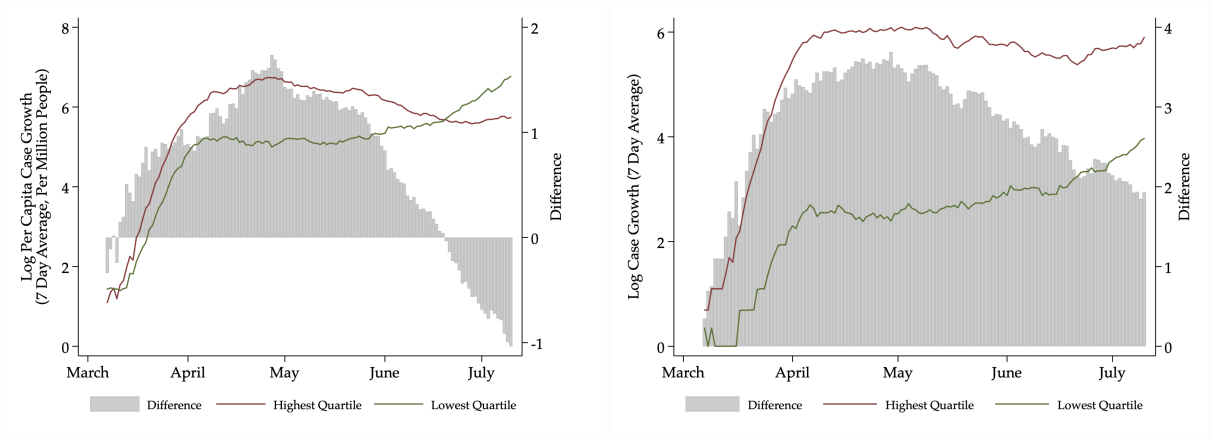
Panel B: First Stage Estimates of Lagged Inflows on SCI

Weighted SCI					33012.669***	32419.067***	0.006***	0.004***
					(516.184)	(534.261)	(0.001)	(0.001)
F					1,442	1,297	133	117

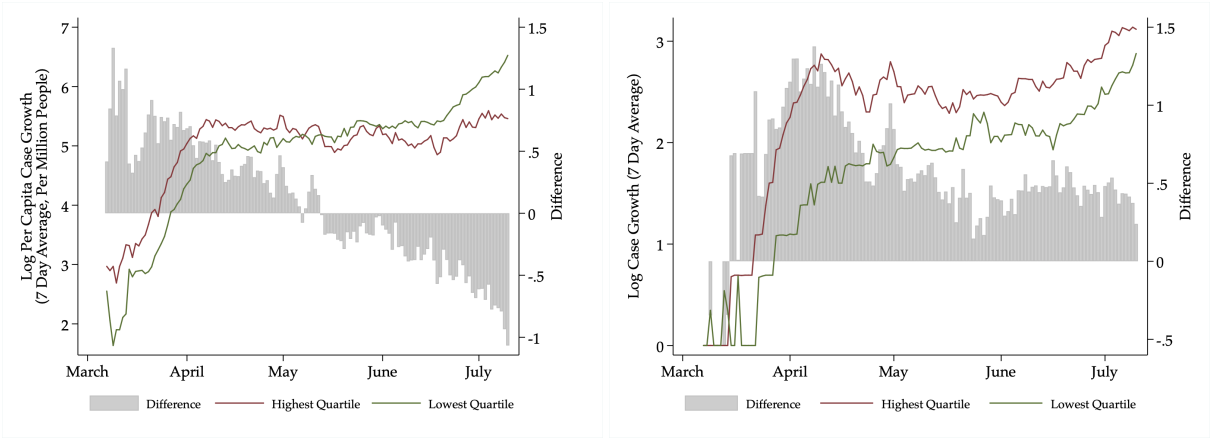
This table replicates table I, using 1-week and 1-month lagged explanatory variables. Columns 1-2 shows the results for total inflow and case growth; columns 3-4 shows results for per capita inflow and case growth. Columns 5-9 is the IV analogue of columns 1-4, where inflow and inflows per capita are instrumented with weighted SCI values, as in 2. The sample period is March 1 through July 13. 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.

FIGURE A5: March Inflows From NYC vs. Log Case Growth

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



Panel B: Micropolitan and Non-Core Areas



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 median log(total cases per capita growth) over time for counties in the fourth quartile and the first quartile of inflows. The two charts on the right show log(total case growth). 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). This figure differs from V by considering case growth, rather than totals.

TABLE A3: March Inflows from NYC vs. Case Growth

	Panel A: Log(New Cases) Against Inflow from NYC			
	OLS		IV	
	(1)	(2)	(3)	(4)
High NYC Inflow	0.800***	0.299***	5.901***	5.262***
	(0.016)	(0.011)	(0.213)	(0.315)
Controls	Y	Y	Y	Y
State × Month × Population Decile FE	Y	Y	Y	Y
N	64,287	63,228	64,287	63,228
	Panel B: First Stage Estimates			
Weighted SCI			0.014	0.142
			(0.0004)	(0.0077)
F			1,364	334

This table shows our baseline regression, where the outcome is the log(New Cases) and the inflow variable is replaced with an indicator equal to 1 if a county was in the top quartile of all counties which received inflows from NYC in March. The sample period is March 1 through July 13. Standard errors are in parentheses, and * denotes 10% significance, ** denotes 5% significance, *** denotes 1% significance.

TABLE A4: Total Inflows from NYC vs. Case Growth

Panel A: COVID-19 Case Growth Against Inflow from New York City

	OLS						IV					
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Inflow from NYC	201.893***	145.378***	146.173***				1701.690***	576.055***	525.346***			
	(29.120)	(22.988)	(23.024)				(90.574)	(40.851)	(38.314)			
Per Capita Inflow From NYC				222385.775***	221671.430***	267572.009***				7750.832***	3010.973***	2728.955***
				(12297.432)	(12610.849)	(13777.545)				(306.637)	(195.477)	(190.074)
High Incoming Cases	4119.744***	4283.269***	4271.320***				1830.249***	3547.601***	3445.972***			
	(129.971)	(137.438)	(138.531)				(369.907)	(187.022)	(197.225)			
High Incoming Cases Per Capita				48.556***	49.001***	48.990***				0.035***	0.044***	0.043***
				(0.438)	(0.443)	(0.443)				(0.001)	(0.001)	(0.001)
Far Indicator	4240.410***	2509.743***	2012.740***	-0.101	0.009	1.325	7228.572***	2624.269***	-6226.212***	-0.015***	-0.000	-0.026***
	(491.841)	(684.991)	(708.570)	(0.541)	(1.114)	(1.121)	(610.141)	(680.683)	(1437.595)	(0.001)	(0.001)	(0.005)
Far Indicator × Inflow from NYC			498.949*						8867.827***			
			(288.965)						(1296.926)			
Far Indicator × Per Capita Inflow From NYC						-305462.360***						5932.824***
						(30595.187)						(1112.899)
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	414,847	399,083	399,083	414,847	399,083	399,083	414,847	399,083	399,083	414,847	399,083	399,083

Panel B: First Stage Estimates of Inflow on SCI

Weighted SCI							6759.300***	8362.031***	9188.484***	0.001***	0.001***	0.001***
							(268.328)	(363.540)	(392.011)	(0.00002)	(0.00002)	(0.00003)
F							635	530	289	1,264	930	467

Columns 1-3 shows our main regression specification 1, with inflows from the five NYC boroughs as the explanatory variable rather than total inflows. Columns 4-5 repeat the exercise using inflow per capita as the outcome variable. Column 7-12 repeats the exercise for columns 1-6, where inflow from New York City is instrumented with the weighted SCI, as in 2. The sample period is March 1 through July 13. 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.

TABLE A5: Impact of Migration on COVID-19 Cases with Clustering at the Commuting Zone Level

Panel A: COVID-19 Case Growth Against Inflow

	OLS						IV					
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Inflow	92.585***	12.022*	8.288				267.821***	197.785**	194.605**			
	(17.086)	(6.678)	(8.082)				(36.036)	(80.534)	(85.408)			
Per Capita Inflow				-771.807***	-151.848	-199.698				-2969.855	261.882	197.772
				(207.943)	(224.998)	(227.826)				(11562.496)	(223.374)	(147.866)
High Incoming Cases	4779.249***	4563.025***	4595.383***				5444.178***	6661.104***	7040.705***			
	(757.515)	(831.002)	(830.135)				(894.825)	(1463.364)	(1628.642)			
High Incoming Cases Per Capita				48.924***	49.372***	49.372***				-0.090	0.063***	0.060***
				(1.637)	(1.649)	(1.649)				(0.543)	(0.013)	(0.009)
Far Indicator	3622.632	2333.182	-1362.127	0.427	-0.119	-1.102	3214.693	1842.587	-18806.066*	0.317	0.008	-0.074
	(3400.618)	(1724.929)	(2869.619)	(2.024)	(1.520)	(2.756)	(3336.820)	(1797.050)	(10336.236)	(1.271)	(0.013)	(0.060)
Far Indicator × Inflow			20.448						113.999**			
			(16.700)						(56.528)			
Far Indicator × Per Capita Inflow						369.540						30.254
						(856.118)						(24.184)
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	414,847	399,083	399,083	414,847	399,083	399,083	414,847	399,083	399,083	414,847	399,083	399,083

Panel B: First Stage Estimates of Inflow on SCI

Weighted SCI							42947.436***	29688.738***	32749.326***	-0.002	0.006	0.003
							(443.427)	(456.984)	(526.249)	(0.006)	(0.005)	(0.005)
F							112	85	74	0	2	3

Columns 1-3 shows our main regression specification 1. Columns 4-6 repeat the exercise using inflow per capita as the outcome variable. Column 7-12 repeats the exercise for columns 1-6, where inflow is instrumented with the weighted SCI, as in 2. The sample period is March 1 through July 13. 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. Estimates in this table have standard errors clustered at the Commuting Zone level.