Origin Country Information and Immigrant Behavior: Evidence from the COVID-19 Pandemic in the U.S.

Stephanie Kang *

November 11, 2022

[Please click here for the latest draft]

Abstract

I exploit the timing of COVID-19 outbreaks across immigrants' origin countries to study their behavioral responses to new developments in their origin countries. By conducting shift-share panel regressions, I find that an increase in the percentage of population infected with COVID-19 in the origin country leads to an increase in the average level of social distancing for the relevant immigrant group in the United States. Further, I perform an event study around the date national emergency was declared in the United States to study the interactive role played by the country of residence. I find that immigrants whose origin countries faced an outbreak before the U.S. increase their level of social distancing immediately after the declaration of national emergency in the U.S. That is, the information from the origin country translates into behavioral outcomes for immigrants when it becomes pertinent in their country of residence. Using Facebook connectedness index and Google search trends, I find that real-time transmission of information through the internet is a likely driving force of my findings.

*Department of Economics, Brown University. Email: stephanie_hamin_kang@brown.edu.

Acknowledgement: I am grateful to Pedro Dal Bó and Brian Knight for research guidance and discussions. I would also like to thank Andrew Foster, Peter Hull, Daniel Björkegren, Stelios Michalopoulos, Chien-Tzu Cheng, Steven Lee, and seminar participants at Brown University for helpful comments. All errors are my own.

1 Introduction

Immigrants' behavior in their country of residence is influenced by their pre-migration experiences and networks from their origin countries. The cultural economics literature documents that immigrants bring cultural norms from their home countries.[1] Parallel work in international economics shows immigrants utilize their networks or the relative informational advantage in their origin countries to establish more international flows such as trade or Foreign Direct Investment (FDI) between their destination country and origin countries.[2] Both literatures focus on the pre-migration qualities the immigrants had brought with them to the current country of residence at the time of immigration. However, little is known about whether new developments in the origin country continue to influence the behavior of immigrants and if the relevance of these developments in the country of residence plays an interactive role in determining the behavioral change.

Studying this has become more pertinent for two reasons. First, updating on new developments in the origin country has become substantially easier recently through improvement in communication technology. Social media platforms, online news outlets, and instant messaging are some of the new media through which immigrants can stay updated with new developments in their origin countries. Second, immigration is increasing across the world due to a variety of factors such as globalization and the movement towards free labor mobility. For instance, in the United States, the foreign-born population accounted for 13.7% of the entire population in 2019.[3] Therefore, understanding if and how the immigrants' ties to their origin countries influence their behaviors would provide a better insight into the behaviors of a substantial subset of the population.

In this paper, I study whether immigrants respond to developments around the COVID-

[1] See Blau (1992), Antecol (2000), Giuliano (2007), Fernandez and Fogli (2009), Alesina et al. (2013), Blau and Kahn (2015), Christopoulou and Lillard (2015), and Barsbai et al. (2017).

[2] See Gould (1994), Head and Reis (1998), Rauch and Trindade (2002), Peri and Requena-Silvente (2010), and Parsons and Vezina (2018) for literature on immigrants and trade; see Guiso et al. (2009), Leblang (2010), Cohen et al. (2015), and Burchardi et al. (2019) for literature on immigrants and FDI.

[3] American Community Survey, 2019

19 pandemic in their origin countries and adjust their social distancing behavior in the U.S. I exploit the quasi-random variation in the timing and evolution of the initial COVID-19 outbreak in the origin country to study this link. I look at how changes in the number of COVID-19 cases and deaths and the implementation of a Non-Pharmaceutical Intervention (NPI) in the origin country affect the social distancing behaviors of immigrants from that country in the U.S. Then, in order to check whether the circumstances in the country of residence play an interactive role in determining the immigrants' behavior, I examine the change in the level of social distancing among various immigrant groups around the date the U.S. declares a national emergency.

The combination of GPS data from Safegraph and the American Community Survey allows me to measure the social distancing level of each immigrant group. The Safegraph dataset consists of a large sample of GPS location data from smartphones. It contains measures of social distancing at the census block group level, which I aggregate to the census tract level to merge with the data from the American Community Survey. The 5-year estimates of the American Community Survey provide information on the fractions of residents in a given census tract born in various origin countries.

Using shift-share panel regressions, I find that immigrants exhibit a significant increase in their level of social distancing when in their origin countries: 1) the cumulative count of confirmed cases increase and 2) an NPI is implemented. The probability of staying completely home increases by 0.757 percentage points among first generation immigrants for every confirmed case per 100,000 people in the origin country. This effect is smaller among second-plus-generation immigrants whose probability of staying completely home increases by 0.206 percentage points. The first generation immigrants also strongly respond to an NPI implementation in their origin countries. The probability of the first generation immigrants staying completely home increases by 0.0536 percentage points when their origin country implements an NPI. The results are consistent throughout robustness checks.

To complement this, I conduct event study analyses to find out when the change in the

social distancing behaviors of immigrants is realized. Using March 13th, 2020, the date the U.S. declared national emergency, as the outbreak date, I test whether the situation in the country of residence plays an interactive role in determining the immigrants' behavior. The increase in social distancing for immigrant communities whose origin countries faced a severe outbreak takes place immediately after the declaration of national emergency in the U.S. That is, the increase in social distancing was especially pronounced when the coronavirus became more threatening for those living in the U.S.

I provide two pieces of evidence suggesting that the real-time transmission of information from the origin country is a possible mechanism behind the results. First, I document that the relative connectedness on Facebook between a U.S. county and an origin country increases with the fraction of immigrants in the county from the corresponding origin country. Second, I show that the relative frequency of COVID-19 related Google searches in the U.S. in the native language of each origin country increases soon after the outbreak in the origin country.

The papers closest to this study are Tian et al. (2020) and Apfel (2020). The former explores the link between residents in Mexico and how their connections to the United States could have influenced their social distancing behavior. I can study a much larger network of residents because I am studying the behavior of the 20 largest immigrant groups in the Untied States. This also allows me to exploit the different timing of outbreaks in the origin countries as opposed to only relying on the outbreak in the United States. The latter conducts a state-level analysis. I am able to show if the effect can be identified at a more granular level thanks to the dataset I have that allows me to conduct the analysis at the census tract level. Relatedly, Bailey et al. (2020) and Valsecchi and Durante (2021) study the importance of social networks and their role in influencing individuals' social distancing behaviors among other determinants of social distancing.[4]

The paper proceeds as follows. In Section 2, I introduce the setting of this study and describe the data sources. Sections 3 and 4 each describe the method and findings pertaining

[4]See Akesson et al. (2020), Allcott et al. (2020)., Bursztyn et al. (2020), Simonov et al. (2020), and Egorov et al. (2021) for papers on determinants of social distancing other than social networks.

to the shift-share panel regression and the event study design respectively. Section 5 explores the potential mechanism behind the link between the information from the origin country and immigrants' behavior is explored in. Finally, the paper ends with concluding remarks in Section 6.

2 Setting and Data

The COVID-19 pandemic offers a quasi-random experiment that allows for studying immigrants' response to new developments in their origin country. Each origin country has been subject to the same source of shock, the COVID-19 virus, but the timing of the initial outbreak was quasi-random.[5] This presents a setting in which I can study if immigrants exhibit a behavioral change following new developments in their origin country.

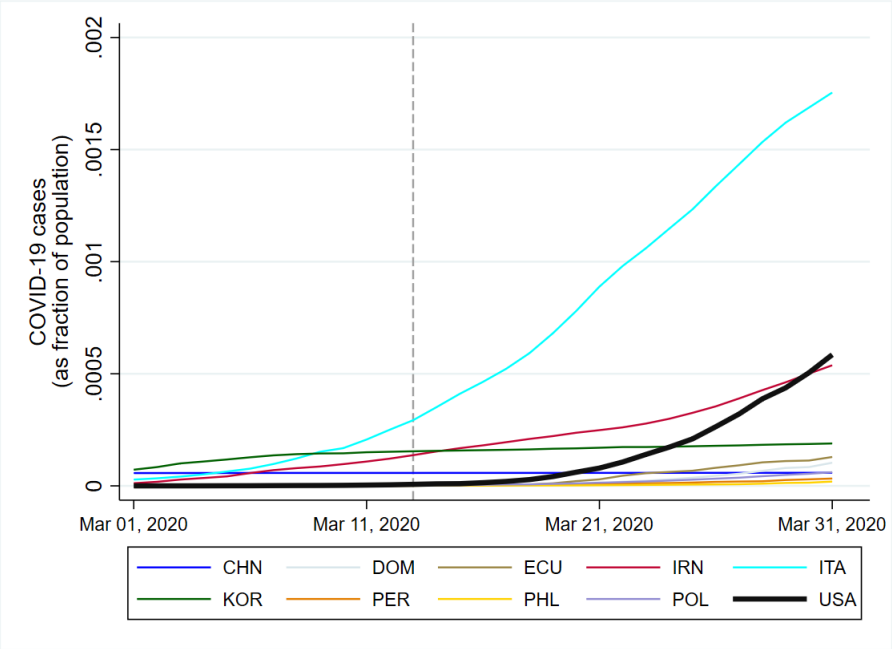
Figure 1 illustrates the variation in the timing of the first wave of outbreak in various countries. It shows the evolution of cumulative confirmed cases as the fraction of population in each country. Countries that differ in multiple dimensions such as China, Iran, Italy, and South Korea, were the first four most severely affected countries before the outbreak in the U.S. This bolsters the idea that the timing of the initial outbreak was quasi-random.

In the main analyses, the key data for analyzing the social distancing behavior of various immigrant groups are obtained by joining the social distancing data created by SafeGraph[6] and the American Community Survey data on the fraction of population born in a given foreign country at the census tract level. Then, for testing the mechanism, I use data from two different sources–the Facebook Connectedness Index and the Google Trends Index to

[5] In this paper, I only focus on the first wave of the coronavirus outbreak–numerous countries had multiple waves–because I am interested in whether having had the exposure to more information on COVID-19 through their origin country had induced a different social distancing behavior for the immigrants as opposed to the natives. Thus, it is crucial that the new developments in the origin country takes place before the outbreak in the U.S. for the design of this study. As the second wave of outbreak and onward happened after the first wave of outbreak in the U.S., I exclude them from my study.

[6] "from SafeGraph, a data company that aggregates anonymized location data from numerous applications in order to provide insights about physical places, via the SafeGraph Community. To enhance privacy, SafeGraph excludes census block group information if fewer than two devices visited an establishment in a month from a given census block group."

Figure 1: Number of Confirmed COVID-19 Cases as a Fraction of Population



The figure shows the evolution of the number of confirmed COVID-19 cases as a fraction of population for the 9 origin countries—China, Dominican Republic, Educador, Iran, Italy, South Korea, Peru, Philippines, and Poland—that are among the 20 origin countries of the studied immigrant groups and had the greatest number of COVID-19 cases on March 13th, 2020, as well as the evolution of that for the U.S. The dashed vertical line marks March 13th, 2020, which is the date when the national emergency was declared in the U.S. The numbers are calculated based on the cumulative confirmed cases from Dong et al. (2020) and the population data from the World Bank (2019).

see if real time transmission of information could be a potential mechanism.

2.1 Social Distancing and Immigrant Groups

I bring two datasets together to infer the level of social distancing for various immigrant groups and the native population. To study the level of social distancing, I use the SafeGraph mobile GPS location data for cellphone users in each census block group. To identify the immigrant clusters among these areas, I impute the percentage of users to be immigrants from a given foreign country by using the American Community Survey.

The SafeGraph data consist of a panel of GPS pings from anonymous mobile devices provided at the census block group level. The dataset contains social distancing metrics

calculated by SafeGraph based on how much time the cellphone users from the census block group are spending at their home. Home is defined as the common nighttime location of each mobile device over a 6 week period to a Geohash-7 granularity, which corresponds to an area of $153m^2$. The data are only collected while the GPS is turned on.

As the SafeGraph dataset only provides the level of social distancing detected for each census block group, but not for each immigrant community of interest, I combine it with the American Community Survey. The ACS 5-year (2015-2019) estimates contain the fraction of census tract population born in each origin country of interest.[7] Census tract is the most granular level at which information on the residents' countries of birth are provided. Thus, to combine the two datasets, I aggregate the social distancing metrics from SafeGraph to the census tract level. Table 1 illustrates the presence of immigrant clusters at the census tract level in the studied sample, which strengthens the validation of the analysis at the census tract level. I also use the microdata for the 5-year estimates from the 2015-2019 American community surveys provided by IPUMS USA to provide a further analysis on the composition of the second-plus-generation "immigrants" in the robustness section.

For my main analyses, I restrict my sample to the four largest Metropolitan Statistical Areas (MSAs) in the U.S: New York, Los Angeles, Chicago, and Dallas. I focus on these four cities for two reasons. First, most immigrant clusters are found in these four MSAs as presented in Table 2. Thus, including more cities would create a control group that is dissimilar to the treatment group. Second, only focusing on the four largest MSAs minimizes the risk of an ecological fallacy. This is because the average area of a census tract is much smaller in densely populated cities as the census tracts are determined by the size of the population living in the area. Having census tracts of smaller area can reduce ecological fallacy arising from social distancing behaviors caused by some geographic traits in a section of the tract that is not occupied by the immigrants from an origin country of interest. The area of each census tract in the New York metropolitan and the Los Angeles metropolitan

[7]The fraction of population born in each origin country is not provided in a single-year response of the American Community Survey.

Table 1: Twenty Largest Immigrant Groups in the Four Largest Metropolitan Statistical Areas

% Born in:	Obs.	Mean	Std. dev.	Min	Max
China	10,651	1.60	4.65	0	60.58
Colombia	10,651	0.42	1.15	0	23.12
Dominican Rep.	10,651	1.23	3.98	0	46.91
Ecuador	10,651	0.58	1.87	0	32.64
El Salvador	10,651	1.04	2.45	0	23.82
Guatemala	10,651	0.66	1.80	0	46.12
Guyana	10,651	0.48	2.21	0	37.20
Haiti	10,651	0.40	1.72	0	29.00
Honduras	10,651	0.29	0.83	0	16.28
India	10,651	1.16	2.84	0	48.50
Iran	10,651	0.35	1.65	0	35.68
Italy	10,651	0.31	0.77	0	11.44
Jamaica	10,651	0.65	2.52	0	32.31
Korea (South)	10,651	0.84	2.58	0	44.93
Mexico	10,651	6.36	9.36	0	49.38
Peru	10,651	0.31	0.87	0	15.63
Philippines	10,651	1.10	2.14	0	35.92
Poland	10,651	0.50	1.67	0	35.24
Taiwan	10,651	0.29	1.01	0	18.73
Vietnam	10,651	0.63	2.60	0	56.86

The table displays the summary statistics on the 20 largest immigrant groups in the 4 largest MSAs in the U.S.–New York, Los Angeles, Chicago, and Dallas–relating to the percentage of residents born in each of the 20 origin countries in a given census tract.

are illustrated in Figures 2 and 3 respectively.

Throughout this paper, the primary independent variable of my analyses is the extensive margin of social distancing - percentage of mobile devices staying *completely* home. The breakdown of this metric according to the MSA and the observation period is shown in Table 3.

Table 2: Distribution of Immigrant Clusters in Large Metropolitan Statistical Areas

Tracts with $\geq 5\%$ residents born in:	Total count	% in 4 MSAs	% in 10 MSAs
China	1787	44%	53%
Colombia	447	33%	86%
Dominican Rep.	1081	69%	84%
Ecuador	349	87%	90%
El Salvador	1325	49%	74%
Guatemala	719	41%	57%
Guyana	286	93%	95%
Haiti	721	37%	83%
Honduras	350	96%	98%
India	2045	29%	43%
Iran	197	87%	90%
Italy	55	15%	59%
Jamaica	664	59%	84%
Korea (South)	531	69%	83%
Mexico	12785	28%	35%
Peru	127	51%	75%
Philippines	1204	39%	40%
Poland	259	88%	89%
Taiwan	143	73%	73%
Vietnam	805	31%	48%

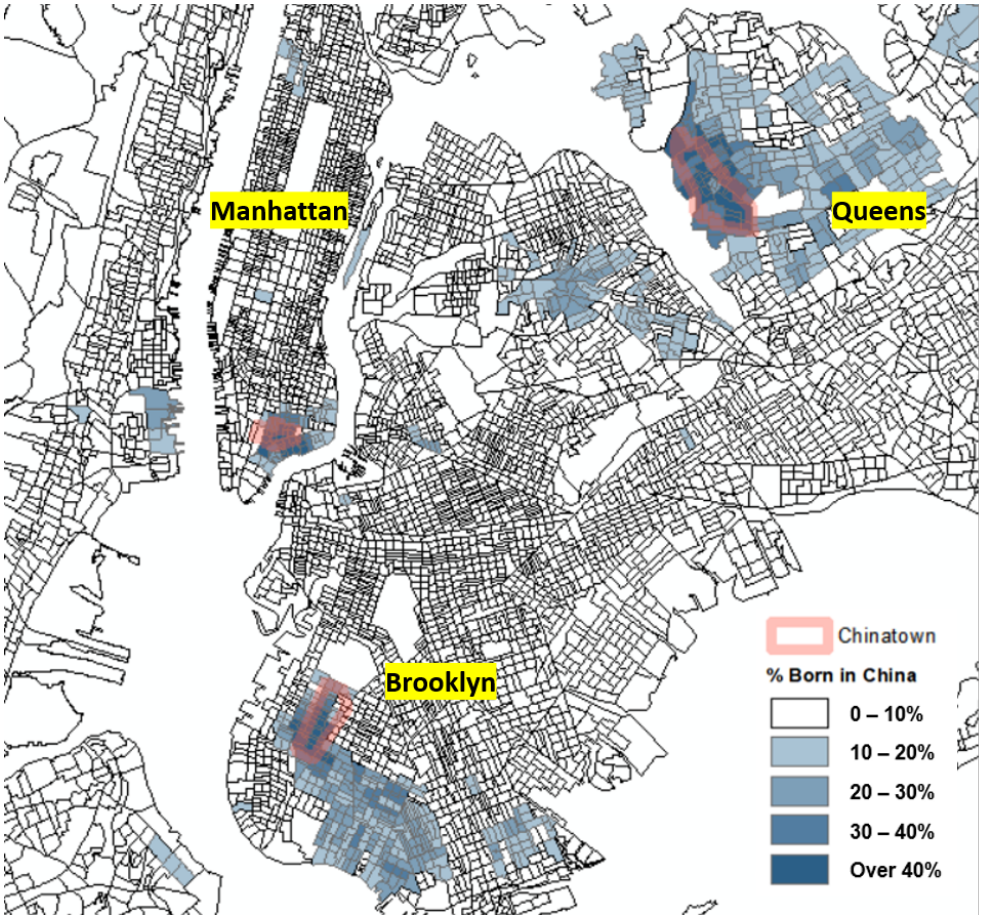
The second column of the table shows the total count of census tracts in the U.S. that have more than 5% of their residents born in each of the given origin countries. The third and fourth columns show the percentage of census tracts located in the 4 and 10 largest MSAs in the U.S. respectively among those counted in the second column. All numbers come from the 5-year (2014-2018) estimates of the American Community Survey.

Table 3: Average Percentage of Mobile Devices Staying Completely Home

MSA:	New York	Los Angeles	Chicago	Dallas
Before 3/13	24.94%	24.10%	25.30%	22.65%
3/13 Onwards	44.59%	40.70%	41.33%	36.79%

The table displays the average percentage of mobile devices staying completely home in a given census tract in each of the four largest metropolitan statistical areas. The table shows these numbers for before and after March 13th, 2020, the declaration of national emergency in the U.S.

Figure 2: Census Tracts in New York Metropolitan Statistical Area

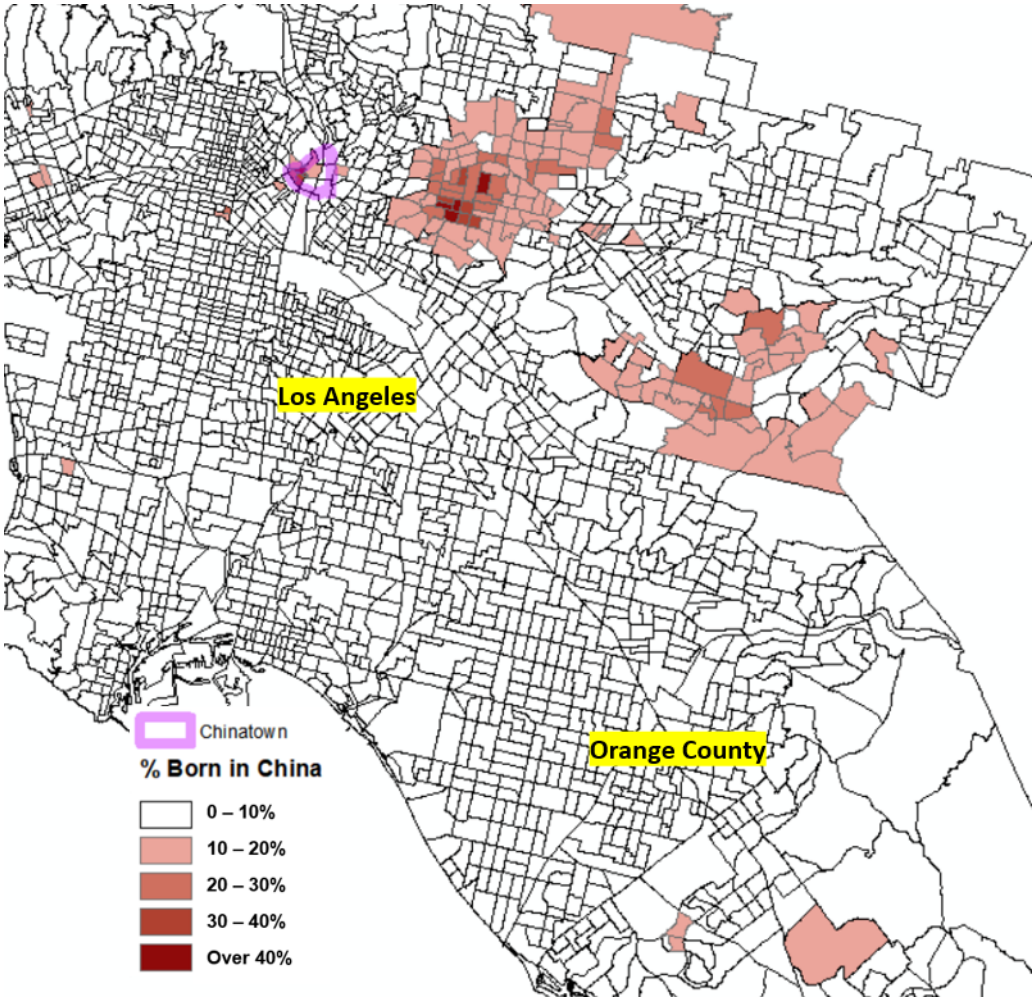


The above map gives a general sense of the area of census tracts in the New York Metropolitan Statistical Area. Moreover, the map demonstrates the existence of immigrant clusters where Chinese immigrant communities were chosen for illustration.

2.2 Measures Related to COVID-19 Outbreak in the Origin Countries

I choose two types of indicators of new developments around the COVID-19 pandemic in the origin country: the severity of the COVID-19 outbreak and the status of NPI in the origin country. I choose these measures because both measures have an objective message and are comparable across countries. They differ in that the former is about the actual outbreak and the latter includes preventive measures taken in the origin country out of caution. Hence, by looking at both I can study whether it is the actual outbreak in the origin country that

Figure 3: Census Tracts in Los Angeles Metropolitan Statistical Area



The above map gives a general sense of the area of census tracts in the Los Angeles Metropolitan Statistical Area. Moreover, the map demonstrates the existence of immigrant clusters where Chinese immigrant communities were chosen for illustration.

affects the social distancing behavior of the immigrants or the information obtained from the actions of the government in the origin country.

The severity of the COVID-19 outbreak in the origin country is measured by the cumulative count of confirmed cases or deaths resulting from the coronavirus. Although there have been allegations on the inaccuracy of these reported numbers in some countries, the numbers themselves convey a clear message. That is, the immigrant can perceive whether there is an outbreak in their origin country and how fast the virus is spreading and claiming lives by

looking at these numbers. This makes the interpretation of the message much more objective than personal accounts of the experience in the origin country, which could be interpreted in numerous different ways, making it more difficult to objectively compare across countries.

The NPI I study is whether the national or a local government decided to enforce school closures. First, this measure has a clear interpretation—schools are either ordered to close or not—unlike other NPIs that could differ in the intensity. For instance, a curfew could be placed for different hours in different countries. Similarly, social gathering limits also varied greatly in the limit on the number of people that can gather. Second, ordering school closures was a popular NPI placed at the beginning of an outbreak or as a preventive measure against an outbreak. Every origin country of the selected immigrant groups to be studied had implemented some form of school closures (including a postponement of the return to school from a break).

2.3 Information Transmission

To study the mechanism driving the results I find, I use two different data sets–the Facebook Connectedness Index and the Google Trends Index. I study the Facebook Connectedness Index for the relative amount of ties between those residing in the U.S. and those in other countries. I use the Google Trends Index to see when the interest in COVID-19 peaked for different language users in the U.S.

2.3.1 Facebook Connectedness Index

The Facebook Connectedness Index is a scaled index of social connectedness between U.S. counties with more than 100 active users and individual countries with at least 50,000 active users. This index, provided by Facebook, gauges the relative probability of a Facebook friendship connection between Facebook users in the U.S. counties and the chosen countries. Facebook first measures the level of social connectedness between a U.S. county a and a country b by dividing the number of actual Facebook friendship links between a and b by

the number of total possible Facebook friendship connections between a and b.

$$Social\ Connectedness_{a,b} = \frac{FB\ Connections_{a,b}}{FB\ Users_a * FB\ Users_b} \tag{1}$$

This measure is then turned into a scaled index to be a value between 1 and one billion.

2.3.2 Google Trends Index

Google Trends provide the relative search frequency of a particular term over a specified period of time in a specified location. Using an unbiased sample of actual Google search requests, the Google Trends index is created by: 1) finding the ratio between the number of searches on a particular term on a give day and the total number of searches on a given day in the given region—a metropolitan area or a country in this paper; and 2) adjusting the maximum value of such ratio to take on the value 100 and the rest of the values commensurately. Thus, the index can only be interpreted as relative search frequency and no statement can be made about the absolute quantity of the search requests. Because the Google Trends index is created for the exact terms specified in the query, I am able to get a sense of when the interest peak in COVID-19 was for each language group, which I can then map to a specific immigrant group based on their native language.

3 Immigrants' Behavioral Response to COVID-19 Developments in Origin Countries

3.1 Shift-Share Panel Regression

To answer whether immigrants respond to real-time information from their origin countries I use a shift-share panel regression. The shift-share design translates the individual-level shock into the census tract level by weighting the treatment from each origin country by the

fraction of residents in the census tract who were born in the given origin country.

As the severity and the composition of origin countries that have implemented an NPI change daily, the treatment at the census tract level shifts everyday according to the daily information update from the origin country weighted by the fraction of immigrants from the corresponding origin country in the tract. Because I assume the timing and intensity of the outbreak in the origin country are orthogonal to the immigrants' social distancing behavior in the U.S., the time-variant treatment can identify the immigrants' response to the COVID-19 outbreak in their origin countries. This relationship is summed up in the following equation:

$$\begin{aligned} y_{tract,t} = \alpha + \beta\big[\sum_{c\in\mathbb{C}} P^c_{tract} \cdot Severity^c_{t-1}\big] + \delta\big[\sum_{c\in\mathbb{C}} P^c_{tract} \cdot NPI^c_{t-1}\big] \\ +\theta_t + \eta_{tract} \times NPI^{US}_t + \epsilon_{tract,t}. \end{aligned} \tag{2}$$

The dependent variable $y_{tract,t}$ is the percentage of mobile devices that are home for the entire duration of observation in *tract* on date t. The first independent variable $\sum_{c\in\mathbb{C}} P^c_{tract} \cdot Severity^c_{t-1}$ denotes a shift-share treatment at the census tract level, created by weighting the shock from each origin country $c \in \mathbb{C}$ on date $t-1$ by the fraction of corresponding immigrant population in the tract, P^c_{tract}. The shock is measured by either the cumulative count of coronavirus cases or deaths depending on the specification. The second independent variable $\sum_{c\in\mathbb{C}} P^c_{tract} \cdot NPI^c_{t-1}$ is a shift-share treatment in which the implementation of the NPI in origin country c date $t-1$ is weighted by the fraction of the corresponding immigrant population in the tract. The variable θ_t denotes date fixed effects and $\eta_{tract} \times NPI^{US}_t$ denotes fixed effects of tract by NPI status in the U.S. That is, I am controlling for census tract fixed effects separately for pre- and post-NPI in the U.S. Finally, $\epsilon_{tract,t}$ represents the unobservables.

3.2 Findings

The severity of the outbreak, measured either by the cumulative number of cases of deaths, was chosen because it is the most salient, yet comparable, measure of an outbreak across countries. However, there were countries that did not experience an outbreak, but still chose to take precautionary steps. As discussions around the COVID-19 virus and the government's stance against it could have reached the general public of these countries and therefore the immigrants from these countries, I also test whether an implementation of an NPI in the origin country could have affected the social distancing behavior among immigrants.

I find that first generation immigrants increase their level of social distancing both with the increase in the severity of the outbreak and with the implementation of an NPI in the origin country. The regression results of equation (2) are reported in Table 4. The each column reports effects from a different combination of the severity and NPI status in the origin countries. In columns 1-4, severity is measured by the number of confirmed COVID-19 cases per 100,000 people in the origin country whereas in columns 5-8 it is represented by the number of confirmed COVID-19 deaths per 100,000 people. These columns demonstrate that results are robust to the choice of measure of severity. Columns 3, 4, 7, and 8 include the severity measure and the NPI status weighted by the fraction of second-plus-generation from the relevant countries in each tract.

The preferred specification, reported in column 4, shows that an average immigrant group increases social distancing by 0.757 percentage points for every confirmed case per 100,000 people in the origin country. Assuming that every immigrant group responds to severity in their origin country in the same way, the interpretation can be tailored to each country. To illustrate, Italy experienced 3.125 new daily cases per 100,000 people on average in the week of 3/8/2020 - 3/14/2020. Thus, according to the model, the probability of staying completely home on average increases by 0.757 * 3.125 = 2.366 percentage points for Italian immigrants in the U.S., holding fixed the level of severity in other origin countries.

Second-plus-generations also exhibit a positive and significant effect, although the mag-

nitude of the effect is almost half of that for the first generation immigrants. An average second-plus-generation immigrant group, or those claiming ancestry from but are not born in the origin country, increases social distancing by 0.206 percentage points for every confirmed COVID-19 case per 100,000 people in the origin country.

The first generation immigrants also have a significant and positive reaction to the implementation of an NPI. An NPI in the origin country increases the probability of staying completely home by 0.0536 percentage points among the immigrants from that origin country. This effect is much smaller than that for the severity of outbreak in the origin countries. I find a negative and significant effect of the implementation of an NPI on the level of social distancing among the second-plus-generations claiming ancestry from corresponding origin countries. However, these estimates are very small in magnitude. The implementation of an NPI in the origin country decreases the probability of staying completely home only by 0.00923 percentage points among corresponding immigrants of second-plus-generation. As the small standard errors indicate a high precision of these estimates, I conclude the effect is close to null. This is not surprising since the information on severity has much more salience than the NPI status, especially in international news.

3.3 Identification

To make an inference at the census tract level, I aggregate the individual level specifications and conduct analyses at the census tract level. Although ideally I would make direct inferences from the individual level specification that shows the relationship between the change in the developments around the COVID-19 pandemic in the origin country and the social distancing behavior of each immigrant, the data only allow for census tract level analyses. Thus, I start from an individual level ideal specification and work towards aggregating it up to the census tract level.

Equation (3) captures the ideal individual level specification.

$$y_{it} = \tilde{\alpha} + \tilde{\beta} Severity_{i,t-1}^{OC} + \tilde{\delta} NPI_{t-1}^{CO} + \zeta_{i \times t} + \tilde{\epsilon}_{it} \tag{3}$$

The dependent variable y_{it} is an indicator variable equal to 1 if individual i stays home at date t. Thus, it shows the extensive margin of social distancing. The independent variable $Severity_{i,t-1}^{OC}$ is the severity of the COVID-19 spread in individual i's origin country at date $t-1$, which is either measured by the cumulative count of coronavirus cases or deaths. The variable NPI_{t-1}^{CO} is an indicator variable equal to 1 if individual i's origin country has implemented the NPI by date $t-1$. The ideal specification would also include the date-by-census tract fixed effects, $\zeta_{i \times t}$. Finally, ϵ_{it} represents the unobservables.

Table 4: Shift-Share Panel Regression Results

	Measure of severity in origin country							
	No. confirmed cases per 100,000 people				No. confirmed deaths per 100,000 people			
Dependent Variable: % completely home	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
$Severity^{OC}$	1.332***	1.310***	0.832***	0.757***	11.74***	11.70***	5.464***	5.170***
(*country of birth*)	(0.0614)	(0.0615)	(0.0540)	(0.0532)	(0.554)	(0.553)	(0.499)	(0.496)
NPI^{OC}		0.0202***		0.0536***		0.0331***		0.0627***
(*country of birth*)		(0.00363)		(0.00505)		(0.00361)		(0.00503)
$Severity^{OC}$			0.198***	0.206***			1.586***	1.646***
(*ancestry*)			(0.00486)	(0.00484)			(0.0407)	(0.0404)
NPI^{OC}				-0.00923**				-0.0146***
(*ancestry*)				(0.00413)				(0.00413)
Constant	34.32***	34.01***	33.77***	33.14***	34.72***	34.19***	34.31***	33.64***
	(0.0343)	(0.0633)	(0.0307)	(0.0772)	(0.0161)	(0.0598)	(0.0164)	(0.0738)
Date FE	yes	yes	yes	yes	yes	yes	yes	yes
Tract $\times NPI^{US}$ FE	yes	yes	yes	yes	yes	yes	yes	yes
N	328708	328708	328708	328708	328708	328708	328708	328708
adj. R^2	0.860	0.860	0.862	0.862	0.860	0.860	0.862	0.862

Standard errors clustered at the census tract level in parentheses * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

This table reports the effect of severity of the COVID-19 outbreak in the origin country, proxied by the number of confirmed cases per 100,000 people, and of the implementation of an NPI in the origin country on the level of social distancing practiced by the relevant immigrant groups. Social distancing is measured by the percentage of mobile devices staying completely home in a given census tract.

The identification assumption for the ideal specification in equation (3) is $\mathbb{E}(\epsilon_{it}|X_{it}) = 0$, where X_{it} is a vector of all regressors in equation (3). As the timing and evolution of the COVID-19 outbreak in each origin country were quasi-random, I argue the change in the social distancing behavior among immigrants as a response to the new developments in their origin countries cannot be attributed to factors unrelated to the evolution of the outbreak in the origin countries. Then, the remaining threats to identification in claiming the information link between the immigrants and their origin countries to be the mechanism behind the behavioral change would be factors related to the developments of the outbreak in the origin countries, but are not information transmission, such as the real threat of disease transmission. However, as travel bans were placed in many countries that faced an outbreak early on, I am able to dismiss this concern.[8]

Aggregating equation (3) to the census tract level presents two issues. To aggregate the individual level specification to the census tract level, I take the individual level specification and aggregate it to the census tract level by taking a sum of the individual level specifications for each census tract and dividing both sides by the number of residents in the given tract. The first issue is that the unobservables might be systematically correlated within the census tracts, which would lead to the aggregate regression to have incorrect standard errors. The second issue is that I can no longer have tract-by-date fixed effects. Notice that when I aggregate equation (3) to the census tract level, the treatment will be at the tract-by-date level. Thus, the multicollinearity disallows controlling for the tract-by-date fixed effects.

I employ some remedies to address the two issues. The first issue can be resolved by clustering the standard errors by census tracts in the aggregate specification. The second issue is more complicated. Leaving out the tract-by-date fixed effects means, I cannot control for the local evolution of the COVID-19 outbreak, which is likely to affect the social distancing behavior. If the evolution of the outbreak is random across different census tracts, the identification remains unthreatened. However, as the spread of the disease could be related to

[8]Foreign nationals who have traveled in China, Iran, and the Schengen area are temporarily suspended from entering the U.S. starting from February 2nd, March 2nd, and March 13th on 2020 respectively.

certain demographic traits that are correlated with the composition of immigrant residents within the census tract, I perform the following two analyses. First, in the preferred specification, I include separate census tract fixed effects for pre- and post-NPI in the U.S. This is to control for the time-invariant characteristics of a given census tract that determines the level of social distancing differently in the pre- and post-NPI periods. Second, as one of the robustness checks, I include date-by-MSA fixed effects to control for the development of the outbreak at the city level.

Implementing the changes, the individual level specification to aggregate boils down to the following:

$$y_{it} = \alpha + \beta Severity^{OC}_{i,t-1} + \delta NPI^{CO}_{t-1} + \theta_t + \eta_{tract} \times NPI^{US}_t + \epsilon_{it} \tag{4}$$

The new variable, NPI^{US}_t, denotes the indicator variable equal to 1 if date t falls on or after March 13th, 2020, which is when the U.S. declared the national emergency. Thus, I am controlling for date fixed effects and census tract fixed effects for pre- and post-NPI in the U.S. I aggregate this hypothetical individual level specification to the census tract level under the assumption that clustering the standard errors at the census tract level, or the level of aggregation, allows me to interpret the coefficients from the aggregated specification in the same way as the hypothetical individual level specification. Thus the unobservables become $\frac{1}{N_{tract}} \sum_i^{N_{tract}} \tilde{\epsilon}_{it} = \tilde{\epsilon}_{tract,t}$. I maintain that once I cluster by census tract I can obtain this condition.

3.4 Robustness Checks

3.4.1 MSA-by-Date Fixed Effects

I run the preferred specification, which is shown in column 4 of Table (4), including the MSA-by-date fixed effects. Recall that in the ideal hypothetical specification at the individual level expressed in equation (3), I control for tract-by-time fixed effects to account for the

local evolution of the COVID-19 outbreak. However, once aggregated up to the tract level, it is impossible to have these fixed effects as they are at the same level as the treatment. To ensure the local development of COVID-19 is not interfering with the results, I include the MSA-by-date fixed effects as a robustness check. The robustness results in Table 5 confirm that the significance and sign of the main results regarding severity and NPI status in the origin country weighted by the fraction of first generation immigrants as well as severity weighted by the fraction of second-plus-generation do not change. The sign and significance for the NPI status weighted by the fraction of second-plus-generation from relevant origin countries in each census tract do change. The coefficient estimate loses significance and becomes positive. However, the interpretation does not change as I have previously explained that the negative effect reported in Table (4) was negligible given the small magnitude of the effect and the high precision of the estimate.

3.4.2 Excluding Immigrant Groups from the Control

I perform the robustness check of excluding an immigrant group one by one from the control group to make sure the results are not driven by one particular immigrant group. To address this issue, As I am trying to identify off of multiple immigrant communities with arguably different characteristics, it is less likely that the set of all immigrant groups whose origin countries faced a severe outbreak would have had a similar behavior before March 13th, 2020. However, if there are particular groups that are biasing the estimates this exercise would show that.

Figure 4 shows that excluding an immigrant group from the treatment does not significantly change the result for the responsiveness of immigrants to the COVID-19 outbreak in their origin country except for when Iranians are excluded from the control group. However, excluding Iranians increases the coefficient, which suggests that the main result reported is a lower bound. As for the responsiveness of immigrants to the implementation of NPI in the

Table 5: Robustness Check: Include MSA-by-Date Fixed Effects

	Measure of Severity:	
	cases per 100,000	deaths per 100,000
Dependent Variable: % completely home	(1)	(2)
$Severity^{OC}$	0.495***	3.423***
$(country\ of\ birth)$	(0.0488)	(0.468)
NPI^{OC}	0.00942*	0.00920*
$(country\ of\ birth)$	(0.00546)	(0.00545)
$Severity^{OC}$	0.134***	0.999***
$(ancestry)$	(0.00541)	(0.0450)
NPI^{OC}	0.00466	0.00285
$(ancestry)$	(0.00475)	(0.00475)
MSA × Date FE	yes	yes
Tract × NPI^{US} FE	yes	yes
N	328708	328708
adj. R^2	0.874	0.874

Standard errors in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

origin country, dropping Italian immigrants from the control group seems to push down the estimate of the effect of among first generation immigrants. Nevertheless, the result remains positive and significant. The robustness check on the coefficient representing the responsiveness of second-plus-generation to the severity of the COVID-19 outbreak in the origin country seems a little bit troubling. The estimate is positive and significant for all specifications, but when the Italian immigrants are excluded from the control group, the coefficient seems to jump significantly. No significant change is detected for the coefficient estimating the effect of NPI in the origin country on the second-plus-generation due to exclusion of any immigrant population from the control group.

One potential explanation for the jump in the estimated effect of severity in the origin

country on the social distancing behavior among second-plus-generation when Italians are excluded from the control group is that people claiming ancestry from Italy are significantly more likely to be distant descents. Table 6 shows the percentage of second-plus-generation of each immigrant group that speaks a language from their origin country. If descents are less likely to know a language from the origin country the further away in generations they are from the ancestors who immigrated, this table shows that Italian descents consist of the largest number of descents furthest in generations from the ancestors who immigrated to the U.S. As it is less likely to be connected to the country of origin as a distant descent, it is not surprising that much smaller effect of the severity of COVID-19 outbreak in the origin country on the social distancing behavior among the descents living in the U.S. was found among Italian descents.

4 Interactive Role of the Country of Residence in Determining Immigrants' Behavioral Response

I hypothesize that the real-time information from the origin countries could lead to a behavioral change among immigrants at two different points of time. The first is upon the arrival of the information from the origin country and the second is when the information becomes relevant in the country of residence. To see if the level of outbreak in the U.S. jointly determines the immigrants' social distancing behavior, I conduct event study analyses around the date national emergency was declared in the U.S.

Figure 4: Robustness Check: Exclude Immigrant Group from Control One by One

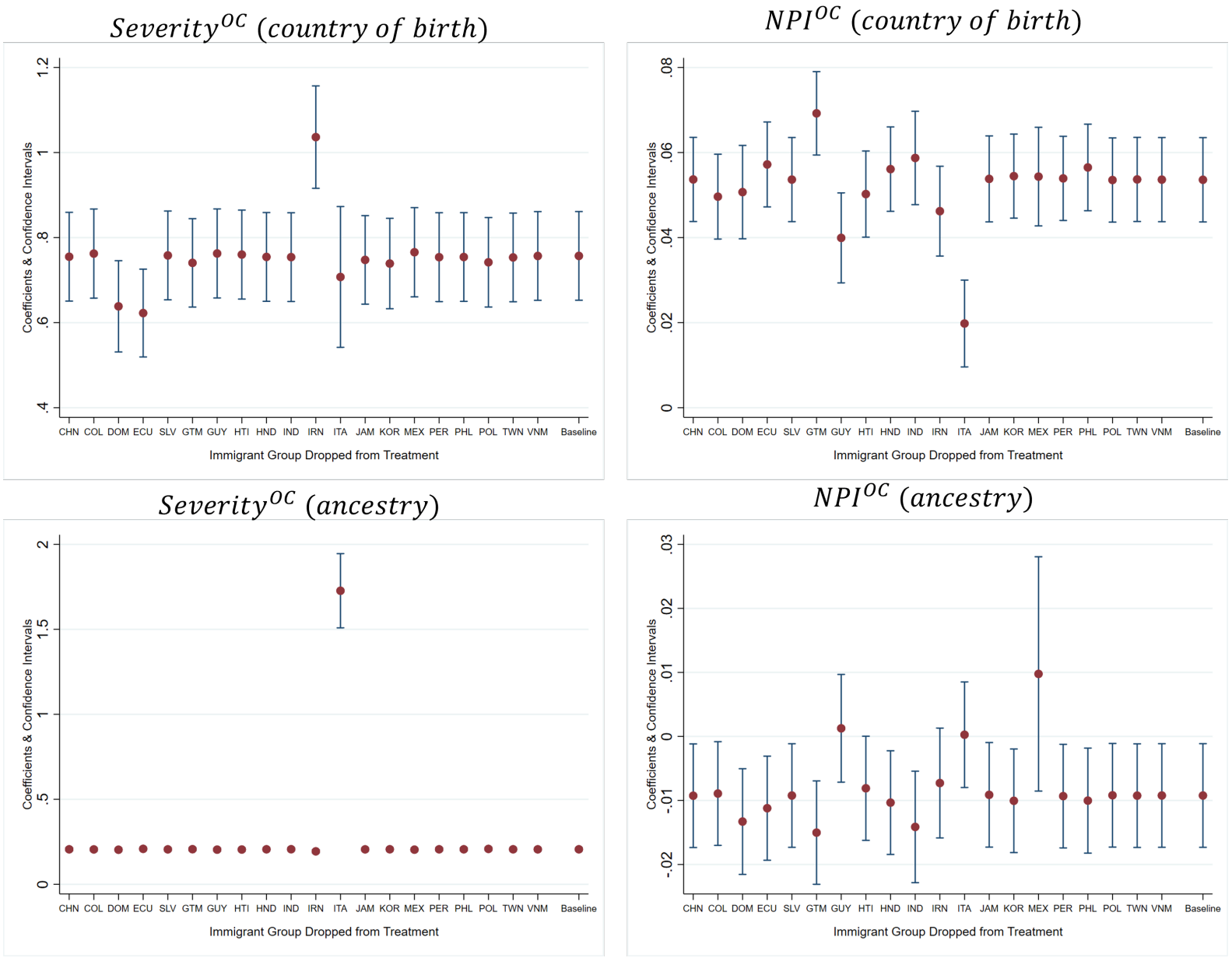


Table 6: Percentage of second-plus-generation of Each Immigrant Group that Speaks the Language of Origin Country

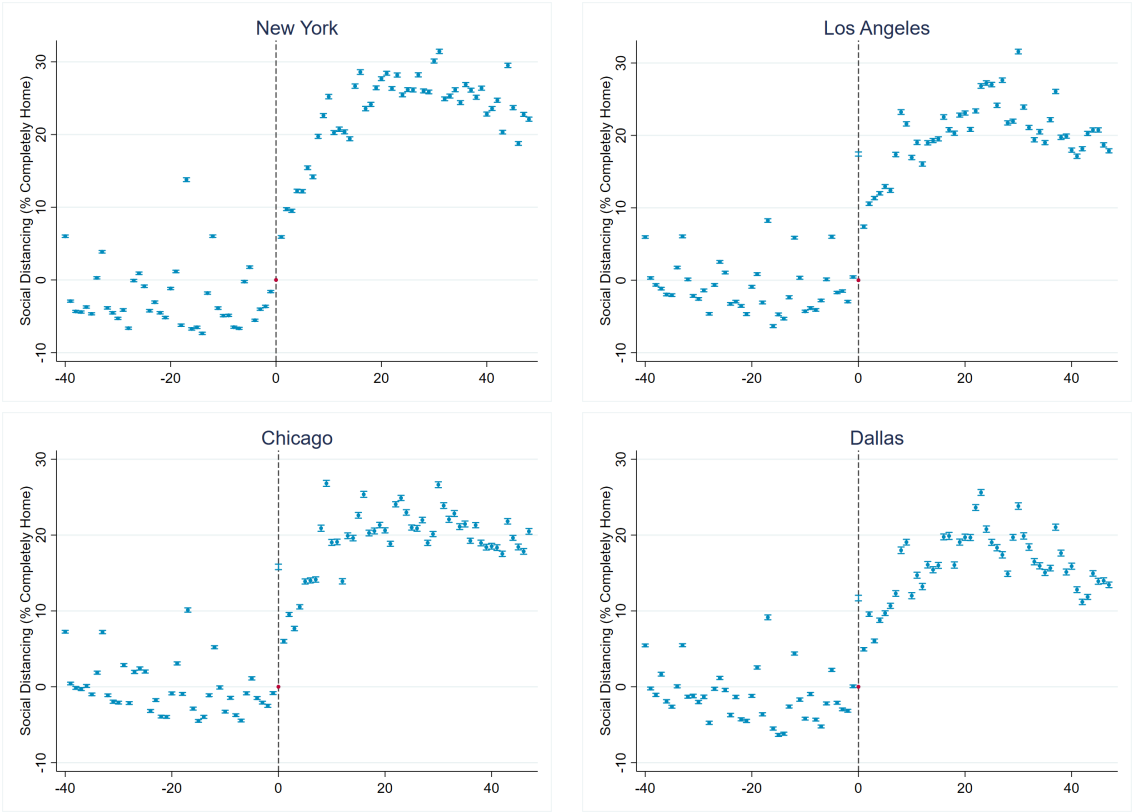
Immigrant group	Number of observations	% speaking language of origin country at home among second-plus-generation*
Chinese	34,314	48.84%
Colombian	4,843	57.77%
Dominican	13,265	65.53%
Ecuadorian	5,590	64.76%
Guatemalan	7,718	69.36%
Haitian	3,662	40.63%
Honduran	3,085	63.66%
Iranian	2,858	33.80%
Italian	152,641	2.71%
Korean	9,735	42.17%
Mexican	178,236	60.37%
Peruvian	3,055	58.10%
Polish	74,690	4.31%
Salvadoran	13,020	71.26%
Taiwanese	2,565	53.53%
Vietnamese	7,750	49.26%
Filipino	15,078	12.87%
Indian	39,995	14.62%

second-plus-generation*: Individuals claiming ancestry from a origin country who were not born in the given origin country
This table reports the percentage of second-plus-generation individuals, of the 20 immigrant groups selected for this study, in the four largest MSAs speaking a language of the corresponding origin country at home.
Guyanese and Jamaican are excluded from this table because the official language in these countries is English and therefore it is difficult to distinguish the language of origin country from the language most commonly spoken in the U.S.
Filipino and Indian are reported separately from the rest because a significant population of the Philippines and India speak English, which again makes it more difficult to distinguish the language of origin country from the language most commonly spoken in the U.S.

I run the event study analysis around March 13th, 2020, or the day the national emergency was declared, because this is the date that COVID-19 became "relevant" for the general public in the U.S. Figure 5 illustrates the social distancing pattern among the general population around the declaration of national emergency. In all four cities, the jump in the level of social distancing can be observed around March 13th. Whether this means the

increase in the disease transmission or in the awareness of COVID-19, March 13th marks an important date in which the general public responded to the pandemic by practicing social distancing. Thus, I conduct an event study around this date to see if the increase in the relevance of COVID-19 in the U.S. has played an interactive role in determining the level of social distancing for various immigrant groups.

Figure 5: Social Distancing Pattern of the General Population
Around Declaration of National Emergency

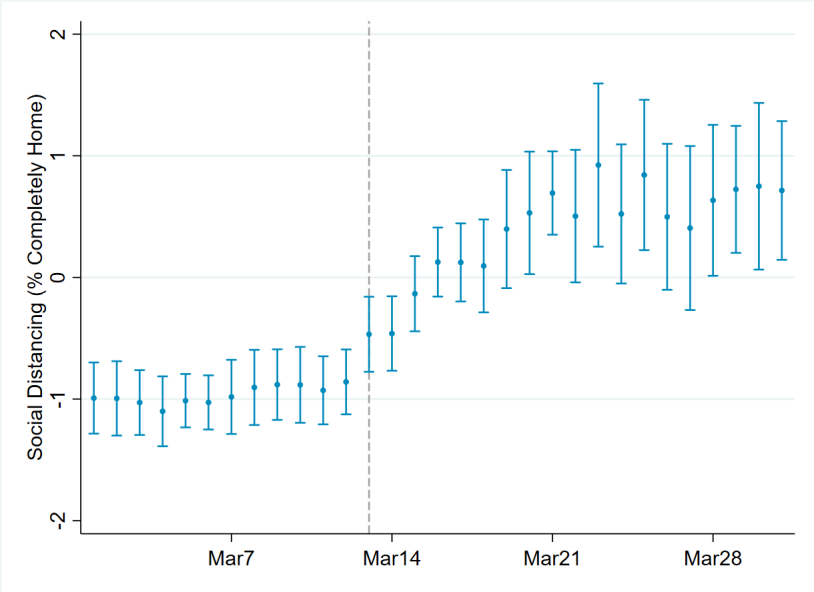


The graphs above plot the coefficient estimates and their 95% confidence intervals from regressing the percentage of mobile devices staying completely home on date dummies after clustering by census tracts. The x-axis shows the number of days since March 13th, 2020, or the day the national emergency was declared.

To get a sense of what the social distancing behavior looked like for immigrant groups whose origin country was one of the first countries to face a COVID-19 outbreak, I plot the relationship between social distancing and the percentage of tract population born in Italy in Figure 6. Each interval is the 95% confidence interval around the regression coefficients

from all cross-section regressions of percent of mobile devices staying completely home on the percentage of tract population born in Italy for each given day after including county fixed effects and clustering the standard errors by county. Note that at best Figure 6 only captures the reduced form relationship and is presented just to illustrate the timing of increase in social distancing for a given immigrant group. In this section I look at whether such effect is a general effect found in the largest immigrant communities and if identify the type of information from the origin country that helps quantify the effect.

Figure 6: Social Distancing Behavior of Census Tracts with Italian Immigrants



The above figure displays coefficients from a cross-section regression of percent of mobile devices staying completely home on the percentage of tract population born in Italy for each given day with county fixed effects and clustered errors clustered by county.

4.1 Method

The event study estimates the immigrants' response to a time-invariant measure of the outbreak in the origin country, which permits analyzing the daily changes in the level of social distancing without any interference from the time-varying treatment. The treatment is again at the census tract level and is a result of weighting the severity in each origin

country by the fraction of population in a given tract from each of those origin countries. However, the treatment does not evolve over time and is fixed at the level on March 1st to make the event study analysis possible. The specification can be found below where the superscripts ES are simply short for Event Study.

$$\begin{aligned} y_{tract,t} = \alpha^{ES} + \sum_{n=-12}^{12} \Big\{ \beta_n \big[\sum_{c \in \mathbb{C}} P^c_{tract} \cdot Severity^c \big] \times NPI^{US}_{t,n} \\ + \delta_n \big[\sum_{c \in \mathbb{C}} P^c_{tract} \cdot NPI^c \big] \times NPI^{US}_{t,n} \Big\} + \theta^{ES}_t + \eta^{ES}_{tract} + \epsilon^{ES}_{tract,t} \end{aligned} \tag{5}$$

4.2 Findings

Could the situation in the country of residence play an interactive role in how the immigrants process the information from their origin countries? The immigrants might defer updating their behavior despite new information they have received from their origin countries depending on the relevance of the information in the country of residence. In the context of the COVID-19 pandemic, immigrants whose origin countries experienced an outbreak before the U.S. could have been more knowledgeable about the fatality and the rate of transmission of the COVID-19 virus than their native counterparts or other immigrant groups in the beginning. However, they might not have had the incentive to alter their behavior when the virus did not pose a real threat to the residents in the U.S. Under this scenario, the greater knowledge on the virus only becomes relevant once the outbreak occurs in the country of residence.

I find that indeed immigrants whose origin countries faced an outbreak before the U.S. increase their level of social distancing significantly more, compared to their native counterparts and other immigrant groups, immediately after the declaration of national emergency in the U.S. Figures 7 and 8 show the evolution of the social distancing behavior before and after the declaration of national emergency in the U.S. of the immigrants whose origin country faced a COVID-19 outbreak before the U.S.

The top figure in Figure 7 and Figure 8 is the evolution of the immigrants' social dis-

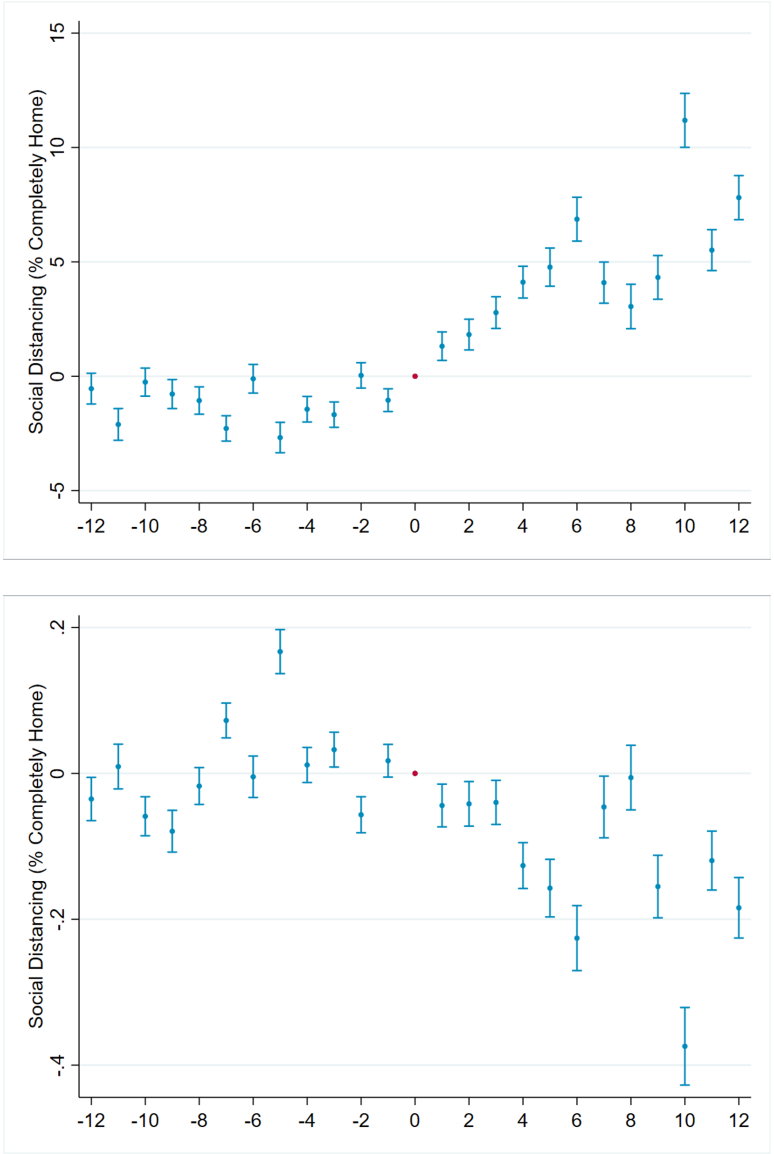
tancing behavior as a response to the severity of COVID-19 outbreak in the origin country. Severity is measured by the number of confirmed COVID-19 cases in Figure 7 and confirmed COVID-19 deaths in Figure 8 per 100,000 people in the origin country. Both figures show a significant increase in the level of social distancing among immigrants whose origin countries faced an outbreak before the U.S.

The bottom figure is the social distancing response among immigrants whose origin country implemented an NPI by March 1st, the date when the severity measures and NPI statuses were fixed for the event study analysis. Both Figure 7 and Figure 8 display a negative effect on social distancing of NPI implementation in the origin country. However, notice that the scale of the y-axis is substantially smaller than that in the top figure in both cases. That is, the negative effect found for NPI implementation is much smaller than the effect found for severity measured by confirmed COVID-19 cases and deaths respectively. The coefficient on NPI implementation should be interpreted with caution as the countries that have implemented an NPI are a mix of countries that faced an actual outbreak and countries that did not face an actual outbreak but implemented the NPI as a precautionary step.

5 Mechanism

I find that real-time transmission of information is a likely mechanism through which the new developments in the origin country induce a behavioral change among immigrants living in the United States. I present two pieces of suggestive evidence. The first is the intensity of social media connections between the immigrants in the U.S. and those living in the origin country. The second is the alignment in the timing of search peak of COVID-19 in a given origin country and in the U.S. in the language of the corresponding country of origin.

Figure 7: Event Study Result (severity: cumulative count of COVID-19 cases in origin country per 100,000 people)

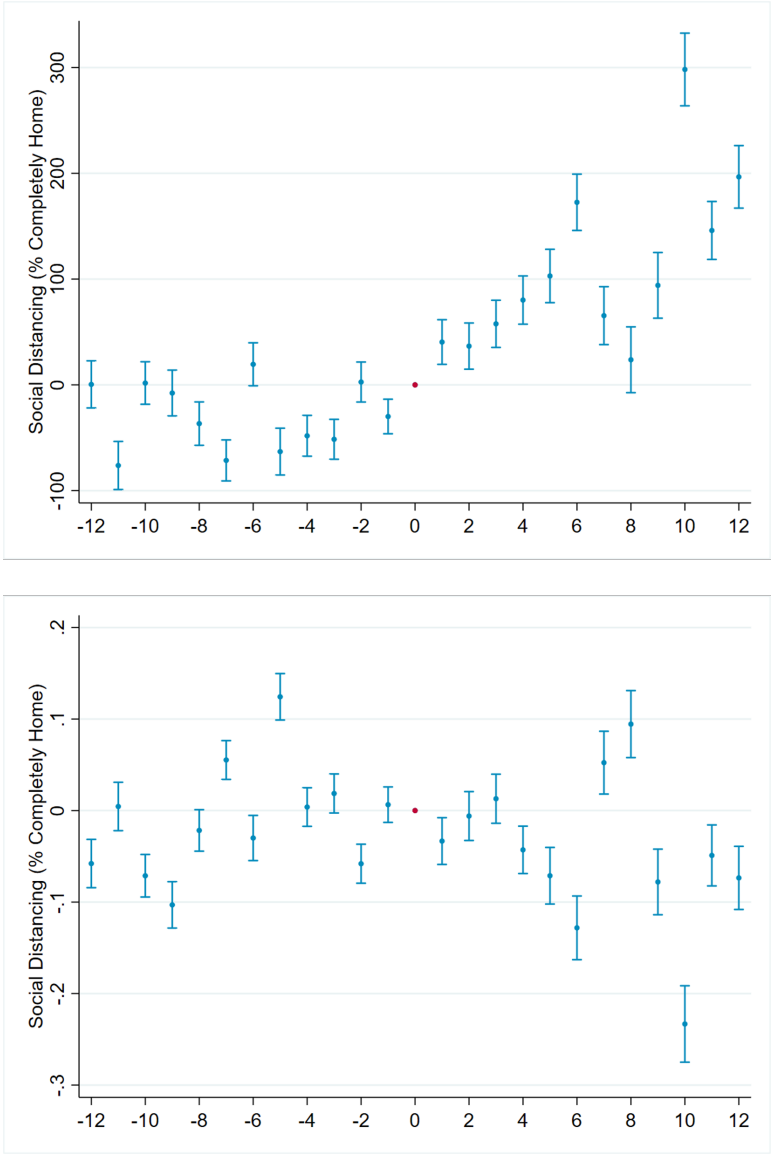


The two graphs display the estimates for the event study coefficients β_n and δ_n respectively from equation (5) where severity of COVID-19 in origin country is measured by the cumulative count of confirmed COVID-19 cases per 100,000 people in the given origin country.

5.1 Facebook Connectedness Index

I hypothesize that the Internet and the social media play a significant role in facilitating the real-time transmission of information from the origin country to immigrants living in the U.S. To check the potential channel through which the immigrants living in the U.S. could

Figure 8: Event Study Result (severity: cumulative count of COVID-19 deaths in origin country per 100,000 people)



The two graphs display the estimates for the event study coefficients β_n and δ_n respectively from equation (5) where severity of COVID-19 in origin country is measured by the cumulative count of confirmed COVID-19 deaths per 100,000 people in the given origin country..

receive real-time information from their origin countries, I look at whether the immigrants are connected to their origin countries through social media. The Facebook Connectedness Index is a scaled index of relative probability of a Facebook friendship connection between Facebook users in the U.S. counties and a given country. I test whether the fraction of population in

a county born in a given country predicts the strength of Facebook Connectedness between that county and the given country, I run an OLS regression specified in Equation (6).

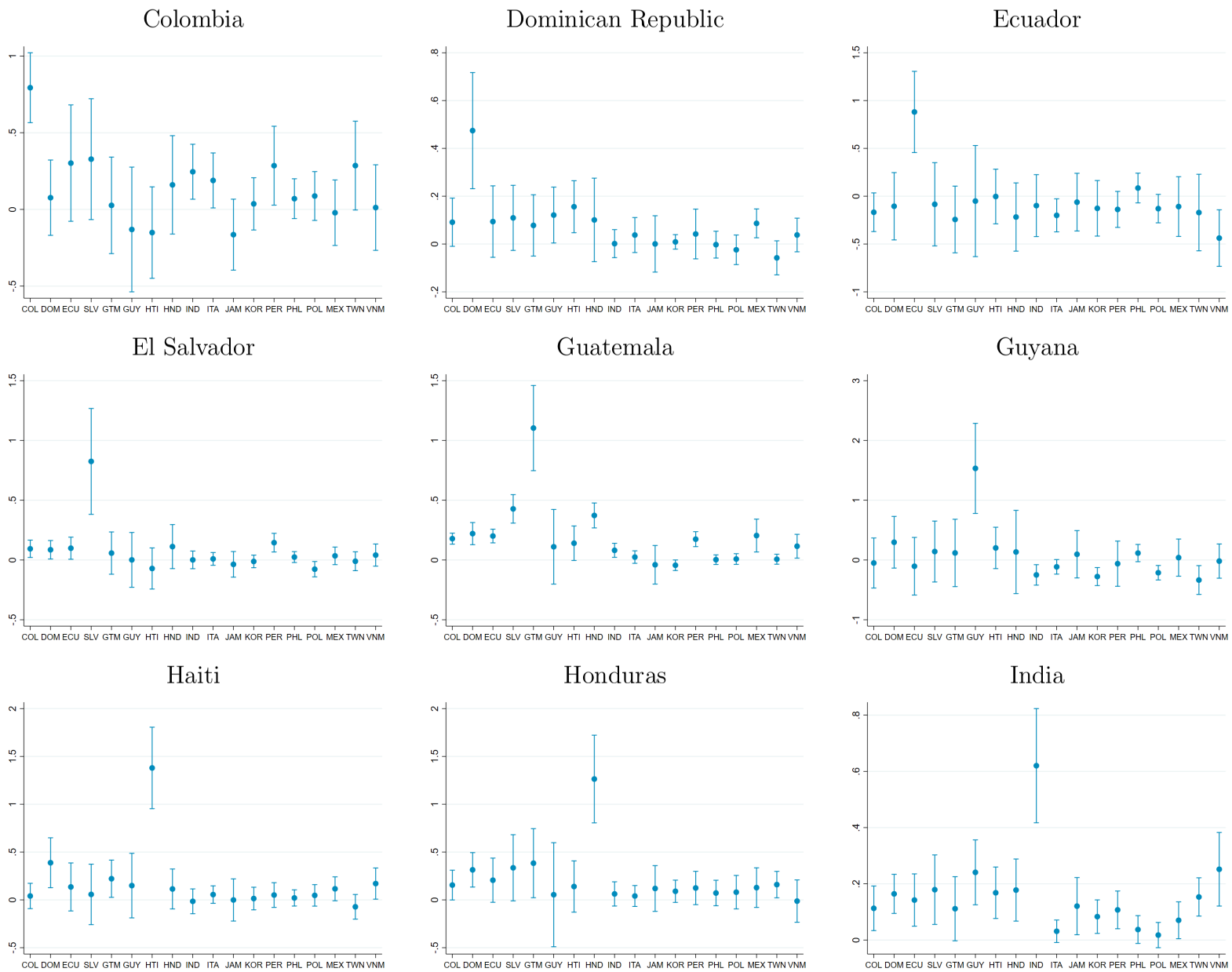
$$log(FC^{oc}_{county}) = \pi_0 + \sum_{c \in \mathbb{C}} \pi^c P^c_{county} + \alpha_{state} + e_{county} \tag{6}$$

Figure 9 and Figure 10 report π^c, or the coefficients representing the relationship between the fraction of population in each county that is born in a given foreign country and the log of the Facebook Connectedness Index for the corresponding pair of the U.S. county and foreign country. The x-axes of the graphs indicate to which immigrant group's presence in each county the coefficient estimate corresponds. The title of each coefficient plot states to which foreign country the Facebook Connectedness Index corresponds. These coefficient plots show that the fraction of population in a given county born in the relevant origin country is indeed the strongest predictor of the Facebook Connectedness between the given county and that origin country.

5.2 Google Trends Index

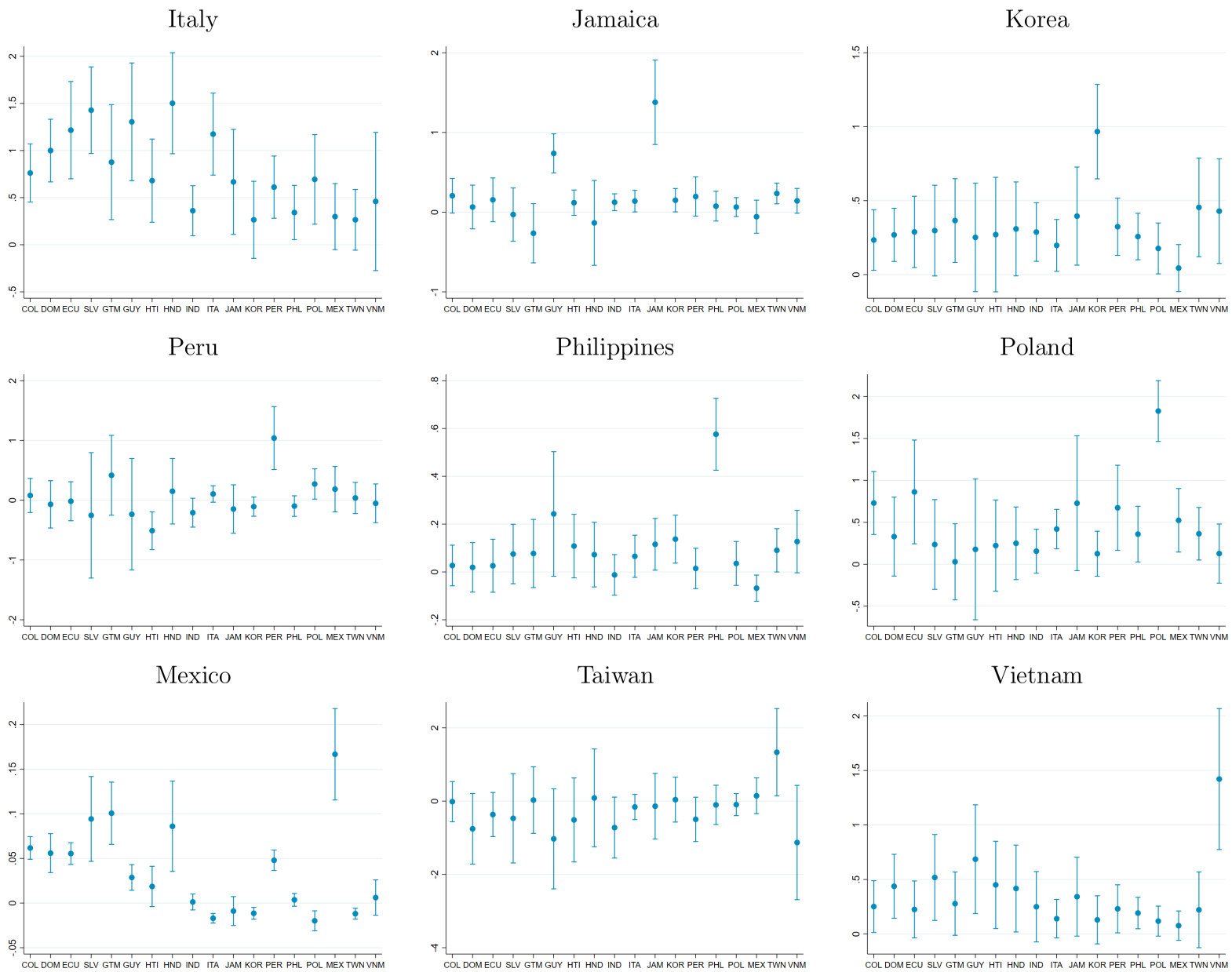
In this subsection, I present suggestive evidence that immigrants follow real-time news from their origin countries. I can compare the peak in the search interest for COVID-19 in an origin country and among corresponding immigrant groups in the U.S. through using Google Trends Index, which allows for identifying the evolution of amount of search requests on COVID-19 made to Google relative to all search requests on a given day.

Figure 9: Facebook Conenctedness Between U.S. Counties and a Origin Country



The coefficient plots above display the coefficient estimates for π^c from equation (6). The x-axes of the plots list to which immigrant group's presence in each county the coefficient estimate corresponds. The title of each coefficient plot states to which foreign country the Facebook Connectedness Index corresponds.

Figure 10: Facebook Conenctedness Between U.S. Counties and a Origin Country



The coefficient plots above display the coefficient estimates for π^c from equation (6). The x-axes of the plots list to which immigrant group's presence in each county the coefficient estimate corresponds. The title of each coefficient plot states to which foreign country the Facebook Connectedness Index corresponds.

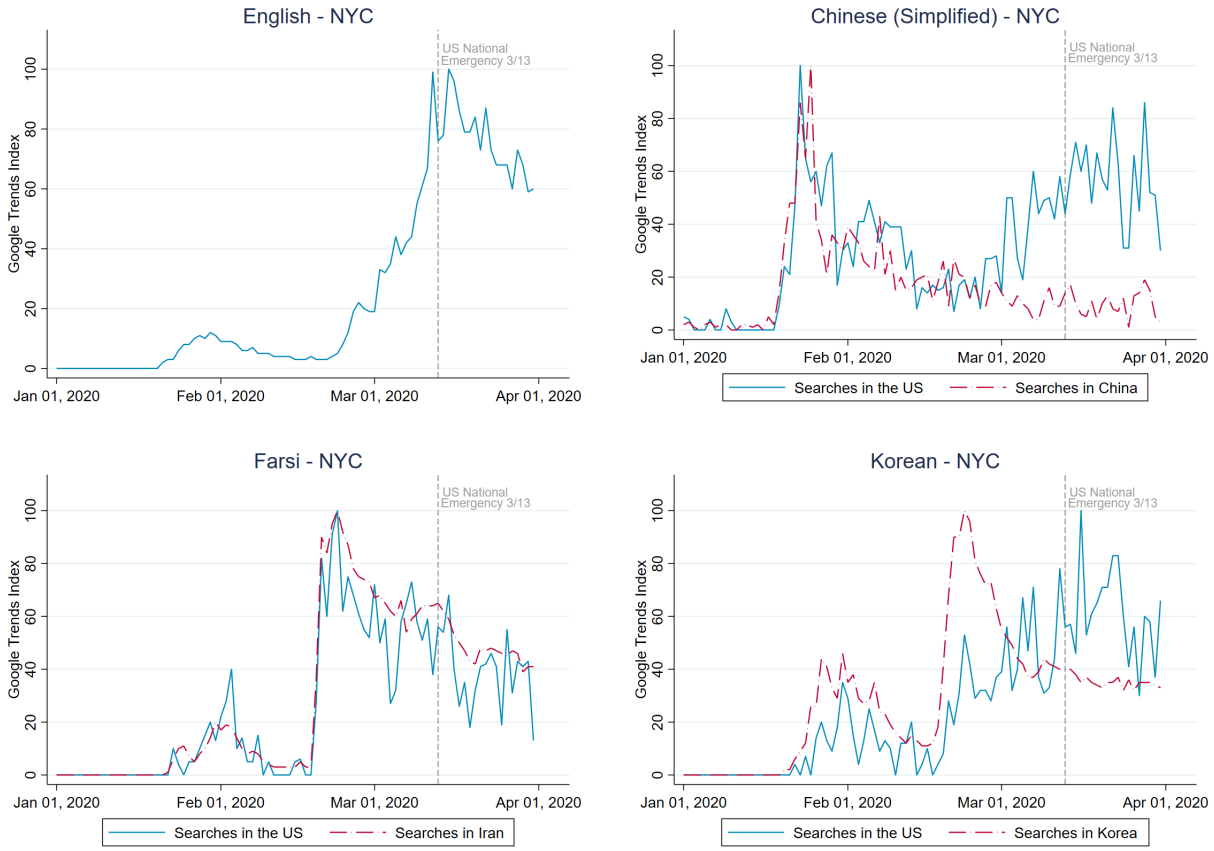
As the Google Trends index is calculated using the actual search requests for the terms specified in the query, I am able to study the relative search frequency for COVID-19 in various languages as long as the term is unique to the given language. For instance, Coronavirus is written in the same way in many languages including English, Spanish, Italian, and so on. However, if a language has a unique writing system or a term for COVID-19 that is unique to that language, I am able to study the evolution of the search frequency for the given language users. Figures 11, 12, and 13 show the search peaks on COVID-19 in China, Iran, Italy, and South Korea, the four countries that faced a severe outbreak before the U.S.[9] The figures compare these search peaks with those in the U.S. and confirm that the search peak in the U.S. presumably of relevant immigrant groups–proxied by the language of the immigrants–coincide with the search peak in the origin country.

6 Conclusion

I find that immigrants respond to real-time information from their origin countries, but when the transfer of information is applicable in the setting of the country of residence. In the context of the COVID-19 pandemic, I find that immigrants living in the U.S. were responsive to the new developments in their origin countries. That is, an increase in the severity of the outbreak and the implementation of an NPI in the origin country were both effective in increasing the level of social distancing among the immigrants whose origin countries were affected. The effect of the new developments in the origin country around the COVID-19 outbreak extended beyond first generation immigrants and had a significant effect on the second-plus-generation as well. Moreover, I find that the increase in the level of social distancing among immigrants whose origin countries were severely affected happens after the declaration of national emergency in the U.S. In other words, the circumstances in the country of residence plays an interactive role in inducing a behavioral change among the

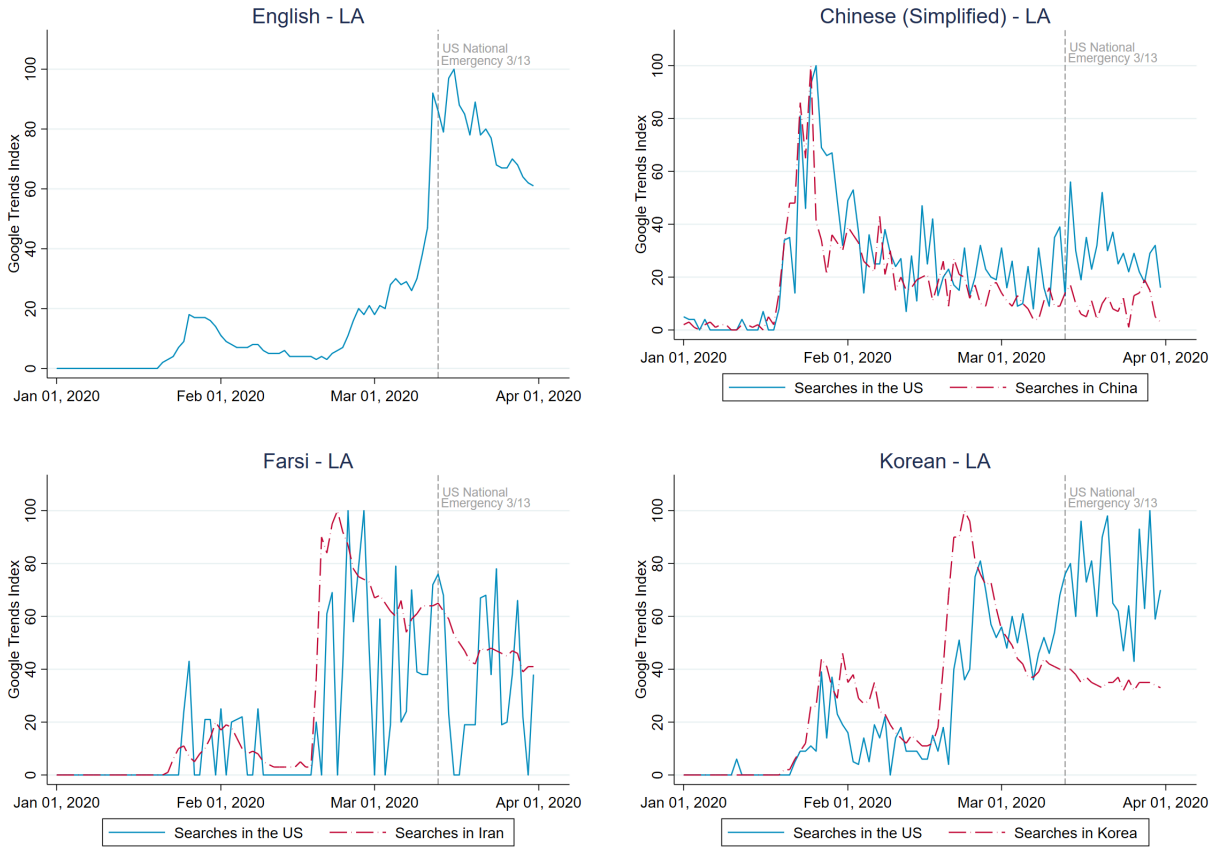
[9]Because Coronavirus is written in the same way in English and Italian, I compare the search frequency of *Coronavirus* in Italy to *Coronavirus Italia* in the US to distinguish search queries made by the Italian speakers in the US from those made by the English speakers.

Figure 11: Comparison in Google Search Trends in New York Metropolitan vs. Origin Country



immigrants. Finally, I show that real-time transmission of information facilitated by the Internet and social media is a likely mechanism behind immigrants' behavioral response to new developments in their origin countries. However, further study could shed light on the role of social media and other channels on the Internet in affecting the immigrants' behavior through changes in their origin country.

Figure 12: Comparison in Google Search Trends in Los Angeles Metropolitan vs. Origin Country



References

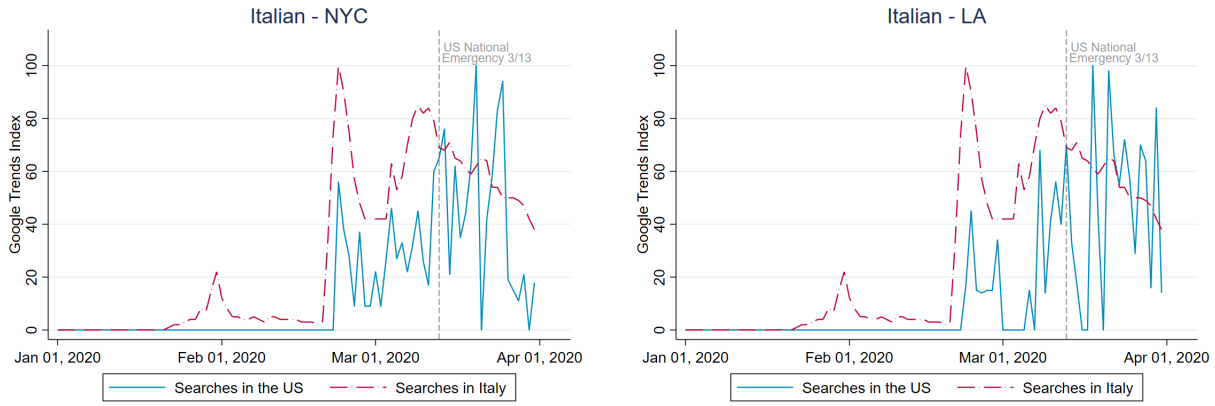
Allcott, Hunt, Levi Boxell, Jacob Conway, Matthew Gentzkow, Michael Thaler, and David Yang (2020). "Polarization and public health: Partisan differences in social distancing during the coronavirus pandemic." Journal of Public Economics. 191.

Akesson, Jesper, Sam Ashworth-Hayes, Robert Hahn, Robert D. Metcalfe, and Itzhak Rasooly (2020). "Fatalism, beliefs, and behaviors during the COVID-19 pandemic." Working Paper.

Alesina, Alberto, Paola Giuliano, and Nathan Nunn (2013). "On the Origins of Gender Roles: Women and the Plough." Quarterly Journal of Economics. 128(2), pp.469–530.

Antecol, Heather (2000). "An examination of cross-country differences in the gender gap in labor force participation rates." Labour Economics. 7(4), pp.409-426.

Figure 13: Comparison in Google Search Trends of *coronavirus Italia* in the United States vs. Origin Country



Apfel, Nicolas (2020). "Information and Migrant Communities in the Pandemic." Working Paper.

Bailey, Michael, Drew Johnston, Martin Koenen, Theresa Kuchler, Dominic Russel, and Johannes Stroebel (2020). "Social Distancing During a Pandemic: The Role of Friends." Working Paper.

Blau, Francine (1992). "The Fertility of Immigrant Women: Evidence from High-Fertility Source Countries." In *Immigration and the Workforce: Economic Consequences for the United States and Source Areas*, edited by George J. Borjas and Richard B. Freeman. University of Chicago Press, pp. 93–134.

Barsbai, Toman, Hillel Rapoport, Andreas Steinmayr, and Christoph Trebesch (2017). "The Effect of Labor Migration on the Diffusion of Democracy: Evidence from a Former Soviet Republic." American Econonic Journal: Applied Economics. 9(3), pp.36–69.

Blau, Francine D., Lawrence M. Kahn (2015). "Substitution between Individual and source country characteristics: social capital, culture, and US labor market outcomes among immigrant women." Journal of Human Capital. 9(4), pp.439–482.

Burchardi, Konrad B., Thomas Chaney, and Tarek A. Hassan (2019). "Migrants, Ancestors, and Foreign Investments." Review of Economic Studies. 86(4), pp.1448-1486.

Bursztyn, Leonardo, Aakaash Rao, Christopher P. Roth, and David H. Yanagizawa-Drott (2020) "Misinformation during a pandemic." Working Paper.

Christopoulou, Rebekka. and Dean R. Lillard (2015). "Is smoking behavior culturally determined? Evidence from British immigrants." Journal of Economic Behavior Organization. 110(0), pp.78–90.

Dong, Ensheng, Hongru Du, and Lauren Gardner (2020). "An interactive web-based dashboard to track COVID-19 in real time." The Lancet Infectious Diseases. 20(5), pp.533-534.

Egorov, Georgy, Ruben Enikolopov, Alexey Makarin, and Maria Petrova (2021). "Divided we stay home: Social distancing and ethnic diversity." Journal of Public Economics, 194.

Fernandez, Raquel, and Alessandra Fogli (2009). "Culture: An Empirical Investigation of Beliefs, Work, and Fertility." American Economic Journal: Macroeconomics, 1(1), pp.146–77.

Giuliano, Paola (2007). "Living Arrangements in Western Europe: Does Cultural Origin Matter?" Journal of the European Economic Association. 5(5), pp.927–52.

Gould, David M. (1994). "Immigrant Links to the Home Country: Empirical Implications for U.S. Bilateral Trade Flows." Review of Economics and Statistics. 76(2), pp.302-316.

Head, Keith and John Ries (1998). "Immigration and trade creation: Econometric evidence from Canada." Canadian Journal of Economics. 31(1), pp.47-62.

Peri, Giovanni and Francisco Requena-Silvente (2010). "The trade creation effect of immigrants: evidence from the remarkable case of Spain." Canadian Journal of Economics. 43(4), pp.1433–1459.

Simonov, Andrey, Szymon K. Sacher, Jean-Pierre H. Dubé and Shirsho Biswas (2020). "The persuasive effect of Fox news: Non-compliance with social distancing during the COVID-19 pandemic. Working Paper.

Tian, Yuan, Maria Esther Caballero, and Brian K. Kovak (2020). "Social Learning along International Migrant Networks." Working Paper.

Valsecchi, Michele and Ruben Durante (2021). "Internal migration networks and mortality in home communities: Evidence from Italy during the Covid-19 pandemic" European Economic Review. 140.

U.S. Census Bureau (2015-2019). American Community Survey 5-Year Estimates.

U.S. Census Bureau (2015-2019). American Community Survey 5-Year Public Use Microdata Sample (PUMS).

World Bank (2019). "Population, Total" *World Development Indicators*.