Online Appendix for

"How Racial Animus Forms and Spreads"

Runjing Lu and Sophie Yanying Sheng

APPENDIX 1. RESCALED GOOGLE SEARCH INDEX

Google Trends reports the search index in either a time series or a cross-sectional format. To construct a panel data set for each media market and time, we need to extract the search index in each media market separately. However, the search index reported by Google Trends is the search rate normalized by the maximum search rate in an extraction and is not comparable across extractions. To build a panel of search indexes that are normalized by the same base, we rescale the search index using the following method.

In a time series extraction of the search index in media market m over period T, the search index in median market m at time t is approximately

$$\text{Search Index}_{mt,T} = 100 \times \frac{\frac{\text{Searches including "chink(s)"}_{mt}}{\text{Total searches}_{mt}}}{\max\limits_{t \in T}\left\{\frac{\text{Searches including "chink(s)"}_{mt}}{\text{Total searches}_{mt}}\right\}} \tag{5}$$

Meanwhile, in a cross-sectional extraction of the search index at time t for all media markets $m \in M$, the search index in media market m at time t is approximately

$$\text{Search Index}_{mt,M} = 100 \times \frac{\frac{\text{Searches including "chink(s)"}_{mt}}{\text{Total searches}_{mt}}}{\max\limits_{m \in M}\left\{\frac{\text{Searches including "chink(s)"}_{mt}}{\text{Total searches}_{mt}}\right\}} \tag{6}$$

If we are willing to assume that the numerators in equations 5 and 6 are the same, then we can calculate the ratio of the two denominators as

$$Ratio_{m.MT} = \frac{\max\limits_{t \in T}\left\{\frac{\text{Searches including "chink(s)"}_{mt}}{\text{Total searches}_{mt}}\right\}}{\max\limits_{m \in M}\left\{\frac{\text{Searches including "chink(s)"}_{mt}}{\text{Total searches}_{mt}}\right\}} = \frac{\text{Search Index}_{mt,M}}{\text{Search Index}_{mt,T}} \tag{7}$$

when both search indexes are non-zero. We can scale the time series search index over period T in each media market $m \in M$ by multiplying it with the corresponding $Ratio_{m,MT}$. The resulting time series are normalized by the same $\max\limits_{m \in M}\left\{\frac{\text{Searches including "chink(s)"}_{mt}}{\text{Total searches}_{mt}}\right\}$. However, Google Trends returns a zero value when the absolute level of search in a given media market

39

and time is below an unreported threshold, under which the rescaling does not work. After extracting cross-sectional search indexes on all possible weeks in the sample period, we can at best back out the rescaled search index for 35 media markets using Huntsville-Decatur (Florence) media market's search rate on March 15, 2020, as the base. Alternatively, we can back out 29 media markets using Wilkes Barre-Scranton media market's search rate on March 29, 2020, and 29 media markets using Buffalo media market's search rate on April 5, 2020, as the base. When combined, these three measures cover 50 media markets.

Note that Google calculates the search index using a random sample of searches, which can be different across extractions. As a result, the numerators in equations 5 and 6 are similar but may not be exactly the same. To the extent that these two numerators are not the same, we may be introducing measurement errors to the dependent variable and attenuating the main effects.

Appendix 2. Additional Figures & Tables
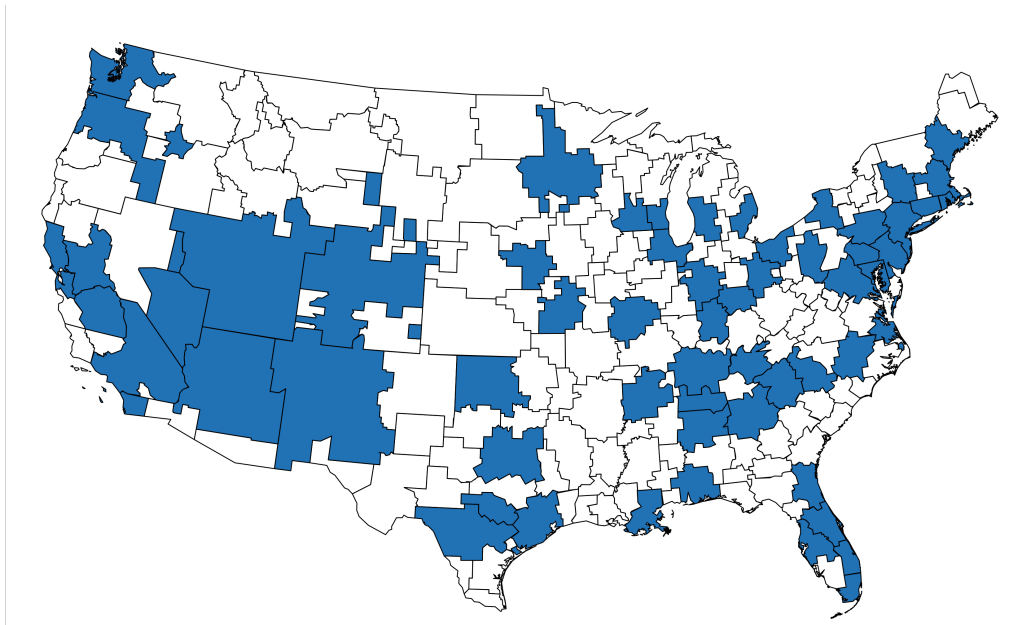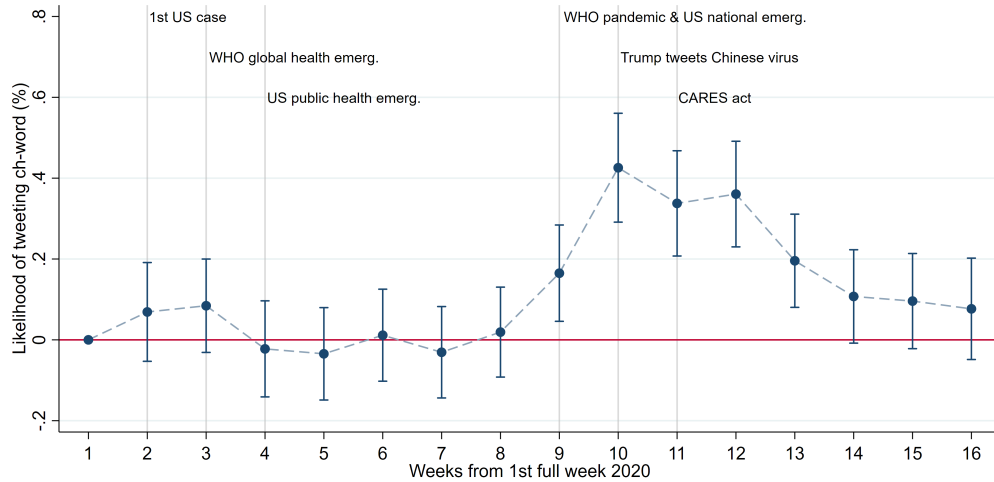

A: Google media market
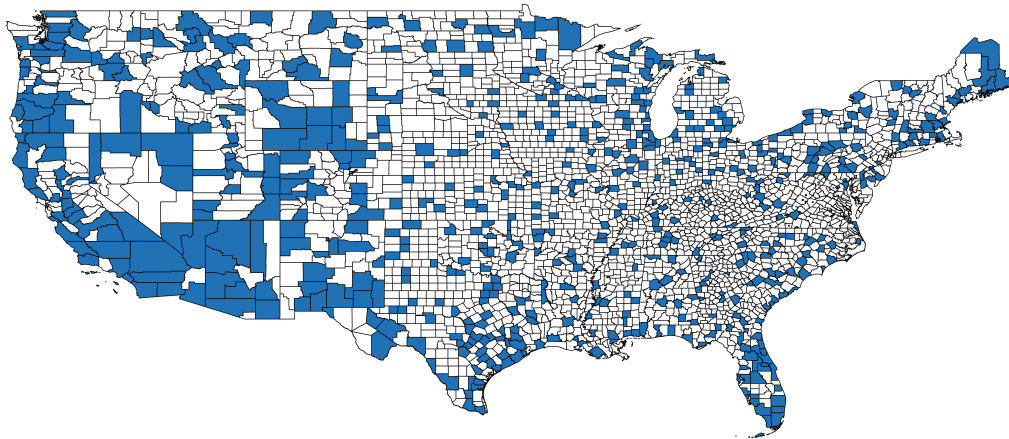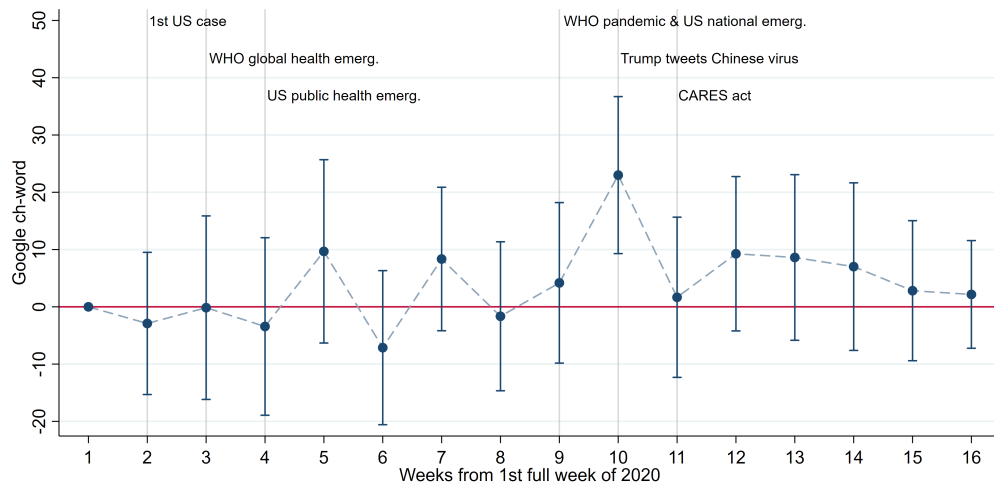

B: Twitter county

Figure A1: Location of Media Markets and Counties with Data

Note: The figure plots the locations of media markets with Google data (panel A) and counties with Twitter data (panel B).

A: Likelihood of tweeting the ch-word

B: Google search index

Figure A2: Timeline of COVID-19 Developments and Evolution of Racial Animus

Note: The figure presents the relationship between the timeline of important COVID-19 developments and the evolution of racial animus in the United States. Panels A and B plot the estimates and 95 percent confidence intervals of the coefficients on the event dummies in equation 2 using user's likelihood of tweeting the ch-word and the racially charged Google search index as the outcome, respectively. Regressions control for week-of-year fixed effects and user fixed effects (panel A) or media market fixed effects (panel B). Standard errors are clustered by user (panel A) or by media market (panel B).

A: Benchmark 3/15/2020

B: Benchmark 3/29/2020

C: Benchmark 4/5/2020
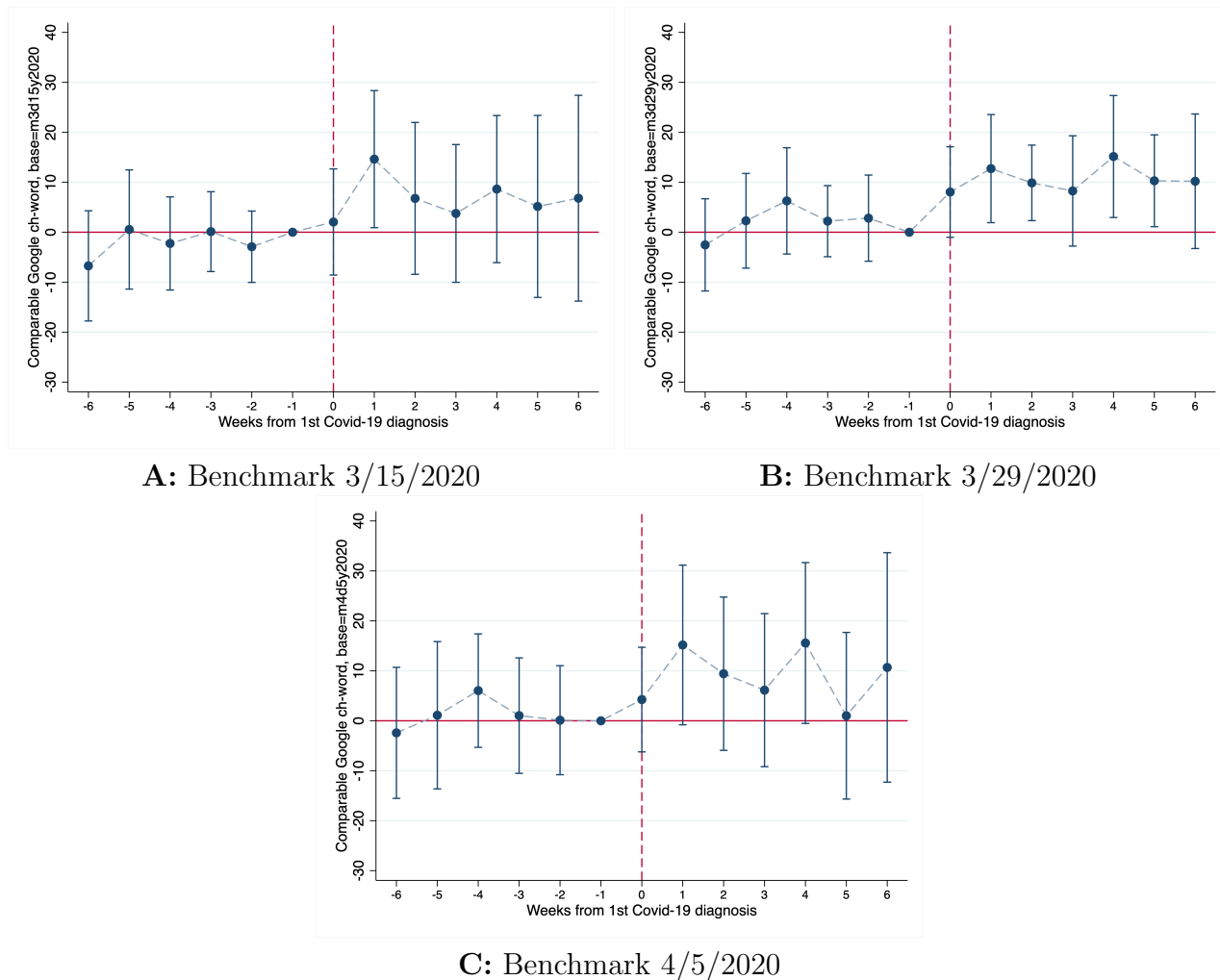
Figure A3: The Effect of the First Local COVID-19 Diagnosis on Racial Animus
Rescaled Google Search Index

Notes: The figure presents the effect of the first local COVID-19 diagnosis on various *rescaled* racially charged Google search indexes. Panels A, B, and C plot the estimates and 95 percent confidence intervals of the coefficients on the event dummies in equation 3 using an area's racially charged Google search rate scaled by Huntsville-Decatur (Florence) media market's search rate on March 15, 2020, by Wilkes Barre-Scranton media market's search rate on March 29, 2020, and by Buffalo media market's search rate on April 5, 2020 as the outcome, respectively. See Appendix 1 for the definitions of these indexes. Specifications mirror those in column (1) of Table 2.
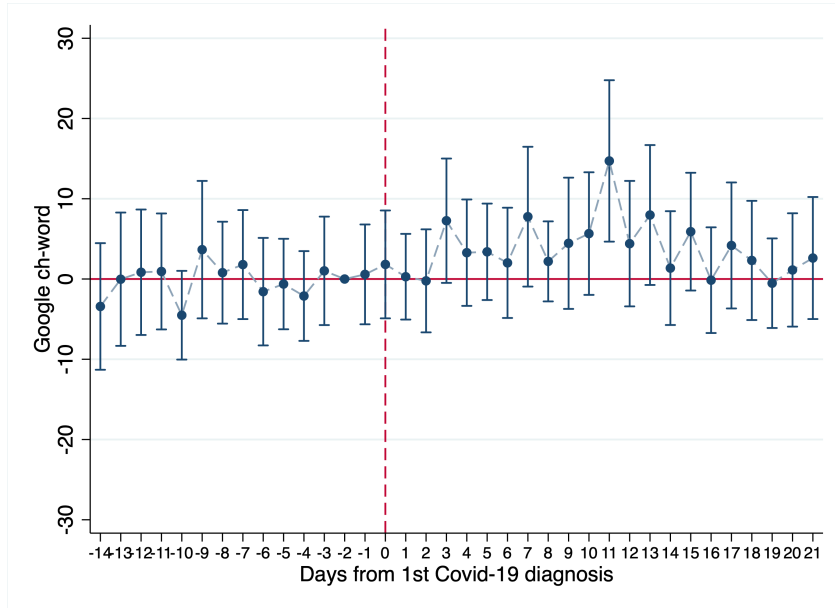
A: Google search index
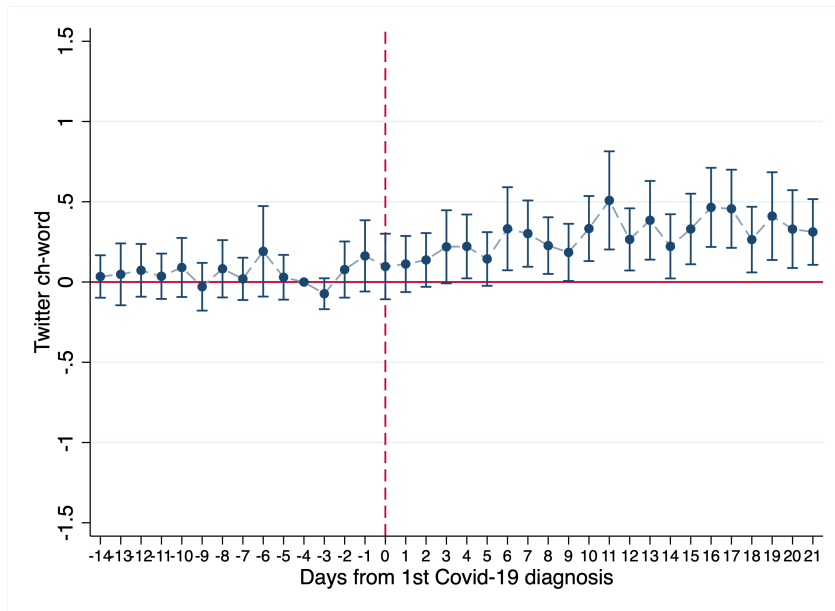
B: Twitter post index

Figure A4: The Effect of the First Local COVID-19 Diagnosis on *Daily* Racial Animus

Note: The figure presents the effect of the first local COVID-19 diagnosis on the *daily* racially charged Google search index and Twitter post index. Panels A and B plot the estimates and 95 percent confidence intervals of the coefficients on the event dummies in equation 3 using the Google search index and the Twitter post index as the outcome, respectively. Regressions control for year-month fixed effects, day-of-week fixed effects, and media market fixed effects (panel A) or county fixed effects (panel B). Standard errors are clustered by media market (panel A) or by county (panel B).
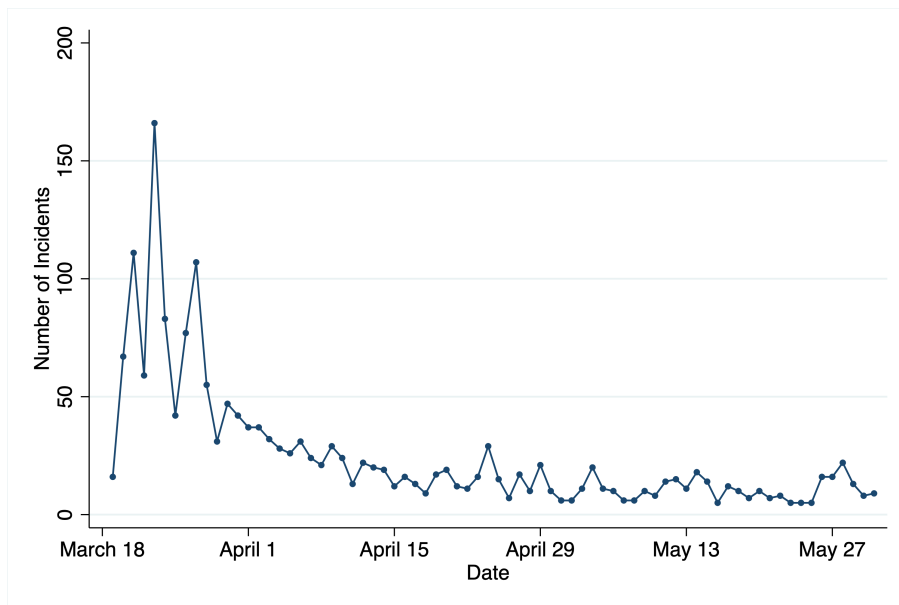
Figure A5: Self-Reported Hate Incidents in the United States

Notes: This figure presents the daily number of hate incidents from AP3CON Stop AAPI Hate Reporting system between March 19, 2020 (start of the data) and September, 2020.
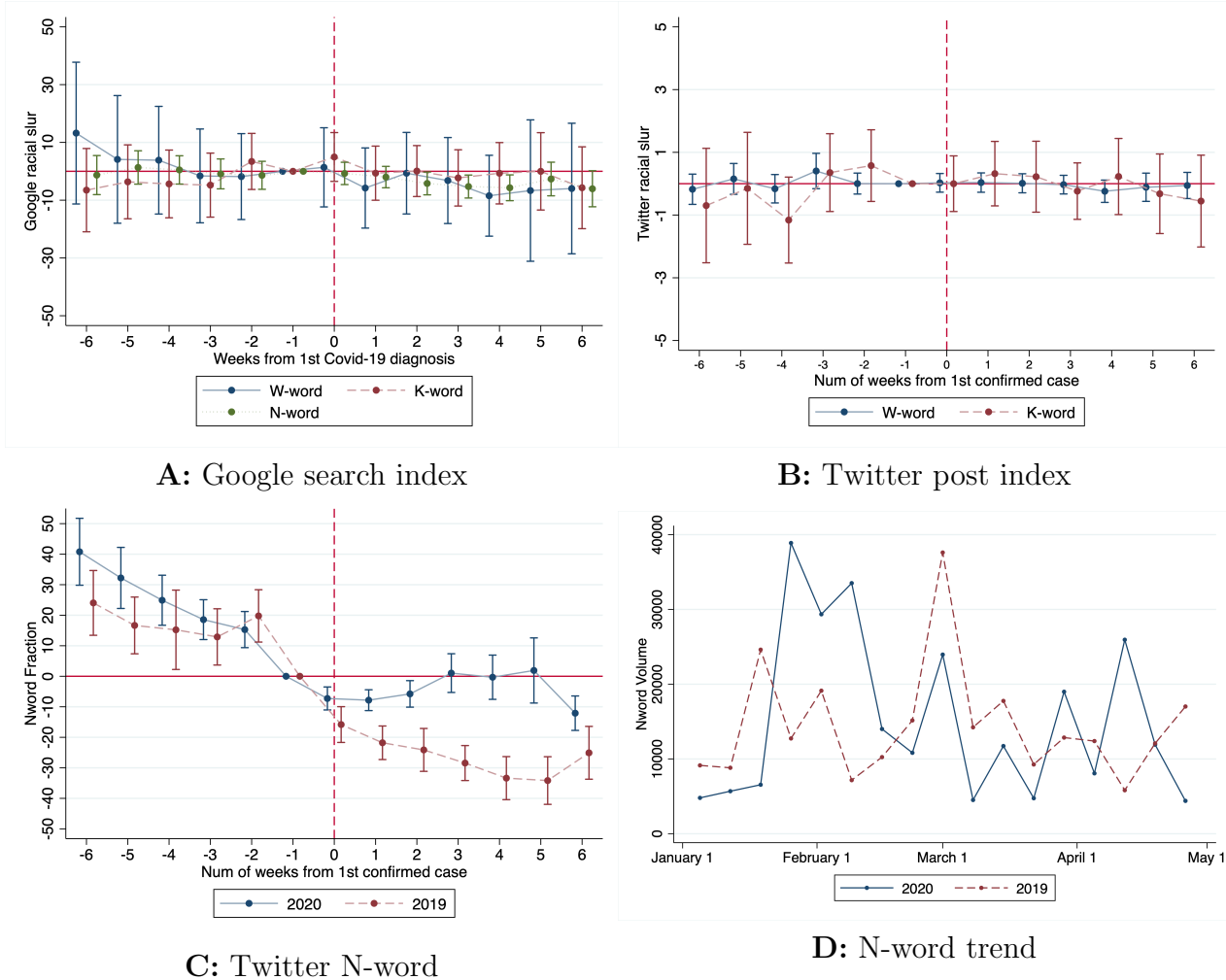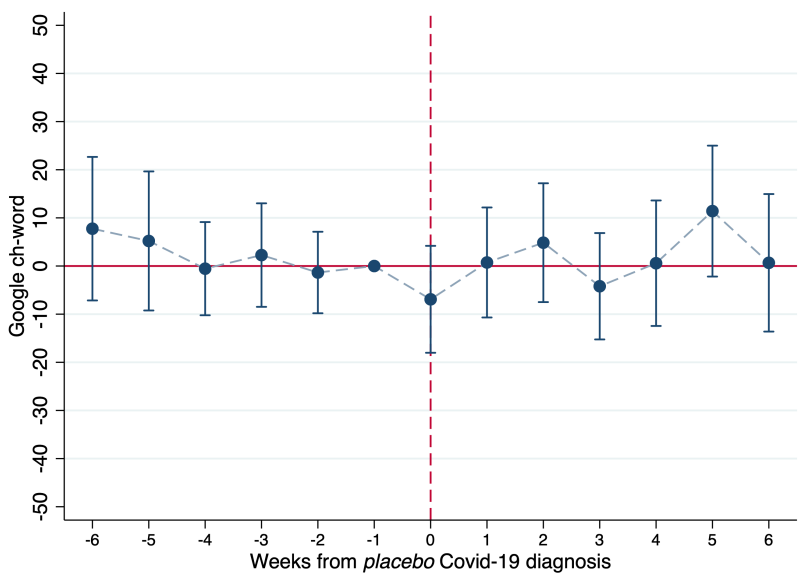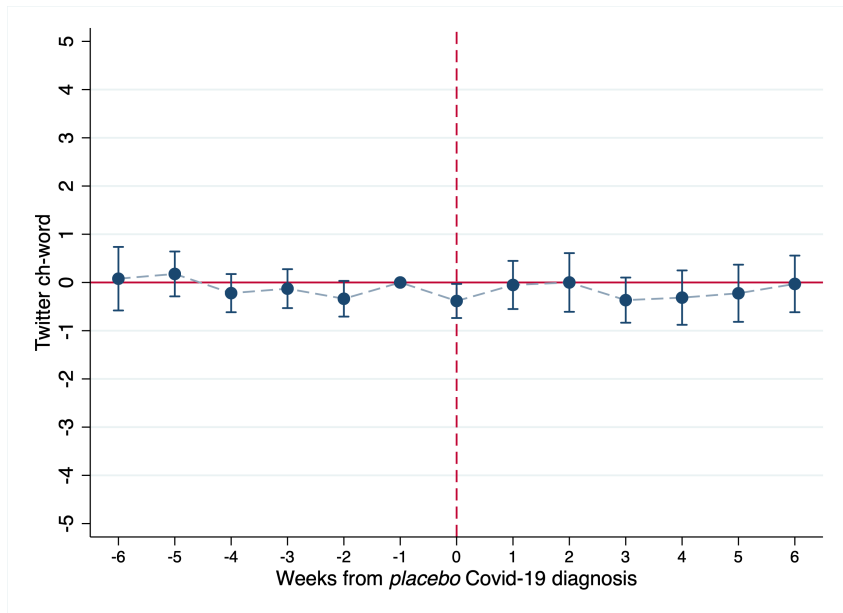
A: Google search index

B: Twitter post index

C: Twitter N-word

D: N-word trend

Figure A6: The Effect of the First Local COVID-19 Diagnosis on Racial Animus against *Non-Asian* Minorities

Notes: The figure presents the effect of the first local COVID-19 diagnosis on racial animus against the Hispanic, Jewish, and African American populations, using the Google search index and Twitter post index for "wetback(s)", "kike(s)", and the n-word as proxies. The indexes are defined following the method outlined in section 2.1. Regression samples for the n-word, k-word, and w-word Google search indexes contain 203, 78, and 27 media markets (panel A). Regression samples for the w-word and k-word Twitter post indexes contain 599 counties (panel B). Estimates of the coefficients and the 95 percent confidence intervals of the event dummies are from estimating equation 3 using the above indexes as outcomes. We include an indicator for the week of January 26, 2020 in the regression for the n-word to control for a spike in its use due to Kobe Bryant's death and MSNBC's anchor using the n-word while reporting the news. We include an indicator for the week of February 23, 2020 in the regression for the k-word to control for a spike in its use due to the Los Angeles Dodgers player Enrique ("Kiké") Hernandez's performance in that week. All other specifications in panels A and B mirror those in column (1) of Table 2 and column (1) of Table 3, respectively. Panel C plots the estimates and 95 percent confidence intervals of the coefficients on event dummies in equation 3 using the Twitter post index for the n-word between November 2019 and April 2020 (blue line) and that between November 2018 and April 2019 (red line) as the outcomes. For the regression using the 2018-2019 data, we replace the date of the first local COVID-19 diagnosis with a placebo date which shares the same day and month as the actual date in 2020 but with the year as 2019. For the regression using 2019-2020 data, we include an indicator for the week of January 26, 2020 to control for Kobe Bryant's death on January 26, 2020 and an indicator for the week of February 9, 2020 to control for an extremely viral video tweet *unrelated* to COVID-19 but mentioning the n-word on February 10, 2020. Panel D plots time trends for the Twitter post index for the n-word in 2020 (blue line) and in 2019 (red line).

46

A: Google search index

B: Twitter post index

Figure A7: The Effect of the First Local COVID-19 Diagnosis on Racial Animus
Placebo Test

Note: The figure presents a placebo test for the effect of the first local COVID-19 diagnosis on the racially charged Google search index and Twitter post index. We replace the date of the first local COVID-19 diagnosis with a placebo date using the same calendar day and month of the actual diagnosis date but changing the year from 2020 to 2019. Panels A and B plot the estimates and 95 percent confidence intervals of the coefficients on the event dummies in equation 3 using the Google search index and the Twitter post index as the outcome, respectively. Specifications in panels A and B mirror those in column (1) of Table 2 and column (1) of Table 3, respectively.

47
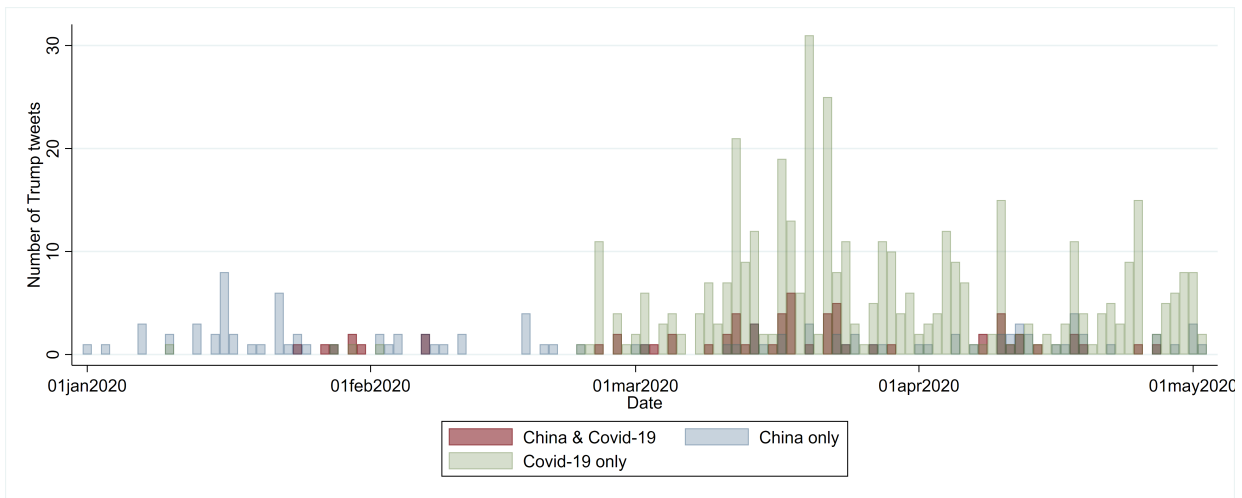
Figure A8: Number of President Trump's Tweets about China or COVID-19

Notes: This figure plots the number of President Trump's tweets by category on each day between January 1, 2020 and May 2, 2020. We categorize the president's tweets that include "china", "chinese", "huawei", "xi", "COVID", "COVID-19", "corona", "coronavirus", "virus", "epidemic", or "pandemic" into three categories: those mentioning only China (China only), only COVID-19 (COVID only), and both China and COVID-19 (China-and-COVID).

Table A1: Sample Selection - Media Markets and Counties with Google and Twitter data

VARIABLES	(1) Google sample	(2) Twitter data	(3) Twitter sample
Log(pop)	0.193***	0.142***	0.142***
	(0.036)	(0.007)	(0.007)
% Asian	0.059**	0.007	0.007
	(0.030)	(0.009)	(0.009)
% Asian^2	-0.002**	-0.001**	-0.001**
	(0.001)	(0.000)	(0.000)
% Male	0.001	-0.001	-0.001
	(0.037)	(0.002)	(0.002)
% 65+	-0.013	-0.002	-0.003
	(0.015)	(0.002)	(0.002)
% BA+	0.009	0.002	0.002*
	(0.007)	(0.001)	(0.001)
% Unemp	-0.003	0.002	0.002
	(0.020)	(0.005)	(0.005)
% Vote share dem-rep	-0.001	0.001***	0.001***
	(0.002)	(0.000)	(0.000)
Hate crime/1m	-0.087***	0.001	0.001
	(0.026)	(0.005)	(0.005)
Intl airport enplanement	0.006*	0.006**	0.006**
	(0.003)	(0.002)	(0.002)
Observations	205	3,111	3,111
R-squared	0.695	0.341	0.343
Outcome mean	0.292	0.205	0.202

Notes: The table presents the sample selection in Google and Twitter data. The data are at the media market level in column (1) and at the county level in columns (2) and (3). The outcome is an indicator for having Google data in column (1), an indicator for having Twitter data in column (2), and an indicator for being in the Twitter regression sample in column (3). Note that all media markets with Google data are in the Google regression sample. "Log(pop)" is the natural log of local population estimates in 2018 from Census Bureau. "%Asian", "% Male", "% 65+", and "% BA+" are the percentage of Asians, males, population 65 years old or over, and population with Bachelor's or above degree in the local area from American Community Survey 2014-2018 five-year average. "%Unemp" is the average monthly local unemployment rate between 2014 and 2018 from the BLS. "% Vote share Dem-Rep" is the difference between the Democratic and the Republican vote shares in the 2012 presidential election from Harvard Dataverse. "Hate crime/1m" is the average annual number of anti-Asian hate crimes per million population between 2014-2018 from UCR. "Intl airport enplanement" is the international airport enplanements in 2016 according to the Federal Aviation Administration. The number of media markets and counties is less than 210 and 3141 due to missing covariates. All regressions control for state fixed effects. Standard errors in parentheses are clustered by media market in column (1) and by county in columns (2) and (3).
** $p <0.01$, ** $p <0.05$, * $p <0.1$.

Table A2: Summary Statistics

VARIABLE	(1) Google Sample	(2) Twitter Sample
Ch-word index		
Weekly	25.213	0.591
	(29.008)	(2.572)
Daily	6.965	0.309
	(20.316)	(2.215)
Hourly	-	0.357
	-	(0.676)
Other indexes (weekly)		
N-word	31.781	33.448
	(25.042)	(65.104)
K-word	34.183	2.970
	(26.750)	(12.705)
W-word	29.452	0.436
	(29.702)	(2.708)
Asian(s)	79.305	138.026
	(10.815)	(199.423)
Other animus measures		
Anti-Asian hate crime/1m	0.037	0.003
	(0.099)	(0.036)
Chinese restaurant visits/1m	26353	23846
	(13148)	(12328)
Total restaurant visits/1m	698341	606620
	(189770)	(220331)
Geographic unit	Media market	County
Unique geo-units	60	641

Notes: The table presents summary statistics for our main regression samples. See section 2.1 for the definitions of Google search index and Twitter post index. "Anti-Asian hate crime/1m" is the monthly anti-Asian hate crimes per million population in a media market between January 2014 and December 2018. "Chinese (or total) restaurant visits/1m" is the monthly visits to Chinese (or all) restaurants per million population in a median market between January 2018 and December 2019. All other variables are measured at the media market×time level in column (1) and at the county×time level in column (2).

Table A3: Timing of the First Local COVID-19 Diagnosis - Weeks from Jan 19, 2020

VARIABLES	(1) Google sample Weeks from Jan192020	(2) Twitter sample Weeks from Jan192020
Log(pop)	-1.499***	-0.673***
	(0.474)	(0.053)
% Asian	0.156	-0.018
	(0.212)	(0.038)
% Asian^2	-0.004	-0.001
	(0.006)	(0.001)
% Male	-1.158***	0.019
	(0.414)	(0.033)
% 65+	-0.031	0.004
	(0.067)	(0.014)
% BA+	0.048	-0.005
	(0.040)	(0.007)
% Unemp	0.623***	0.061
	(0.212)	(0.044)
% VS dem-rep	-0.022**	-0.000
	(0.009)	(0.002)
Hate crime/1m	-0.850	-0.023
	(0.790)	(0.034)
Intl airport enplanement	0.035**	-0.028
	(0.015)	(0.017)
Observations	60	630
R-squared	0.984	0.646
Outcome mean	5.983	8.12

Notes: The table presents the relationship between the timing of the first local COVID-19 diagnosis and the characteristics of the local area. The data are at the media market level in column (1) and at the county level in column (2). The outcome is the number of weeks from the week of the first diagnosis in the United States, i.e., the week of January 19, 2020. See note to Table A1 for variable definitions. The number of observation in column (2) is smaller than 641 due to missing covariates. All regressions control for state fixed effects. Standard errors are clustered by media market in column (1) or by county in column (2).
*** $p <0.01$, ** $p <0.05$, * $p <0.1$.

Table A4: Number of Media Markets and Counties by Week of the First COVID-19 Local Diagnoses

Date of Sunday	(1) Media markets	(2) Counties
Jan 19, 2020	2	
Jan 26, 2020	4	5
Feb 9, 2020	1	1
Feb 16, 2020	1	2
Feb 23, 2020	1	1
Mar 1, 2020	20	28
Mar 8, 2020	31	148
Mar 15, 2020		229
Mar 22, 2020		139
Mar 29, 2020		58
Apr 05, 2020		16
Apr 12, 2020		7
Apr 19, 2020		4
Total	60	641

Notes: The table presents the number of media markets and counties in our main regression samples by the week of their first local COVID-19 diagnoses.

Table A5: Characteristics of First-Time Ch-word Users and Control Users

	Ch-word users		Control users	
	Mean	SD	Mean	SD
Panel A: User characteristics				
Account years	6.309	3.652	5.708	3.220
Followers	4103.629	57537.343	1330.260	9672.603
Followings	1440.742	3509.394	788.713	1920.440
Panel B: Prob. (re)tweet/reply/mention				
During pandemic:				
COVID conspiracy	0.017	0.131	0.004	0.063
Before pandemic:				
Anti-minority content	0.164	0.371	0.011	0.104
Anti-Asian user	0.866	0.341	0.228	0.419
Trump	0.561	0.496	0.157	0.364
McCarthy	0.137	0.344	0.004	0.063
McConnell	0.024	0.154	0.002	0.045
Pelosi	0.227	0.419	0.015	0.122
Schumer	0.178	0.383	0.010	0.101
Fox	0.282	0.450	0.027	0.161
CNN	0.388	0.487	0.045	0.207
CBS	0.115	0.319	0.009	0.094
N users	3,033		3,000	

Notes: This table presents the characteristics of first-time ch-word users and control users. Panel A reports information from Twitter user profiles. Panel B reports a user's likelihood of mentioning certain keywords in their tweets or interacting with certain users. "During pandemic" and "Before pandemic" refer to the period between January 21, 2020 and May 2, 2020 and that before January 21, 2020, respectively. See note to Figure 4 for definitions of the remaining variables.

Table A6: Predictors of First-time Ch-word Users - Twitter Activity

VARIABLES	(1) First-time ch-word user
Anti-Asian user	0.558***
	(0.013)
Anti-minority	0.200***
	(0.012)
COVID consp.	0.004
	(0.046)
Trump	0.055***
	(0.014)
McCarthy	0.056***
	(0.016)
McConnell	-0.078**
	(0.031)
Pelosi	0.050***
	(0.019)
Schumer	0.010
	(0.020)
CBS	0.039**
	(0.017)
CNN	0.123***
	(0.018)
Fox	0.042**
	(0.019)
Account years	-0.017***
	(0.002)
Log(followers)	-0.001
	(0.004)
Log(followings)	-0.022***
	(0.005)
Observations	6,033
R-squared	0.465
Outcome mean	.502

Notes: This table reports the regression coefficients plotted in Figure 4 panel A. See note to the figure for variable definitions and regression specifications.

*** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Table A7: Predictors of First-time Ch-word Users - User Profile Keywords

VARIABLES	(1) First-time ch-word user
artist	-0.171***
	(0.043)
business	-0.088*
	(0.052)
can	0.085*
	(0.050)
dad	-0.028
	(0.053)
enthusiast	-0.084
	(0.054)
fan	-0.087**
	(0.034)
father	-0.057
	(0.051)
games	-0.247***
	(0.054)
good	0.057
	(0.052)
hehim	0.034
	(0.038)
husband	-0.096*
	(0.054)
just	-0.043
	(0.035)
life	-0.058
	(0.036)
like	-0.067
	(0.045)
love	-0.060*
	(0.032)
lover	-0.092**
	(0.041)
mom	-0.055
	(0.046)
music	-0.044
	(0.044)
new	0.142***
	(0.046)
one	-0.038
	(0.053)
opinions	0.011
	(0.051)
politics	0.143***
	(0.051)
proud	-0.009
	(0.050)
retired	-0.018
	(0.050)
sheher	0.025
	(0.037)
sports	-0.076
	(0.049)
things	0.012
	(0.047)
time	-0.088*
	(0.048)
trump	0.291***
	(0.049)
twitch	-0.311***
	(0.041)
wife	-0.163***
	(0.049)
world	-0.018
	(0.048)
writer	0.008
	(0.042)
Account years	0.009***
	(0.002)
Log(followers)	0.036***
	(0.005)
Log(followings)	0.020***
	(0.007)
Observations	5,266
R-squared	0.072
Outcome mean	.502

Notes: This table reports the regression coefficients plotted in Figure 4 panel B. See note to the figure for variable definitions and regression specifications.

*** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Table A8: Examples of President Trump's Tweets about China or COVID-19

Category	Post	Date
China-only	"Years from now, when we look back at this day, nobody's going to remember nancy's cheap theatrics, they will remember though how president trump brought the Chinese to the bargaining table and delivered achievements few ever thought were possible." @ingrahamangle @foxnews	1/17/20
China-only	The Wall Street Journal editorial board doesn't have a clue on how to fight and win. Their views on tariffs & trade are losers for the U.S., but winners for other countries, including China. If we followed their standards, we'd have no country left. They should love sleepy joe!	4/11/20
COVID-only	The coronavirus is very much under control in the USA. we are in contact with everyone and all relevant countries. CDC & World Health have been working hard and very smart. Stock market starting to look very good to me!	2/24/20
COVID-only	I am fully prepared to use the full power of the federal government to deal with our current challenge of the coronavirus!	3/11/20
China-and-COVID	Just received a briefing on the Coronavirus in china from all of our great agencies, who are also working closely with china. we will continue to monitor the ongoing developments. We have the best experts anywhere in the world, and they are on top of it 24/7!	1/30/20
China-and-COVID	I will be having a news conference today to discuss very important news from the FDA concerning the Chinese Virus!	3/18/20
China-and-COVID	Just finished a very good conversation with President Xi of China. Discussed in great detail the Coronavirus that is ravaging large parts of our planet. China has been through much & has developed a strong understanding of the virus. We are working closely together. Much respect!	3/22/20

Notes: This table presents examples of President Trump's tweets mentioning China and/or COVID-19. We manually categorize all President Trump's tweets between January 1, 2020 and May 2, 2020 that contain any of the words "china", "chinese", "huawei", "xi", "covid", "covid-19", "corona", "coronavirus", "virus", "epidemic", or "pandemic" into three categories: those mentioning only China (China-only), only COVID-19 (COVID-only), and both China and COVID-19 (China-and-COVID).

Table A9: Relationship between Racial Animus Nationwide and Tweets from Politicians and National News Outlets

VARIABLES	(1) McCarthy	(2) McConnell	(3) Pelosi	(4) Schumer	(5) CBS	(6) CNN	(7) Fox
Panel A: Twitter post index							
China-and-COVID(t)	0.0328				-0.0131	0.0278*	0.0283
	(0.0594)				(0.0237)	(0.0155)	(0.1858)
China only(t)	0.0076			0.0446	-0.0203	0.0027	
	(0.0201)			(0.0391)	(0.0205)	(0.0132)	
Covid only(t)	0.0170	-0.0213*	-0.0087	-0.0304	0.0063***	-0.0012	0.0235
	(0.0109)	(0.0128)	(0.0226)	(0.0236)	(0.0018)	(0.0015)	(0.0223)
Observations	123	123	123	123	123	123	123
R-squared	0.5118	0.5088	0.4916	0.5057	0.5383	0.5090	0.5074
Outcome mean	.3445	.3445	.3445	.3445	.3445	.3445	.3445
Panel B: Log(hate incidents)							
China-and-COVID(t)	-0.5740*				-0.0175	-0.1768	0.2794
	(0.3287)				(0.1326)	(0.1365)	(0.2438)
China only(t)	-0.0090				0.2764	0.4743**	
	(0.0734)				(0.1714)	(0.2041)	
Covid only(t)	0.0426	-0.0081	0.0077	-0.0745	0.0031	0.0075	-0.0310
	(0.0553)	(0.0265)	(0.1218)	(0.0932)	(0.0052)	(0.0059)	(0.0690)
Observations	45	45	45	45	45	45	45
R-squared	0.8522	0.8198	0.8196	0.8236	0.8348	0.8611	0.8264
Outcome mean	3.1932	3.1932	3.1932	3.1932	3.1932	3.1932	3.1932

Notes: The table presents the relationship between the number of the tweets of major politician and news outlets about COVID-19 or China and racial animus nationwide. The outcome variable in panel A is the daily number of ch-word tweets per 100,000 "the" tweets nationwide between January 1, 2020 and May 2, 2020. The outcome variable in panel B is the natural log of the daily number of anti-Asian hate incidents nationwide from AP3CON Stop AAPI Hate Reporting system between March 19 and May 2, 2020. See note to Table 5 for definitions of the independent variables. All regressions control for year-week fixed effects and day-of-week fixed effects. Standard errors are clustered by date.
** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.