

# Social Media, Public Sentiment, and Social Distancing: Evidence from 10 Million Tweets

George Tyler

January 13, 2021

## Abstract

Does social media predict risk-taking behaviour? I investigate this question in the context of COVID-19 by exploiting a large panel of tweets. Using inferred and explicit geolocation data embedded in the tweets, I study the extent to which public expressions of sentiment such as fear, anger, and optimism influence social distancing, as measured by GPS-located smartphone data. In this 2000-word excerpt, I motivate the research design, conduct a literature review, and present descriptive statistics on a portion of the dataset.

## 1 Introduction

### 1.1 Overview

The early stages of the COVID-19 pandemic saw an unprecedented shift in behaviour for most citizens of the United States. In a short period of time, a large number changed their habits of working, socialising, and travelling. They did so both as a result of government restrictions in the form of non-pharmaceutical interventions (NPIs) and as a private response to the spread of the pandemic. Economists have taken interest in how citizens formed these behaviour changes, and the role that beliefs and risk attitudes played in determining the response to public policy. A new way to measure belief formation and public sentiment is with social media, an increasingly common platform for expression of opinion. It is plausible that those who express less risk-averse sentiment towards COVID online will be inclined to respond in a more lax fashion to social distancing and other public health regulation. In this dissertation, I investigate the role of online expressions of risk attitude on public behaviour in the early months of the pandemic. Specifically, I study whether a measure of risk-seeking sentiment on Twitter is linked to decreased social distancing behaviour, aggregated at the county/week level.

A key vector for expressing sentiment is social media, with Twitter and Facebook’s suite of products<sup>1</sup> being the most widely-adopted, each platform having over 80 million monthly active users in the US. A survey by the Pew Research Foundation indicates that 22% of US adults use Twitter, with 42% of these using it on a daily basis (Perrin & Anderson, 2019). On Twitter, users can share their own text, with the option to link to a website; alternatively, they can ‘retweet’ another user’s text or link. Users can also use ‘hashtags’ in their tweet, which connects their tweet to a particular topic. If the user has allowed it, Twitter also records the location of the tweet; and it is also possible for the user to set their location on their profile. In this way, it is possible to create a panel of geographically-located tweets about a particular topic.

I exploit GeoCov19, a dataset of 524 million geolocated tweets, to measure the local public sentiment on COVID in the US. The tweets cover the period from 1st February to 1st May, the period I focus on. The particular subset of the data I use contains X tweets in total; X are exactly geolocated (the user has provided a GPS location), and X are inferred from the location tab in the user’s profile. The tweets were collected using the Twitter API, querying for a random sample of 1% of the tweets containing any of a list of 800 COVID-related keywords. I also use anonymous smartphone location data, collected by the company SafeGraph, as a measure of the extent of social distancing in an area. I present two measures

---

<sup>1</sup>Facebook, Facebook Messenger, Instagram, and WhatsApp

of social distancing at county level: first, the median minutes spent at home during 8am-6pm; second, the proportion of measured devices that stayed at home all day (SafeGraph, Inc., 2020). Demographic controls are also acquired and presented from the American Community Survey and the 2010 US census.

I use dictionary-based text analysis to assess the level of risk sentiment in a tweet. In the absence of a lexicon of risk preference, the NRC Emotion Lexicon (Mohammad & Turney, 2013) is used. This is a widely-used mapping of English words to eight basic emotions (anger, fear, anticipation, trust, surprise, sadness, joy, and disgust). Starting from a set of tweets that mention COVID, I assign tweets containing fear-associated words to a risk-averse sentiment. Other more sophisticated methods of text analysis, and their use in Economics, are discussed in the Methods section.

The justification for using social media as a measure for risk appetite is as follows. The textual content of social media posts broadly reflects the poster’s opinion of a topic: for example, in response to the January 31st news of a China-US travel ban, a Twitter user may use positive

This research contributes to the recent economics literature seeking to explain the disparities in social distancing in the early stages of the pandemic in the US. In particular, partisanship has been shown to be a significant factor on the practice of social distancing: Allcott et al. (2020), Barrios and Hochberg (2020), and Painter and Qiu (2020) show that areas with more Republicans engaged in less social distancing, are associated with lower perceptions of risk of the pandemic, and exhibited less remote transactions. Ananyev et al. (2020) and Simonov et al. (2020) also measure the causal effect of the right-wing Fox News network on social distancing during the pandemic. This paper builds on Barrios and Hochberg (2020) in particular: by measuring risk perception,

This paper takes a similar approach to Stephens-Davidowitz (2014), which uses Google search data to proxy an area’s racial animus, and uses this to estimate the Obama vote share. I use geolocated Twitter sentiment to proxy the local attitude to COVID in a given week, and test to see if this predicts social distancing practice.

This dissertation also contributes to the rapidly-expanding field of text analysis in economics, and presents an example of how rich sentiment data encoded in social media communication can inform insights into public behaviour. This topic is mature in finance, where social media sentiment is used to predict stock market reactions, and central bank statements coded according to their dovishness or hawkishness.

The argument of the dissertation rests on the following assumptions: first, that social media data is a valid proxy for local risk appetite, and that fear-associated language in COVID-related tweets is an effective estimator of the risk appetite encoded in the tweet. These assumptions, including methods to further address them, are discussed in the dissertation.

## 2 Literature Review

### 2.1 Misinformation and media bias

### 2.2 Digital trace datasets

### 2.3 COVID-19 policies and social distancing

## 3 Data

## 4 Methods

The primary specification is the following:

$$\log(y_{it}) = \alpha_{it} s_{it} + \mu_i + \gamma_t + X_{it}$$

## 5 Results

## 6 Discussion

1006 words in main body, excluding headers and bibliography.

## References

- Allcott, H., Boxell, L., Conway, J., Gentzkow, M., Thaler, M., & Yang, D. (2020, April). *Polarization and Public Health: Partisan Differences in Social Distancing during the Coronavirus Pandemic* (Working Paper No. 26946). National Bureau of Economic Research. Cambridge, MA. (Cit. on p. 2).
- Ananyev, M., Poyker, M., & Tian, Y. (2020). The safest time to fly: Pandemic response in the era of Fox News. *Covid Economics*, (49) (cit. on p. 2).
- Barrios, J., & Hochberg, Y. (2020, April). *Risk Perception Through the Lens of Politics in the Time of the COVID-19 Pandemic* (Working Paper No. 27008). National Bureau of Economic Research. Cambridge, MA. Retrieved January 8, 2021, from <https://www.nber.org/system/files/working-papers/w27008/w27008.pdf>. (Cit. on p. 2)
- Mohammad, S. M., & Turney, P. D. (2013). Crowdsourcing a Word–Emotion Association Lexicon. *Computational Intelligence*, 29(3), 436–465. Retrieved January 13, 2021, from <https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1467-8640.2012.00460.x> (cit. on p. 2) eprint: <https://onlinelibrary.wiley.com/doi/pdf/10.1111/j.1467-8640.2012.00460.x>
- Painter, M., & Qiu, T. (2020). Political Beliefs affect Compliance with COVID-19 Social Distancing Orders. *SSRN Electronic Journal*. Retrieved January 8, 2021, from <https://www.ssrn.com/abstract=3569098> (cit. on p. 2)
- Perrin, A., & Anderson, M. (2019, April 10). *Share of U.S. adults using social media, including Facebook, is mostly unchanged since 2018*. Pew Research Center. Washington, DC. Retrieved January 6, 2021, from [https://www.pewresearch.org/wp-content/uploads/2019/04/FT\\_19.04.10.SocialMedia2019\\_topline\\_methodology.pdf](https://www.pewresearch.org/wp-content/uploads/2019/04/FT_19.04.10.SocialMedia2019_topline_methodology.pdf). (Cit. on p. 1)
- SafeGraph, Inc. (2020). *Social Distancing Metrics*. SafeGraph. Retrieved January 7, 2021, from <https://docs.safegraph.com/docs/social-distancing-metrics>. (Cit. on p. 2)
- Simonov, A., Sacher, S., Dubé, J.-P., & Biswas, S. (2020). *The Persuasive Effect of Fox News: Non-Compliance with Social Distancing During the COVID-19 Pandemic* (Working Paper No. 27237). National Bureau of Economic Research. Cambridge, MA. Retrieved January 7, 2021, from <http://media.mediapost.com.s3.amazonaws.com/uploads/FoxNewsPaper.pdf>. (Cit. on p. 2)
- Stephens-Davidowitz, S. (2014). The cost of racial animus on a black candidate: Evidence using Google search data. *Journal of Public Economics*, 118, 26–40. Retrieved January 12, 2021, from <https://linkinghub.elsevier.com/retrieve/pii/S0047272714000929> (cit. on p. 2)