Social Media, Risk Perception, and Social Distancing: Evidence from 15.3 Million Geolocated Tweets

George Tyler

March 16, 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 influence social distancing, as measured by GPS-located smartphone data.

`9043 words in main body, excluding headers and bibliography.`

Contents

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 more risk-averse sentiment towards COVID online will be inclined to respond in a stricter fashion to social distancing and other public health regulations. In this dissertation, I study the impact of local expressions of risk attitude on public behaviour in the early months of the pandemic. Specifically, I study whether a measure of risk-averse sentiment on Twitter is linked to increased social distancing behaviour at the county/week level.

This dissertation contributes to two strands of the recent economics literature on the COVID-19 pandemic. First, it investigates the relationship between partisanship and risk preference. Previous papers posit that political preference influences social distancing through risk preference; I contend that *local* risk preference is a separate factor to political partisanship, and has an independent impact on social distancing. More broadly, the paper investigates the relationship between risk preference and economic behaviour, and presents a novel example of economic inference from social media using text analysis.

A key vector for expressing sentiment is social media, with Twitter and Facebook's suite of products[1] 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 (Qazi et al., 2020), 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 33.36 million tweets in total; a small subset are exactly geolocated (the user has provided a GPS location), while most are inferred from the location tab in the user's profile. The tweets were collected using the Twitter Streaming API, querying for 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 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.

[1] Facebook, Facebook Messenger, Instagram, and WhatsApp

I use dictionary-based text analysis to assess the level of risk sentiment in a tweet. More sophisticated methods of text analysis like latent factor modelling are discussed in the Methods section. 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. The base unit of analysis is the county-week; as such, I measure the proportion of tweets that contain fearful language in each county and week.

It is plausible that social media is a valid measure for risk appetite. The intuition is that the textual content of a social media post broadly reflects the poster's current opinion of a topic: for example, in response to the first confirmed US COVID death on February 26th, a user might express fearful sentiment, or a neutral sentiment. This opinion of the topic, particularly their level of fear, maps to a user's broader expectations about the course of the pandemic: while other emotions like joy, anticipation, and trust may rely on the context of the discussion, expressions of fear are plausibly consistent in mapping to risk-averse sentiment. When restrictions are implemented, users who initially formed pessimistic expectations may be more inclined to adhere more to them than a user who formed optimistic or netural expectations. The key aspect to the data is that the Twitter conversations provide a real-time insight into local sentiment as NPIs are implemented; this sentiment changes both in response to current, local experience *and* broader partisan interpretations of current events.

The primary econometric specification is a panel model with county and week fixed effects;

$$Y_{it} = \alpha + \beta r_{it} + \mu c_{it} + \tau_i + \delta_t + X_{it}\gamma + \epsilon_{it}$$

where Y_{it} is a vector of social distancing metrics, βr_{it} is the risk perception measure, (i.e. the proportion of total tweets containing fearful language), μc_{it} the number of COVID cases, $\tau_i + \delta_t$ county and week-level fixed effects, and $X_{it}\gamma$ demographic controls.

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, which shows that online risk perception is predicted by Trump voter share: by measuring risk perception with a high-frequency geolocated dataset, my approach controls for political alignment and assesses the effect of risk perceptions on their own. In essence, I measure expressions of sentiment regarding COVID risk, and given this data I ask whether local risk sentiment predicts social distancing behaviour beyond political affilitation. Second, this research relates to the recent economics literature around heterogeneous-agent epidemiological models, which endogenise individual behaviour – including social distancing – into the effective reproductive number $R(t)$. These recent models, such as Acemoglu, Makhdoumi, et al. (2020), Brotherhood et al. (2020), and Eichenbaum et al. (2020), assume that preferences over risk are predictive of social distancing behaviour; this paper looks to empirically confirm this key assumption.

This dissertation also contributes to the rapidly-expanding field of text analysis in economics, and presents an example of how the rich sentiment data encoded in social media communication can inform insights into public behaviour. This topic is particularly mature in finance – where sentiment data from public company documents, news media, and social media have been shown to predict stock market reactions (Bollen et al., 2011) – and monetary economics, where central bank statements, coded according to their attitude to inflation, predict fluctuations in Treasury securities (Gentzkow et al., 2019; Lucca & Trebbi, 2009). On the topic of empirical economics, this paper takes a similar approach – by using online data to predict local sentiment – as 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.

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. It is also important to note a possible selection effect in the dataset: tweets about COVID may attract a greater level of fear-related language and not reflect an individual's true opinion about social distancing and other preventative measures. I address these assumptions and drawbacks and discuss methods to alleviate them in the Results section.

2 Literature Review

2.1 Text analysis and sentiment mining in Economics

Context and applications of Text Analysis in economics; technical section describing the various methods used in the literature is in the methods section.

2.2 Usage of 'Digital Trace' and geolocation datasets in economics

General; description of previous uses of geolocation data

2.3 COVID-19: Empirical estimates of drivers of social distancing behaviour

Summary of empirical research on COVID The primary aim of this dissertation is to investigate the role of anxiety and risk preference in determining social distancing behaviour. There has been a large quantity of work done to investigate the determinants of social distancing, both on an empirical basis and a modelling basis. To inform my econometric specification, it is necessary to account for all possible determinants and causal pathways around risky COVID behaviour. As such, the next two sections summarise the literature on behavioural determinants, starting by addressing the empirical work.

Glaeser et al. (2020) documents the variations in SafeGraph mobility across the United States and estimates the effect of mobility reduction in moderating COVID spread. They find that mobility reduction significantly decreases COVID cases, and that the strength of the effect is heterogeneous across different locations. They find a clear importance of mobility reduction to reducing COVID spread. Several factors affect the ability to reduce mobility, and often these differentiate across income. For example, Wright et al. (2020) shows that areas with low economic endowments complied less with stay-at-home orders, and a one-time cash transfer significantly increased social distancing in these areas. Another factor affecting the ability to socially distance is internet access: Chiou and Tucker (2020) show that high-speed internet access accounts for much of the income effect on mobility changes. The correlation between income and internet access shows that ability to self-isolate not only depends on the sector of employment, but also the ability to work remotely in general.

A central issue to the dissertation is whether observed mobility changes are a result of policy or a function of individual decisions. The general consensus in the economics literature is that mobility changes pre-date stay-at-home orders, meaning that they are largely attributable to individual responses. Establishing this is a key factor in the identification of fear as a significant factor: since individual responses are important to mobility, it is important to understand and measure the drivers of individual mobility preferences, like anxiety.

Goolsbee and Syverson (2021) is one of a number of papers that addresses the question of mobility as a function of policy or individual action, and is also a broader overview of the empirical factors of social distancing behaviour. They isolate the difference between *fear-driven* voluntary social distancing and policy-driven social distancing by using SafeGraph footfall data on businesses close to a jurisdictional border, where there is policy variation across the border. They find that legal orders only account for a small share of changes in footfall, indicating that most social distancing was in response to *individual* fear. They connect social distancing practice to fear specifically, as the drop in footfall is strongly correlated with local COVID deaths, and increases with the previous busyness of the establishment. In other words, consumers respond to experiencing the *local* impact of COVID, and in consequence avoid places which are perceived to be high-density. Maloney and Taskin (2020) uses Google mobility data for the United States to support the conclusion that voluntary distancing accounted for most of the changes in mobility, though (as would be expected) policy orders do have a smaller but still significant impact. They find that the results are consistent across all but the lowest income group, where COVID-risky jobs are concentrated.

The interaction between political affiliation and social distancing in the United States is a complex and important factor. Donald Trump initially downplayed the seriousness of the virus and subsequently the imposition of restrictions was made a partisan issue. Baccini and Brodeur (2021) found that during March 2020 Democratic state governors were 50% more likely to impose a stay-at-home order, and they did so more quickly in response to cases above a certain threshold[2]. In addition to affecting the likelihood of a stay-at-home order, it has also been shown that political affiliation affects voluntary decisions to socially distance. Barrios and Hochberg (2020) is an example of this, investigating individual risk perception and social distancing from a partisan standpoint, written in early to mid-March. Using Google searches for COVID as a proxy for risk perception, they find that these Google searches decline strongly in counties that voted Trump in the 2016 election. They also find that Trump counties exhibit a muted mobility change in response to COVID cases, and comply less with stay-at-home orders. Painter and Qiu (2020) also documents this result. Allcott et al. (2020) shows that in addition to observed risk and policy, partisanship plays an important role in driving distancing behaviour. They also show through a survey that beliefs about the likelihood of contracting COVID are divided along partisan lines[3]. Another aspect to the partisanship divide in distancing is exposure to COVID-sceptical media. Ananyev et al. (2020) and Simonov et al. (2020) use the same identification strategy, exploiting variation of the position of COVID-skeptical Fox News on the TV dial to conclude that areas with greater exposure to Fox have a muted change in mobility. Bursztyn et al. (2020) also uses Fox to show that exposure to COVID-skeptical content decreases risk perception: in late February 2020, viewers of COVID-downplaying *Hannity* changed their mobility later than viewers of a more cautious show on the same network. This research on partisanship and media makes clear several salient points. First, distancing behaviour varied markedly along partisan lines during February and March. Second, the behaviour change was a function of individual beliefs about COVID risk; third, media consumption *in addition to pure partisanship* directly affected these

[2] They control for disease spread by including deaths in the state as a covariate in their OLS specification.

[3] This result is generalised by Pástor and Veronesi (2020), who find that Democrats are more risk averse than Republicans.

risk beliefs. This body of research shows, then, that political partisanship colours risk perceptions and affects behavioural choices. However, risk perceptions are moderated not only by political beliefs but also by media consumption and other confounding factors like occupation, income, and health risk level. Individual risk perception is important on its own standing and depends on separate local factors as well as partisanship; with this in mind, this dissertation measures and investigates the impact of risk perception *in the abstract* in determining mobility.

Past research in health economics has also addressed the determinants and motivations of health behaviours during epidemics and pandemics. Galizzi and Wiesen (2018) and **The bottom of p2 of Campos-Mercade has more on the determinants of health behaviour; see also the !SPATIAL HETEROGENEITY IN HEALTH BEHAVIOUR! that C-M mention on page 8. THIS SPATIAL HETEROGENEITY IS A KEY MOTIVATOR FOR THE DISS.**

2.4 COVID-19: Models of risk perception and social distancing

At the outbreak of the pandemic, many economists augmented the standard epidemiological 'Susceptible-Infected-Recovered' (SIR) models with economics-derived models of human behaviour. Murray (2020), writing as an epidemiologist, identifies

2.4.1 Standard SIR model

Kermack and McKendrick (1927) introduced the SIR model and it remains the basis for most modern epidemiological models; this explanation follows Avery et al. (2020). Each member of the population can be in one of three 'Susceptible, Infected, Recovered' states; therefore, at each time period we have

$$S(t) + I(t) + R(t) = 1 \tag{1}$$

, where the population is normalized to 1. A susceptible individual can only move to the infected state through contact with an infected individual, and an infected individual can only move to the recovered state. The recovered state, in the base version, includes *both* those who have recovered and those who have died: they share the key characteristics of being noninfectious and not susceptible to future infection. Three key parameters govern the rate of transitions between these states: γ, the recovery rate, represents the probability per unit time for an individual to move from Infected to Recovered; R_0, the basic reproductive number, is the "number of people an infectious person would infect over the course of their disease in a fully susceptible population" (Avery et al., 2020, p. 81). R_0 therefore stands in an inverse relationship with the recovery rate: $R_0 = \beta/\gamma$, where β is the expected number of contacts an individual makes per unit time in normal circumstances. Assuming that the individual infects every contact, the states evolve according to

$$\dot{S}(t) = -S(t)I(t)R_0\gamma \tag{2}$$

$$\dot{I}(t) = S(t)I(t)R_0\gamma - \gamma I(t) \tag{3}$$

$$\dot{R}(t) = \gamma I(t) \tag{4}$$

. For epidemiologists, then, R_0 is seen as a ‘compound parameter’ of both the virus’ natural infectivity[4] and the expected number of in-person interactions during pre-pandemic life (Avery et al., 2020, p. 84): government policies and behaviour change therefore have the effect of reducing the R_0 parameter to some variable R_0^t. This gives rise to the key aim of ‘flattening the curve’: achieving a state where – given the current susceptible fraction of the population – the expected number of people that a contagious individual infects over the course of their illness is below 1:

$$R_t \equiv R_0^t S(t) < 1 \tag{5}$$

. R_t is known as the effective reproductive number.

Avery et al. (2020) sorts the contributions of economic research to the basic SIR model into three basic categories: pointing out the endogeneity of the reproductive number, adapting the model to heterogeneity of different subpopulations, and adapting SIR models to policy-relevant issues like social distancing compliance. I now set out the insights that model-based COVID economic research gives to the topic in this dissertation: the relationship between risk preferences, local demographic characteristics, and social distancing behaviour.

2.4.2 Models of endogenous social distancing

In the base SIR model, R_0^t is endogenous, since individuals adjust their exposure to others in response to the state of the epidemic. A priority for this dissertation is to establish the characteristics and determinants of social distancing; a large body of literature has emerged in economics discussing this issue.

Toxvaerd (2020) models individuals as making non-cooperative, forward-looking decisions to engage in costly social distancing by solving a tradeoff against beneficial social behaviour and the risk of infection. Toxvaerd finds that equilibrium social distancing depends on the threshold infection probability, which is determined by the aggregate disease prevalence; this entails that individuals react to higher prevalence

[4] Which itself is governed by the expected length of contagiousness $\frac{1}{\gamma}$ and the transmissibility, which we assume to be 1.

by distancing more, mitigating the flow rate between $S(t)$ and $I(t)$. In this model, individuals assess the value of becoming infected as equal to the expected discounted lifetime utility of being in the infected state. The model considers a homogenous population and so does not take demographics into account.

Farboodi et al. (2020) also models the tradeoff of exposure against health risks. Using the SafeGraph dataset, they find that social activity levels fall before imposition of mandatory measures. This yields their key observation: desire to avoid illness is a key determinant of social distancing, meaning that there is a strong laissez-faire reduction in social activity. This cost-response reduction produces the majority of social distancing behaviour, but not enough for optimal pathogen suppression; as such, social distancing orders are recommended. They implement quadratic matching with random search, introduced in Diamond and Maskin (1979), to model social interactions. Individuals are split into the S, I, R states, where the level of chosen social activity is the same among all individuals in that state. Disease transmission is therefore a function of β, the number in the susceptible and infectious states, and the social activity level of each of those states. Individuals choose their level of social activity, taking the external social activity level and the number of infected as a given. The model, however, abstracts from subpopulation heterogeneity.

Eichenbaum et al. (2020) also point out that exposure comes either from purchasing consumption goods, from working, and from random interactions such as touching surfaces: this shows the importance of income and mode of work for disease transmission.

Chernozhukov et al. (2021) estimate a structural equations model, which incorporates voluntary social distancing into a causal framework. This framework decomposes the change in COVID caseload into three drivers: direct effects of policies (e.g. mask mandates), behaviour changes due to policies (e.g. stay-at-home-orders), and individual behaviour changes. In their model, individuals respond to global information – which is represented by month dummies – and they respond to local information – which is represented by the local growth rate and total cases. In other words, they take information to be broadly equivalent to lagged health outcomes. Individual-level distancing behaviours are a function of policies, information, and observed confounders. These observed confounders are at the state level and include the demographic characteristics of population, area, unemployment and poverty rates, percentage of people at risk of illness, and governor's party. Their empirical estimation shows that log case growth (at national and county level), stay-at-home orders, and business closure policies have a significant effect on mobility; they find that including case growth in the specification is a moderately better proxy for information than deaths growth. Finally, they find additional evidence that individuals respond voluntarily, in response to information about COVID, rather than as a forced response to policies.

2.4.3 Subpopulation heterogeneity

An important contribution that economists made was to incorporate multi-population SIR models. By including separate populations, it is possible to account for the significant heterogeneity in risk with respect to age. In respect of the social distancing risk decision, this variation in the cost of infection is theorised to generate a disparate response in different age brackets, in addition to the usual demographic and occupational variation in mobilty patterns with age. A key component of multi-group SIR models is the expansion of R_t from a single population average to a matrix with a measure of R_t for the interactions between subpopulations. In theory, this allows for policies to be targeted at a local and risk-group level; however, in practice this policy approach has not in general been implemented. Favero et al. (2020) calibrate an SIR model with nine age brackets and three occupation sectors. In this setting, the decision to risk exposure is a function of the actual probability of infection, the perceived importance of the activity, and the perceived cost of infection. They find that perceived cost is an important factor to reduce risky behaviour. Another multi-population model is Acemoglu, Chernozhukov, et al. (2020), which advocates for targeted age-dependent policies as a significant utility increase over blanket policies. In an optimal-control model with state variables S, I, R in three age categories, they derive the Pareto frontier between economic loss and deaths, finding that only targeting the old population is nearly as effective as a fully-differentiated policy, while both perform significantly better on the tradeoff space than blanket policies.

Brotherhood et al. (2020) includes age heterogeneity in an augmented SIR model, and also explicitly models individual behavioural choice. They also include imperfect information of infection status for symptomatic individuals, yielding an important role for testing. This paper derives social interactions and economic behaviour from a time-allocation utility framework. Individuals gain utility from consumption, outside-home (risky) leisure, and at-home leisure. They allocate time between at-home work, outside-home work, at-home leisure, and outside-home leisure. Infection risk rises with cumulative time spent outside home, so as infection risk rises, individuals imperfectly substitute to tele-work and at-home leisure, both of which are more costly than their outside-home counterparts. The calibrated model predicts that behavioural adjustment of the old is the most significant factor in reducing deaths, but that risky behaviour by the young can in fact reduce transmission. However, the reduction in hospital capacity caused by this risk-taking negates the effect and increases deaths in the long run. The model also predicts that blanket stay-at-home orders are only effective when lasting longer than 6 months.

Finally, economists have interpreted the decision to reduce mobility as a function of *social preferences* (Fehr & Schmidt, 1999) in addition to self-interest. Campos-Mercade et al. (2021) use this approach, under the notion that prosocial motivations are a determinant of physical distancing behaviour. They use an incentivised game to measure prosociality, where participants can expose others to risk for a payoff, and simultaneously collect a health behaviour survey. They conclude that individuals are generally unlikely exposure others to risk for personal gain even at a high level of payoff; to the extent that prosociality varies, however, they find that it predicts compliant health behaviours like mask wearing and social distancing.

In summary, there are many empirical and model-based factors that drive the social distancing decision. These include individual health risk, income, political affiliation,

3 Text analysis: an overview

In the literature review I summarised the current uses of Text Analysis in the economic literature. In this section, I present a detailed look at the inference problem for text analysis of English documents, and the various types of approaches possible to tackle it. The section introduces concepts that are used in the later Methods section, and also serves as an introduction to the field of text analysis for economics.

3.1 The inference problem

For economists, the fundamental concept is that relevant information exists within a corpus of text. This information can be represented as a low-dimensional variable, relevant to a model of economic decisionmaking: for example, interest rate expectations or risk aversion. However, this low-dimensional variable is expressed in text, a noisy, extremely high-dimensional format: a list of documents which are n words long, drawn from a vocabulary of size p, has a dimension of p^n. Given this computational constraint, the central challenge is to isolate the latent variable from the high-dimensional noise in a robust manner. Gentzkow et al. (2019) present a useful notation for a two-step process of this extraction: first, the raw text $\mathcal{D}$ is mapped to an array of tokens, $\mathbf{C}$. These tokens – usually in the form of words, phrases, or sentences – are the fundamental units of analysis. Second, the token array is mapped to the outcome array $\hat{\mathbf{V}}$, which is an estimate of the latent variable $\mathbf{V}$. $\hat{\mathbf{V}}$ is then used in the final analysis. In these two mappings, there are two challenges: first, to include only tokens relevant to the variable in $\mathbf{C}$; second, to accurately estimate $\mathbf{V}$ using $\mathbf{C}$.

3.2 Mapping $\mathcal{D}$ to C: text pre-processing

In Gentzkow et al. (2019)'s notation, the vector $\mathcal{D}$ is the corpus of text to be analysed, consisting of documents $\mathcal{D}_i$): a central bank announcement, for example. $\mathbf{C}$ is a numerical matrix, where the columns correspond to tokens, and rows correspond to documents.[5] $\mathbf{C}_{ij}$ is a scalar representing the

[5] It is sometimes called the 'Document Term Matrix (DTM)'

term frequency: how many times each token appears in a document. In other words, **C** is a frequency matrix, with each member of the matrix representing the counts in a document of a particular token. Each column is an element of the vocabulary – the set of unique tokens in the whole corpus – and so the number of columns equals the size of the vocabulary. Representing raw text in this fashion is a central part of the *information extraction* problem (Manning & Schütze, 1999, p. 529). Ultimately, we wish to discard all tokens which do not convey information relevant to the latent variable, and quantify the amount of information each token conveys.

The first decision is to define the token. The simplest method is to assign tokens to single words, or 'unigrams'. This method, known as the 'bag-of-words', reduces the dimensionality of the document by the maximum amount by ignoring any dependence between the tokens. However, it simplifies language to a large extent, as (by definition) word order, grammar, ambiguous terms ("hard", "line"), and modifiers ("not") are lost in the mapping. These problems can be moderated by instead considering n consecutive words to be a token: "in the beginning" maps to ("in.the", "the.beginning". Since the words overlap, the size of **C** increases exponentially, requiring a corresponding increase in computational power. **C** also becomes more sparse, requiring more observations. However, even considering pairs of words yields a significant improvement in extracting meaning (Cheng et al., 2006), as we expand the remit to two-word phrases. As computational power has increased, it is possible to efficiently include a wider context than 2- or 3-grams by using the word embeddings technique.

There are several steps that can reduce the size of the vocabulary – the number of columns of **C** – without a great degree of information loss. When the bag of words method is used, **C** is a word frequency matrix. It will therefore reflect the structure of English in that each document will have a large frequency of structural words like "from", "the", "could". These are called stop-words, and do not convey meaning; the first step of a text analysis is to remove these words by filtering on a list. There are standard lists, but it is important to take the domain of the corpus into account when filtering stop-words; for example, Twitter has domain-specific stop-words "@", "#", but filtering punctuation risks removing emoticons (":)"). A closely related step is converting all words to lowercase, on the general assumption that word meaning does not depend on whether it starts a sentence (Denny & Spirling, 2018). However, in the context of social media, this may be problematic: emotion and intensity of sentiment is often expressed by various forms of capitalisation. Internet-targeted sentiment analysers usually attempt to take this into account. Words are also expressed in English in different forms ("laughing, laughs"), so grouping words by their stem ("laugh") is also done (Porter, 1980). These filtering steps are very effective in reducing the dimensionality of **C**, but the problems with word ambiguity and word order cannot be avoided.

Maximising information by stop-word removal is often extended by applying term frequency weighting to the matrix. The linguistic principle is that the amount of information a word conveys about a sentence is the inverse of its frequency in the wider text. Stop-words occur most frequently in language, and so convey the least amount of information about the content of a document; conversely, tokens with a lower frequency will convey more information about the content of a document (for example, place names or precise terminology like 'monopsony'). The most amount of information will be given by terms that appear in a small number of the total set of documents – they have a low document frequency – but are repeated in the document being considered – a high term frequency. Dividing these two frequencies gives the "term frequency - inverse document frequency" metric, which represents the amount of information about the document a given token conveys (Manning et al., 2008, p. 100).

For economists, a vitally important concern in these pre-processing steps is reproducibility. The decisions made during this initial process can significantly affect the final result: the necessity of intensive data transformation combined with the inherent variability of the process leads nearly inevitably to the charge that an analysis could lead to a different result if different preprocessing steps had been taken. This contingency of the analysis upon the data preparation is called the "forking paths" problem (Gelman & Loken, 2014), and with the observational data common in economics not much can be done to alleviate it: replication and pre-registration are often unviable. Aside from fully specifying the details of the data pre-processing, Denny and Spirling (2018) propose calculating the distance between alternative document term matrices as an ad-hoc metric.

3.3 Methods of estimating $\hat{\mathbf{V}}$

Once the document term matrix has been created, the final step is to estimate the variable. There are a number of methods; the primary factors in deciding the estimation procedure are the computational power available and the theorised characteristics of the latent variable. A central component, particularly for economists, in choosing the technique is the direction of causality. This can flow either way in a text analysis: the information in the text encoded in $\mathbf{C}$ can cause the variable of interest $\hat{\mathbf{V}}$, but in other cases the text is a function of the latent variable. In other words, approaches might model either $p(\mathbf{v}_i|\mathbf{c}_i)$, or $p(\mathbf{c}_i|\mathbf{v}_i)$; the discussion so far has centred on the latter in the form of latent variable modelling, where an unobserved outcome variable *generates* the text; but text analysis can also work by mapping text frequencies to observed outcomes. To illustrate this, contrast Jegadeesh and Wu (2013), where positive words in company report filings (C cause higher stock market returns V, with Bandiera et al. (2020), where the latent variable 'CEO behaviour' V generates text in their diaries (C. Approaches incorporating $p(\mathbf{v}_i|\mathbf{c}_i)$ as the structural form usually take the form of text regressions, and are called discriminative models; models of $p(\mathbf{c}_i|\mathbf{v}_i)$ are broadly termed generative models (Jurafsky & Martin, 2009, p. 81). Dictionary-based methods are the simplest, and do not directly involve statistical analysis; we begin with an overview of these.

3.3.1 Dictionary-based methods

The first, and simplest, way to perform inference on a latent variable is to use a simple dictionary matching technique. This is the most well-established method used in the economics and finance literature, largely because it is computationally simple and easy to implement. In a nutshell, the method is to use a known function $f(\cdot)$ to stipulate $\hat{\mathbf{v}}_i = f(\mathbf{c}_i)$. No information is inferred from the language corpus; rather, known attributes of $\mathbf{c}_i$, established by linguistic and domain-specific research, are leveraged. This is the method that I take, which is explained in detail in section 5.1. Beyond computational ease, the an advantage of this technique is that it avoids 'model stacking', which is the case if $\hat{\mathbf{V}}$ is subsequently used in the primary analysis. This approach is particularly useful if the variable of interest has no prior observed outcomes (which excludes supervised learning approaches, for example). However, it accordingly must rely heavily on well-established and often domain-specific information about the mapping $f(\cdot)$. When using a dictionary-based approach, the researcher must justify the relevance of this mapping to her application, as general dictionaries may not transfer well to particular contexts.

3.3.2 Text regressions: regularised OLS under high dimensionality

Text regressions are a popular method of inference in cases where we wish to estimate an outcome based on the content of a text corpus. In a text regression, we predict $\mathbf{v}_i$ from $\mathbf{c}_i$ in the normal way, by using ordinary least squares (Gentzkow et al., 2019, p. 541). Assuming a linear form, we approximate the conditional expectation with

$$E[\mathbf{V}_i|\mathbf{C}_i = c] = \beta^T c = \sum_{k=1}^{K} \beta_k c_k$$

, and the corresponding estimator is

$$\hat{\beta}^{OLS} = \underset{\beta}{\operatorname{argmin}} \sum_{n=1}^{N} (V_i - \mathbf{C}_i^T \beta)^2 = (\mathbf{C}^T\mathbf{C})^{-1}\mathbf{C}^T V_i$$

. $\mathbf{C}^T\mathbf{C}$ is an $N \times K$ matrix. As K approaches N, each covariate adds the same amount of variance (σ^2/N) to the estimator, regardless of the true coefficient of the covariate. So as K increases, the precision of the estimator decreases (Davidson & MacKinnon, 2004, p. 101). If we have extremely high dimensionality and $K > N$, as may be the case when each covariate β_k corresponds to a token in a large text vocabulary, we cannot invert the matrix $\mathbf{C}^T\mathbf{C}$ and obtain the OLS estimate. The general method of addressing this is to introduce a penalty term for each additional covariate; we minimise

$$\hat{\beta}^{POLS} = \underset{\beta}{\operatorname{argmin}} \sum_{n=1}^{N} (V_i - \mathbf{C}_i^T \beta)^2 + \lambda(||\beta||_q)^{\frac{1}{q}}$$

, where $||\beta||_q$ is the L^q distance function[6] (Athey & Imbens, 2019, p. 696). Choices of q correspond to different methods: the LASSO method has $q = 1$, for example (Tibshirani, 1996). The effect of this penalisation is to shrink β_k towards zero with increasing λ, thereby maximising the precision of the estimator. The process for the econometrician is then to choose the value of λ, the 'tuning parameter', according to some measure of model fit such as the Akaike or Schwartz Information Criterion.

3.3.3 Generative Language models

Recently, machine learning methods have seen an expansion in the academic economics literature (Athey & Imbens, 2019). Industrial applications of text analysis are usually characterised by access to an extremely large, high-frequency corpus: search data is a salient example. In this case, large language models are used, modelling grammar and structure within text using supervised and unsupervised 'generative' models. A supervised model starts with a dataset where the outcome $\mathbf{v}_i$ is labelled; the model 'trains' on this ground-truth dataset and then is assessed by its predictive accuracy on unseen datasets. The canonical example is email spam detection: given a dataset of emails labelled with a binary spam/not spam variable, a supervised model[7] generates a probability that subsequent emails are spam. For economists, supervised language models have been applied to sentiment detection in financial discussion, or measuring the political bias of news outlets (Groseclose & Milyo, 2005).

Unsupervised models do not begin with labelled data; rather, the model classifies documents into categories according to prior assumptions about the characteristics of the latent variable. Instead of treating the document term matrix as *a priori* observational data, this approach sees terms as the output of a latent generative process, using a model of $p(\mathbf{c}_i|\mathbf{v}_i)$. The model therefore infers abstract latent variables, such as 'topics', from the text. Economists have used topic modelling to measure the impact of greater transparency on the content of Federal Open Market Committee statements (Hansen et al., 2018).

4 Data

4.1 GeoCoV19 and geolocation inference of Twitter datasets

The primary dataset of tweets I present is a subset of GeoCoV19, a project which collected tweets relating to COVID-19 between February 1 and May 1 2020. The Twitter Streaming API provides a live filter function for a number of keywords and hashtags, and returns all tweets matching any of the search terms. The terms were chosen to cover a broad base of COVID-related talk, including searches for symptoms

[6] $L^q = \sum_{k=1}^{K} |\beta_k|^q$. To illustrate, L^2 is therefore the familiar Euclidean metric and L^1 is the Manhattan (taxicab) metric.

[7] Specifically, a naive Bayes model.

(e.g.'breathing difficulties'), behaviour (e.g. '#masks4all'), and popular hashtags (e.g. '#IStayHome, #FlattenTheCurve') in addition to central discussion topics like 'coronavirus'. The full list of search terms used is in a separate appendix. The dataset was collected with multilingual use in mind, but a majority of the tweets were based in the US; tracking with the fact that the US makes up most of Twitter's user base. In total, 524 million tweets were collected during the time period; during the first three weeks of February this number was lower, reflecting lower general interest, and increasing to around 6.4 million per day during March and April. We filter for tweets that are in English and are geotagged to originate from the United States. This primary dataset is supplemented with exact-geolocation Tweets from two other similar datasets, Banda et al. (2021) and Lamsal (2020). The collection method for both datasets is broadly similar to GeoCoV19, using the Streaming API and a set of COVID-related keywords. Of the final dataset, around 150,000 have exact geolocation embedded in the tweet. This is due to the fact that geo-tagging is an explicit option that needs to be set for each tweet, involving activating location data on the mobile app. Another option for geo-tagging is to select a place from a search box; yielding a 'place' in the metadata. Both of these types of metadata involve accurate locations, but make up a small proportion of the total geolocated tweets. A third method of geolocating tweets is used to identify most of the locations: when activating a Twitter account the user is strongly encouraged to set their location in their profile. Although it is a free field (generated Place suggestions are included, but are optional), most users set this to their current location; this metadata is included with every tweet. The maintainers of the dataset then employ a toponym extraction approach to elicit the location of the location field. The text of the user location field is first cleaned of non-text characters and symbols. Candidates are then created from the remaining unigrams (single words) and bigrams (pairs of adjacent words), ensuring that two-word place names like 'Los Angeles' are included. Groups of three or more words are not considered. Each remaining candidate is filtered against a list of stopwords (see section 3),

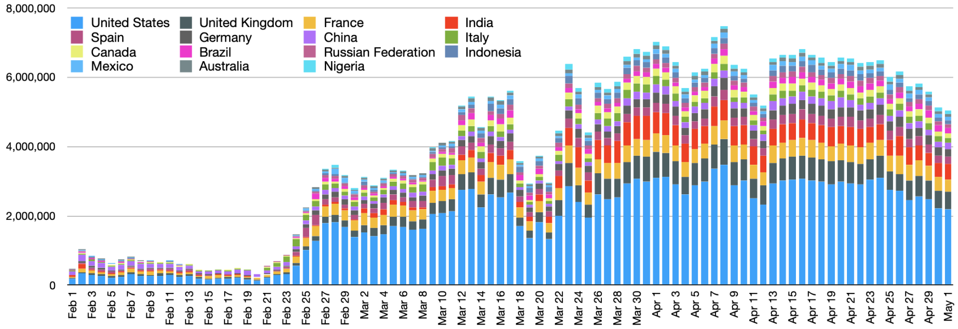
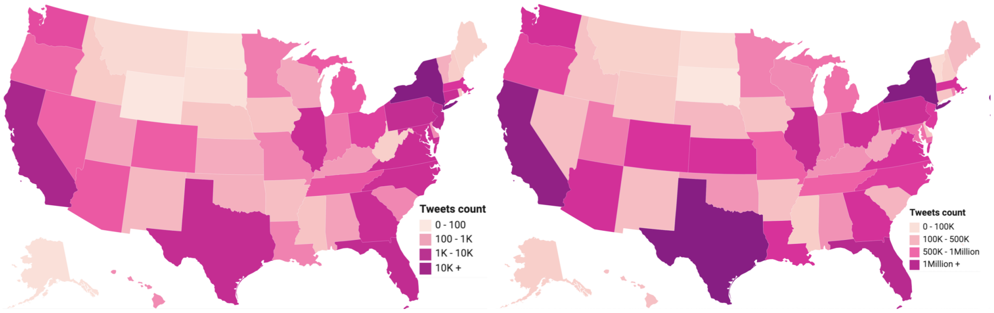


Figure 1: Daily distribution by country of GeoCoV19 tweets, Feb 1st to May 1st, 2020 (Qazi et al., 2020)

and against the 'World Cities Database'[8], an index of 3.1 million worldwide place names, covering 141,989 locations in the US. The remaining candidates are sent as one query to Nominatim, the OpenStreetMap search engine, yielding a best-attempt geolocation: the procedure works best when state and place name is given. Cross-checking the procedure with GPS-geolocated tweets, the dataset shows good coverage and accuracy across US counties, and so makes a panel approach viable. The maintainers of the dataset also presented locations derived from the text of the tweet itself using the same procedure, but we filter these tweets out of the final dataset due to low accuracy. A drawback to this gazeteer approach is that, since users can set their profile location freely, users from other countries or states could masquerade as Americans in particular locations, or the classification process could mis-classify a foreign (particularly English) place as a US location due to sharing a name – for example, York County, Maine. In order to account for this we remove counties from consideration that have a share of total tweets significantly higher than their population share of the US: Earth, Texas, is the most prominent example of this. Other than misclassification, deliberate setting of the profile location to a US location is possible. This may be a concern for large, well-known cities like Los Angeles and New York, but it is less likely that a given user will set their location to a less well-known American county. Finally, users may move county and not change their location.



(a) USA: Organic geotagged tweets

(b) USA: User distribution

Figure 2: Geographic distribution of GeoCoV19 tweets and users (Qazi et al., 2020)

Twitter's terms of service restrict the large-scale sharing of Tweet datasets; hence public-facing tweet datasets can only be made available 'dehydrated', with only the universal identifiers (a long number that maps to the tweet in Twitter's database) given, instead of the tweet text and assorted metadata. In order to use the dataset to analyse tweet content, researchers wishing to use the dataset must apply and be accepted for a Developer account. This gives access to a password called an API key, which is used to query the Twitter API with the identifiers to 'rehydrate' and gain access to the full tweet text and metadata. Since the tweets are delivered from the servers at download time, this procedure

[8] Available at https://www.kaggle.com/max-mind/world-cities-database?select=worldcitiespop.csv

entails that a proportion of Tweets – those identified as containing misinformation or those sent by users whose accounts have since been suspended or deleted – will be unavailable on request from the service when the dataset is ‘rehydrated’. This presents a selection problem for topics like misinformation, where Twitter enacts a stringent and continuous policy. A prominent concern for my data might be the suspension of Donald Trump’s prolific Twitter account; his tweets were a touchstone for right-wing US online conversations. This does not present an issue to my data, however, because tweets interacting with tweets from a deleted account are still available from the API. The deletion rate ultimately encountered during the rehydration process was around 20%, reflecting the normal removal of machine-generated content.

GeoCoV19 was made available as a dehydrated dataset, containing the tweet identifiers and the inferrred geolocation information. The tweets were rehydrated in December 2020 using Hydrator, an open-source tool made available for academic research by the digital archive organisation Documenting the Now (Summers, 2020). Each tweet is delivered in the Javascript Object Notation (JSON) format: this is a common file standard used to deliver arbitrarily nested data. Each entry contains a ‘tree’, which corresponds to a key-value pair. Each value can itself contain a list of named objects; so, unlike a CSV file, the data is not a ‘flat’ table. An example from the Developer documentation is reproduced below: note the nesting and the metadata delivered with the tweet text. The final dataset contained 63,654,120 tweets and measured approximately 207 GB.

Listing 1: Example Tweet object JSON file (Twitter, Inc., 2021)

```
{
  "created_at": "Thu Apr 06 15:24:15 +0000 2017",
  "id_str": "850006245121695744",
  "text": "1\/ Today we\u2019re sharing our vision for the future of the
      Twitter API platform!\nhttps:\/\/t.co\/XweGngmxlP",
  "user": {
    "id": 2244994945,
    "name": "Twitter Dev",
    "screen_name": "TwitterDev",
    "location": "Internet",
    "url": "https:\/\/dev.twitter.com\/",
    "description": "Your official source for Twitter Platform news,
        updates & events. Need technical help? Visit https:\/\/
        twittercommunity.com\/ \u2328\ufe0f #TapIntoTwitter"
  },
  "place": {
  },
  "entities": {
    "hashtags": [
    ],
    "urls": [
      {
        "url": "https:\/\/t.co\/XweGngmxlP",
        "unwound": {
          "url": "https:\/\/cards.twitter.com\/cards\/18ce53wgo4h\/3xo1c",
          "title": "Building the Future of the Twitter API Platform"
        }
      }
    ],
    "user_mentions": [
    ]
  }
}
```

After the dataset downloaded, a Python script was used to parse and un-nest the JSON files, select the relevant variables, and write them to a CSV file. This file was then joined by matching the UIDs in the CSV with the geolocation dataset in GeoCoV19, yielding a large CSV-format dataset containing the tweet text and metadata from Twitter, and inferred geolocation information from GeoCoV19. Language processing was then performed on the tweet text, as is detailed in section 5.1.

4.2 SafeGraph: Geolocated smartphone data for measuring social distancing

A second central dataset is provided by SafeGraph, a company which usually collects data on commercial footfall, but made their dataset available for academic research in light of the pandemic. The social distancing dataset was downloaded in January 2020 using the provided API key. The data was parsed and loaded into R using the SafeGraphR package (Huntington-Klein, 2020). The SafeGraph dataset consists of over 45 million anonymised smartphone GPS pings located to an accuracy of a 150m square location (SafeGraph, Inc., 2020). Since the data was collected in 2019 and 2020, year-on-year change can also be presented. The data is presented at district level, but for the purposes of the analysis I aggregated this to a county-level daily metric. For each device on each day, the most common night-time location is determined from the previous 6 weeks. This home location is used to determine the number of devices in a county which leave their home, the length and time of day they leave and return, the district they travel to, and the points of interest they visit. Although this is a sensible method of determining a device's 'home', it may be a source of measurement error if an individual works a night shift, does not bring their smartphone with them when they leave the house, or sleeps regularly at a partner's house (Chiou & Tucker, 2020). However, I believe that the long lead time of 6 weeks' determination, the large sample size, and the rarity of this occurence mitigate the size of the potential measurement error. There is also a chance for selection bias: the data does not represent those Americans who do not own a GPS-enabled smartphone, and does not represent those who refuse to allow GPS tracking on their smartphones. We assume, with evidence, that these omitted populations are mean-zero with respect to social distancing mobility behaviour: 81% of US adults own a smartphone and 96% own a cellphone (Pew Research Center, 2019), and (Athey et al., 2017) indicates that the decision to share personal information has a large random element (Chiou & Tucker, 2020). SafeGraph also address the issue of sampling bias[9] by comparing, for each census block, the observed proportion of all devices to that census block's proportion of total US population[10]. They assess this sampling bias at the county and demographic level and do not find a significant in observed proportion to that indicated by the census data. From this rich data, I derive two primary variables: the median minutes spent at home during 8am-6pm and the proportion of measured devices that stayed at home all day. It should be noted that the differential anonymisation means that the exact sum of devices is inaccurate; although there is a possibility to bias the proportions I calculate, this only arises in areas which are sparsely populated. These areas usually do not yield enough geo-located tweets to be included in the final analysis, so the problem is largely circumvented.

4.3 Other datasets

4.3.1 Oxford COVID-19 Government Response Index

An ongoing feature of the US response to COVID-19 is the differential policies enacted in each state. These policies had varying effects on the extent of social distancing practised in the state and sub-state areas. In order to account for this variation, it is necessary to report the nature and strength of the policy in place in each county at each time point. Petherick et al. (2020) provide the *de facto* standard index for tracking government responses. This dataset, originating from the Blavatnik School of Government at Oxford University, reports 19 indicators of government response in containment/control, economic support, and health categories. The containment/control variables are of most interest; these include indicators for school/workplace closures, public transport closures, stay at home requirements, internal/international movement restrictions, restrictions on gathering size, and public event cancellation. These are coded on an ordinal scale (usually 0 to 2, but sometimes 0 to 4), measuring the severity or intensity of the policy. These indicator variables are then aggregated into a 'policy stringency' index, a numeric measure of the general severity of restrictions on movement in a governmental area. This index is created by taking the ordinal values of each indicator, rescaling each by the maximum value of the scale, to create a score between 0 and 100. Although this approach inevitably masks substantial subtleties in the context of each policy, they crucially provide a comparable index. The dataset is available at the state level for the US, meaning that it does not cover county-level differences in policy. However, this does not present a problem for the analysis; practically all non-federal COVID containment policies were

[9] `https://colab.research.google.com/drive/1u15afRytJMsizySFqA2EPlXSh3KTmNTQ#sandboxMode=true`

[10] That is, if New York County has 3.14% of the US population, they would expect to find 3.14% of their total observed devices in New York County

enacted by blanket state-level orders.

4.3.2 American Community Survey and Census

Demographics also have a significant bearing on social distancing behaviour; an area with more elderly people will display less movement than a younger district, due both to baseline movement patterns and differing shielding behaviours during the pandemic. In the main specifcation, we account for this with county fixed effects; but this data is also useful for exploring patterns of sentiment and social distancing among broader demographics. Therefore, I include data from the American Community Survey (ACS) and the 2011 US Census. This includes variables like population density, income, and age distribution. I match the Public Use Microdata Sample of the ACS to the county level, the smallest level available.

4.4 Descriptive statistics

5 Methods

5.1 Text analyses of Twitter data

5.1.1 NRC Emotion Lexicon

Sentiment analysis is the task of inferring the author's opinion of a subject from the text they write about that subject. Due to major commercial incentives for accurate sentiment inference from internet companies, this area has seen major advancement over the last 15 years. The NRC Emotion Lexicon is a widely-used tool in this field[11]; the project aims to answer questions like "Is the author happy with, angry at, or fearful of the target?" (Mohammad & Turney, 2013). The lexicon uses the notion that the emotions expressed in a text are a function of the choices of words: that is, words like 'gloomy' indicate sadness in a text, 'delightful' indicates joy, and so on. The lexicon is composed of a term and the emotion mapped to the term; each term can be mapped to multiple emotions, or none at all. There are eight basic emotions that each word is mapped to: anger, anticipation, disgust, fear, joy, sadness, surprise, and trust; these correspond to Plutchik (1980)'s widely-used taxonomy of emotions. The final lexicon contains 14,182 terms annotated with their corresponding emotions. This is large, covering a very wide portion of common English vocabulary. It is also very simple, as it is a binary, one-to-many mapping from term to emotions. This entails that the lexicon takes no account of *intensity* of emotion. It also deals with ambiguous words – such as the vernacular usage of 'sick', which connotes a positive emotion – in a simple way, by assigning every connected emotion to the term. It is therefore possible that a term

[11] For example, Mohammad and Turney (2013) – the paper describing the NRC lexicon – has been cited over 1,300 times.

could connote every one of the eight emotions. Finally, the lexicon only considers unigrams, which means phrases and multi-word contexts are omitted, and negation is not considered.

Table 1: Sample from NRC Emotion Lexicon (Mohammad & Turney, 2013)

term	anger	anticipation	disgust	fear	joy	sadness	surprise	trust
aback	0	0	0	0	0	0	0	0
abacus	0	0	0	0	0	0	0	1
abandon	0	0	0	1	0	1	0	0
abandoned	1	0	0	1	0	1	0	0
abandonment	1	0	0	1	0	1	1	0

Following the download from the Twitter API, the text content and UID of each tweet was separated from the metadata. This uncleaned dataset contains 34.7 million tweets, totalling 7 GB of data. This dataset consists of both tweets and retweets, a function where a user 'forwards' somebody else's tweet to their own followers. In order to isolate the possible effect of retweeting another person's opinion, which in some cases may differ from one's own, I filter the full dataset for only original tweets; this 'original-only' dataset contains around 6.4 million tweets. NRC sentiment analysis was performed on the full and original-only datasets; due to computational constraints the VADER analysis was performed solely on the original-only dataset. The NRC sentiment analysis uses the same code for both datasets and the pre-processing steps are identical regardless of dataset or analysis technique.

The first cleaning step is to remove links and username mentions from the text, and remove any tweets which are blank. Stop-word removal was not necessary, since the lexicon only matches sentiment-laden terms. Next, the `sentimentR` package (Rinker, 2015, August 16/2018) is used to split the text into sentences, as we compute the sentiment at the sentence level. Next, the text is split into its constituent tokens using the `tidytext` package (Silge & Robinson, 2017). The resulting document term matrix is then matched against the NRC lexicon and aggregated at the sentence level; we therefore obtain a binary factor variable for whether words connoting the different emotions appear in each sentence. The 'fear' emotion is the variable of interest; the fear factor variable is aggregated back up to the tweet level[12], giving a binary factor for whether words associated with fear appear in the tweet text. Tf-idf scaling was not performed.

5.1.2 VADER Rule-Based Sentiment Approach

Although the NRC approach has the advantage of yielding emotion scores, allowing the analysis to pick out the 'fear' dimension of each tweet text, the dictionary-based approach lacks sophistication. In

[12] This sentence-level analysis is an artifact of the R packages used to conduct the text-cleaning process; there is no change by aggregating back up to the tweet level after the sentiment analysis is done.

particular, in addition to the problems with ambiguity, omission of multi-grams, and lack of negation mentioned above, the lexicon is not *domain-specific*. On Twitter, as with all online platforms, emojis (for example) take on a significant role for communication; additionally, the platform has its own manner of communication which is different to other forms of writing. This means that certain emotions may be mis-allocated, or – in the case of emojis – omitted altogether. This motivates the use of an alternative text analysis tool, which is designed for sentiment analysis on the Twitter platform. The 'Valence Aware Dictionary for sEntiment Reasoning' (VADER) is a popular tool for Twitter-based sentiment analysis, and addresses these concerns (Hutto & Gilbert, 2014).

VADER is considered to be the gold-standard for social media sentiment lexicons, and performs equally well as human raters at matching sentiments. Instead of emotions, however, VADER reports sentiment as a *polarity* score. In the broadest sense, polarity (sometimes also referred to as 'liking' or 'valence') is a binary measure of the attitude of the author towards the topic of a text, and so is either positive, neutral, or negative. In addition, the polarity score can also reflect the intensity (sometimes also referred to as 'activeness' or 'arousal') of the emotion; compare 'exceptional' to 'okay'. VADER takes into account the intensity of valence, and so reports its polarity index on a continuous scale. When using VADER, the polarity can be reported at the text level, but this text polarity score is always an aggregate function of the term polarities. This aggregate function might include, for example, term frequency / inverse document frequency (tf-idf) weighting (as discussed in section 3.2), perhaps in addition to some transformation which takes into account the context of each term.

VADER is also able to identify the contextual meaning of terms by using a pre-trained sentiment classifier, and additionally incorporates five rules extracting meaning from word order and grammar. These are punctuation, capitalisation, degree modifiers (e.g. 'extremely', 'marginally'), 'but' as a signal of polarity shift, and a sophisticated negation detector. This final rule examines the trigram before a sentiment-laden term (above a certain absolute value) and determines if it has been negated; this rule achieves 90% accuracy. Crucially, all the lexical features VADER includes are calibrated and verified to identify sentiment on the Twitter platform. In general, VADER performs as well or better than cutting-edge, compute-heavy sentiment extractors on Twitter data.

The R package `vader` (Roehrick, 2020) is used to perform the sentiment analysis. Due to the extensive memory requirements of the analysis process, the uncleaned tweets dataset was split into 64 files of 100,000 tweets each and the code run on a virtual machine on the Google Cloud Compute platform. The same cleaning pre-processing was performed: removing links, usernames, and hashtags. The sentiment analysis calculates the polarity of each word in the sentence on a continuous scale, and reports the individual word scores, the compound polarity score (normalised to a continuous $[-1, 1]$ scale from the sum of the individual sentiment scores), and the adjusted proportion of the text that is positive (and likewise for negative and neutral. Note that the adjusted proportion of the score accounts for both text length and text intensity, reflecting a higher score for higher intensity and similar results regardless of length. An example of the output from the VADER program is in appendix A.

5.2 Empirical model

Given this county-level measure of tweet sentiment, we now incorporate it into an empirical model of social distancing.

5.2.1 Inference with Panel Fixed Effects

5.2.2 Specifications and hypotheses

6 Results

6.1 Robustness checks

7 Discussion and conclusion

References

Acemoglu, D., Chernozhukov, V., Werning, I., & Whinston, M. (2020, May). *Optimal Targeted Lockdowns in a Multi-Group SIR Model* (w27102). National Bureau of Economic Research. Cambridge, MA. Retrieved October 22, 2020, from http://www.nber.org/papers/w27102.pdf. (Cit. on p. 11)

Acemoglu, D., Makhdoumi, A., Malekian, A., & Ozdaglar, A. (2020, July). *Testing, Voluntary Social Distancing and the Spread of an Infection* (w27483). National Bureau of Economic Research. Cambridge, MA. Retrieved January 20, 2021, from http://www.nber.org/papers/w27483.pdf. (Cit. on p. 5)

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 pp. 4, 7).

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 pp. 4, 7).

Athey, S., Catalini, C., & Tucker, C. (2017, June). *The Digital Privacy Paradox: Small Money, Small Costs, Small Talk* (w23488). National Bureau of Economic Research. Cambridge, MA. Retrieved March 5, 2021, from http://www.nber.org/papers/w23488.pdf. (Cit. on p. 21)

Athey, S., & Imbens, G. W. (2019). Machine learning methods that economists should know about. *Annual Review of Economics*, *11*, 685–725 (cit. on p. 16).

Avery, C., Bossert, W., Clark, A., Ellison, G., & Ellison, S. F. (2020). An Economist's Guide to Epidemiology Models of Infectious Disease. *Journal of Economic Perspectives*, *34*(4), 79–104. Retrieved March 6, 2021, from https://pubs.aeaweb.org/doi/10.1257/jep.34.4.79 (cit. on pp. 8, 9)

Baccini, L., & Brodeur, A. (2021). Explaining Governors' Response to the COVID-19 Pandemic in the United States. *American Politics Research*, *49*(2), 215–220. Retrieved March 12, 2021, from http://journals.sagepub.com/doi/10.1177/1532673X20973453 (cit. on p. 7)

Banda, J. M., Tekumalla, R., Wang, G., Yu, J., Liu, T., Ding, Y., Artemova, K., Tutubalina, E., & Chowell, G. (2021, January 3). *A large-scale COVID-19 Twitter chatter dataset for open scientific research - an international collaboration*. Zenodo. Retrieved January 6, 2021, from https://zenodo.org/record/4414856. (Cit. on p. 17)

Bandiera, O., Prat, A., Hansen, S., & Sadun, R. (2020). Ceo behavior and firm performance. *Journal of Political Economy*, *128*(4), 1325–1369 (cit. on p. 14).

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 pp. 4, 7)

Bollen, J., Mao, H., & Zeng, X. (2011). Twitter mood predicts the stock market. *Journal of Computational Science*, *2*(1), 1–8. Retrieved January 13, 2021, from https://linkinghub.elsevier.com/retrieve/pii/S187775031100007X (cit. on p. 5)

Brotherhood, L., Kircher, P., Santos, C., & Tertilt, M. (2020). *An Economic Model of the COVID-19 Epidemic: The Importance of Testing and Age-Specific Policies* (Discussion Paper No. 13265). IZA Institute of Labor Economics. Bonn. (Cit. on pp. 5, 11).

Bursztyn, L., Rao, A., Roth, C., & Yanagizawa-Drott, D. (2020, June). *Misinformation During a Pandemic* (w27417). National Bureau of Economic Research. Cambridge, MA. Retrieved March 12, 2021, from http://www.nber.org/papers/w27417.pdf. (Cit. on p. 7)

Campos-Mercade, P., Meier, A. N., Schneider, F. H., & Wengström, E. (2021). Prosociality predicts health behaviors during the COVID-19 pandemic. *Journal of Public Economics*, *195*, 104367. Retrieved March 12, 2021, from https://www.sciencedirect.com/science/article/pii/S0047272721000037 (cit. on p. 11)

Cheng, W., Greaves, C., & Warren, M. (2006). From n-gram to skipgram to concgram. *International journal of corpus linguistics*, *11*(4), 411–433 (cit. on p. 13).

Chernozhukov, V., Kasaha, H., & Schrimpf, P. (2021). Causal Impact of Masks, Policies, Behavior on Early Covid-19 Pandemic in the U.S. *Journal of Econometrics*, *220*(1), 23–62. Retrieved March 7, 2021, from http://arxiv.org/abs/2005.14168 (cit. on p. 10)

Chiou, L., & Tucker, C. (2020, April). *Social Distancing, Internet Access and Inequality* (w26982). National Bureau of Economic Research. Cambridge, MA. Retrieved March 12, 2021, from http://www.nber.org/papers/w26982.pdf. (Cit. on pp. 6, 20, 21)

Davidson, R., & MacKinnon, J. G. (2004). *Econometric theory and methods*. Oxford University Press. (Cit. on p. 15).

Denny, M. J., & Spirling, A. (2018). Text preprocessing for unsupervised learning: Why it matters, when it misleads, and what to do about it. *Political Analysis*, *26*(2), 168–189 (cit. on pp. 13, 14).

Diamond, P. A., & Maskin, E. (1979). An equilibrium analysis of search and breach of contract, I: Steady states. *The Bell Journal of Economics*, 282–316 (cit. on p. 10).

Eichenbaum, M., Rebelo, S., & Trabandt, M. (2020, March). *The Macroeconomics of Epidemics* (w26882). National Bureau of Economic Research. Cambridge, MA. Retrieved January 20, 2021, from http://www.nber.org/papers/w26882.pdf. (Cit. on pp. 5, 10)

Farboodi, M., Jarosch, G., & Shimer, R. (2020, April). *Internal and External Effects of Social Distancing in a Pandemic* (w27059). National Bureau of Economic Research. Cambridge, MA. Retrieved January 20, 2021, from http://www.nber.org/papers/w27059.pdf. (Cit. on p. 10)

Favero, C. A., Ichino, A., & Rustichini, A. (2020). *Restarting the Economy While Saving Lives Under COVID-19* (Discussion Paper DP14464). Centre for Economic Policy Research.

London. Retrieved March 11, 2021, from https://www.ssrn.com/abstract=3580626. (Cit. on p. 11)

Fehr, E., & Schmidt, K. M. (1999). A theory of fairness, competition, and cooperation. *The quarterly journal of economics*, *114*(3), 817–868 (cit. on p. 11).

Galizzi, M. M., & Wiesen, D. (2018, March 28). Behavioral Experiments in Health Economics. *Oxford Research Encyclopedia of Economics and Finance*. Oxford University Press. Retrieved March 13, 2021, from https://oxfordre.com/economics/view/10.1093/acrefore/9780190625979.001.0001/acrefore-9780190625979-e-244. (Cit. on p. 8)

Gelman, A., & Loken, E. (2014). The Statistical Crisis in Science. *American Scientist*, *102*(6), 460–466 (cit. on p. 14).

Gentzkow, M., Kelly, B., & Taddy, M. (2019). Text as Data. *Journal of Economic Literature*, *57*(3), 535–574. Retrieved January 13, 2021, from https://pubs.aeaweb.org/doi/10.1257/jel.20181020 (cit. on pp. 5, 12, 15)

Glaeser, E., Gorback, C., & Redding, S. (2020, July). *How Much does COVID-19 Increase with Mobility? Evidence from New York and Four Other U.S. Cities* (w27519). National Bureau of Economic Research. Cambridge, MA. Retrieved October 22, 2020, from http://www.nber.org/papers/w27519.pdf. (Cit. on p. 6)

Goolsbee, A., & Syverson, C. (2021). Fear, lockdown, and diversion: Comparing drivers of pandemic economic decline 2020. *Journal of Public Economics*, *193*, 104311. Retrieved March 6, 2021, from https://linkinghub.elsevier.com/retrieve/pii/S0047272720301754 (cit. on p. 6)

Groseclose, T., & Milyo, J. (2005). A measure of media bias. *The Quarterly Journal of Economics*, *120*(4), 1191–1237 (cit. on p. 16).

Hansen, S., McMahon, M., & Prat, A. (2018). Transparency and deliberation within the FOMC: A computational linguistics approach. *The Quarterly Journal of Economics*, *133*(2), 801–870 (cit. on p. 16).

Huntington-Klein, N. (2020). *SafeGraphR* (Version 0.3.0). https://safegraphinc.github.io/SafeGraphR/. (Cit. on p. 20)

Hutto, C., & Gilbert, E. (2014). Vader: A parsimonious rule-based model for sentiment analysis of social media text. *Proceedings of the International AAAI Conference on Web and Social Media*, *8*(1) (cit. on p. 24).

Jegadeesh, N., & Wu, D. (2013). Word power: A new approach for content analysis. *Journal of financial economics*, *110*(3), 712–729 (cit. on p. 14).

Jurafsky, D., & Martin, J. H. (2009). *Speech and language processing: An introduction to natural language processing, computational linguistics, and speech recognition* (2nd ed). Pearson Prentice Hall. (Cit. on p. 14)

OCLC: 213375806.

Kermack, W., & McKendrick, A. (1927). A Contribution to the Mathematical Theory of Epidemics. *Proceedings of the Royal Society of London. Series A, Containing Papers of a Mathematical and Physical Character*, *115*(772), 700–721 (cit. on p. 8).

Lamsal, R. (2020). Design and analysis of a large-scale COVID-19 tweets dataset. *Applied Intelligence*. Retrieved January 13, 2021, from http://link.springer.com/10.1007/s10489-020-02029-z (cit. on p. 17)

Lucca, D. O., & Trebbi, F. (2009, September 17). *Measuring Central Bank Communication: An Automated Approach with Application to FOMC Statements* (w15367). National Bureau of Economic Research. Retrieved January 13, 2021, from https://www.nber.org/papers/w15367. (Cit. on p. 5)

Maloney, W., & Taskin, T. (2020, May). *Determinants of Social Distancing and Economic Activity during COVID-19: A Global View*. World Bank, Washington, DC. Retrieved March 12, 2021, from http://hdl.handle.net/10986/33754. (Cit. on p. 7)

Manning, C. D., Raghavan, P., & Schutze, H. (2008). *Introduction to Information Retrieval*. Cambridge University Press. Retrieved February 15, 2021, from http://ebooks.cambridge.org/ref/id/CBO9780511809071. (Cit. on p. 14)

Manning, C. D., & Schütze, H. (1999). *Foundations of statistical natural language processing*. MIT Press. (Cit. on p. 13).

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 pp. 4, 22, 23)

_eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/j.1467-8640.2012.00460.x

Murray, E. J. (2020). Epidemiology's Time of Need: COVID-19 Calls for Epidemic-Related Economics. *Journal of Economic Perspectives*, *34*(4), 105–120. Retrieved January 20, 2021, from https://pubs.aeaweb.org/doi/10.1257/jep.34.4.105 (cit. on p. 8)

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 pp. 4, 7)

Pástor, Ľ., & Veronesi, P. (2020). Political Cycles and Stock Returns. *Journal of Political Economy*, *128*(11), 4011–4045. Retrieved March 12, 2021, from https://www.journals.uchicago.edu/doi/abs/10.1086/710532 (cit. on p. 7)

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. (Cit. on p. 3)

Petherick, A., Kira, B., Hale, T., Phillips, T., Webster, S., Cameron-Blake, E., Hallas, L., Majumdar, S., Tatlow, H., Boby, T., & Angrist, N. (2020, December 10). *Variation in government responses to COVID-19* (Working Paper No. 2020/32). Blavatnik School of Government. Oxford. www.bsg.ox.ac.uk/covidtracker. (Cit. on p. 21)

Pew Research Center. (2019). *Demographics of Mobile Device Ownership and Adoption in the United States*. Retrieved March 5, 2021, from https://www.pewresearch.org/internet/fact-sheet/mobile/. (Cit. on p. 21)

Plutchik, R. (1980, January 1). A general psychoevolutionary theory of emotion. In H. Kellerman (Ed.), *Theories of Emotion* (pp. 3–33). Academic Press. Retrieved March 15, 2021, from https://www.sciencedirect.com/science/article/pii/B9780125587013500077. (Cit. on p. 22)

Porter, M. F. (1980). An algorithm for suffix stripping. *Program* (cit. on p. 13).

Qazi, U., Imran, M., & Ofli, F. (2020). GeoCoV19: A dataset of hundreds of millions of multilingual COVID-19 tweets with location information. *ACM SIGSPATIAL Special*, *12*(1), 6–15. Retrieved January 13, 2021, from https://dl.acm.org/doi/10.1145/3404820.3404823 (cit. on pp. 3, 17, 18)

Rinker, T. (2018). *Sentimentr*. Retrieved March 15, 2021, from https://github.com/trinker/sentimentr. (Cit. on p. 23)

Roehrick, K. (2020, September 7). *Vader: Valence Aware Dictionary and sEntiment Reasoner (VADER)* (Version 0.2.1). Retrieved March 16, 2021, from https://CRAN.R-project.org/package=vader. (Cit. on p. 24)

SafeGraph, Inc. (2020). *Social Distancing Metrics*. SafeGraph. Retrieved January 7, 2021, from https://docs.safegraph.com/docs/social-distancing-metrics. (Cit. on pp. 3, 20)

Silge, J., & Robinson, D. (2017). *Text mining with R: A tidy approach* (First edition). O'Reilly. (Cit. on p. 23)

OCLC: ocn993582128.

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 pp. 4, 7)

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. 5)

Summers, E. (2020). *Hydrator*. https://github.com/docnow/hydrator. (Cit. on p. 19)

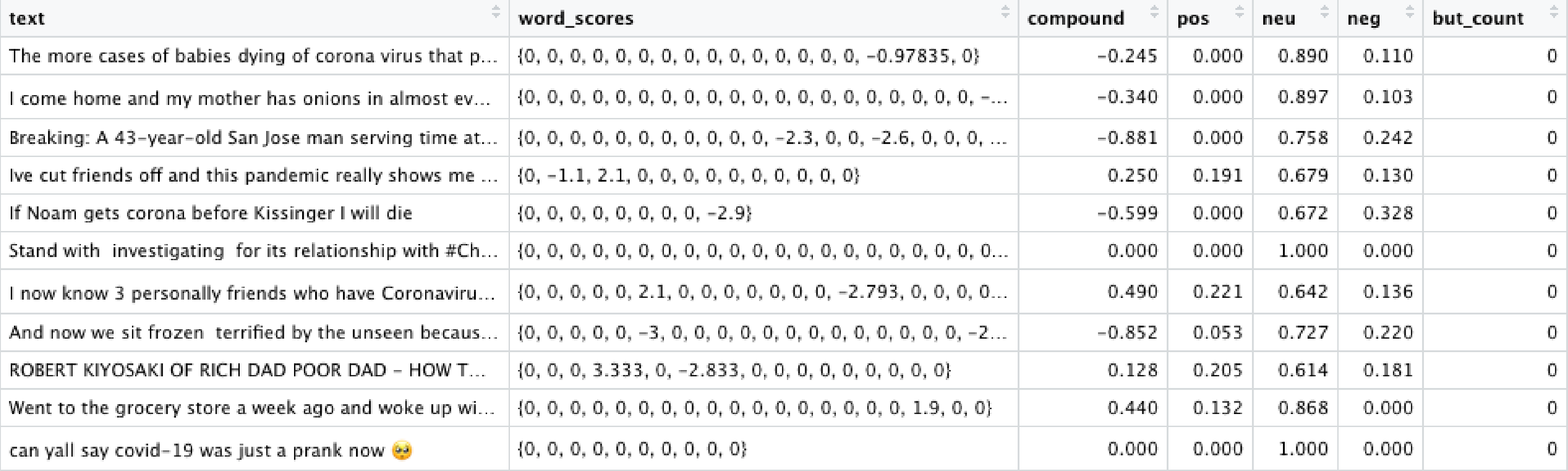
Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society: Series B (Methodological)*, *58*(1), 267–288 (cit. on p. 16).

Toxvaerd, F. (2020). *Equilibrium Social Distancing* (Working Paper No. 2020/08). Institute for New Economic Thinking. Cambridge. (Cit. on p. 9).

Twitter, Inc. (2021). *Data dictionary: Standard v1.1*. Twitter Developer Documentation. Retrieved March 4, 2021, from https://developer.twitter.com/en/docs/twitter-api/v1/data-dictionary/overview. (Cit. on p. 19)

Wright, A. L., Sonin, K., Driscoll, J., & Wilson, J. (2020). Poverty and economic dislocation reduce compliance with COVID-19 shelter-in-place protocols. *Journal of Economic Behavior & Organization*, *180*, 544–554. Retrieved March 12, 2021, from https://www.sciencedirect.com/science/article/pii/S0167268120303760 (cit. on p. 6)

A Additional Figures



text	word_scores	compound	pos	neu	neg	but_count
The more cases of babies dying of corona virus that p…	{0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, -0.97835, 0}	-0.245	0.000	0.890	0.110	0
I come home and my mother has onions in almost ev…	{0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, -…	-0.340	0.000	0.897	0.103	0
Breaking: A 43-year-old San Jose man serving time at…	{0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, -2.3, 0, 0, -2.6, 0, 0, 0, …	-0.881	0.000	0.758	0.242	0
Ive cut friends off and this pandemic really shows me …	{0, -1.1, 2.1, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0}	0.250	0.191	0.679	0.130	0
If Noam gets corona before Kissinger I will die	{0, 0, 0, 0, 0, 0, 0, 0, -2.9}	-0.599	0.000	0.672	0.328	0
Stand with investigating for its relationship with #Ch…	{0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0…	0.000	0.000	1.000	0.000	0
I now know 3 personally friends who have Coronaviru…	{0, 0, 0, 0, 0, 2.1, 0, 0, 0, 0, 0, 0, 0, -2.793, 0, 0, 0, 0…	0.490	0.221	0.642	0.136	0
And now we sit frozen terrified by the unseen becaus…	{0, 0, 0, 0, 0, -3, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, -2…	-0.852	0.053	0.727	0.220	0
ROBERT KIYOSAKI OF RICH DAD POOR DAD - HOW T…	{0, 0, 0, 3.333, 0, -2.833, 0, 0, 0, 0, 0, 0, 0, 0, 0}	0.128	0.205	0.614	0.181	0
Went to the grocery store a week ago and woke up wi…	{0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 1.9, 0, 0}	0.440	0.132	0.868	0.000	0
can yall say covid-19 was just a prank now 🥺	{0, 0, 0, 0, 0, 0, 0, 0, 0, 0}	0.000	0.000	1.000	0.000	0

Figure 3: Example sentiment analysis output of VADER package.