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

Z0983692*

April 8, 2021

Abstract

How much does public sentiment affect social distancing? I investigate this question by exploiting a very large panel of tweets. Using geolocation data embedded in the tweets, I create an index of public sentiment towards COVID-19 at the county level, and use this index to estimate the extent to which local sentiment influences social distancing. Using a two-way fixed-effects design over two separate measures of public sentiment and two separate measures of social distancing, I find consistent evidence that an increase in negative or fearful sentiment about COVID-19 decreases the level of community mobility.

`13196 words in main body, excluding headers and bibliography.`

*I thank my advisor, Angel Hernando-Veciana, for his upbeat encouragement and his interest in my research project. I thank my family for their unwavering support (both on FaceTime and in person), and finally I thank my dog Truffle for her excellent research assistance and willingness to take lunchtime walks.

Contents

List of Figures

List of Tables

1 Introduction

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 drastically changed their habits of working, socialising, and travelling: however, even when faced with the same objective information, the extent to which they did so varied substantially by area. Economists have taken an interest in what caused this county-level variation in social distancing: because the risk of infection is a function of the behaviour of other people in your neighbourhood, understanding what drives neighbourhood behaviour is crucial for minimising disease spread. I show that public sentiment towards COVID is a central determinant of local social distancing. Sentiment varies individually, but also does so at a local level: this dissertation looks to prove that this local public sentiment is a key factor in determining mobility in a given area.

A common medium for expressing sentiment is social media: by calculating the sentiments expressed online in a particular area, we can construct a time-varying index of public sentiment. I aggregate individual expressions of risk attitude towards COVID into a common measure of public sentiment. Under the assumption that this aggregate of individual online sentiments is representative of the sentiment in the area, this dissertation tests the impact of local expressions of health sentiment on public behaviour in the early months of the pandemic. Specifically, I study whether more risk-averse sentiment towards COVID on Twitter is linked to increased social distancing behaviour at the county/week level.

This dissertation contributes to the recent economics literature on the COVID-19 pandemic. Primarily, it investigates the relationship between mobility and health sentiment; broadly, therefore, it contributes to the empirical literature investigating the various determinants of social distancing among counties. Previous papers have established that changes in social distancing pre-dated policies: as such, understanding determinants of behaviour (rather than reactions to policies) is important. I show that risk preferences play a central role in determining mobility, and moreover that local sentiment has a separate and significant impact on social distancing. More broadly, the paper presents a novel example of economic inference from social media using text analysis. In short, the dissertation establishes a link between an index of local sentiment towards COVID-19 and the level of social distancing in that area. This has useful policy applications: measuring consumer sentiment is a common barometer for macroeconomic activity, and this dissertation contributes new methods and data towards creating an analogous measure of local COVID sentiment, which remains important for economic policy in the near future.

I chose Twitter as the platform for measuring public sentiment. Twitter has over 80 million monthly active users in the US, meaning that 22% of US adults use the platform, with 42% of these using it on a daily basis (Perrin & Anderson, 2019). I exploit GeoCov19 (Qazi et al., 2020), a dataset of 524 million geolocated tweets, to construct the index of local public sentiment. The tweets cover the period from

1st February to 1st May. 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 are targeted towards measuring COVID sentiment: they were collected by searching for tweets containing any of a list of 800 COVID-related keywords. I use anonymous smartphone location data, collected by the company SafeGraph, as a measure of the extent of social distancing in an area (SafeGraph, Inc., 2020).

Starting from GeoCoV19's US-located tweets, I use two methods to assess the level of risk sentiment: a dictionary-based text analysis and a more sophisticated, social media-calibrated sentiment extractor. In the first measure, I assign tweets containing fear-associated words to a risk-averse sentiment; in the second, I assign a score of positive or negative sentiment to each tweet. The base unit of analysis is the county-week; as such, the two measures of health sentiment are the proportion of COIVD-related tweets that contain fearful language, and the average negativity expressed in the COVID-related tweets, in each county and week. Since the first measure is more specific, and plausibly relates to risk preferences, it is the measure of interest; for robustness, I show that the second, alternative measure of sentiment produces comparable econometric results.

It is plausible that social media is a valid measure for health risk sentiment. 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. The key aspect to the data is that the Twitter conversations provide a real-time insight into local sentiment; this sentiment then acts as a determinant of social distancing.

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 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 empirically confirms 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, sentiment mining, and social media in Economics

Ultimately, text is the most popular medium for encoding and recording information. As such, it is the medium for a very large amount of data which could be useful to economic research. The limiting factor to text has been the fact that this information is recorded in a noisy, high-dimensional format; however, recent advances in processing techniques have made it possible to extract the relevant information from text. This section reviews the previous applications of sentiment mining in economic research, putting the dictionary-based sentiment mining used in this dissertation in context.

Broadly speaking, sentiment is the disposition of an entity (either an individual or a collective) towards a topic (Algaba et al., 2020). This disposition has a *polarity*: in the general sense, this corresponds to whether the disposition is positive or negative, but can also reflect domain-specific concepts like Dovish/Hawkish (Picault & Renault, 2017) or Democrat/Republican (Gentzkow & Shapiro, 2010). In this way, polarity defines the qualitative notion of disposition as a quantitative measure; this process is described in detail in section 3.

For economists, quantitatively defining sentiment helps to measure economic decision-making: sentiments in personal communications might convey information about the author's expectations and preferences, and sentiments expressed in instituational communications like company filings and news reports might convey information about fundamental economic factors. The concept of polarity is particularly suited to measuring and predicting the impact of confidence in financial markets and other macroeconomic entities; as such, sentiment analysis is a particularly well-established tool in finance research. Antweiler and Frank (2004) and Tetlock (2007) are prominent examples: both use sentiment polarity as a proxy for investor confidence, analysing news articles and internet message boards respectively. A closely-related application is measuring latent sentiment in the population. For example, Baker et al. (2016) show that uncertainty about fiscal, regulatory, and monetary policy has a significant impact on economic activity, and measure this uncertainty by creating an index of the sentiment expressed in newspaper articles. This conception of sentiment is the closest analogue to the health sentiment I estimate.

Economists have also causally linked sentiment expressed in texts to observed behaviours. Stephens-Davidowitz (2014) is a prominent example of linking county-level variation in sentiment to behaviour, as are Choi and Varian (2012) and Saiz and Simonsohn (2013). Ginsberg et al. (2009) does not measure sentiment, but is one of many examples which use Google searches for influenza symptoms to predict local outbreaks of disease; this dissertation extends this project by treating sentiment as a local-level health indicator.

Social media has been used as an indicator for sentiment, particularly in finance. Similar to the empirical strategy of this dissertation, Affuso and Lahtinen (2019) uses the sentiment of geolocated Tweets to measure investor sentiment in different regions; Bollen et al. (2011) is a particularly widely-cited example of using Twitter sentiment to predict stock market returns. Daas and Puts (2014) connects Twitter and Facebook sentiment data to Dutch consumer confidence indices, while K. Müller and Schwarz (2020) uses geolocated Facebook posts to construct a measure of anti-refugee sentiment, which strongly predicts hate crimes. Finally, Gorodnichenko et al. (2018) finds that geolocated tweets about Brexit are a strong predictor of Leave vote-share.

In conclusion, sentiments embedded in social media and other forms of text are a powerful tool in uncovering the determinants of individual behaviour. This dissertation asserts that Twitter sentiment can predict social distancing behaviour: the next section reviews the economic literature on the dynamics and determinants of COVID-19 related behaviour.

2.2 COVID-19: Individual and demographic determinants of 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. 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. Fan et al. (2020) finds systemic differences in SafeGraph-measured mobility along lines of gender, income, and political affiliation at the county level; they also find significant spatial heterogeneity in beliefs and preferences about COVID-19.

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[1]. 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[2]. 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 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.

More general research has also been done on health behaviours in the context of the COVID-19 pandemic, which does not include mobility as an observed outcome; rather, these papers identify individual-

[1]They control for disease spread by including deaths in the state as a covariate in their OLS specification.
[2]This result is generalised by Pástor and Veronesi (2020), who find that Democrats are more risk averse than Republicans.

level determinants of pandemic-related health behaviour. First, S. Müller and Rau (2021) uses a survey to elicit preferences over risk, time, trust, and honesty; this is linked to self-reported policy compliance (stay-at-home behaviour, social distancing, and testing), perception of the crisis (panic buying and reported fear of COVID-19), and pre-pandemic social altrusim (fare-dodging, agreement for compulsory measles vaccines). They find that risk aversion, patience, and social responsibility predicts greater social compliance and stay-at-home behaviour, and observe that individuals exhibiting present bias engage in more panic buying. They also find that risk tolerance significantly decreases a subject's reported fear of COVID-19. Another survey-based study is Thunström et al. (2021). The authors report that fear for one's own health, extraversion, proclivity for policy compliance, and age are primary determinants of whether respondents would attend a costly COVID test. Briscese et al. (2020) also study determinants of policy compliance, and find that expectations about policies serve as a reference point for future compliance efforts. Respondents are less likely to comply with a stay-at-home order if the announced duration is longer than expected, and individuals grow more impatient over time. Finally, Zettler et al. (forthcoming) finds that individual perceptions in the form of personality traits affect reported compliance behaviours: they find that prosocial traits are a significant predictor of compliant health behaviour[3].

Past research in health economics has also addressed the determinants and motivations of health behaviours. First, individual risk preferences matter: Anderson and Mellor (2008) measures risk preference and health behaviour on a large scale with a lottery choice experiment and survey, finding that risk aversion has a negative association with risky health behaviours like smoking and drinking[4]. Second, time preferences matter for determining health behaviour: Chapman and Coups (1999) observe in a randomised trial that individuals with lower discount rates with respect to money and health are more likely to take up a free influenza vaccine, with Chesson et al. (2006) and Harrison et al. (2010) reaching equivalent results for smoking and risky sexual behaviour respectively[5]. Finally, personality traits can be a determinant of health behaviour: Rustichini et al. (2016) find that personality traits and indiviudal preferences are strongly related and predict health behaviours[6].

We have established that several dimensions of individual preference predict social distancing behaviours. There is also some substantial evidence supporting the notion that these preferences vary by region. Falk et al. (2018) conducted the 'Global Preference Survey' of time and risk preference, altruism, and trust over 80,000 people in 76 countries. They found heterogeneity across countries, and even greater within-country heterogeneity. The heterogeneity in preferences tracked with individual characteristics like age and gender, but a substantial proportion of the variation is attributed purely to a region effect,

[3] See also, e.g. Almås et al. (2019) and Herrmann et al. (2008)
[4] See also, e.g. Dohmen et al. (2011).
[5] See also, e.g. DellaVigna and Malmendier (2006), Schilbach (2019), Stutzer and Meier (2016), and Sutter et al. (2013).
[6] See also, e.g. Booth-Kewley and Vickers (1994) and Strickhouser et al. (2017).

perhaps in the form of culture and bio-geographic variables. As mentioned above, Fan et al. (2020) also supports the notion of spatial heterogeneity, documenting county-level substantial variation in preferences and COVID-related beliefs.

In summary, the general health economics literature and COVID-related work has established that time preferences, risk preferences, pro-social tendencies, and fear of COVID are strong individual determinants of pandemic compliance behaviours. Additionally, there is strong evidence for spatial heterogeneity in these individual preferences and traits. This last point in particular is key for the identification of the empirical specification, since I argue that the research on spatial heterogeneity of preferences allows for an empirical design that measures individual characteristics on the aggregate level.

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

Models of the relationship between human behaviour and viral spread are necessary for understanding the relationship between individual-level health decisions and community-level behaviours and health outcomes. Many economists have looked to augment epidemiological models with robust models of human behaviour; this section summarises the models of social distancing relevant to my empirical specification.

2.3.1 Standard SIR model

Kermack and McKendrick (1927) introduced the SIR model and it remains the basis for most modern epidemiological models; a mathematical exposition is in appendix C. In short, both government policies and behaviour change have the effect of reducing the R_0 infectivity 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:

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

, where $S(t)$ is the fraction of the population susceptible to infection. 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.

2.3.2 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 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 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. 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 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.

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.

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.

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 to expose 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. The empirical framework uses the theoretical insight set out in this section that the aggregate of individual preferences generates a certain level of social distancing behaviour.

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 summarise the inference problem for text analysis of English documents, and the approach I take to tackle it.

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.[7] $\mathbf{C}_{ij}$ is a scalar representing the *term frequency*: how many times each token appears in a document. In other words, $\mathbf{C}$ 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

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

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). I use a simple dictionary-based method for emotion extraction, and an augmented rule-based method for polarity analysis[8].

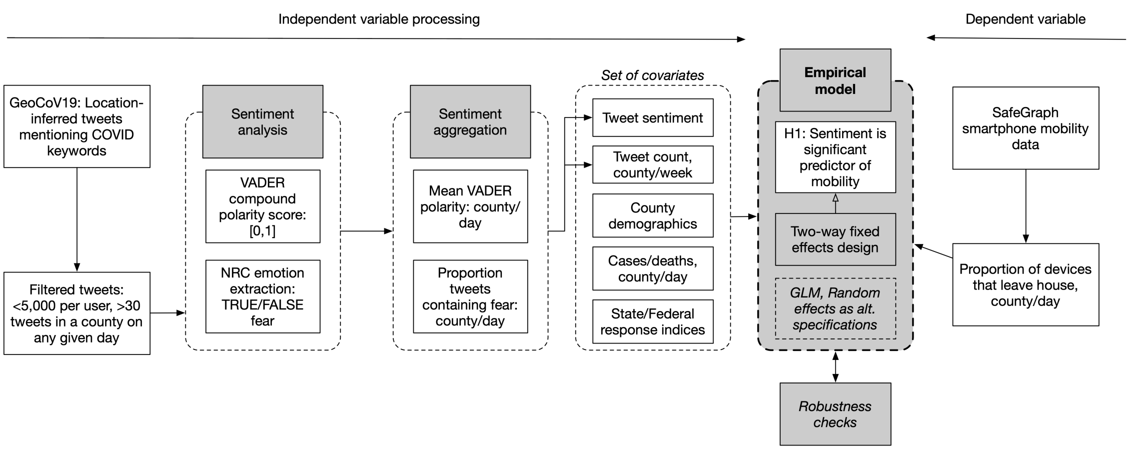
Dictionary matching is the simplest way to perform inference on a latent variable in text. It is also the most well-established method used in the economics and finance literature. In a nutshell, the method is to use a known function $f(\cdot)$ to stipulate $\hat{\mathbf{v}}_i = f(\mathbf{c}_i)$. No information is inferred from the language corpus; rather, known attributes of $\mathbf{c}_i$, established by linguistic and domain-specific research, are leveraged. This

[8] The first corresponds to the NRC approach; the second, the VADER approach.

is the method that I take, which is explained in detail in section 5.4. Beyond computational ease, 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.

4 Theoretical and Empirical Framework

Figure 1: Overview of data collection, processing, and estimation.



4.1 Central Argument and Empirical Specficiation

In this dissertation, the causal relationship of interest is the impact of public sentiment on social distancing. Suppose that risk preferences vary by time and region, a notion backed up by the Global Preferences Survey (Falk et al., 2018) and applied to the COVID-19 pandemic by Fan et al. (2020). Suppose further that a portion of individual-level risk preferences is time variant and dependent on external factors. Call this sentiment, and that this sentiment is expressed by what that individual says on social media; as such, those with risk-averse sentiment will post more negatively about COVID-19. The location of these users indicates a level of risk preference that is shared among those in the local area: call this public sentiment[9]. With a sufficiently large sample of tweets about COVID-19, the aggregate of individual sentiments is an estimator for the latent variable of public sentiment. This public sentiment arises because

[9]Specifically, public sentiment *towards COVID-19*.

complex combinations of local factors – factors such as the trust in that particular area's local government, the pattern of supply shortages in that area's supermarket, the quality of public transport, and the commuting practises of the area – all contribute to the shared local perception of pandemic risk. In summary, this shared perception can be estimated by the local chatter on social media, and I measure it by creating a county-level daily index of social media sentiment towards COVID-19. We are interested in the relationship between this county index and the level of social distancing that an area practises. The causal relationship is that individual sentiment towards COVID-19 causes different levels of mobility; we observe this relationship at the county/day level under the assumption that the corresponding aggregates of individual-level sentiments and mobility stand in an analogous relationship.

To that end, I estimate the following specification, where a vector of social distancing metrics in county c of state s, on date t and week w is a function of

$$SD_{ct} = \beta_1 Sent_{ct} + \beta_2 StateStringency_{st} + \beta_3 NatStringency_t + \beta_4 Deaths_{ct} + \gamma_c + \delta_w \tag{2}$$

. The sentiment effect is captured by β_1, policy effects by β_2 and β_3, and local health risk and disease spread by β_4. This primary specification captures the local effects of sentiment, as discussed above; in order to account for the Februrary selection effect discussed in section 5.1, where far lower COVID tweets observed in Februrary could indicate only those fearful of the virus tweeting about it, we include the number of tweets observed in the county on that week as a covariate. Including this possible confounder in the regression allows for the effect of low observations to be separated from the true effect of tweet sentiment. As such, our primary hypothesis has the first and second coefficients as the effect of interest such that

$$H_0 : \beta_1 = \beta_2 = 0$$

$$H_1 : \beta_1 \neq \beta_2 \neq 0$$

, where β_1 has the appropriate sign according to its metric, $\beta_2 > 0$, and the coefficients are jointly and individually significant.

The third and fourth coefficients cover time-varying factors like the imposition of policy orders and the spread of the virus. β_4 represents not only viral spread, but can act as a proxy for perceived local health risk: this follows Chernozhukov et al. (2021), in whose model COVID death counts represent the time-varying local and national beliefs on the seriousness of the pandemic. It is important to note that a central identifying assumption in the two-way fixed effects design is that there is no source of time-varying, unobserved heterogeneity: thus it is important to verify that the correct specification is chosen for these time-varying confounders. Finally, γ_c is a vector of county-level fixed effects, capturing all constant demographic variables like percent above 65, income, commuting patterns, etc.; δ_w is a vector of week fixed effects, capturing among others the national-level disease spread and health risk perception. All regressions have two-way White heteroscedasticity-robust standard errors, clustered at the county and week level following (Abadie et al., 2017).

4.1.1 Variables

We have estimated two measures of tweet sentiment. Both of the measures estimate the sentiment of individual tweets, $\hat{\mathbf{V}}$, from the raw text $\mathbf{C}$. For the NRC lexicon, this is a binary measure: 'tweet contains fear-related words' is true or false. For each county in each day, a proportion of the tweets contain fear-related words. This proportion is the reported variable. Similarly, the VADER measure is a bounded index over $[-1, 1]$, but does not arise directly from count data. The mean of this measure is taken over each county and day, which forms the aggregation procedure. AtHome is an indicator variable for whether the state had imposed a stay-at-home order; and NatStringency reports the the Oxford stringency index, which is normalised to between 0 and 100. Finally, the deaths in each county are included as a time-varying factor; as discussed above, these offer a less biased estimate of the spread of the virus than reported cases, which vary by region and do not include asymptomatic cases. The dependent variables are two mobility measures: the median time spent at home during working hours (expressed as a percentage), and the proportion of measured devices that stay at home all day.

4.2 Inference with Panel Fixed Effects

The chosen model is a two-way, time and location fixed effects design. This empirical model is very common with a panel dataset measured at the county level, as the location fixed effect flexibly controls for all unobserved, time-invariant factors in each county. Moreover, the time fixed effect accounts for time-varying factors that are common to all counties – like the progression of the virus on the world stage, and the level of 'shared' sentiment towards COVID across the U.S.

The fixed effects model is characterised by the linear model

$$E[Y_{it}|A_i, X_{it}, t] = \alpha + \lambda_t + A_i'\gamma + X_{it}'\beta + \rho D_{it} \tag{3}$$

, where Y_{it} is the outcome, α the intercept, λ_t the time effect, $A_i'\gamma$ unobserved, fixed confounders, $X_{it}'\beta$ observed time-varying confounders, and ρD_{it} the variable of interest (Angrist & Pischke, 2009, p. 222). The assumptions are that the effect of ρ is additive and constant, and that the unobserved fixed effects do not vary with time[10]. In our case, both assumptions are plausible. This directly implies the model

$$Y_{it} = \alpha_i + \lambda_t + \rho D_{it} + X_{it}'\beta + \epsilon_{it}, \text{, where} \tag{4}$$

$$\alpha_i \equiv \alpha + A_i'\gamma \tag{5}$$

. In this way we estimate the (potentially large number of) unobserved confounders as the *fixed effect* α_i, which enters into the model as a coefficient on an indicator for each region, and also include a *time effect* λ_t.

4.2.1 Identification

With these effects accounted for, we are able to obtain a consistent estimator for the overall, shared effect of our variable of interest, under the assumption that the time-varying confounders are adequately specified. Note that although we have specified a causal relationship of interest – that risk-averse sentiment causes greater social distancing – we do not have random assignment or exogenous variation of risk-averse sentiment, the treatment variable. In order to establish causality in a fixed effects design we must establish strict exogeneity of social media sentiment and social distancing. In particular, there should be plausibly no simultaneity or reverse causality: mobility cannot affect social media sentiment, and the two should not be jointly determined. Conversely, we can say that Stable Unit Treatment Value assumption is satisfied: there are plausibly no unit spillovers of sentiment, and given that sentiments expressed on a given week are acted on in that same week(a plausible notion), there are no time spillovers either. Unfortuantely, reverse causality is perhaps possible: social distancing might cause frustration which is expressed in social media sentiment; however, this is likely to be very highly correlated with time and state stringency, both of which we control for. Despite this, we cannot rule out the existence of there being a reverse causal relationship.

In conclusion, although I go some way towards establishing a causal link, because of the lack of exogenous variation and our inability to rule out reverse causality, we must conclude that causal conclusions are not warranted. The objective of this research design is therefore to establish a plausible correlation between sentiment and social distancing.

[10]That is, the omission of a time subscript on $A_i'\gamma$ is accurate.

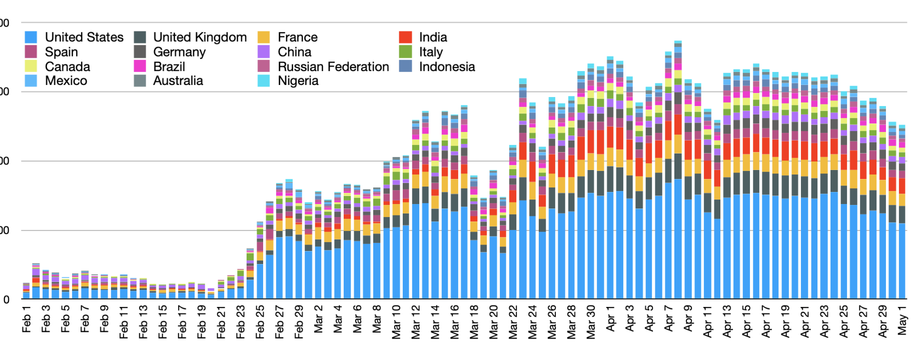
5 Data

5.1 GeoCoV19 and geolocation inference of Twitter datasets

People use Twitter to share text, images, and links to websites, often using hashtags to link their tweets to topics. The retweet function enables them to copy another user's tweet to their profile and share it with their followers. 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.

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. As such, we are able to observe the population of tweets about the search terms; an identification assumption is that this population is representative of the population of the US as a whole. This is plausible, but a drawback to the design is that it could select for age and technology access. The terms were chosen to cover a broad base of COVID-related talk, including searches for symptoms (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 listed at `https://crisisnlp.qcri.org/covid19`. 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. As we observe the population of tweets, there is no attrition or missing data, and imputations were not necessary. Note that the volume of tweets captured

Figure 2: Daily distribution by country of GeoCoV19 tweets, Feb 1st to May 1st, 2020 (Qazi et al., 2020)

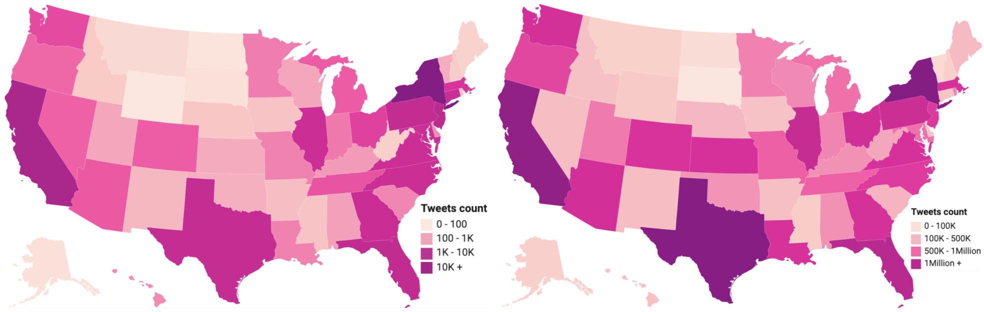


in the dataset increase very substantially over February and early March, levelling off in mid-March. This gives rise to a potential selection problem, that risk-seeking individuals do not tweet about COVID in February and bias the dataset. This concern is addressed in the main specification, which is discussed in section 4.

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), and against the 'World Cities Database'[11], 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.

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

Figure 3: Geographic distribution of GeoCoV19 tweets and users (Qazi et al., 2020)



(a) USA: Organic geotagged tweets (b) USA: User distribution

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 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.

Next, two data cleaning steps were performed. First, it was necessary to drop accounts containing machine-generated content. On Twitter, these are 'spambots': accounts that tweet tens of thousands of times per day. The text generated by these accounts does not accurately reflect the local area sentiment, and so must be dropped. We do this by filtering the tweet dataset by user ID, and dropping all tweets by accounts which have posted more than 5,000 tweets over the three-month period. This trims the outlier accounts while keeping the large proportion of the dataset.

Language processing was then performed on the tweet texts, as is detailed in section 5.4.

5.2 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[12] by comparing, for each census block, the observed proportion of all devices to that census block's proportion of total US population[13]. 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

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

[13] 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

problem is largely circumvented.

5.3 Other datasets

5.3.1 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 enacted by blanket state-level orders.

5.3.2 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.

5.3.3 COVID Case and Death Counts

COVID-19 case and death counts are provided by the Johns Hopkins University dashboard Dong et al. (2020); we report cases and death counts both cumulatively and daily. In the main econometric specification, deaths are used over reported cases; it is widely accepted that there was a substantial amount of measurement error in the early pandemic, with under-reporting due to low testing and variation in testing capacity over times and regions. Deaths are more accurately and consistently reported, although some measurement error may remain, particularly in February, where some individuals may have died in hospital without being tested for COVID-19. COVID deaths which occur outside of hospital in the early pandemic are also unlikely to have been counted; however, both of these populations are likely to be very small.

5.4 Text analysis: approach taken

5.4.1 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[14]; 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 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.

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

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

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

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[15], giving a binary factor for whether words associated with fear appear in the tweet text. Tf-idf scaling was not performed.

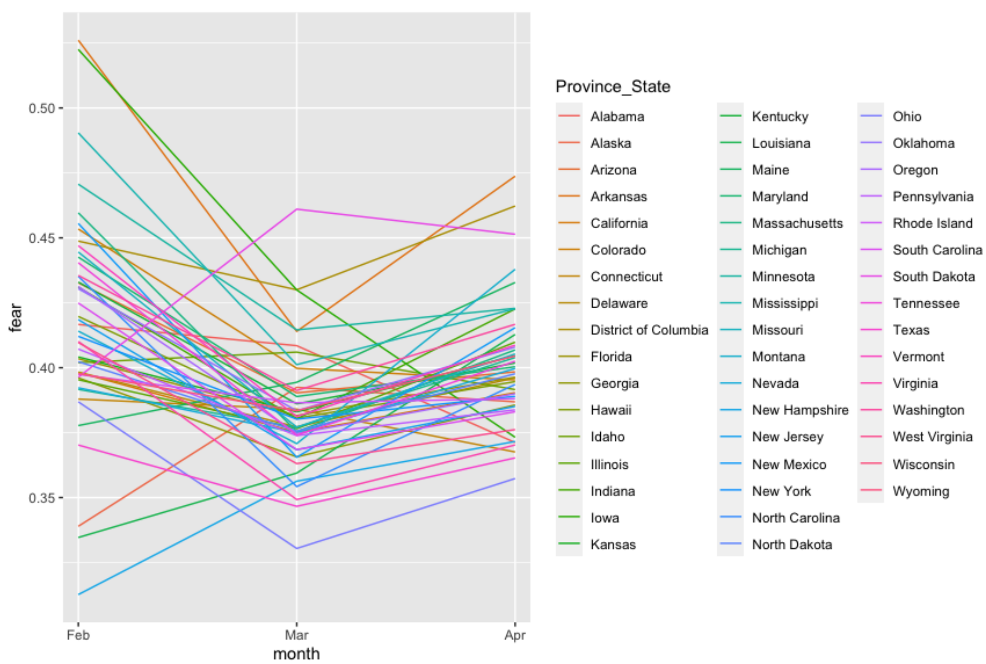
5.4.2 Tweet aggregation: Central Limit Theorem applied to repeated Bernouilli trials

For both sentiment estimation measures, the sentiment scores for the tweets were aggregated to the county level. The NRC score was aggregated by calculating the proportion of tweets in a county on a given day that contained fearful language. If we assume that each tweet is an independent 'draw' from the local population, this proportion follows a binomial distribution, as the variable represents the number of successes – tweets containing fearful language – on a sequence of independent Bernoulli trials (tweets in a location on a particular day). We know from the Central Limit Theorem that a binomial distribution

[15] 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.

approximates a normal distribution with repeated trials; however, for this to be the case, the number of trials on a given day has to be sufficiently large. This means that on days where the number of observed tweets is low, the proportion of fearful tweets in the observed sample will not be an accurate estimator for the population value. Take the extreme example: in a county where there is one observed tweet on a given day, the value of the proportion can only be 0 or 1. With repeated sampling, the proportion will converge to the population proportion, but this N has to be sufficiently large. This effect gives rise to a shape of the histogram of county-day proportions for the whole dataset, where we observe an increase in density at 0 and 1. The central limit theorem does not specify a lower bound sample size to alleviate this issue, but 30 is a common approximation. Hence, in order to rectify the inaccurate estimators for small-n county/days, we exclude county/days where the number of tweets is below 30. In doing so, we discard around 8.1% of the data. This procedure, however, means that the panel is unbalanced based on tweet count; this possible selection effect is discussed in section 6.2. The histogram and quantile plots for the fear proportion variable are figures 6 and 7 in appendix B: we observe that, although the tails are heavy, the distribution is a fair approximation to normal.

Figure 4: Proportion of fearful tweets, averaged at state/month level.



As we observe in figure 4, the variation in sentiment is greater earlier in the pandemic in February than later on in March and April. The figure indicates that expressed fear *decreased* on the aggregate during March, followed by an increase in April. This may be dubious, as we might expect fearful tweets to increase during the critical March period: however, this graph does not take covariates and county-level effects into account. The variation in sentiment observed across states in February is evidence counting against the existence of an early-stage selection effect, where only those fearful of the virus tweet about it early on: if this were the case, then we would expect that all states' fear levels should be high in February,given that the virus had not substantially spread anywhere in the US at that point.

5.4.3 VADER Rule-Based Sentiment Approach

A key identifying assumption of the research design is that the sentiment procedure accurately reflects the latent variable of 'local sentiment'. Since sentiment is fundamentally unobserved this assumption is difficult to robustly test, but a strong signal for accuracy is if the results can be replicated using different ways of measuring sentiment. Additionally, the NRC method has certain drawbacks: although it has the advantage of yielding emotion scores, allowing the analysis to pick out the 'fear' dimension of each tweet text, it lacks sophistication. In 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, Hutto and Gilbert (2014)) is a popular tool for Twitter-based sentiment analysis, and both addresses our sophistication concerns and allows us to test the validity of the identifying assumption.

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 is a binary measure of the disposition of the author towards the topic of the text, and is either positive, neutral, or negative. The polarity score can also reflect the intensity of the disposition; for example, 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 B.

The aggregation procedure for the VADER package is broadly similar to the NRC lexicon: the mean for the individual tweet sentiment is calculated at the county/day level, with counts below thirty dropped from the dataset. Visually, the compound measure is symmetrical around 0, with other modes at around ± 0.5. The distribution of the county/day mean sentiment is skew, centered around 0.1[16].

5.5 Descriptive statistics

The primary dependent variable is '% devices at home'; that is, the proportion of devices on a given county-day which stay completely at home. It has a mean of 37%; as expected, it is strongly correlated with time. . Tweets per week and device count have a very large range, since these track with county population (which is very variable, unlike voting districts: Los Angeles county has 10 million residents, for example). However both variables show little per capita variation, pointing to consistent sampling across the whole of the US: in this way, sampling bias on these variables is unlikely. Deaths and cases are path-dependent, as New York City, California, and Seattle acquired serious outbreaks earlier than other areas. Data is reported for all 3,142 counties over the 90 days between February 1 and May 1 2020.

[16]See figures 8, 9 in Appendix B.

Table 2: Descriptive statistics, all US counties, Feb 1 - May 1 2020

	Mean	Std Dev	Min	Max
Deaths, cumulative	58.6	211.6	0	4710
Cases, cumulative	1374.9	3701.8	0	53032
% 65+	13.5	3.6	6.5	51.6
Persons per HH	2.5	0.23	2	4
HH Income (000s)	55.77	12.73	20.97	122.24
% Below poverty level	16.3	4.4	3.6	39.2
Population per sq. mile	2593	5560	0.2	69468
National Stringency Index	60.89	22.78	0	72.69
State Stringency Index	56.45	25.68	0	87.96
Tweets per week	10546	14560	41	55263
Tweets per week per capita	0.02	0.04	0.0002	0.39
Device Count	52913	59662	72	344342
% Devices at home	37.46	8.96	11.50	63.11
Minutes spent away from home	38889	68469	11	803271
% Fearful tweets (NRC)	39.3	5.9	0	1
Negative sentiment score (VADER)	0.07	0.01	0	0.25

All variables reported at county level. Stringency indices bounded between 0 and 100. Household income as measured in 2013 US dollars, unadjusted.

6 Results

6.1 Main Specification

The basic message of the main specification can be seen in figure 10 in appendix B, which shows that counties with a fear sentiment above the median value have 3 percentage points more devices staying at home. In the main specification as in equation 2, there are two sentiment measures: `fearprop`, the NRC-derived proportion of fearful tweets, and `negscore`, the VADER-derived negativity score of individual tweets[17]. The two dependent variables are the median minutes spent outside home, and the proportion of devices that do not leave home all day. Models 1 and 2 report the NRC sentiment measure, while 3 and 4 report the VADER sentiment measure; models 1 and 3 report the median minutes mobility measure, while 2 and 4 report the proportion-home measure. In this specification, the sentiment scores are aggregated to the county-day level, as discussed in section 5.4.2; after data trimming, we estimate the model on a balanced panel of 16,474 observations over 597 counties and 14 weeks. All regression output was computed with the R package `fixest` (Bergé, 2018). First, we estimate the most parsimonious model, a linear model with OLS and no fixed effects. For space, I present a trimmed single model of the four models, reporting proportion fearful tweets and proportion of devices at home as the sentiment

[17] `negscore` reports results as a number between 0 and 1, so the higher the score, the more negative the tweet.

and social distancing variables[18]: for the full model, including coefficients for all controls, see table 5 in the appendix. We find that the sentiment score is a significant predictor of mobility, with the correct

Table 3: OLS Results for the effect of fearful tweets on social distancing

Dependent Variable:	Proportion devices at home	
Variables		
(Intercept)	0.1865***	0.0625***
	(0.0024)	(0.0079)
Proportion fearful tweets	0.0522***	0.0484***
	(0.0053)	(0.0048)
Nat Stringency	0.0010***	0.0012***
	(5×10^{-5})	(4.54×10^{-5})
State Stringency	0.0015***	0.0014***
	(4.63×10^{-5})	(4.21×10^{-5})
Deaths	-0.0001***	-5.47×10^{-5}***
	(6.24×10^{-6})	(5.81×10^{-6})
Cases	1.33×10^{-5}***	8.18×10^{-6}***
	(4.17×10^{-7})	(3.89×10^{-7})
Controls	No	Yes
Fit statistics		
Observations	16,474	16,474
R^2	0.61708	0.68826
Adjusted R^2	0.61697	0.68805

Normal standard-errors in parentheses
*Signif. Codes: ***: 0.01, **: 0.05, *: 0.1*

Trimmed specficiation, reporting proportion fearful tweets and NRC fear sentiment variable. With and without controls.

hypothesised sign (greater proportion of fearful tweets increases the proportion of devices at home); this verifies that a relationship does exist and gives the basis for a more sophisticated fixed effects design. Adding controls to the equation does not change the sign or significance of the treatment variable, but does reduce the standard error. This is an encouraging sign for the robustness of the specification. With linear OLS, however, there may be unobserved, time-invariant factors which may confound mobility and the sentiment measure; we use a two-way fixed effects design to account for this.

We see that in models 2, and 4, the sentiment measure is a significant predictor of mobility at the 10% level, while in model 1 sentiment is significant at 1 %. In model 3, the negativity score has a larger standard error, with a corresponding p-value of 0.148. Since we use OLS fixed effects, the parameters can readily be interpreted: a 100% proportion of fearful tweets leads to on average 16,667 fewer minutes spent outside in a county each day, and leads to a 0.94 percentage point increase in the proportion of devices staying home during work hours. Similarly, a county exhibiting completely negative tweets corresponds

[18]This corresponds to model 2 of the main specification.

Table 4: Two-way Fixed Effects OLS Results for the effect of tweet sentiment on social distancing

Model:	(1)	(2)	(3)	(4)
Variables				
fear_prop	-16,667.2***	0.0094*		
	(5,196.4)	(0.0048)		
neg_score			-46,448.1	0.0656*
			(30,243.4)	(0.0344)
NAT_Stringency	-1,138.2***	0.0029***	-1,157.0***	0.0029***
	(252.0)	(0.0003)	(251.7)	(0.0003)
STA_Stringency	-218.5**	0.0005*	-220.0**	0.0005*
	(94.12)	(0.0002)	(94.37)	(0.0002)
death	119.7**	-6.93×10^{-5}**	119.7**	-6.92×10^{-5}**
	(50.80)	(2.45×10^{-5})	(50.90)	(2.45×10^{-5})
confirmed	-12.95**	8.43×10^{-6}***	-12.94**	8.43×10^{-6}***
	(4.562)	(2.34×10^{-6})	(4.573)	(2.34×10^{-6})
Fixed-effects				
week	Yes	Yes	Yes	Yes
FIPS	Yes	Yes	Yes	Yes
Fit statistics				
Observations	16,474	16,474	16,474	16,474
Adjusted R^2	0.72062	0.85262	0.72038	0.85272
Within Adjusted R^2	0.19660	0.16258	0.19588	0.16312
Log-Likelihood	-192,287.8	31,513.9	-192,295.1	31,519.3
BIC	390,546.9	-57,056.5	390,561.5	-57,067.3
RMSE	28,374.8	0.03573	28,387.4	0.03571
Size of the 'effective' sample	16,474	16,474	16,474	16,474
F-test	70.203	156.21	70.118	156.33
F-test (projected)	777.40	616.96	773.89	619.44
Wald (joint nullity)	7.7500	130.97	7.5517	136.04

Dependent variables: (1), (3) median minutes not at home; (2), (4) proportion devices home all day
Two-way (week & FIPS) standard-errors in parentheses
*Signif. Codes: ***: 0.01, **: 0.05, *: 0.1*

Both NRC and VADER sentiment measures, county aggregated.

to a 6.5% reduction in proportion of devices leaving the home. These effect sizes, while relatively small in comarison to the Stringency Index policy variables[19], remain significant predictors for key mobility metrics. In all cases, increases in policy stringency correspond to decreased mobility, as we expect. The effect sizes are small for deaths and cases because the range of the variable is large. Concerningly, however, the cumulative death count in a county appears to have a significant effect corresponding to risk-seeking behaviour, where an additional death corresponds to an increase in minutes spent outside the home and a decrease in share of devices remaining at home. I hypothesise that confirmed cases and deaths measure the same effect, the beliefs updating with the progression of the virus; as such, including it in addition to deaths cancels out the deaths effect. Indeed, dropping the observed cases causes the coefficient sign on deaths to flip and indicate risk-averse behaviour, as it captures the entire effect. The models explain between 72% - 85% of the overall variation, and between 16% and 19 % of the variation

[19]Which are reported on a 0 to 100 scale.

within county groups over time. The F tests and Wald tests for nullity of the parameters are all highly significant.

6.2 Robustness checks

In order to verify that the results are robust, we estimate several different alternative specifications. All use the same fixed-effects OLS estimation procedure, and the regression tables referred to are in appendix B. First, the main sentiment measure uses a dataset which omits all 'retweets' from a user. This was done in order to more accurately capture the user's own risk preference, as one is be more likely to retweet inflammatory and emotional content than write it oneself. As such, when we estimate a model which uses a sentiment measure that includes retweets, we observe a drop in predictive power of the variable of interest. As in table 9, the emotion-based sentiment performs poorly, while the polarity-based sentiment is significant at the 10% level. Assuming the model is correctly specified, this indicates that the overall polarity is more consistent than emotion between tweets and retweets. This implies that individuals are likely to retweet things that conform to their overall disposition, but with differing levels of emotion.

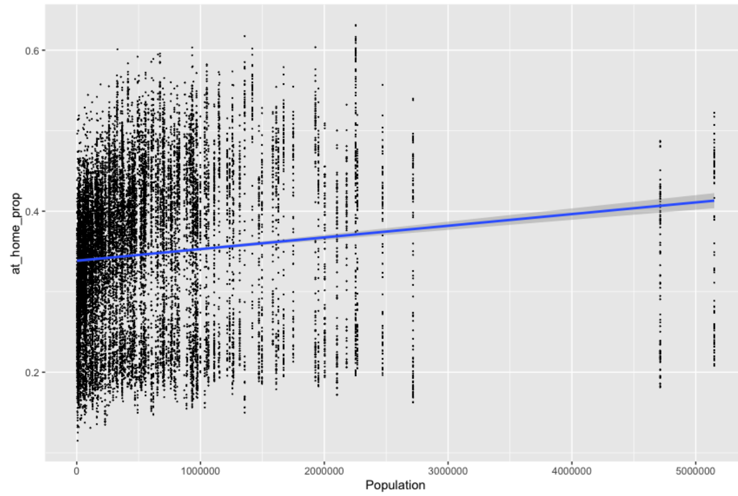
As discussed in section 5.4.1, a potentially problematic feature of the tweet data is that the number of tweets observed in a county track with the county population. This is a problem because the aggregation procedure for the sentiment index required a minimum of 30 observations for unbiasedness, so all county/days with tweets lower than this number were dropped. This means that the panel we created for the primary specification is unbalanced, with data missing not at random. This may present a selection effect, since population, the variable causing this selection, may be correlated with the dependent variable of mobility. The selection effect can be ignored if we can show that the variable driving the selection is uncorrelated with the dependent variable (Wooldridge, 2010, p. 552). On visual inspection of figure 5, a positive relationship does exist between the proportion staying at home and the county population (meaning that higher-population counties have *decreased* mobility); however, it appears that this is primarily driven by the two large county outliers. These correspond to the highly dense Los Angeles County and Cooke County, Illinois[20]. Discounting these highly dense outlier cities, we can tentatively say that mobility is largely independent of county population. As a consequence, an unbalanced panel could be accomodated. Indeed, if population outliers had a major effect on the predictive power of sentiment, removing the counties with the top and bottom 5 % of the population[21] would negate the effect. However, running the main specification after trimming these counties shows no change in the predictive power of sentiment on mobility; see table 7.

The Heckman correction procedure would be another option for obtaining unbiased parameter es-

[20] i.e., Chicago

[21] This process is known as 'Winsorising' the data.

Figure 5: Scatterplot of county population against daily mobility score.



The vertical lines in the diagram correspond to counties with mobility scores observed at different dates.

timates (Wooldridge, 1995), as would weighting the samples by population; and the final option is to avoid aggregating the sentiment scores and run the empirical specification on the individual level. That is, the base unit is an individual tweet rather than the county/day. This yields a much larger, and balanced panel of the individual observations. However, the main argument, that sentiment and risk preference varies according to region, turns on the notion of aggregating the sentiment measure from individual tweets to the county-day level. As such, estimating a model with individual tweets as the base unit does not make sense, as the dependent variable is measured at the county, rather than individual, level. However, if user fixed effects are included rather than county fixed effects, inference is possible, since user-level fixed effects also encapsulate all of the time-invariant properties of the county which the user lives in. Indeed, the sentiment measures maintain predictive power with an individual-level, user fixed-effects design: see table 8.

Another concern is that there is treatment heterogeneity, where significant results from subgroups drive the overall results. Namely, the link between mobility and sentiment in the early pandemic could be driven by the states where the virus spread first: states where the virus spread first had both a higher negative sentiment and lower mobility, and that this ‘early’ effect is confounder of the two variables. To investigate this, we split the sample into the first and second 6 weeks; however, as in table 6, the results do not change, indicating that there is no early effect.

7 Discussion and conclusion

Economic research into the response to COVID-19 was wide-ranging and speedy, and quickly established that the amount that individuals changed their behaviour varied significantly across regions and time. It also established that individual decisionmaking, rather than a response to policy, determined the majority of the behaviour change, and that several factors were key in determining both the ability and the proclivity to socially distance.

This dissertation adds to that literature by demonstrating a new source for measuring a key determinant of economic behavior in a pandemic: health risk preference. Using a novel and very large dataset, alongside the notion of regional variation in risk preference, this dissertation has argued that online communication can be an important indicator of the behavioural response to the spread of a pandemic. Under the assumption that social media posts indicate this risk preference in the form of sentiment, this disseration has used two different methods of extracting risk preferences from the posts. Both methods are a significant predictor of mobility at the county level, when estimated in both a linear model and a two-way fixed effects model. This result is robust to several alternative specifications and is largely uniform across counties and time periods; a number of possible sources of bias were also considered and addressed. This leads to the conclusion that social media sentiment can predict the extent of local mobility. However, the sentiment measures are not consistently significant predictor variables, unlike policy and COVID case variables, and the estimation procedure involves a number of data prrocessing steps. Additionally, the research design does not seek to establish a causal link, and rests on the assumption that the sentiment of social media posts are determined in part by risk preferences, and that this risk preference consists of a local and individual part. More research is needed to investigate whether these results replicate in different scenarios, to establish a firm link between risk preference and social media sentiment, and whether social media data can be used to predict other sorts of economic variables.

It is important to understand the implications of this result. The key source of data for this research is a novel method of obtaining geolocated communications expressing personal opinions. Access to the data was approved by Twitter after a lengthy application process, and these communications are entirely in the public domain. However, the public nature of the communications are likely not realised by the vast majority of users. While it must be stressed that this particular application of the Twitter data was approved and ethical, the notion that online public communications can be used to predict local behaviour is powerful. Particularly salient is the fact that the sentiment measure used in this dissertation is far less sophisticated than a large number of current techniques. This research is a timely reminder that applying text analysis to internet communications can be a powerful tool for a variety of economic actors.

References

Abadie, A., Athey, S., Imbens, G. W., & Wooldridge, J. (2017). *When should you adjust standard errors for clustering?* (Tech. rep.). National Bureau of Economic Research. (Cit. on p. 19).

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. 13)

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)

Affuso, E., & Lahtinen, K. D. (2019). Social media sentiment and market behavior. *Empirical Economics*, *57*(1), 105–127. Retrieved March 18, 2021, from https://doi.org/10.1007/s00181-018-1430-y (cit. on p. 7)

Algaba, A., Ardia, D., Bluteau, K., Borms, S., & Boudt, K. (2020). Econometrics Meets Sentiment: An Overview of Methodology and Applications. *Journal of Economic Surveys*, *34*(3), 512–547. Retrieved January 13, 2021, from https://onlinelibrary.wiley.com/doi/abs/10.1111/joes.12370 (cit. on p. 6)

_eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/joes.12370

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. 9).

Almås, I., Cappelen, A. W., & Tungodden, B. (2019). Cutthroat Capitalism versus Cuddly Socialism: Are Americans More Meritocratic and Efficiency-Seeking than Scandinavians? *Journal of Political Economy*, *128*(5), 1753–1788. Retrieved March 17, 2021, from https://www.journals.uchicago.edu/doi/10.1086/705551 (cit. on p. 10)

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. 9).

Anderson, L. R., & Mellor, J. M. (2008). Predicting health behaviors with an experimental measure of risk preference. *Journal of Health Economics*, *27*(5), 1260–1274. Retrieved March 17, 2021, from https://www.sciencedirect.com/science/article/pii/S0167629608000714 (cit. on p. 10)

Angrist, J. D., & Pischke, J.-S. (2009). *Mostly harmless econometrics: An empiricist's companion*. Princeton University Press. (Cit. on p. 19)
OCLC: ocn231586808.

Antweiler, W., & Frank, M. Z. (2004). Is all that talk just noise? The information content of internet stock message boards. *The Journal of finance*, *59*(3), 1259–1294 (cit. on p. 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. 25)

Athey, S., & Imbens, G. W. (2019). Machine learning methods that economists should know about. *Annual Review of Economics*, *11*, 685–725 (cit. on p. 59).

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. 11, 58)

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. 9)

Baker, S. R., Bloom, N., & Davis, S. J. (2016). Measuring Economic Policy Uncertainty*. *The Quarterly Journal of Economics*, *131*(4), 1593–1636. Retrieved March 18, 2021, from https://academic.oup.com/qje/article/131/4/1593/2468873 (cit. on p. 7)

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. 16).

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. 9)

Bergé, L. (2018). Efficient estimation of maximum likelihood models with multiple fixed-effects: The R package FENmlm. *CREA Discussion Papers*, (13) (cit. on p. 32).

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 pp. 6, 7)

Booth-Kewley, S., & Vickers, R. R. (1994). Associations between Major Domains of Personality and Health Behavior. *Journal of Personality*, *62*(3), 281–298. Retrieved March 17, 2021, from https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1467-6494.1994.tb00298.x (cit. on p. 10)

_eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/j.1467-6494.1994.tb00298.x

Briscese, G., Lacetera, N., Macis, M., & Tonin, M. (2020, March). *Expectations, reference points, and compliance with COVID-19 social distancing measures* (w26916). National Bureau of Economic Research. Cambridge, MA. Retrieved March 17, 2021, from http://www.nber.org/papers/w26916.pdf. (Cit. on p. 10)

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, 13).

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. 9)

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. 13)

Chapman, G. B., & Coups, E. J. (1999). Time Preferences and Preventive Health Behavior: Acceptance of the Influenza Vaccine. *Medical Decision Making*, *19*(3), 307–314. Retrieved March 17, 2021, from http://journals.sagepub.com/doi/10.1177/0272989X9901900309 (cit. on p. 10)

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. 15).

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 pp. 12, 18)

Chesson, H. W., Leichliter, J. S., Zimet, G. D., Rosenthal, S. L., Bernstein, D. I., & Fife, K. H. (2006). Discount rates and risky sexual behaviors among teenagers and young adults. *Journal of Risk and Uncertainty*, *32*(3), 217–230. Retrieved March 17, 2021, from https://doi.org/10.1007/s11166-006-9520-1 (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. 8, 25)

Choi, H., & Varian, H. (2012). Predicting the Present with Google Trends. *Economic Record*, *88*(s1), 2–9. Retrieved March 18, 2021, from https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1475-4932.2012.00809.x (cit. on p. 7)

_eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/j.1475-4932.2012.00809.x

Daas, P., & Puts, M. (2014). *Social media sentiment and consumer confidence* (Working Paper No. 5). European Central Bank. Retrieved March 18, 2021, from https://data.europa.eu/doi/10.2866/11606. (Cit. on p. 7)

Davidson, R., & MacKinnon, J. G. (2004). *Econometric theory and methods*. Oxford University Press. (Cit. on p. 59).

DellaVigna, S., & Malmendier, U. (2006). Paying Not to Go to the Gym. *American Economic Review*, *96*(3), 694–719. Retrieved March 17, 2021, from https://www.aeaweb.org/articles?id=10.1257/aer.96.3.694 (cit. on p. 10)

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. 15, 16).

Dohmen, T., Falk, A., Huffman, D., Sunde, U., Schupp, J., & Wagner, G. G. (2011). Individual Risk Attitudes: Measurement, Determinants, and Behavioral Consequences. *Journal of the European Economic Association*, *9*(3), 522–550. Retrieved March 17, 2021, from https://doi.org/10.1111/j.1542-4774.2011.01015.x (cit. on p. 10)

Dong, E., Du, H., & Gardner, L. (2020). An interactive web-based dashboard to track COVID-19 in real time. *The Lancet Infectious Diseases*, *20*(5), 533–534. Retrieved March 20, 2021, from https://www.thelancet.com/journals/laninf/article/PIIS1473-3099(20)30120-1/abstract (cit. on p. 27)

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, 12)

Falk, A., Becker, A., Dohmen, T., Enke, B., Huffman, D., & Sunde, U. (2018). Global Evidence on Economic Preferences. *The Quarterly Journal of Economics*, *133*(4), 1645–1692. Retrieved March 17, 2021, from https://academic.oup.com/qje/article/133/4/1645/5025666 (cit. on pp. 10, 17)

Fan, Y., Orhun, A. Y., & Turjeman, D. (2020, May). *Heterogeneous Actions, Beliefs, Constraints and Risk Tolerance During the COVID-19 Pandemic* (w27211). National Bureau of Economic Research. Cambridge, MA. Retrieved October 22, 2020, from http://www.nber.org/papers/w27211.pdf. (Cit. on pp. 8, 11, 17)

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. 12)

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. 13)

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. 13).

Gelman, A., & Loken, E. (2014). The Statistical Crisis in Science. *American Scientist*, *102*(6), 460–466 (cit. on p. 16).

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. 6, 14, 58)

Gentzkow, M., & Shapiro, J. M. (2010). What Drives Media Slant? Evidence From U.S. Daily Newspapers. *Econometrica*, *78*(1), 35–71. Retrieved March 18, 2021, from https://

onlinelibrary.wiley.com/doi/abs/10.3982/ECTA7195 (cit. on p. 6)
_eprint: https://onlinelibrary.wiley.com/doi/pdf/10.3982/ECTA7195

Ginsberg, J., Mohebbi, M. H., Patel, R. S., Brammer, L., Smolinski, M. S., & Brilliant, L. (2009). Detecting influenza epidemics using search engine query data. *Nature*, *457*(7232), 1012–1014. Retrieved March 18, 2021, from https://www.nature.com/articles/nature07634 (cit. on p. 7)

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. 8)

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. 8)

Gorodnichenko, Y., Pham, T., & Talavera, O. (2018, May). *Social Media, Sentiment and Public Opinions: Evidence from #Brexit and #USElection* (w24631). National Bureau of Economic Research. Cambridge, MA. Retrieved March 18, 2021, from http://www.nber.org/papers/w24631.pdf. (Cit. on p. 7)

Groseclose, T., & Milyo, J. (2005). A measure of media bias. *The Quarterly Journal of Economics*, *120*(4), 1191–1237 (cit. on p. 59).

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. 60).

Harrison, G. W., Lau, M. I., & Rutström, E. E. (2010). Individual discount rates and smoking: Evidence from a field experiment in Denmark. *Journal of Health Economics*, *29*(5), 708–717. Retrieved March 17, 2021, from https://www.sciencedirect.com/science/article/pii/S0167629610000780 (cit. on p. 10)

Herrmann, B., Thöni, C., & Gächter, S. (2008). Antisocial Punishment Across Societies. *Science*, *319*(5868), 1362–1367. Retrieved March 17, 2021, from https://science.sciencemag.org/content/319/5868/1362 (cit. on p. 10)

Huntington-Klein, N. (2020). *SafeGraphR* (Version 0.3.0). https://safegraphinc.github.io/SafeGraphR/. (Cit. on p. 25)

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. 30).

Jegadeesh, N., & Wu, D. (2013). Word power: A new approach for content analysis. *Journal of financial economics*, *110*(3), 712–729 (cit. on p. 16).

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. 16)
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. 11).

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. 6)

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. 8)

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. 16)

Manning, C. D., & Schütze, H. (1999). *Foundations of statistical natural language processing.* MIT Press. (Cit. on p. 14).

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. 27, 28)
_eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/j.1467-8640.2012.00460.x

Müller, K., & Schwarz, C. (2020). Fanning the Flames of Hate: Social Media and Hate Crime. *Journal of the European Economic Association*. Retrieved March 18, 2021, from https://doi.org/10.1093/jeea/jvaa045 (cit. on p. 7)

Müller, S., & Rau, H. A. (2021). Economic preferences and compliance in the social stress test of the COVID-19 crisis. *Journal of Public Economics*, *194*, 104322. Retrieved March 17, 2021, from https://www.sciencedirect.com/science/article/pii/S0047272720301869 (cit. on p. 10)

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. 9)

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. 9)

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. 4)

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. 26)

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. 25)

Picault, M., & Renault, T. (2017). Words are not all created equal: A new measure of ECB communication. *Journal of International Money and Finance*, *79*, 136–156. Retrieved March 18, 2021, from https://www.sciencedirect.com/science/article/pii/S0261560617301808 (cit. on p. 6)

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. 27)

Porter, M. F. (1980). An algorithm for suffix stripping. *Program* (cit. on p. 15).

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. 4, 21, 23)

Rinker, T. (2018). *Sentimentr*. Retrieved March 15, 2021, from https://github.com/trinker/sentimentr. (Cit. on p. 28)

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. 31)

Rustichini, A., DeYoung, C. G., Anderson, J. E., & Burks, S. V. (2016). Toward the integration of personality theory and decision theory in explaining economic behavior: An experimental investigation. *Journal of Behavioral and Experimental Economics*, *64*, 122–137. Retrieved March 17, 2021, from https://www.sciencedirect.com/science/article/pii/S2214804316300337 (cit. on p. 10)

SafeGraph, Inc. (2020). *Social Distancing Metrics*. SafeGraph. Retrieved January 7, 2021, from https://docs.safegraph.com/docs/social-distancing-metrics. (Cit. on pp. 5, 25)

Saiz, A., & Simonsohn, U. (2013). Proxying for unobservable variables with internet document-frequency. *Journal of the European Economic Association*, *11*(1), 137–165. Retrieved March 18, 2021, from https://academic.oup.com/jeea/article-lookup/doi/10.1111/j.1542-4774.2012.01110.x (cit. on p. 7)

Schilbach, F. (2019). Alcohol and Self-Control: A Field Experiment in India. *American Economic Review*, *109*(4), 1290–1322. Retrieved March 17, 2021, from https://www.aeaweb.org/articles?id=10.1257/aer.20170458 (cit. on p. 10)

Silge, J., & Robinson, D. (2017). *Text mining with R: A tidy approach* (First edition). O'Reilly. (Cit. on p. 28)

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 p. 9)

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 pp. 6, 7)

Strickhouser, J. E., Zell, E., & Krizan, Z. (2017). Does personality predict health and well-being? A metasynthesis. *Health Psychology*, *36*(8), 797–810. Retrieved March 17, 2021, from http://doi.apa.org/getdoi.cfm?doi=10.1037/hea0000475 (cit. on p. 10)

Stutzer, A., & Meier, A. N. (2016). Limited Self-control, Obesity, and the Loss of Happiness. *Health Economics*, *25*(11), 1409–1424. Retrieved March 17, 2021, from https://onlinelibrary.wiley.com/doi/abs/10.1002/hec.3227 (cit. on p. 10)
_eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/hec.3227

Summers, E. (2020). *Hydrator*. https://github.com/docnow/hydrator. (Cit. on p. 23)

Sutter, M., Kocher, M. G., Glätzle-Rützler, D., & Trautmann, S. T. (2013). Impatience and Uncertainty: Experimental Decisions Predict Adolescents' Field Behavior. *American Economic Review*, *103*(1), 510–531. Retrieved March 17, 2021, from https://www.aeaweb.org/articles?id=10.1257/aer.103.1.510 (cit. on p. 10)

Tetlock, P. C. (2007). Giving Content to Investor Sentiment: The Role of Media in the Stock Market. *The Journal of Finance*, *62*(3), 1139–1168. Retrieved March 18, 2021, from http://doi.wiley.com/10.1111/j.1540-6261.2007.01232.x (cit. on p. 7)

Thunström, L., Ashworth, M., Shogren, J. F., Newbold, S., & Finnoff, D. (2021). Testing for COVID-19: Willful ignorance or selfless behavior? *Behavioural Public Policy*, *5*(2), 135–152. Retrieved March 17, 2021, from https://www.cambridge.org/core/product/identifier/S2398063X20000159/type/journal_article (cit. on p. 10)

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. 59).

Toxvaerd, F. (2020). *Equilibrium Social Distancing* (Working Paper No. 2020/08). Institute for New Economic Thinking. Cambridge. (Cit. on p. 12).

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. 24)

Wooldridge, J. M. (1995). Selection corrections for panel data models under conditional mean independence assumptions. *Journal of econometrics*, *68*(1), 115–132 (cit. on p. 36).

Wooldridge, J. M. (2010). *Econometric analysis of cross section and panel data*. MIT press. (Cit. on p. 35).

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. 8)

Zettler, I., Schild, C., Lilleholt, L., Kroencke, L., Utesch, T., moshagen, m., Böhm, R., Back, M., & Geukes, K. (forthcoming). The role of personality in COVID-19 related perceptions, evaluations, and behaviors: Findings across five samples, nine traits, and 17 criteria. *Social Psychological and Personality Science*. Retrieved March 17, 2021, from https://osf.io/pkm2a (cit. on p. 10)

A Regression Tables

Full regression tables are listed in landscape at the beginning of the next page.

Table 5: Main Specification, OLS, with and without controls

Dependent Variables:	homepct		dwelltime		homepct		dwelltime	
Model:	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Variables								
(Intercept)	0.1986***	0.1865***	98,041.7***	94,559.9***	0.0674***	0.0625***	31,143.1***	29,371.1***
	(0.0025)	(0.0024)	(1,984.5)	(1,882.7)	(0.0079)	(0.0079)	(6,524.1)	(6,558.8)
meanneg	0.0817***		-49,024.6***		0.1652***		-34,204.6**	
	(0.0220)		(17,292.0)		(0.0199)		(16,540.4)	
NAT_StringencyIndex	0.0010***	0.0010***	-1,280.5***	-1,256.8***	0.0013***	0.0012***	-1,164.5***	-1,148.5***
	(5.14×10^{-5})	(5×10^{-5})	(40.37)	(39.41)	(4.66×10^{-5})	(4.54×10^{-5})	(38.64)	(37.73)
STA_StringencyIndex	0.0015***	0.0015***	112.9***	107.4***	0.0014***	0.0014***	45.61	42.55
	(4.64×10^{-5})	(4.63×10^{-5})	(36.52)	(36.50)	(4.21×10^{-5})	(4.21×10^{-5})	(34.95)	(34.95)
death	-0.0001***	-0.0001***	-10.15**	-10.48**	$-5.48 \times 10^{-5***}$	$-5.47 \times 10^{-5***}$	-10.82**	-10.90**
	(6.25×10^{-6})	(6.24×10^{-6})	(4.917)	(4.916)	(5.82×10^{-6})	(5.81×10^{-6})	(4.830)	(4.830)
confirmed	$1.33 \times 10^{-5***}$	$1.33 \times 10^{-5***}$	1.381***	1.399***	$8.17 \times 10^{-6***}$	$8.18 \times 10^{-6***}$	-0.4783	-0.4745
	(4.18×10^{-7})	(4.17×10^{-7})	(0.3287)	(0.3287)	(3.89×10^{-7})	(3.89×10^{-7})	(0.3228)	(0.3228)
meanfear		0.0522***		-3,910.6		0.0484***		-3,461.6
		(0.0053)		(4,217.1)		(0.0048)		(4,024.8)
% 65+					-0.0001	-6.6×10^{-5}	-716.7***	-730.6***
					(0.0001)	(0.0001)	(104.3)	(104.1)
Persons per HH					0.0094***	0.0115***	26,441.6***	26,225.7***
					(0.0021)	(0.0021)	(1,744.5)	(1,748.8)
HH income					$1.7 \times 10^{-6***}$	$1.66 \times 10^{-6***}$	0.0793	0.0848*
					(6.15×10^{-8})	(6.15×10^{-8})	(0.0511)	(0.0511)
% Below poverty level					0.0002	0.0002	-145.6	-139.5
					(0.0002)	(0.0002)	(128.3)	(128.4)
Pop. per sq. mile					$7.19 \times 10^{-7***}$	$7.52 \times 10^{-7***}$	2.164***	2.157***
					(9.47×10^{-8})	(9.45×10^{-8})	(0.0785)	(0.0785)
Tweets per week					$1.13 \times 10^{-6***}$	$1.13 \times 10^{-6***}$	1.395***	1.394***
					(1.16×10^{-7})	(1.15×10^{-7})	(0.0959)	(0.0959)
Fit statistics								
Observations	16,474	16,474	16,474	16,474	16,474	16,474	16,474	16,474
R^2	0.61519	0.61708	0.28535	0.28504	0.68767	0.68826	0.35368	0.35355
Adjusted R^2	0.61508	0.61697	0.28514	0.28482	0.68747	0.68805	0.35325	0.35311

Normal standard-errors in parentheses
*Signif. Codes: ***: 0.01, **: 0.05, *: 0.1*

Table 6: Main Specification, first 6 weeks and second 6 weeks ('Early' effect)

Dependent Variables:	home_devices/devices		non_home_time		home_devices/devices		non_home_time	
Model:	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Variables								
meanneg	0.0800*		-93,125.2*		0.1033**		-8,463.4**	
	(0.0410)		(46,126.6)		(0.0403)		(2,386.6)	
NAT_StringencyIndex	0.0029***	0.0028***	-1,292.6***	-1,252.3***	0.0034***	0.0034***	-753.8***	-753.1***
	(0.0006)	(0.0007)	(194.3)	(205.7)	(0.0001)	(0.0001)	(14.68)	(14.29)
STA_StringencyIndex	-2×10^{-5}	-2.34×10^{-5}	-2.535	1.955	0.0005**	0.0005**	-12.69	-12.69
	(0.0004)	(0.0005)	(176.5)	(178.8)	(0.0002)	(0.0002)	(27.80)	(27.82)
death	-0.0027	-0.0027	3,578.2	3,611.5	-1.53×10^{-5}	-1.55×10^{-5}	11.14*	11.16*
	(0.0015)	(0.0015)	(2,159.2)	(2,130.2)	(8.6×10^{-6})	(8.6×10^{-6})	(5.150)	(5.167)
confirmed	0.0005**	0.0005***	-694.8***	-698.4***	2.53×10^{-6}**	2.54×10^{-6}**	-1.300**	-1.302**
	(0.0001)	(0.0001)	(185.1)	(183.2)	(8.19×10^{-7})	(8.19×10^{-7})	(0.5186)	(0.5199)
meanfear		0.0056		-13,327.2**		0.0067		-1,011.9
		(0.0063)		(5,441.5)		(0.0052)		(899.6)
Fixed-effects								
week	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
FIPS	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Fit statistics								
Observations	4,676	4,676	4,676	4,676	11,798	11,798	11,798	11,798
R^2	0.52838	0.52735	0.87499	0.87467	0.84939	0.84897	0.85953	0.85947
Within R^2	0.25636	0.25474	0.13004	0.12780	0.11860	0.11611	0.15068	0.15033

Models (1-4) first 6 weeks, (5-8) second 6 weeks
Two-way (week & FIPS) standard-errors in parentheses
*Signif. Codes: ***: 0.01, **: 0.05, *: 0.1*

Table 7: Main specification, no retweets, winsorising by population

Dependent Variables:	completely_home_device_count/device_count		median_non_home_dwell_time	
Model:	(1)	(2)	(3)	(4)
Variables				
meanneg	0.0902**		-43,126.1	
	(0.0303)		(24,813.8)	
NAT_StringencyIndex	0.0029***	0.0029***	-1,113.2***	-1,096.2***
	(0.0003)	(0.0003)	(244.8)	(243.7)
STA_StringencyIndex	0.0005**	0.0005**	-90.75	-90.62
	(0.0002)	(0.0002)	(86.75)	(86.77)
death	-7.93×10^{-5}**	-7.97×10^{-5}**	55.38	55.54
	(2.78×10^{-5})	(2.78×10^{-5})	(32.37)	(32.40)
confirmed	9.25×10^{-6}***	9.28×10^{-6}***	-7.177**	-7.196**
	(2.41×10^{-6})	(2.4×10^{-6})	(2.572)	(2.572)
meanfear		0.0097**		-11,017.1**
		(0.0041)		(3,671.2)
Fixed-effects				
week	Yes	Yes	Yes	Yes
FIPS	Yes	Yes	Yes	Yes
Fit statistics				
Observations	14,339	14,339	14,339	14,339
Log-Likelihood	28,197.8	28,185.8	-161,851.1	-161,850.3
BIC	-51,390.2	-51,366.0	328,707.6	328,706.0
RMSE	0.03386	0.03389	19,313.0	19,311.9
Size of the 'effective' sample	14,339	14,339	14,339	14,339
F-test	178.52	178.17	72.605	72.616
F-test (projected)	594.30	588.65	557.09	557.46

Two-way (week & FIPS) standard-errors in parentheses
*Signif. Codes: ***: 0.01, **: 0.05, *: 0.1*

Table 8: Individual level, no retweets, meanfear,neg, UIDFIPS

Dependent Variables:	completely_home_device_count/device_count		median_non_home_dwell_time	
Model:	(1)	(2)	(3)	(4)
Variables				
meanfear	0.0903***		-56,954.4***	
	(0.0114)		(16,497.4)	
NAT_StringencyIndex	0.0014***	0.0014***	-1,382.6***	-1,347.0***
	(0.0001)	(0.0001)	(254.9)	(247.1)
STA_StringencyIndex	0.0015***	0.0016***	136.0	93.81
	(0.0002)	(0.0002)	(144.3)	(147.3)
death	-4.58×10^{-5}*	-4.74×10^{-5}**	155.1***	156.1***
	(2.45×10^{-5})	(2.41×10^{-5})	(54.66)	(55.20)
confirmed	3.45×10^{-6}**	3.55×10^{-6}**	-14.00***	-14.06***
	(1.59×10^{-6})	(1.53×10^{-6})	(2.980)	(3.024)
neg		0.0020***		-893.3*
		(0.0004)		(495.5)
Fixed-effects				
user_id	Yes	Yes	Yes	Yes
FIPS	Yes	Yes	Yes	Yes
Fit statistics				
Observations	3,827,258	3,826,961	3,827,258	3,826,961
Log-Likelihood	7,891,306.1	7,869,349.1	-45,251,783.4	-45,255,721.8
BIC	-4,975,095.2	-4,931,478.9	101,311,083.9	101,318,662.8
RMSE	0.03078	0.03095	33,011.6	33,075.9
Size of the 'effective' sample	3,827,258	3,826,961	3,827,258	3,826,961
F-test	32.629	32.217	14.422	14.349
F-test (projected)	2,076,756.3	2,046,626.6	523,871.3	519,392.1

Two-way (user_id & FIPS) standard-errors in parentheses
*Signif. Codes: ***: 0.01, **: 0.05, *: 0.1*

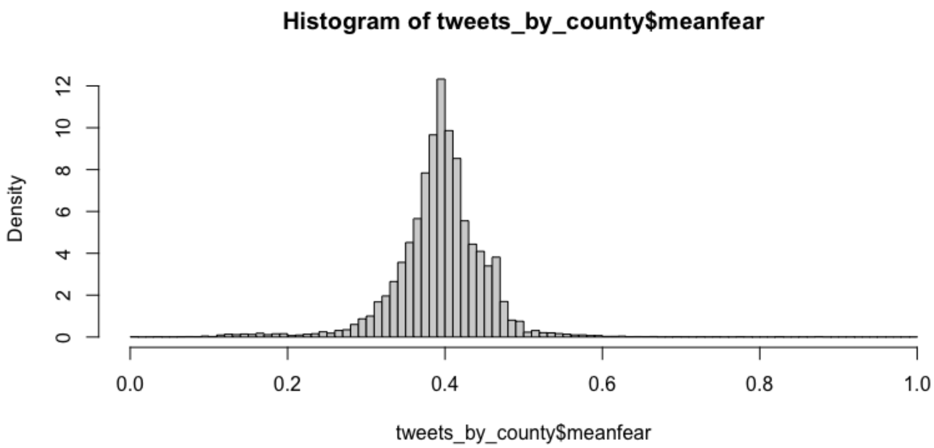
Table 9: Main Specification, including retweets

Dependent Variables:	non_home_dwell_time	home_devices/devices	non_home_dwell_time	home_devices/devices
Model:	(1)	(2)	(3)	(4)
Variables				
meanfear	333.2	0.0209		
	(3,984.5)	(0.0139)		
meanneg			-5,885.2*	0.0143*
			(3,215.6)	(0.0071)
NAT_StringencyIndex	-561.6***	0.0029***	-567.3***	0.0029***
	(109.5)	(0.0003)	(111.6)	(0.0003)
STA_StringencyIndex	-134.6**	0.0003	-136.0**	0.0003
	(61.23)	(0.0002)	(61.77)	(0.0002)
death	127.7**	-0.0001**	127.2**	-0.0001**
	(52.45)	(3.44×10^{-5})	(52.37)	(3.41×10^{-5})
confirmed	-14.34***	1.22×10^{-5}***	-14.29***	1.21×10^{-5}***
	(4.565)	(3.09×10^{-6})	(4.559)	(3.06×10^{-6})
Fixed-effects				
week	Yes	Yes	Yes	Yes
FIPS	Yes	Yes	Yes	Yes
Fit statistics				
Observations	43,258	43,258	42,683	42,683
Log-Likelihood	-493,421.0	79,625.0	-487,071.3	78,687.3
BIC	996,887.1	-149,204.9	984,068.4	-147,448.6
RMSE	21,753.5	0.03840	21,860.3	0.03829
Size of the 'effective' sample	43,258	43,258	42,683	42,683
F-test	109.43	198.68	109.40	199.60
F-test (projected)	2,094.9	1,525.4	2,062.4	1,504.3

Two-way (week & FIPS) standard-errors in parentheses
*Signif. Codes: ***: 0.01, **: 0.05, *: 0.1*

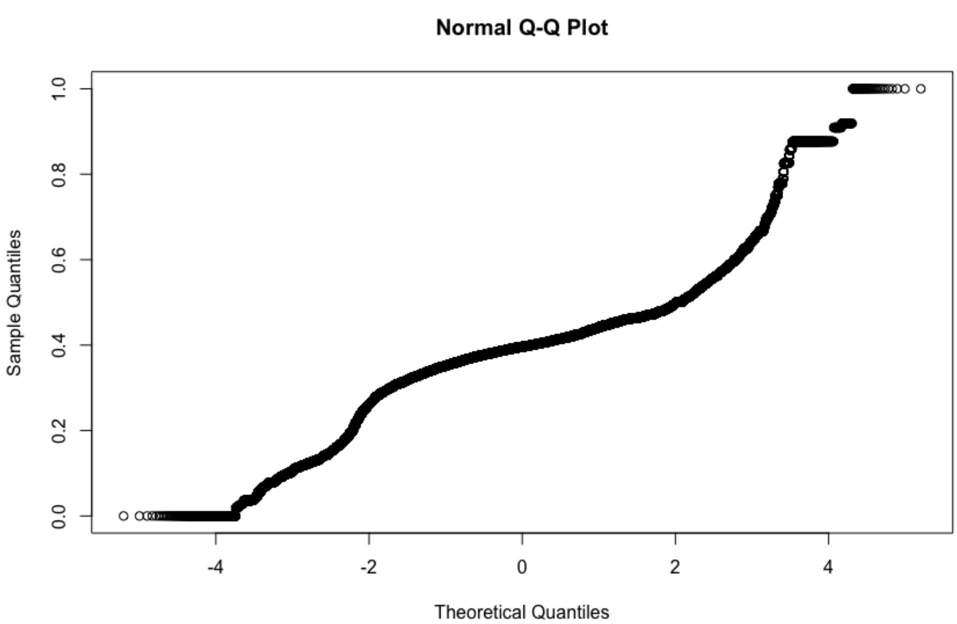
B Additional Figures

Figure 6: Histogram, proportion of tweets containing fearful language by county/day.



Variable presented after trimming county/day buckets with $n < 30$

Figure 7: Normal distribution Quantile-Quantile plot, proportion of tweets containing fearful language by county/day.



Variable presented after trimming county/day buckets with $n < 30$

Figure 8: Histogram, individual VADER sentiment scores, compound positive/negative.

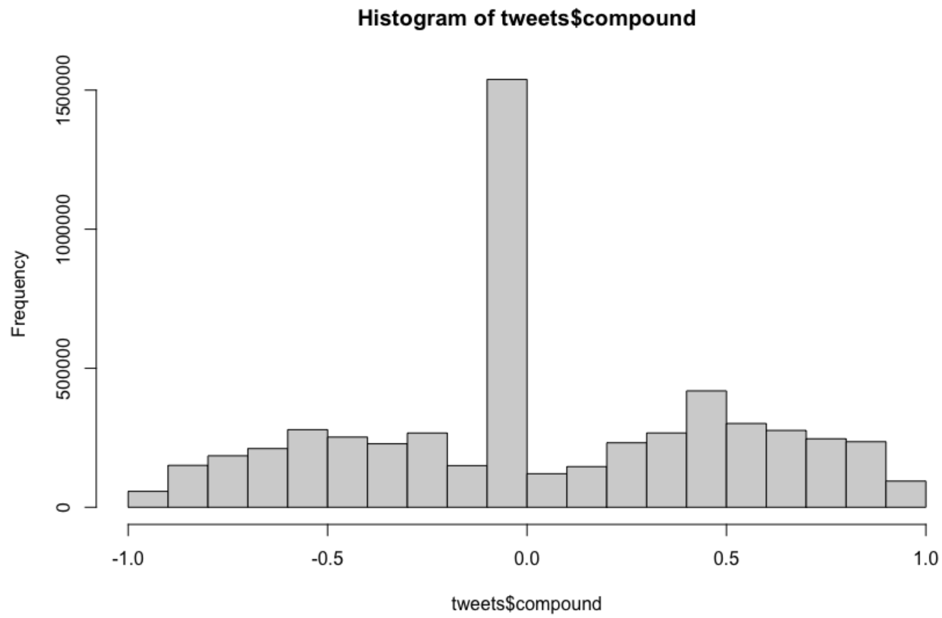


Figure 9: Histogram, mean VADER sentiment scores, compound positive/negative.

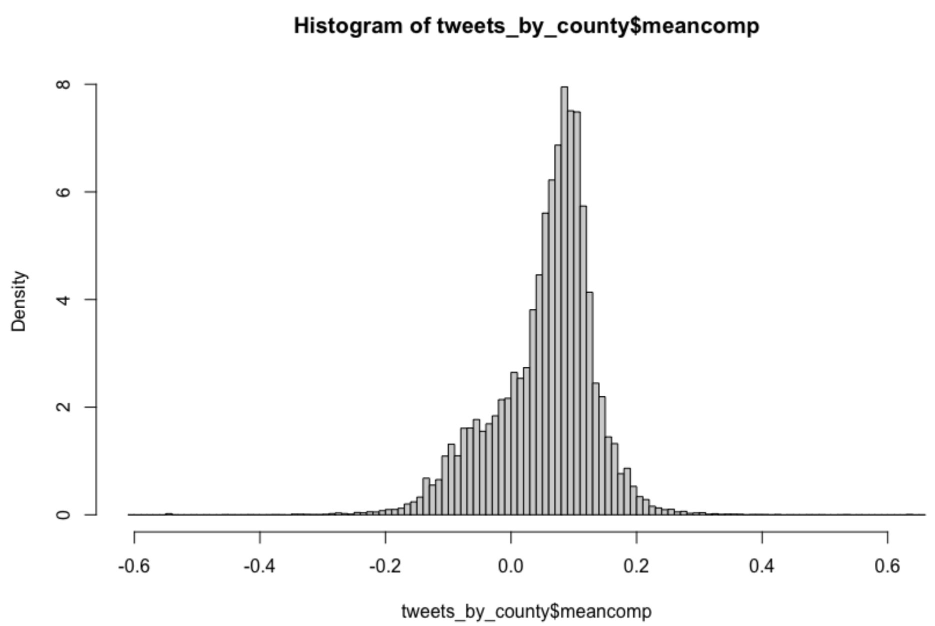
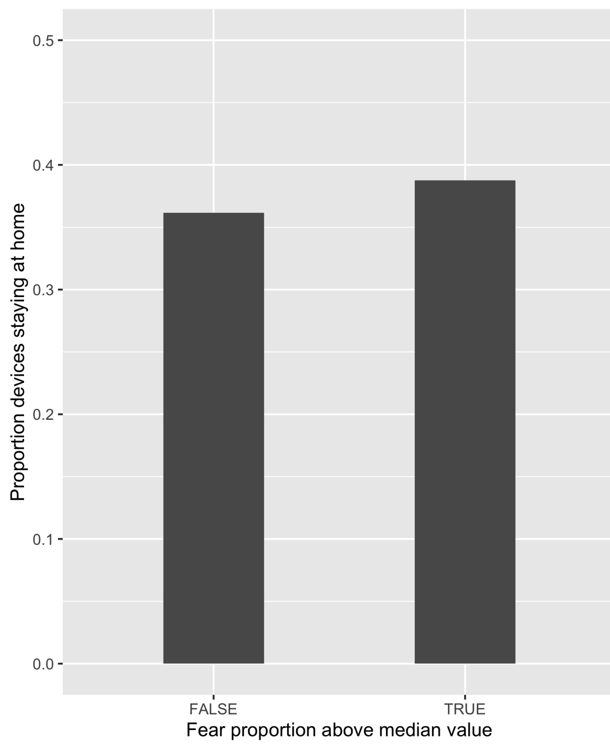


Figure 10: Bar chart, comparing mobility in low-fear counties with high-fear counties.



Low/high fear defined as above or below median proportion of fearful tweets.

Figure 11: Example sentiment analysis output of VADER package.

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

C SIR Model

Each member of the population can be in one of three 'Susceptible, Infected, Recovered' states; therefore, at each time period we have

$$S(t) + I(t) + R(t) = 1 \tag{6}$$

, where the population is normalized to 1. 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

$$\dot{S}(t) = -S(t)I(t)R_0\gamma \tag{7}$$

$$\dot{I}(t) = S(t)I(t)R_0\gamma - \gamma I(t) \tag{8}$$

$$\dot{R}(t) = \gamma I(t) \tag{9}$$

. For epidemiologists, then, R_0 is seen as a 'compound parameter' of both the virus' natural infectivity[22] and the expected number of in-person interactions during pre-pandemic life (Avery et al., 2020, p. 84).

D Text analysis: other methods of estimating $\hat{V}$

D.1 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

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

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

, 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[23] (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.

D.2 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[24] 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

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

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).